Research on Differential Type STT-MRAM Cell for Accelerator Based on Low-Power Digital CiM Architecture

Wang Yongcheng[1,3], Li Tao[1,2] , and Tetsuo Endoh[1,2,3]

[1]School of Engineering, Tohoku University
[2]Center for Innovative Integrated Electronic Systems, Tohoku University
[3]Graduate Program in Spintronics, Tohoku University

1. Introduction

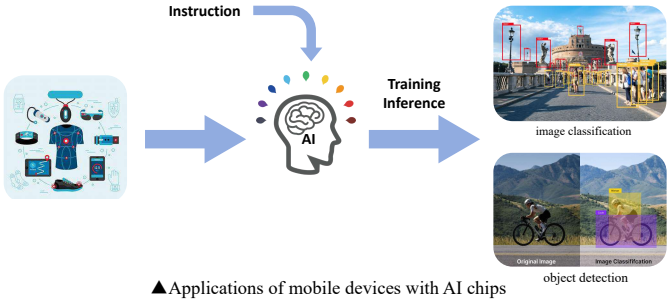
Recent artificial intelligence (AI) technologies make mobile devices more and more intelligent. However, the tremendous amount of data and frequent data transfer between memories and processing elements during processing bring high power consumption to AI chips. Using the Computing in Memory (CiM) architecture to design AI chips is an effective way to reduce power consumption. Spin-Transfer Torque Magnetoresistive Random Access Memory (STT-MRAM) is one of the most promising emerging verified technologies because of its non-volatile, high-speed, high-density, and high-reliability characteristics compared to conventional non-volatile memories. However, most designs of STT-MRAM-based CiM architecture are suffering from the limited accuracy and access speed due to the limited TMR. The goal of this research is to design a differential type STT-MRAM-based CiM cell that can avoid these problems for AI chips.

2. Computing-in-Memory (CiM)

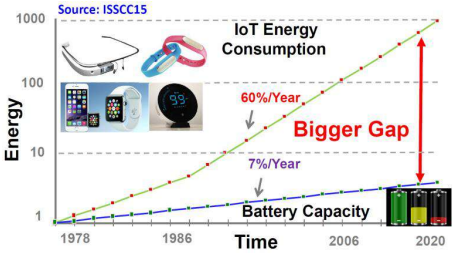
The application of AI chips makes mobile devices more intelligent, but they also consume more power. In particular, the rate of increase in power consumption is faster than the rate of increase in battery capacity of mobile devices, and this has already become a limitation to the development of mobile devices. Therefore, low power consumption of AI chips, which is a very important part of power consumption of mobile devices, needs to be researched.

One of the reasons for the high-power consumption of AI chips is the traditional von Neumann architecture. In this architecture, memory and processing elements are separated, and data must constantly move between the two parts during the process of computation. This data transfer consumed a large amount of power. To solve this problem, one solution is the Computing-in-Memory (CiM) architecture.

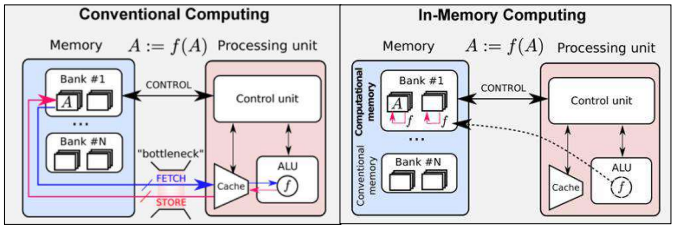
CiM is a new chip architecture in which operations proceed in memory. This architecture can save power consumption for data transfer. Now CiM has become a very hot research topic in the hardware design for AI chips.



▲Applications of mobile devices with AI chips



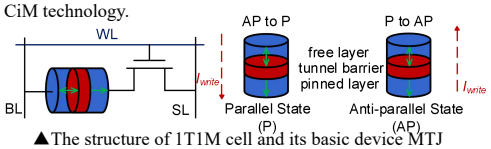
▲Gap of battery capacity and energy consumption of IoT devices



▲Von Neumann architecture ▲CiM architecture

3. STT-MRAM

CiM research can utilize a variety of memory types, but recently, non-volatile memory-based CiM research has gained significant attention. Among non-volatile memories, STT-MRAM is considered the most promising new type of memory because of its advantages, such as high speed, high density, low power consumption, and high reliability. We can achieve more low-power goals if we combine STT-MRAM and CiM technology.



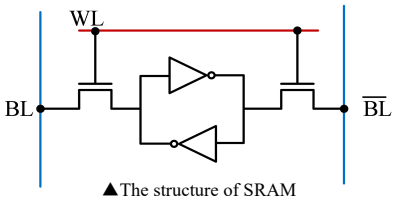
▲The structure of 1T1M cell and its basic device MTJ

	Existing			Emerging		
Memories	SRAM	DRAM	NAND Flash	ReRAM	PCRAM	STT-MRAM
Cell size (F^2)	50-120	6-10	5	6-10	4-19	6-20
Read time (ns)	≤ 2	30	10^3	1-20	≈ 2	1-20
Write time (ns)	≤ 2	50	10^6	50	10^2	≈ 10
Read energy (fJ)	1-5	100	10^6	1000	10	10-20
Write energy (fJ)	1	1000	10^6	1000	100	100-200
Endurance (cycles)	10^{15}	10^{15}	10^5	10^5	10^{10}	10^{15}
Non-volatility	NO	NO	YES	YES	YES	YES

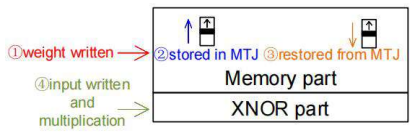
▲ Benchmark of STT-MRAM and other mainstrem memories

4. Proposed STT-MRAM-based DCiM architecture

To design a high speed and high accuracy CiM cell, it is important to ensure stable data storage within the cell. One way to realize this is by making the cell bistable structure to form a latch, like that of SRAM. STT-MRAM based CiM cell with a bistable structure is called differential type STT-MRAM. One common way to realize the logic computing is by putting the logic gate into the cell. We write the weights of deep neural networks into the cell and store them in the MTJ in advance. When the cell receives the input data, it restores the weight from the MTJ to the storage nodes, multiplies it with the inputs, and outputs the results outside.



▲The structure of SRAM



▲Structure and working steps of proposed differential type STT-MRAM based CiM cell.

5. Conclusion

This brief proposes a differential type STT-MRAM based CiM cells to solve the limited accuracy and access speed due to the limited TMR.

Acknowledgements

This research is funded by Next-generation Novel Integrated Circuits Centers (X-NICS) and Pioneering Research Support Project.