8T SRAM Based In-Memory DAC for AI acceleration

Abhash Kumar,

Department of Electronics and Communication Engineering,

National Institute of Technology Patna,

Email: abhash2205@gmail.com

Abstract— **The present-day dominant von-Neumann based computing architecture involves separate computing and storage units where the data has to be continuously shuttled back and forth between physically separate memory and computing cores leading to the well-known *memory-wall* problem especially for data intensive application such as artificial intelligence (AI). Digital-to-analog conversion (DAC) and analog-to-digital conversion (ADC) is one of the frequently used operations to carry out AI/ML based algorithms inside any processing core. In this paper, we have presented an 8T SRAM based in-memory DAC to overcome the von- Neumann bottleneck.**

Keywords— ***Artificial Intelligence (AI), bottleneck, DAC, SRAM, von-Neumann***

I. INTRODUCTION

Most of the recent day processors works on von-Neumann architecture [1] where the program and data units are separately spatially, i.e., stored at different locations on the chip. This significantly increases delay and energy consumption due to frequent to and fro data transfer between these two cores, especially in case of data intensive applications such as in AI/ML. By facilitating computations within memory, both energy efficiency and throughput are expected to improve manifolds [2]. AI/ML operations are often preferred in analog domain to avoid complexity but digital world offers many advantages over analog world, such as ease of storage, processing etc. So, in this work, we will present an in-memory DAC inside 8T SRAM cell for AI acceleration.

II. 8T SRAM CELL FOR IN-MEMORY DAC

The 8T SRAM cell is shown in Fig. 1 (*abhash_6tsram* and two read access transistors SC1, SC2 making an 8T SRAM cell configuration). It has same write operations as that of 6T SRAM cells, but for read operations different port is used by activating RWL line (connected to DC voltage source V2). An 8T-SRAM, without modifying its basic circuit structure and bits stored, can behave as DAC unit. Consider an array of 4 cells connected as shown in Fig. 1. Under normal memory operations, the source terminal of SC1 (and also of SC3, SC5, SC7) is grounded but for DAC operation SL (source line) of same row are all connected to *vin* (voltage source V3). Then, the current flowing through these read access transistors will be proportional to their conductance (G), i.e.,

$$I \propto G \qquad (1)$$

e.g., when logic '1' is stored in memory cell, conductance of SC1 is significant, but when logic '0' is stored in memory cell, the conductance of SC1 is almost zero. Further, the conductance (G) depends on (W/L) ratio of mosfet. Hence, for SC1 to SC8 in Fig. 1,

$$G \propto (W/L) \qquad (2)$$

Hence, properly sizing the (W/L) ratios of these mosfets in each column we can obtained our proposed DAC.

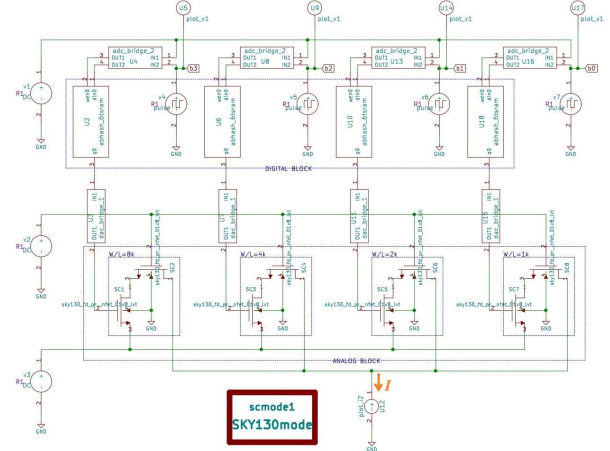Now consider that the ratio of (W/L) of mosfets used in column 1 through 4 is 8: 4: 2: 1 as shown in Fig. 1 (we will

Fig. 1. Final implemented circuit for in-memory DAC implementation.

keep the sizes of mosfet, e.g., SC1 and SC2 same in each column),

$$\left(\frac{W}{L}\right)_{SC1,SC2} = 2^3 k; \qquad \left(\frac{W}{L}\right)_{SC3,SC4} = 2^2 k;$$

$$\left(\frac{W}{L}\right)_{SC5,SC6} = 2^1 k; \qquad \left(\frac{W}{L}\right)_{SC7,SC8} = 2^0 k; \qquad (3)$$

where, k is any constant. Hence, using equation (1), (2), and (3); the total output current I (see Fig. 1(a)) will be:

$$I \propto (2^3 b_3 + 2^2 b_2 + 2^1 b_1 + 2^0 b_0) = w$$

where, $w = (2^3 b_3 + 2^2 b_2 + 2^1 b_1 + 2^0 b_0)$ is analog equivalent of digital weight $w = \{b_3, b_2, b_1, b_0\}$ stored in SRAM cell array. Hence, SRAM cell can be exploited to carry out in-memory DAC.
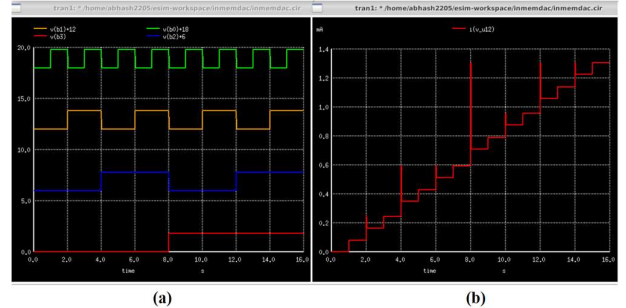
III. FINAL WAVEFORMS

(a)   (b)

Fig. 2. Plot showing (a) digital 4-bit data $\{b_3 b_2 b_1 b_0\}$ stored in SRAM cell, (b) output current, I (simulated in **eSim**) proportional to the analog equivalent of weight w stored within SRAM.

REFERENCES

[1] J. Von Neumann, The computer and the brain. Yale University Press, 2012.

[2] S. Jeloka, N. B. Akesh, D. Sylvester, and D. Blaauw, "A 28nm configurable memory using push-rule 6T bit cell enabling logic-in-memory", IEEE J. Solid-State Circuits, vol. 51, no. 4. Pp. 1009-1021, Apr. 2016.

[3] A. Jaiswal, I. Chakraborty, A. Agrawal and K. Roy, "8T SRAM Cell as a Multibit Dot-Product Engine for Beyond Von Neumann Computing," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 11, pp. 2556-2567, Nov. 2019.