A Digitally Controlled Analog kernel for Convolutional Neural Networks

Malik Summair Asghar[1,2],Muhammad Junaid[1],Hyung Won Kim[1]
[1]School of Electronics Engineering,
Chungbuk National University,
Cheongju, South Korea.
summair@, junaid@, hwkim@chungbuk.ac.kr

Saad Arslan[3], Syed Asmat Ali Shah[2]
[2]Dept. of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad Campus, Pakistan.
[3]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Pakistan.
saad.arslan@comsats.edu.pk, asmat.ali.shah@gmail.com

2021 18th International SoC Design Conference (ISOCC) | 978-1-6654-0174-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISOCC53507.2021.9613851

Abstract—**Convolutional Neural Networks benefit from higher accuracy at the cost of higher hardware requirements and power consumption. The multiply and accumulate unit ((MAC), which perform the convolution operation inside a CNN, consumes a significant amount of power consumption. In this work we propose a mixed-signal approach for implementing analog MAC unit that can replace the digital MAC units in CNNs. The Analog MAC unit architecture is constituted from binary weighted current steering DAC circuit and capacitors. A digital parallel interfaced is designed to provide input image and filter values to the MAC unit. To realize an input layer for the CNN a low power ADC is then employed at the output to convert the final value back to digital value. When a 3×3 convolution is performed the analog MAC unit offers area and power consumption benefits.**

Keywords; Convolutional Operation, Analog Multiplier, Neural Network Accelerator, Convolutional Neural Network

I. INTRODUCTION

Convolutional Neural Networks (CNNs) deliver superior performance at the expense of growing hardware cost and power consumption. CNNs realize convolution operation using multiply-and-accumulate (MAC) units, which consume most of the power. There is a trend towards the exploration of low power and high-performance neural processing units or neural accelerators. Recent studies have focused on developing accelerators for CNNs [1], which attempt to improve the area, power consumption and delay [2]. Some researchers are exploring mixed-signal approaches for CNNs [3], where some are integrating the analog compute units directly with the image sensor [4].

In digital domain the convolutional operation is performed by using multiplier and adders. For the convolution to be performed the number of multipliers depend upon the number of filter values. Moreover, a large number of adders are required to integrate the output of multipliers. Thus, digital MAC units occupy huge area along with higher power consumption. This area and power constraint drifted the researcher's interest to find the new paradigm of Analog kernels for CNN, which can not only perform convolution but can occupy very less area and consume less power.

In this work we present and implementation of mixed signal MAC units for CNNs. a digital parallel interfaced has been employed to directly provide the image and filter values to the MAC units. these MAC units can be integrated to realize any layer of a large-scale CNN.

Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT).

II. ANLOG KERNAL AS MULTIPLY AND ACCUMULATE

A. Archtecture of Convolutional unit (CU)

Fig. 1 explain the overall architecture of the mixed-signal Convolutional Unit (CU) to replace a conventional digital CU. The CU contains n-multipliers, for convolutional layer with filter size $n = H_{filter} \times W_{filter}$ The output of the n-multipliers is summed together by an accumulator circuit. Finally, the output of the accumulator is converted to digital using an ADC.

B. Multiply and Accumulate Unit

The analog multiply and accumulate unit circuit diagram is shown in fig. 2. The implemented 4-bit binary weighted current steering DAC circuit is shown in Fig. 1(a). The left current steering DAC with operand A generates an output current I_A. This then generates an equivalent base voltage for second DAC with operand B, resulting in an output current $1_{A\times B}$. An n-multiplier accumulator circuit is implemented and is shown in

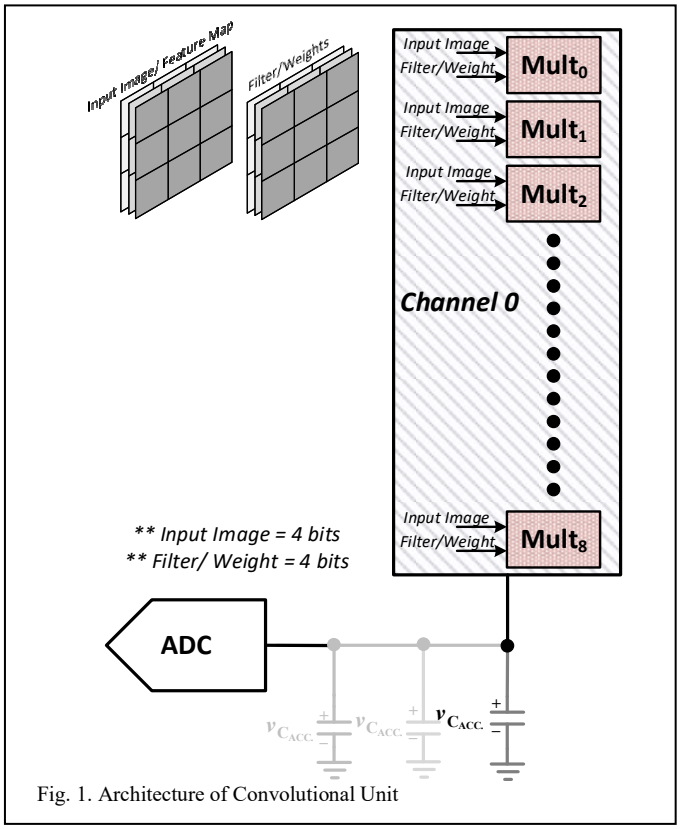

Fig. 1. Architecture of Convolutional Unit

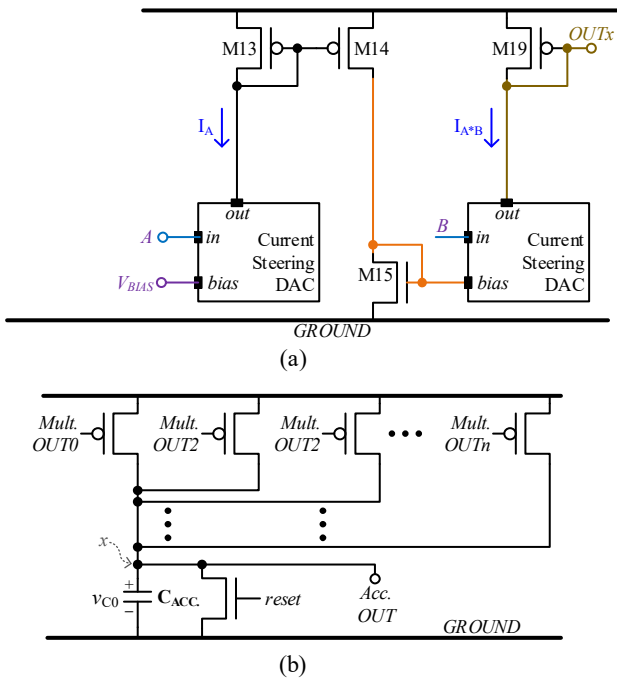

Fig. 2. Multiply and Accumulate Circuits, a) Binarry weighted Current steering DAC MAC circuit, and b) Current mirror based Accumulator.

Fig. 1(b). The accumulator circuit takes output voltage of n-multipliers and generates equivalent current which is summed up and charged upon an accumulation capacitor.

C. Digital Parallel Interface

A 32-bit parallel interface is implemented to control the analog MAC circuit. The Master side of our parallel interface will initially write Filter weights on filter memory, it will write 3x3 data in 7x7 image memory for RGB in 3 memories. Then we will send 3x3 Image and Filter Data to Analog Circuit. Figure 3 represents the architecture of our controller.

III. Simulation Results

The simulation results are shown in figure 4. For simulation, the three channels of input image and filter have maximum value of 15 applied to the one CU. At start accumulation signal the first channel's value will be provided

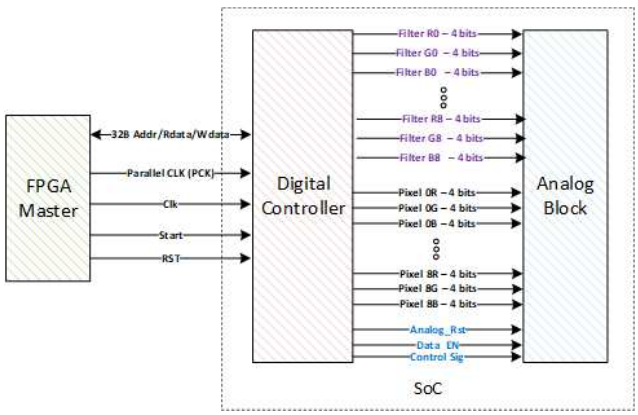

Fig. 3. Block Diagram of Digital parallel Interface.

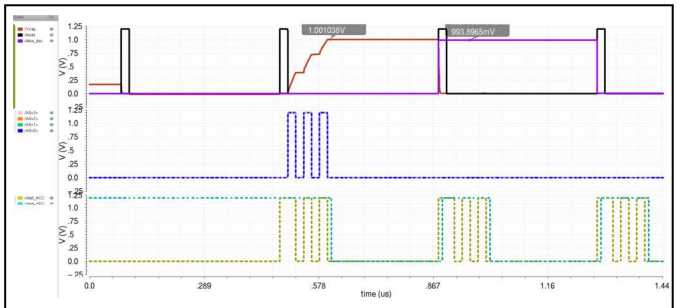

Fig. 4. Simulation results for 1 CU.

to the CU and the equivalent voltage will be charged upon the capacitor (red wave form). Then second and third channels value will be provided respectively, and the capacitor voltage will be charged, respectively. Once accumulation is finished then start ADC signal will start analog to digital conversion. The digital value is shown in purple wave form.

IV. Conclusion

This work proposes a mixed signal implementation of the digitally controlled Analog kernel for convolutional neural networks. The analog kernels can perform multiply and accumulate task much efficiently by consuming very less area and power consumption as compared to their digital counterparts. A low power ADC, is employed to convert analog values to digital values for further processing, can offer replacement of Analog kernel inside digital CNNs. In future, this implementation can be extended to realize a full-scale CNN with max pooling and relu functions implemented in Analog domain.

Acknowledgment

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2020-0-01077, Development of Intelligent SoC having Multimodal IOT Interface for Data Sensing, Edge computing analysis and Data sharing.) and supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-01304, Development of Self-learnable Mobile Recursive Neural Network Processor Technology)

1. References

[1] S.-S. Park and K.-S. Chung, "CENNA: Cost-Effective Neural Network Accelerator," *MDPI Electronics*, vol. 9, no. 1, p. 134, Jan. 2020.

[2] H. Kwon, A. Samajdar, and T. Krishna, "MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects," SIGPLAN Not., vol. 53, no. 2, pp. 461–475, Nov. 2018.

[3] J. Zhu, Y. Huang, Z. Yang, X. Tang and T. T. Ye, "Analog Implementation of Reconfigurable Convolutional Neural Network Kernels," *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Bangkok, Thailand, 2019, pp. 265-268, doi: 10.1109/APCCAS47518.2019.8953177.

[4] J. Choi, S. Lee, Y. Son, and S. Y. Kim, "Design of an Always-On Image Sensor Using an Analog Lightweight Convolutional Neural Network," *MDPI Sensors*, vol. 20, no. 11, p. 3101, May 2020.