# 國立清華大學 電機工程學系

# 實作專題研究成果摘要

# A 16-kb 40-nm 1T1MTJ STT-MRAM Macro with Near-Memory Multiply-Accumulate Functionality and Single-Cap Offset-Canceled Sense Amplifier for Security-Aware Mobile Device

# 40nm 製成之 16-Kb 單電容偏壓消除感測放大器結合近記憶體乘積累加運算功能之自旋轉移力矩式記憶體

專題領域：系統組

組　　別：A217

指導教授：張孟凡教授

組員姓名：高宇勝、劉芷芸、白鈞丞

研究期間：2021 年 7 月 1 日至 2022 年 4 月底止，計 10 個月

# I. Abstract

Non-volatile memory(NVM), spin-transfer torque magnetoresistive random access memory (STT-MRAM) is being developed to realize nonvolatile working memory because it provides high-speed accesses, high endurance, and CMOS-logic compatibility. Furthermore, 1T1MTJ STT-MRAM is a promising candidate for next-generation high-density embedded non-volatile memory.

We present a comprehensive device of 16-Kb STT-MRAM with two-bit inputs two-bit weights near memory computing simultaneously for high performance applications. However, 1T1MTJ STT-MRAM suffers from limited sensing margin and high power consumption.

In order to save the power, we have a self –MAC termination to turn off the read mechanism when the inputs are all 0. In this way, about a half reduction in power consumption is corresponded.

As for the small difference between the high-resistance state (RAP) and the low-resistance state (RP), which leads to the small sensing margin, we apply a single-capacitor offset-canceled sense amplifier and self-generated voltage reference. Compared to those with multiple capacitors, our sensing amplifier reduces the area overheads successfully due to only one capacitor. And since the reference cells are located in the same row as the selected cells, row-wise PVT variations are well tracked and compensated.

Moreover, compared to the conventional STT-MRAM, which aimed at protecting NVM data that are susceptible to reverse-engineering attacks resulting in the leakage of secured keys, we further address its decoder to realize the self-protection by virtue of 6T-XOR based memory protector.

We proposed the modification to STT-MRAM that is implemented by the area-saving sensing amplifier, power-saving mechanism, and protector with area-saving XOR gates. The sensing amplifier canceled the offset for 50%, and the data protector do the securing with only 2% area overhead. To sum up, this work achieves optimized access-time of 2.98ns, and average power of 0.428mW, and Monte analysis got 95% accuracy. The achievements can be seen in Fig.1.

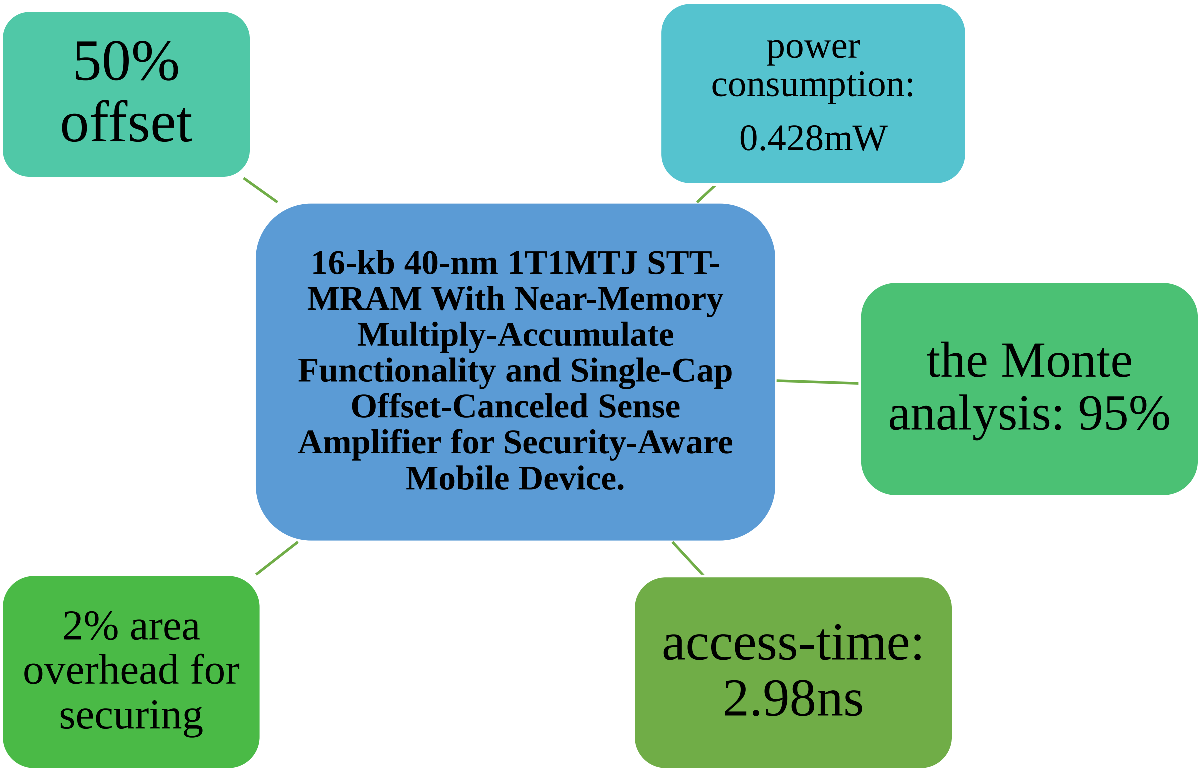

50% offset
power consumption:
0.428mW
16-kb 40-nm 1T1MTJ STT-MRAM With Near-Memory Multiply-Accumulate Functionality and Single-Cap Offset-Canceled Sense Amplifier for Security-Aware Mobile Device.
the Monte analysis: 95%
2% area overhead for securing
access-time: 2.98ns

Fig.1. Achievements of this project

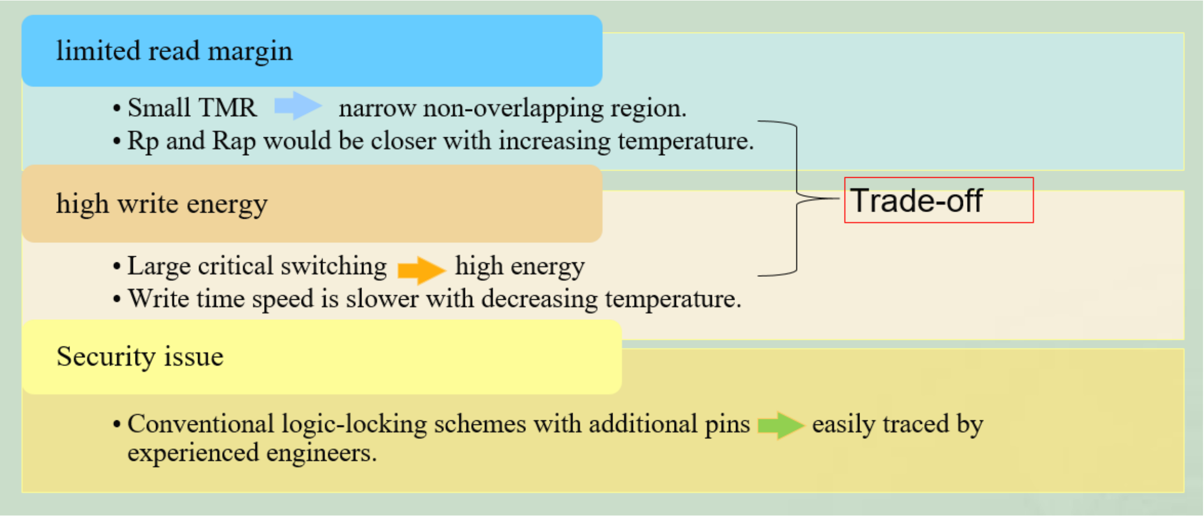
# II. Background and Purpose

MRAM has been considered as a promising replacement of SRAM and DRAM in the cache and memory system design thanks to many advantages, including fast read-write operation, non-volatility, low leakage power, capability of integration to current semiconductor process, SRAM comparable read performance and read energy consumption, higher density than SRAM, better scalability than conventional CMOS technologies, and good CMOS compatibility. STT-MRAM is an advanced type of MRAM device. STT-MRAM enables higher densities, low power consumption and reduced cost compared to regular devices (conventional MRAM). Literally, STT stands for Spin-Transfer Torque, which means the spin of the electrons is flipped using a spin-polarized current. The different direction of current flow generates the parallel resistance state(Rp) and the antiparallel resistance state(Rap). The MTJ consists of one fixed layer, one tunneling barrier layer, and one free layer. If the current flows from free layer to fixed layer, the spin polarization of the free layer will become the same as the fixed layer. The parallel resistance keeps relatively low and we call it Rp. On the contrary, if the current is from fixed layer to free layer, it is Rap. Compared with NOR flash, STT-MRAM has improved performance, higher endurance, and lower energy. However, STT-MRAM still suffers from the following challenges. Compared with NOR flash, STT-MRAM has improved performance, higher endurance, and lower energy. However, STT-MRAM still suffers from the following challenges.

First of all, the read margin is limited because the tunnel magnetoresistance ratio (TMR) is relatively small resulting in the quite narrow non-overlapping region. Therefore, it is difficult to set reference voltage to distinguish Rp voltage from Rap voltage. Moreover, the resistance distribution is influenced by different temperatures. With increasing temperature, the distribution of Rp and Rap would be even closer.

Secondly, the high write energy and long write latency due to large critical switching current bring design challenges. If we decrease the temperature to match the sensing margin, the write time speed would be slower than that in lower temperature. As a result, a fixed write time that ensures successful write for all conditions wastes a significant amount of energy at typical conditions.

Additionally, in terms of the security, conventional logic-locking schemes with the pins, write control, addresses, and data IO well defined. The additional pins used for the input of KEY data can disclose the usage of logic-locking features and the data may be traced by experienced engineers.



limited read margin

- Small TMR → narrow non-overlapping region.
- Rp and Rap would be closer with increasing temperature.

high write energy

- Large critical switching → high energy
- Write time speed is slower with decreasing temperature.

Trade-off

Security issue

- Conventional logic-locking schemes with additional pins → easily traced by experienced engineers.

Fig.2. Challenges of MRAM

# III. Proposed Sensing method

Fig.3 is the block diagram of our designed circuit and the flowchart. There are three parts in this program, including the security process, the memory macro, and the MAC function. There are following characteristic in our design:

(1) The self-generated row-wise reference cell are installed on the same WL as the MRAM cell array for tracking the PVT variation in every row.

(2) Offset-canceled sensing amplifier with single capacitor for better sensing margin, smaller area, and optimized the access-time.

(3) Self-MAC-termination is set after the output of the sensing amplifier, and does the calculation on 2-bit weight and 2-bit input. It can turn down the clampers, which is the discharging mechanism for BL, so the results of MAC remain the same, and the power can be saved.

(4) 6T-XOR based memory protector is installed before the memory macro. It contributes to select either the right or the wrong address depending on the keys the users gave. Its objective is to protect the data to be traced.

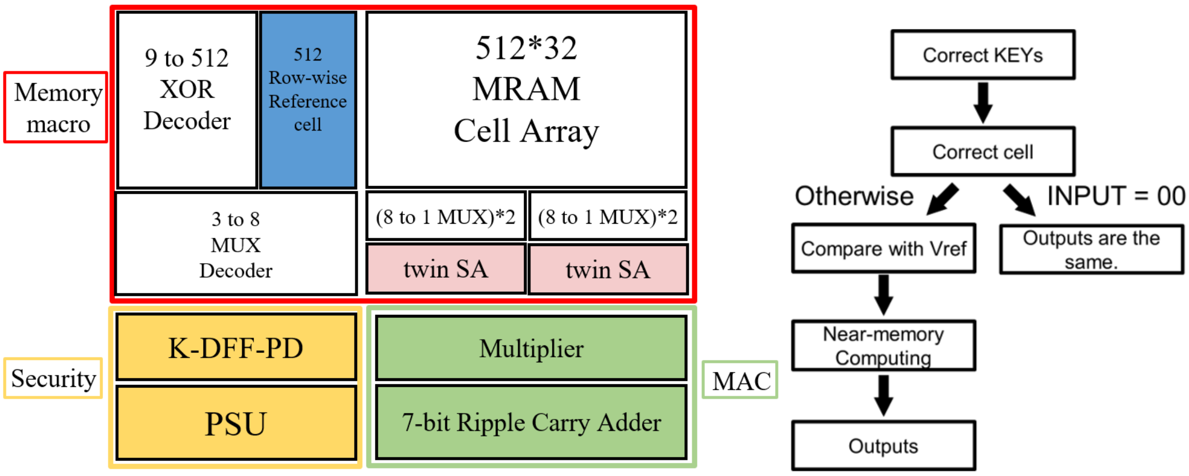
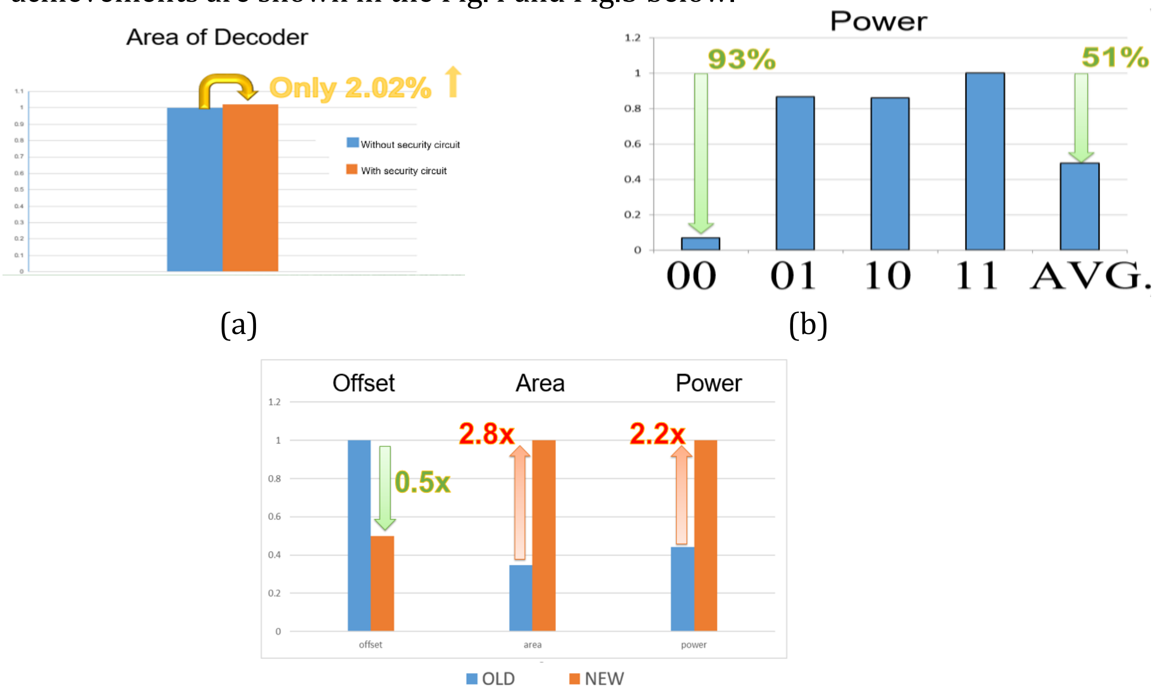


Fig.3. Overview of the proposed whole macro an its flow chart

# IV. Conclusion

After our efforts on this program, we fulfill some simple operations near the memory. In order to decrease the time of data transfer, near-memory computing is essential. In addition, for the sake of protecting the data, the data protector adds to the memory. This work achieved a 16-kb 40-nm 1T1MTJ STT-MRAM with Near-Memory Multiply-Accumulate Functionality and Single-Cap Offset-Canceled Sense Amplifier for Security-Aware Mobile Device, which cancels the offset for 50%, protects the data with only 2% area overhead, has an access-time of 2.98ns and average power of 0.428mW and Monte analysis got 95% accuracy. The overall achievements are shown in the Fig.4 and Fig.5 below.



(a)

(b)

(c)

Fig.4. performance of (a) XOR-based pre-decoder (b) self-MAC-termination (c) offset-canceled amplifier

| **Temperature(℃)** | **25** | **125** | **-40** |
|---|---|---|---|
| **access time(ns)** | **2.98** | **2.75** | **3.18** |
| **power(mW)** | **0.428** | **0.437** | **0.439** |

Fig.5. performance of the overall circuit

# V. 心得感想

經過近一整年在張孟凡教授實驗室的扎實訓練，我們從一開始的大量閱讀 paper 並在有限時間內上台做簡報、將重點彙整給大家從中彼此學習，過程中教授與學長們也常常給我們突破性的提點，讓我們在學習新電路時更容易上手。接下來選擇自己感興趣的記憶體類別加深著墨，閱讀更多相關 paper，從模仿中學習電路的原理與作動，了解電路設計在取捨(trade-off)的精隨。

在這次專題中，我們實現了近記憶體的乘加運算與省功耗的成果。尤其 MRAM 部分，從一開始的 array 建置與原理、decoder 等 peripheral circuits 逐步形成可以獨立運作之記憶體架構，經過多個 switches 之 sensing amplifier，最後，增加乘加運算與記憶體內自我保護機制等功能，慢慢形成整塊電路的完整架構，一開始，我們設計的架構感測裕度(sensing margin)太小，以致無法正確讀取記憶體內所存之單元存值、輸入訊號延遲等等問題都曾經使我們困頓，透過不斷重複閱讀相關文獻與嘗試推理一步一步解決問題。

最後，整個專題學習過程非常感謝張孟凡老師與實驗室學長姊的協助，提供實驗室的資源，在我們模擬電路時，點出我們沒觀察到的現象與可能面臨的問題，在我們遇到無法解決的問題時，陪伴我們一起討論 waveform 不合理的原因，使我們在最後專題呈現時能夠獨立解決所遇到的困難。在這段專題過程中，我們遭遇到許多困難，並深知我們對於相關領域的不足，也期許整個專題過程的所學能對未來研究所有進一步的幫助與學習。