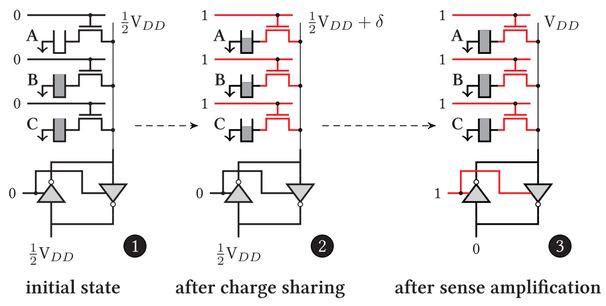
IDEC Chip Design Contest

A 1T1C DRAM Compute-in-Memory BNN Accelerator

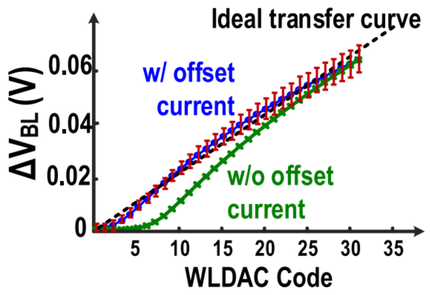
Jihwan Cho, Dalta Imam Maulana, Minkyu Yang, Changjoo Park, and Wanyeong Jung
School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST)
E-mail: {jihwancho, wanyeong}@kaist.ac.kr

Motivation

- Compute-in-memory (CIM) accelerators using SRAM and non-volatile memory (NVM) suffer from data transfer costs due to limited memory capacity
- DRAM CIM → data proximity & high cell density
- DRAM CIM **Challenge 1**: data is destroyed after computation
- DRAM CIM **Challenge 2**: output non-linearity of analog CIM

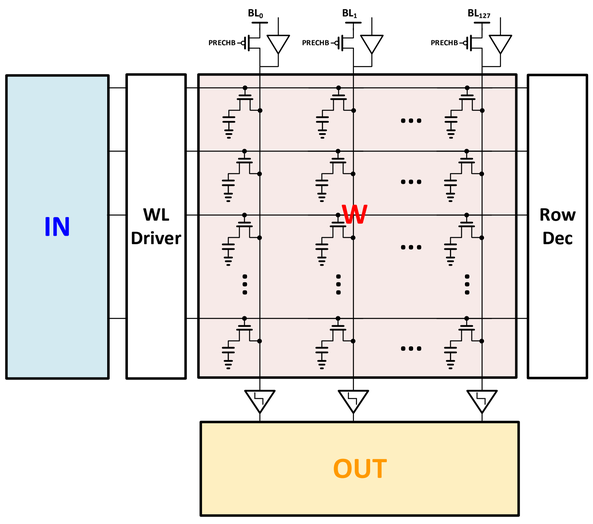


Challenge 1
V. Seshadri, MICRO 2017



Challenge 2
Jintao Zhang, JSSCC 2017

1T1C CIM BNN Accelerator

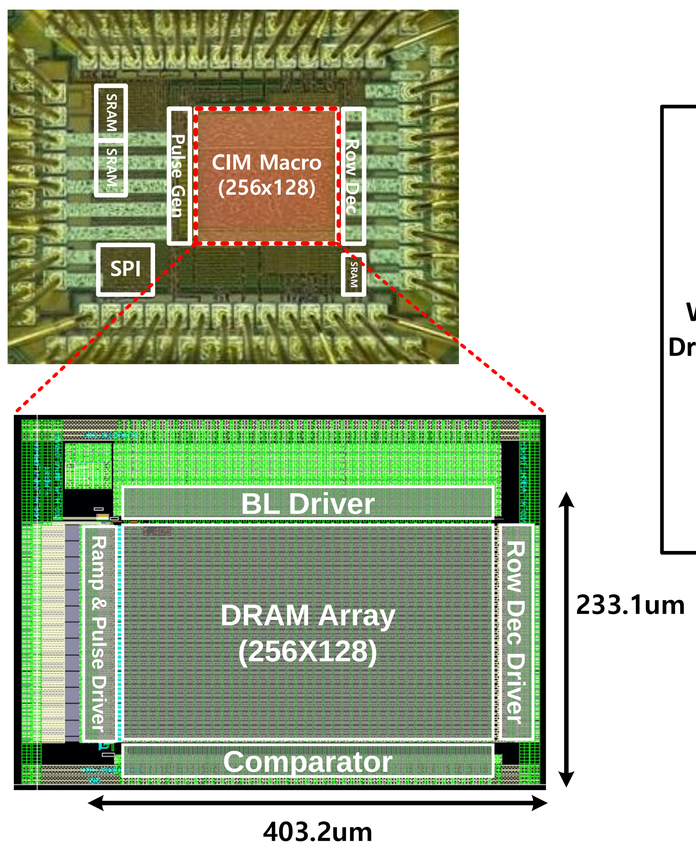


Block diagram of 1T1C CIM BNN accelerator

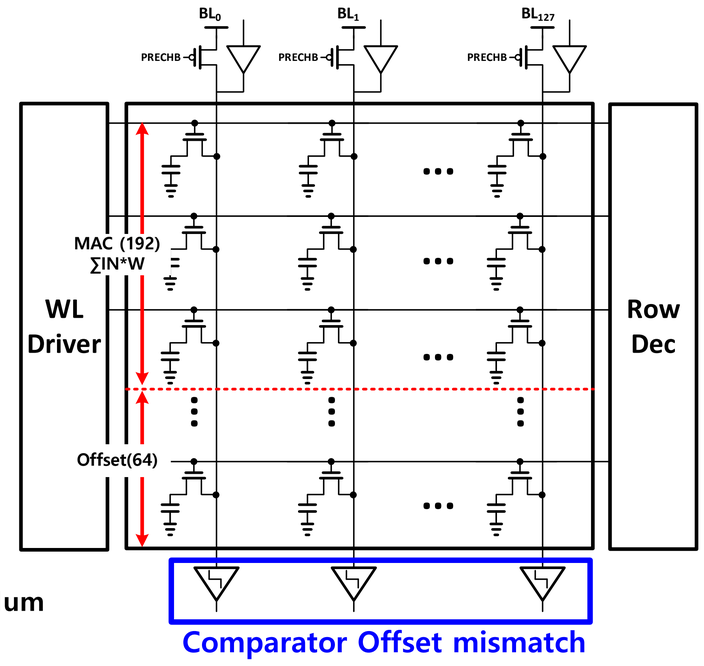
- Analog CIM using 1T1C DRAM cell
- Whole array cells performs binary MAC in parallel
- 1-bit weight (1/0) is stored in each cell, and 1-bit input (+1/-1) is fed to WL
- The stored weight is reused multiple times unlike data destructive DRAM-CIM
- Comparator is used for BL readout and perform binary activation so that reduces ADC overhead

Measurement Results

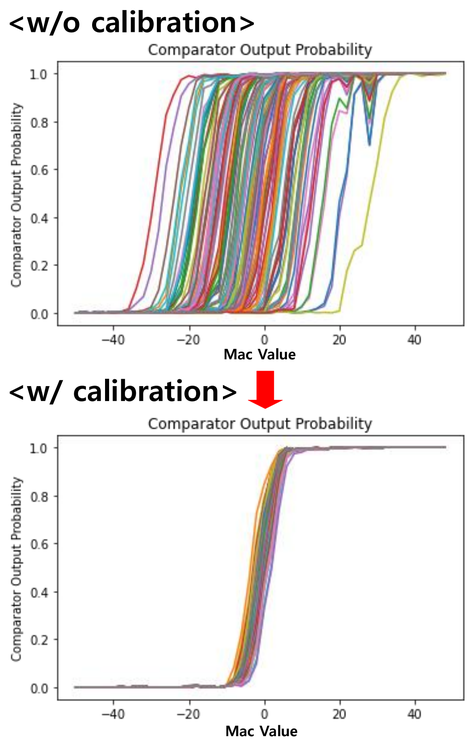
Chip SUMMARY	
Technology	28nm CMOS
Chip area	1.26mm^2
Macro area	0.094mm^2
Macro capacity	32.7 Kb
Supply Voltage	0.8V - 0.9V
CLK Freq	1.01 GHz
Bit precision	1b (IN : +1/-1) (W: +1/0)
Throughput	4.14 TOPS
TOPS/W (CIM Macro)	3281.9
TOPS/W (System)	568.2

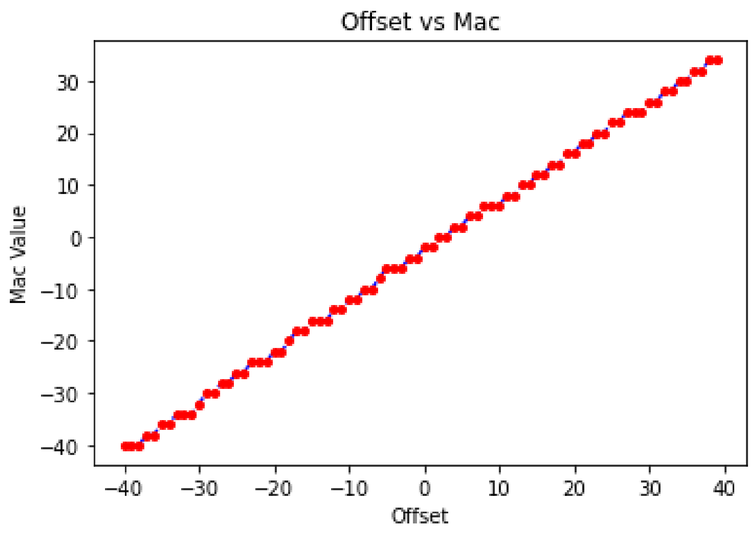


Chip summary and Die photo with CIM Macro Layout



Comparator offset self-calibration using offset cells





Accuracy and linearity of computation output

- The test chip is fabricated in a 28 nm CMOS technology including a 256x128 1T1C DRAM CIM array, SRAM for data storage, and a digital controller.
- Total chip area is 1.26 mm^2 and CIM macro occupies 0.094mm^2
- Offset mismatches between 128 BL Comparators are self-calibrated using offset calibration cells.
- Mac computation output is measured using comparator threshold and it is highly accurate and linear as shown in the right bottom plot.
- The CIM accelerator achieves energy efficiency of 568.2 TOPS/W for system and 3281.9 TOPS/W for CIM macro.

Conclusion

- Analog CIM BNN accelerator using 1T1C DRAM array is fabricated in 28 nm CMOS technology.
- The accelerator enables data-reusable and linear DRAM CIM while achieving high throughput and energy efficiency.
- Energy efficiency of 568.2 TOPS/W for system and 3281.9 TOPS/W for CIM macro is achieved.

Acknowledgement The chip fabrication and EDA tool were supported by the IC Design Education Center(IDEC), Korea.

반도체설계교육센터 IC DESIGN EDUCATION CENTER