D20-030 **DESIGN GROUP**

作品名稱

應用於安全相關功能之具近記憶體運算高讀取帶寬磁阻式記憶體巨集

A 22nm 1Mb 1024b-Read and Near-Memory-Computing Dual-Mode STT-MRAM Macro for Security-Aware Mobile Devices

隊伍名稱

得獎的是 And the Winner Goes to

隊長

邱硯晟 清華大學電機工程研究所

隊員

李俊穎 清華大學電機工程研究所
黃筱喻 清華大學電子工程研究所
鄧士新 清華大學電機工程研究所

作品摘要

許多具有安全相關功能的裝置使用雜湊演算法（SHA algorithm）與進階加密標準（advanced encryption standard，AES）等演算法將儲存於非揮發性記憶體內的資料進行加密。而為了實現這些功能，記憶體需要具備低延遲、高帶寬的特性。隨著新型態的非揮發性記憶體發展漸趨成熟，其中以磁阻式記憶體（spin torque transfer– magnetic random access memory，STT–MRAM）最具有競爭力，成為下世代記憶體。本研究將使用新型態記憶體—磁阻式記憶體，並結合安全相關應用設計周邊電路，以實現高帶寬、低延遲並具有安全應用價值近記憶體運算之巨集（如圖一）。

磁阻式記憶體巨集在設計及應用層面有三個考量：（1）在電路上並排擺列許多記憶體感測放大器達到較高的 IO 數以降低資料拿取時間會導致大量的面積消耗以及大量的峰值電流；若使用較少的感測放大器，以時序上連續性來達成讀取雖然可以降低峰值電流以及面積消耗，但其總讀取時間相比於並排擺列感測放大器更長，讀取帶寬也較低。（2）在整合型晶片上，若記憶體巨集的峰值電流較高，會影響整體晶片的電源供應，進而導致對電源供應敏感度較高的電路產生錯誤。（3）在傳統的架構中，記憶體與邏輯單元分別開來，若要完成一次讀取與位移、旋轉的運算會有很長的延遲（共需約兩個周期：記憶體讀取 + 旋轉、位移邏輯運算）

本作品提出多位元電流感測放大器（multibit current–mode SA：MB–CSA），以達成低讀取延遲、低峰值電流並實現高讀取帶寬。此外，本作品也提出近記憶體計算單元以達成位移、旋轉的運算功能，加速含有相關運算的演算法。本作品在 22 奈米製程實現 1Mb 磁阻式記憶體巨集，並具有高帶寬及近記憶體計算的特點。在電源供應為 0.85 伏特時，高達 1024 位元資料可在 2.75 奈秒平行讀出。為目前已知非揮發性記憶體中最高讀取帶寬者（42.67GB/s）。讀取所消耗能量 ERD 為 0.23pJ/b。而本作品為首度提出近記憶體計算的磁阻式記憶體巨集。降低 33.3% 的邏輯運算單元面積消耗且在 1 位元位移的運算上僅有 170ps 的延遲。

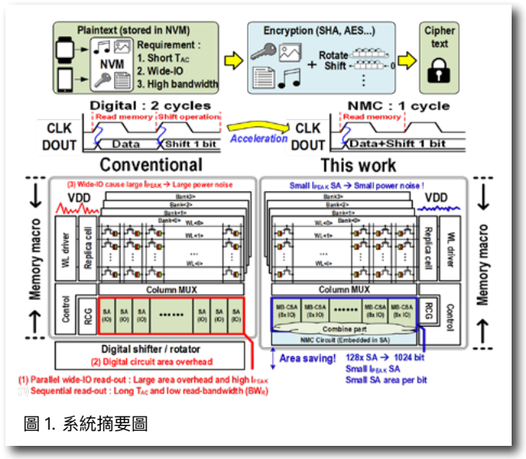
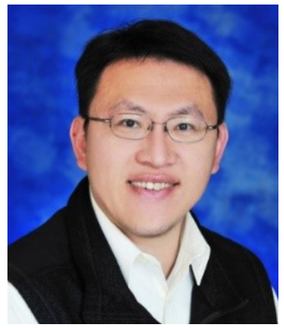


圖 1. 系統摘要圖

指導教授

張孟凡 清華大學電機工程學系

- 現任清華大學電機工程學系特聘教授，曾於工業界服務 10 年以上（台積電、積丞科技等）。
- **研究領域：**記憶體內運算電路設計（深度學習人工智慧晶片應用）、記憶體積體電路設計、使用記憶體只安全與區塊鏈電路設計、仿生人工智慧晶片之憶阻器電路設計、低功率及低電壓積體電路設計、自旋電路與非揮發邏輯電路設計、新奈米元件之電路設計。



Abstract

Many security-aware mobile devices, using the secure hash algorithm (SHA) or the advanced encryption standard (AES) for data encryption, require short readaccess time (tAC) and wide-IO from nonvolatile memory (NVM) for high-read bandwidth and SHA/AES shift/rotate functions. STT-MRAM is the major on-chip NVM for advanced process nodes; however, it requires small-offset sense amplifiers (SAs) for robust reads, against a small TMR-ratio, at the expense of large area overhead and read energy (ERD). As Fig. 1 shows, designing STTMRAM macros for security-related applications imposes three main challenges. (1) Using a large number of SAs for wide parallel-IO readout to achieve a short tAC, but this results in a high peak current (IPEAK) and a large area overhead. Using fewer SAs for sequential wide-IO readout reduces IPEAK and area overhead, but imposes long tAC and a low read bandwidth (BWR). (2) MRAM macros with a high IPEAK degrade the supply (VDD) integrity of the chip, often leading to failure in noisesensitive blocks on the same chip. (3) A conventional memory-logic-separated scheme imposes a long latency (2 cycles: wide-IO memory read + flip-flop (FF) shift/rotate) for NVM-based security logic operations. This paper presents a multibit current-mode SA (MB-CSA) for a high BWR with a short tAC and a low IPEAK. Also presented is a near-memory computing (NMC) unit with a 1-cycle access, to speed up computing for security applications. This work resulted in a 22nm 1Mb STT-MRAM macro with dual-mode operations: wide-IO memory and NMC. The proposed 1Mb macro demonstrates the largest number of data-out operations (1024b) with a tAC of 2.75ns using a 0.85V supply. In memory mode, this device outperformed all reported NVM macros in terms of BWR (42.67GB/s) and ERD (0.23pJ/b). This work also presents the first MRAM macro with NMC functionality, a 33.3% reduction in logic area, and only a 170ps latency, after NVM access, for 1b shift/rotate operations.

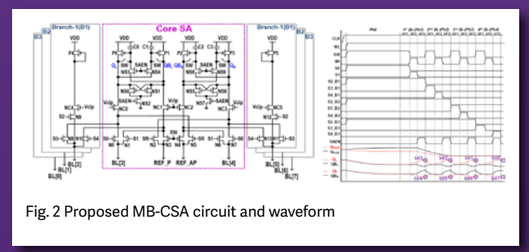


Fig. 2 Proposed MB-CSA circuit and waveform