D14-051

A Normally-Off Instant-On Stress-Relieved Nonvolatile TCAM for Filter-Based Search Engines Used in Energy-Efficient Big-Data Processing

應用於低功耗大數據處理之過濾式搜尋引擎之關鍵核心－可快速開機常關型之非揮發性三元內容循址記憶體

隊伍名稱

節能小小兵 / TCAM, We can!

隊長

黃莉悅 清華大學電機工程學系

隊員

楊耿豪 交通大學資訊科學與工程研究所
林建呈 清華大學電機工程學系
李岳陞 清華大學電機工程學系

作品摘要

隨著嵌入式系統與網路的快速發展，世界各處無時無刻產生大量資料，新興電子應用（如：智慧型手機、平板、無線感測網絡、辨識系統以及物聯網（Internet of Things）等）有著需處理大量資料、待機時間長之特性，並有低耗能的需求；我們認為當未來大數據（Big-Data）時代中，雲端運算（Cloud computing）進行多筆資料傳輸、複雜運算耗能之大，若增加一過濾器（Filter）如：三元內容循址記憶體（Ternary Content Addressable Memory，TCAM）於各種平台介面間，可先將接收到的資料進行快速比對過濾，就能大幅減少高耗能後傳處理的資料量，達到省能效用。

然而此過濾器需適應這些新興應用的特色－1.容量需求大 2.待機時間長；傳統靜態隨機存取器（Static Random Access Memory，SRAM）- based TCAM 需要兩組SRAM儲存三種資料、為16顆電晶體組成的記憶胞，使得系統面積有限的情況下，傳統SRAM-based TCAM容量小；此外隨製程微縮SRAM待機漏電耗能的頭疼問題亦發生於TCAM，一般解決方法為將資料搬動到另外一塊非揮發性記憶體（Nonvolatile memory，NVM）內，然而這個作法不但開關機需搬動資料耗能大，且受限於傳輸介面的輸入輸出端數（IO）導致開關機資料搬動時間隨容量而增加。

本計畫欲結合新興NVM進行改良，新興NVM中電阻式記憶體（ReRAM）為相當具有潛力的非揮發記憶體元件，其特色為高密度、高低阻態比值大、可快速隨機存取並且不會有長時間儲存後阻值漂移的問題，然而讀取ReRAM時，若為了較快讀取速度或較大感測邊界而施加較大讀取電壓，使ReRAM長時間兩端跨壓大，則可能造成阻值漂移，嚴重時會造成資料讀取錯誤，稱之為讀取干擾（Read disturb）；因此使用ReRAM做高速電路設計還須考慮Read disturb造成的資料可靠度問題。

本計畫提出一電阻電容過濾式降壓（RC-filtered stress-decoupled，RCSD）4T2R非揮發性三元內容循址記憶體（NV-TCAM），於記憶胞層級結合ReRAM，達到：

1. 4T2R記憶胞為傳統SRAM-based TCAM（16T）面積的1/4倍，即同樣面積條件下可達到4倍的容量。
2. 解決待機漏電耗能問題，並可直接關機、開機只需等待周邊邏輯電路恢復功能，達到快速、低耗能開關機特色。
3. 創新的電阻電容延遲（RC-delay）搜尋方式，可使搜尋時ReRAM的「壓力」（流經ReRAM的電流對時間積分）比起目前國際論文已發表之架構小約7倍。
4. 實體晶片128WL*32bits （4kb） RCSD-4T2R NV-TCAM的搜尋時間為目前國際最短量測數據。

2014

指導教授

張孟凡 / 清華大學電機工程學系

成功大學電機學士，美國賓州州立大學電機碩士，交通大學電子博士，2006 年進入清華大學電機工程學系任教至今。具備 10 年以上產業界經驗，任教至今曾獲得國科會吳大猷先生紀念獎及中央研究院年輕學者研究著作獎。

研究領域

奈米與下世代記憶體電路設計，低功耗低電壓電路、3D-IC 電路、憶阻器電路。

指導教授

陳添福 / 交通大學資訊科學與工程研究所

臺灣大學資工學士，美國華盛頓大學資工碩士，美國華盛頓大學資工博士，現職於交通大學資訊科學與工程研究所教授。曾任 Wang Computer Ltd 任軟件工程師及 Microprocessor Research Lab, Intel, USA 客座教授。

研究領域

系統架構，SOC 設計，嵌入式系統軟件。

Abstract

With the rapid growth of the internet and embedded systems, large amounts of data are created everywhere in the world simultaneously. Emerging Electronic Applications (ex. Smart Phones, PADs, wireless sensing networks, recognition systems and Internet of things) have characteristics including large amounts of data waiting to be processed, long standby time, and need for low energy. We believe that in the era of big data, energy caused by the large amount of data transfers and complex computations Cloud computing generates could be greatly reduced by introducing a filter like Ternary Content Addressable Memory (TCAM) between different platforms. This kind of memory can compare and filter incoming data, thus reducing the amount of data sent to following stages and save power.

However, this kind of filter needs to fulfill the needs of these emerging applications - 1. Large Capacity 2.long standby time. Traditional Static Random Access Memory (SRAM)-based TCAM need two pairs of SRAM cells (16transistors) to save 3 states. Under the same area constraint, SRAM-based TCAM has a small capacity; In addition, the leakage problem caused by SRAM process scaling also appears in TCAM. Conventional solutions are to move the stored data to a block of Nonvolatile memory (NVM), but this method not only requires large energy for moving data, but is also a slow process limited by the IO amount of the interface.

This project plans to improve current solutions by combining emerging NVM with CMOS process. Resistive Random Access Memory (ReRAM) is a very promising non-volatile memory, with high density, distinctive states (high R-Ratio) , fast random access, and good data retention time. However, a large voltage may be placed across the ReRAM for fast access, which may cause the ReRAM resistance to drift, and could lead to read failure under serious conditions. This situation (Read Disturb) must be put into consideration when designing high-speed ReRAM circuits.

This project proposes a RC-filtered stress-decoupled, (RCSD) 4T2R nonvolatile Ternary Content Addressable Memory (NV-TCAM), which combines ReRAM with CMOS technology at the cell level to achieve

1. 4T2R Cell is 1/4 the area of SRAM-based TCAM(16T), which gives 4 times the capacity in the same area
2. Solves standby-leakage problem, and gives instant power-on/off that only needs to wait for peripheral logic recovery, achieving fast and low energy switching
3. New RC-delay searching method reduces the voltage stress across the ReRAM during search by a factor of 7 compared to presented papers.
4. Measured results of a 128WL*32bits (4kb) RCSD-4T2R NV-TCAM Chip has achieved the fastest search time in internationally.

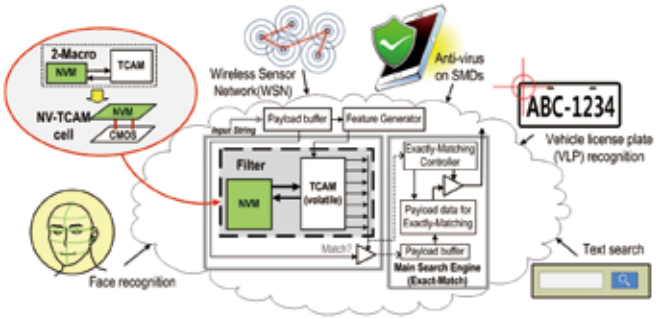

Fig.1 > Local Filter used in Big-Data processing

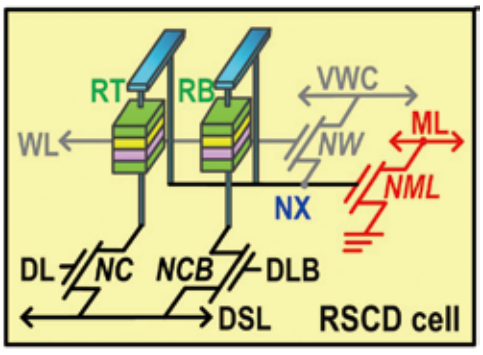

Fig.2 > Proposed cell structure