MRAM DTCO and Compact Models

Jeehwan Song*, Jian-Ping Wang, and Chris H. Kim

Dept. of ECE, University of Minnesota
*Now with TSMC, San Jose, CA

This work was supported in part by C-SPIN, one of the six SRC STARnet Centers, through MARCO and DARPA and in part by the NSF/SRC E2CDA Program.
Spin Transfer Torque MRAM Basics

- Magnetic Tunnel Junction (MTJ) is the storage element of STT-MRAM
- Thermal stability factor ($\Delta$) determines the retention time while anisotropy field ($H_K$) determines the energetic preference of the magnetization vector
- Key features: Low operating voltage, good CMOS compatibility, high speed, high density (<20F^2), zero static power, and high endurance.
- Applications: non-volatile memory (eflash replacement in 22nm), cache (SRAM replacement)
• Initial set of compact models and design rules for a preliminary standard cell library and entire optimization loop is repeated to obtain a satisfactory set of the device and design rule parameters

• MTJ compact model is a critical component of the overall MRAM DTCO flow.
Outline

- Introduction: DTCO and STT-MRAM
- MTJ SPICE Models
- STT-MRAM Array Level Evaluation
- State-of-the-art STT-MRAM Circuits
- Conclusions
MTJ SPICE Model

- Landau-Lifshitz-Gilbert (LLG) equation solved using SPICE
- The MTJ SPICE models are available at mtj.umn.edu.

Numerical form:
\[
\frac{1 + \alpha^2}{\gamma} \cdot \frac{d\vec{M}}{dt} = -\vec{M} \times \overline{H_{\text{eff}}} - \alpha \cdot \vec{M} \times (\overline{M} \times \overline{H_{\text{eff}}}) + A_{\text{st}} \cdot \vec{M} \times (\overline{M} \times \overline{M}_p), \quad A_{\text{st}} = \frac{hPJ}{2e\tau M_z}
\]

Circuit implementation (y-coordinate):

HSPICE script (y-coordinate):

H. Sun, Nature 2003 (IBM)

J. Kim, et. al., CICC, 2015.
MTJ SPICE Model Verification

- Model parameters can be tuned to match experimental data.
- Variability effects of both the MTJ ($W, L, t_F, RA$) and CMOS ($W, L, V_{th}, T_{ox}$) can be incorporated.

J. Kim, et. al., CICC, 2015.
Simulated Write and Read Delay Distributions

- Includes realistic variation for both MTJ (i.e. W, L, tF, RA) and CMOS (i.e. transistor W, L, Vth, tox)
- Top: Write delay distribution is narrower for a higher write voltage due to faster switching.
- Bottom: TMR ratio \(\frac{(R_{AP} - R_P)}{R_P}\) impact on read sensing margin.

J. Kim, et. al., CICC 2015
Incorporating Stochastic Switching Behavior

- Thermal fluctuation causes switching time to vary each write operation
- Can be incorporated into the model using the initial angle parameter (e.g. Monte Carlo) and time-varying thermal field

\[ \rho_{\text{initial}}(\theta) = \frac{\exp(-\Delta \sin^2 \theta)}{\int_0^{\pi/2} \sin \theta \exp(-\Delta \sin^2 \theta) d\theta} \]

\( \Delta = \) Thermal stability factor, \\
\( \theta = \) Initial angle between magnetization and Z-axis

I. Ahmed, et al., JxCDC 2017
Outline

- Introduction: DTCO and STT-MRAM
- MTJ SPICE Models
- STT-MRAM Array Level Evaluation
- State-of-the-art STT-MRAM Circuits
- Conclusions
**“Dennard” Style Scaling Study for STT-MRAM Caches**

- First order scaling analysis based on MTJ macromodels
- $\Delta$ for a target retention time is set by adjusting free layer thickness and MTJ anisotropy.
- STT-MRAM’s scalability based on a constant $J_{c0} \cdot RA/V_{DD}$ scaling scenario

### Thermal stability ($\Delta$)

$\Delta \propto H_k \cdot t \cdot w^2$

### Technology scaling

- Diameter: $w \rightarrow \alpha w$
- Thickness: $t \rightarrow t/\alpha^2$

### $\Delta$ and $J_{c0}$ adjustment with technology scaling

- $\Delta \rightarrow \Delta$
- $J_{c0} \rightarrow J_{c0}/\alpha^2$

### $J_{c0} \cdot RA$ scaling

- $RA \rightarrow \alpha^2 \cdot \beta RA$
- $J_{c0} \cdot RA/V_{DD}$: constant

---

K. Chun, et al., JSSC 2013

*based on MTJ macromodels, before MTJ SPICE models became available

$\alpha$: dimension scaling factor

$\beta$: voltage scaling factor

$J_{c0}$: Critical switching current density

RA: Resistance area product

---

Device/material parameters, design rules

MTJ, CMOS models

Monte Carlo, Evaluate PPA

MRAM circuit design (layout, read/write)
**“Dennard” Style Scaling Study for STT-MRAM Caches**

- **Left:** Sub-array architecture of STT-MRAM cache (~3 times smaller than 6T-SRAM cache)
- **Right:** Simulation setup includes process variation in the memory cells and SA circuit as well as variation of wire resistances, capacitances, reference biases, and supply levels.

### Power Supply Noise

- **-10%** to account for supply noise

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit-cell</td>
<td>Device mismatches</td>
</tr>
<tr>
<td>Parasitic capacitance (C_{BL})</td>
<td>(\sigma/\mu=5%): each (\mu) are calculated based on sub-array size</td>
</tr>
<tr>
<td>Resistance area product</td>
<td>(\sigma/\mu=5%)</td>
</tr>
<tr>
<td>Sense Amplifier (S/A)</td>
<td>I-applying and V-sensing method (AP direction read) + Voltage S/A : (I_{REF}) (\sigma/\mu=2.5%), S/A pair mismatches</td>
</tr>
<tr>
<td>Reference cell</td>
<td>Reference cell averaging scheme with MTJ replica cells</td>
</tr>
<tr>
<td>Write threshold current</td>
<td>(\sigma/\mu=5%)</td>
</tr>
</tbody>
</table>

* Mismatches are based on inverse square root relationship of devices' areas.
* Based on historic data, we assume \(\sigma_V/F\) is constant with tech. scaling
* \(\mu(C_{BL})\) is assumed to be scaled proportional to scaling factor.

K. Chun, et al., JSSC 2013
6σ BL Sensing and Write Delay Trends

- Sensing delay decreases with scaling, and with a higher TMR ratio.
- Write delay becomes worse due to the lower drive current (in planar CMOS).
- The trends follow basic circuit intuition (i.e. read and write always have a conflict).

Outline

- Introduction: DTCO and STT-MRAM
- MTJ SPICE Models
- STT-MRAM Array Level Evaluation
- State-of-the-art STT-MRAM Circuits
- Conclusions
Write Verify Scheme

- MRAM cell is repetitively written until correct value is verified → significant write error rate reduction at the expense of high write energy and long write time
- More suitable for NVM applications than cache memory
- Write driver has programmable write strength

H. Noguchi, et al., ISSCC 2016 (Toshiba 65nm)  
L. Wei, et al., ISSCC 2019 (Intel 22nm)
Array with Shared SL Voltage

- Common source line for compared area and reduced parasitic resistance
  - Improves read and write margins with a modest layout area overhead.
  - Inhibit voltage is applied on the unselected BL
  - Negative voltage applied to unselected WLs to suppress the BL leakage current

Y.-D. Chih, et al., ISSCC 2020 (TSMC 22nm), C. Kim, et al., ISSCC 2015 (Samsung)
Readout Circuit

- Clamp transistor trimming circuit
  - Clamps BL voltage to prevent read disturbance
  - Trimming circuit to remove offset

- 1T4MTJ reference cell provides a stable \((R_P+R_{AP})/2\) reference value without causing read disturbance

- Half-\(V_{DD}\) detection circuit improves the sensing margin by extending the signal development time

Y.-D. Chih, et al., ISSCC 2020 (TSMC 22nm)
Conclusions

• This invited paper covers various aspects of MRAM DTCO including device, circuit, and architecture considerations.
  – SPICE MTJ device model
  – Array level STT-MRAM PPA evaluation
  – Scalability and variability studies
  – Novel MRAM read/write circuit, array, and bit-cell layout design techniques can improve the yield of the large arrays and must be accounted for in DTCO flow

This work was supported in part by C-SPIN, one of the six SRC STARnet Centers, through MARCO and DARPA and in part by the NSF/SRC E2CDA Program.