Future cache design

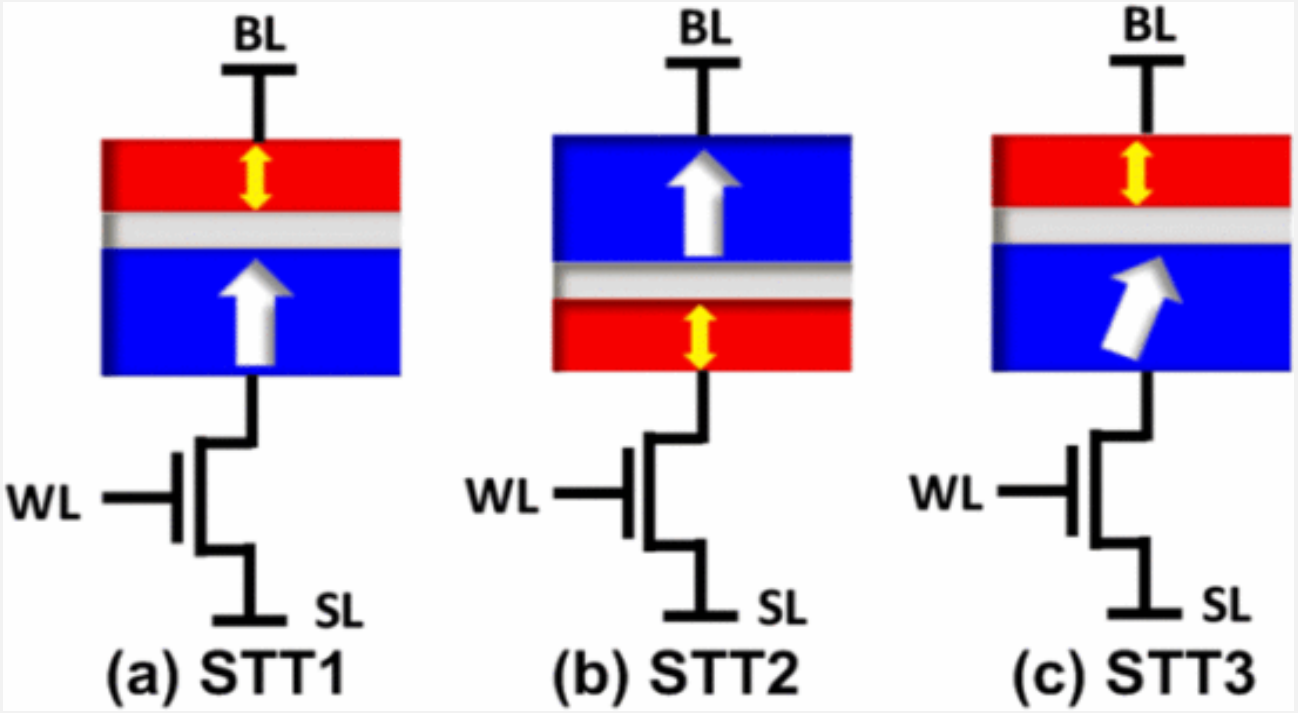
--STTMRAM

Introduction

- The ever-increasing gap between processor speed and main memory latency has driven the demand for larger on-chip caches in processors.
- Traditionally, on-chip caches in modern processors are implemented using static random access memories (SRAM).
- However, limited scalability, susceptibility to soft errors and high leakage power of SRAM pose challenges to high-density on-chip cache implementation. In order to address the limited scalability of SRAMs, several recent processors have adopted embedded dynamic RAM (EDRAM) in lower level caches.
- Among various candidates, spin-transfer torque magnetic RAM (STT MRAM) is considered as a promising technology that can offer desirable memory attributes such as high endurance, non-volatility, soft error immunity, zero standby power and high integration capability

- A conventional STT MRAM cell comprises of a magnetic tunnel junction (MTJ) and an access transistor in series
- The MTJ contains a pinned layer and a free layer separated by a dielectric layer (e.g. MgO).

- A read operation is performed by sensing resistance difference of the two binary states
- A write operation is performed by passing a current (I_W) through the bitcell that exceeds a critical current (I_C). The direction of (I_W) determines the final magnetization of the free layer (*i.e.*, parallel or anti-parallel states of the MTJ)


BL
WL
SL
(a) STT1
BL
WL
SL
(b) STT2
BL
WL
SL
(c) STT3

STT MRAM Cache vs. SRAM Cache

- A cache comprises of multiple arrays for storing tags and data bits. In conventional on-chip caches, both the tag and data arrays are implemented using SRAM. On the other hand, in the proposed STT MRAM caches, the tag arrays are implemented using SRAM and data arrays are implemented using STT MRAM.
- This is due to the fact that the write latency of STT MRAM may not be suitable for tag array operation, which requires frequent and fast updates of status bits and history bits
- From simulator It is shown that STT MRAM caches have a much higher integration density than SRAM cache.

Cache Utilization and Energy Consumption

- The contribution of active and leakage energy to total energy consumption is different for SRAM and STT MRAM-based caches.
- The leakage energy in an STT MRAM cache is smaller than an SRAM cache even with 4 times larger capacity (at iso-area).
- On the other hand, the dynamic energy for a write operation is higher in an STT MRAM cache compared to an SRAM cache.
- It is important to note that the total energy dissipation in a cache depends on factors such as cache access patterns (number of read and write operations) and cache utilization (number of times a processor accesses the cache per unit cycle).

THANK YOU