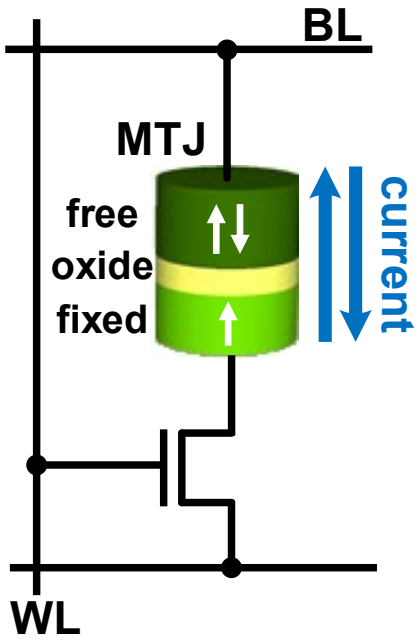
MRAM DTCO and Compact Models

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

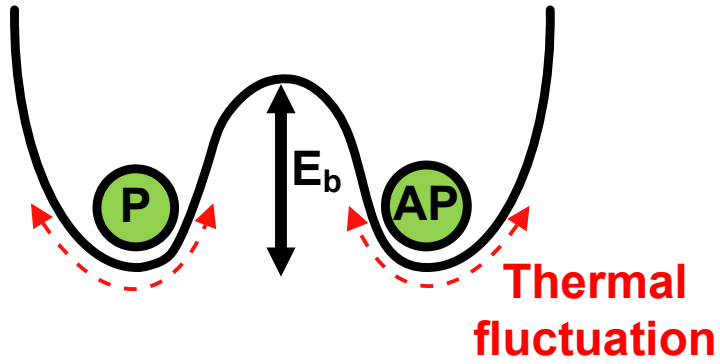
Dept. of ECE, University of Minnesota
**Now with TSMC, San Jose, CA*

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

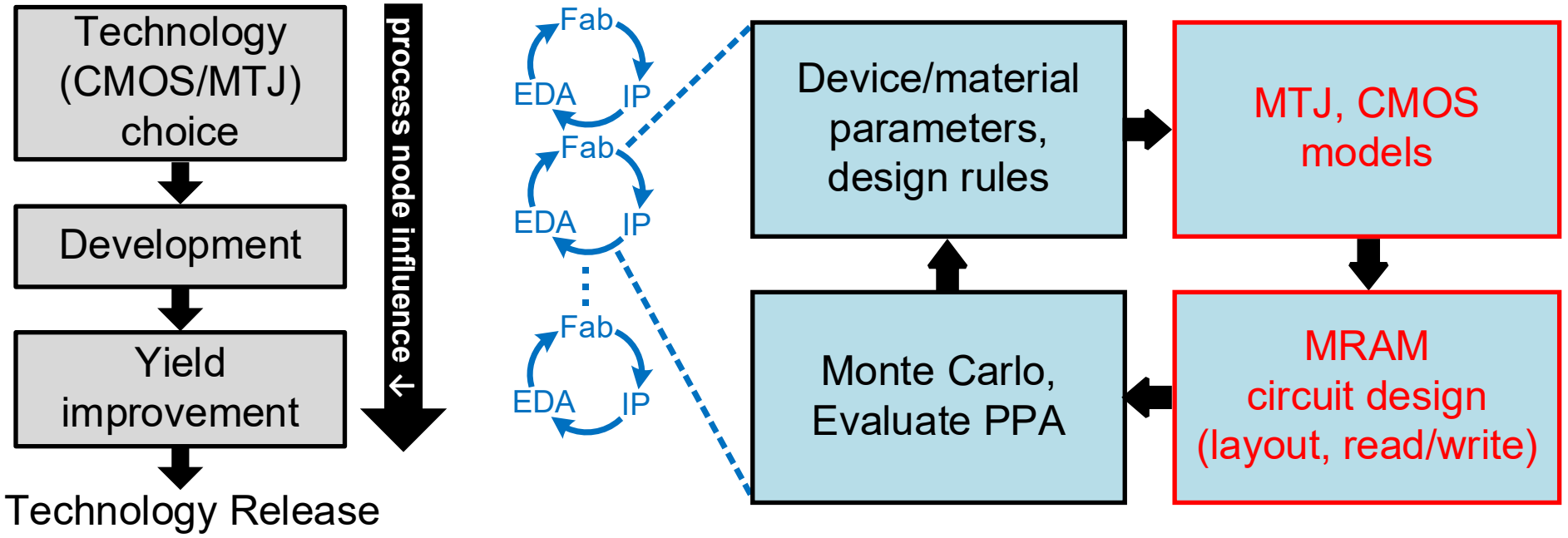


$$\Delta = \frac{E_b}{k_B T} = \frac{H_k M_s V}{2 k_B T}$$



- Magnetic Tunnel Junction (MTJ) is the storage element of STT-MRAM
- Thermal stability factor (Δ) 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)

DTCO for STT-MRAM

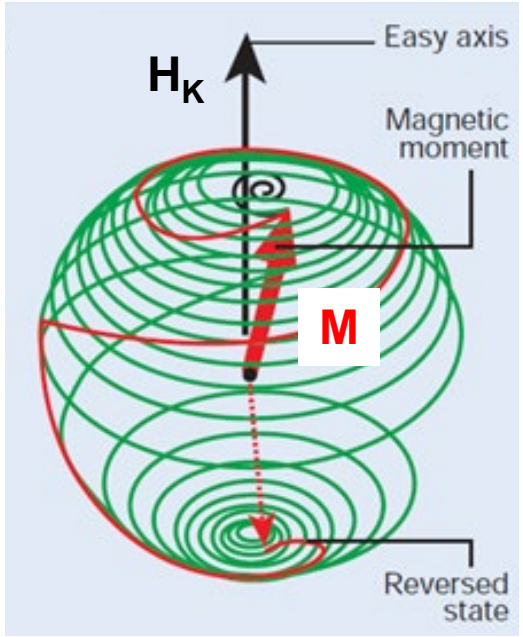


- Initial set of compact models and design rules for a preliminary standard cell library and entire optimization loop is repeated to obtain a satisfactory set of the device and design rule parameters
- MTJ compact model is a critical component of the overall MRAM DTCO flow.

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



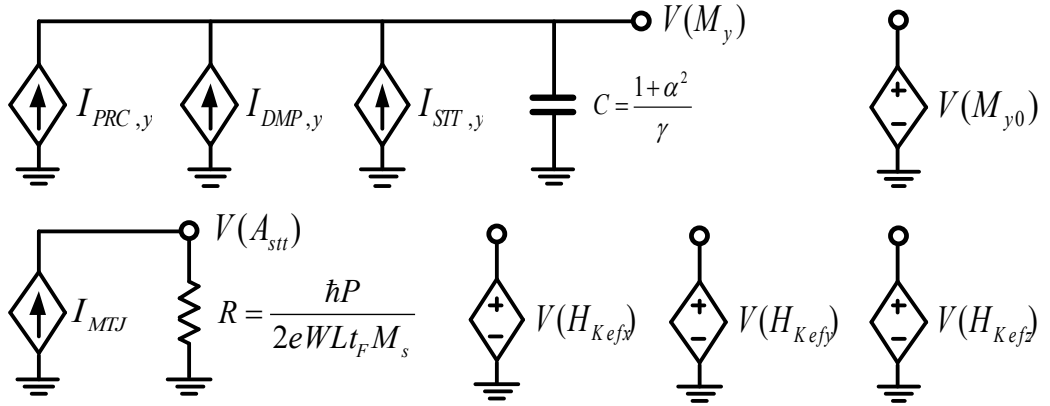
J. Sun, Nature 2003 (IBM)

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

Numerical form:

$$\frac{1+\alpha^2}{\gamma}\cdot\frac{d\overline{M}}{dt} = \underbrace{-\overline{M}\times\overline{H}_{Keff}}_{\text{Precession}} \underbrace{-\alpha\cdot\overline{M}\times(\overline{M}\times\overline{H}_{Keff})}_{\text{Damping}} + \underbrace{A_{stt}\cdot\overline{M}\times(\overline{M}\times\overline{M}_p)}_{\text{Spin torque}},\ A_{stt}=\frac{\hbar PJ}{2et_F M_s}$$

Circuit implementation (y-coordinate):

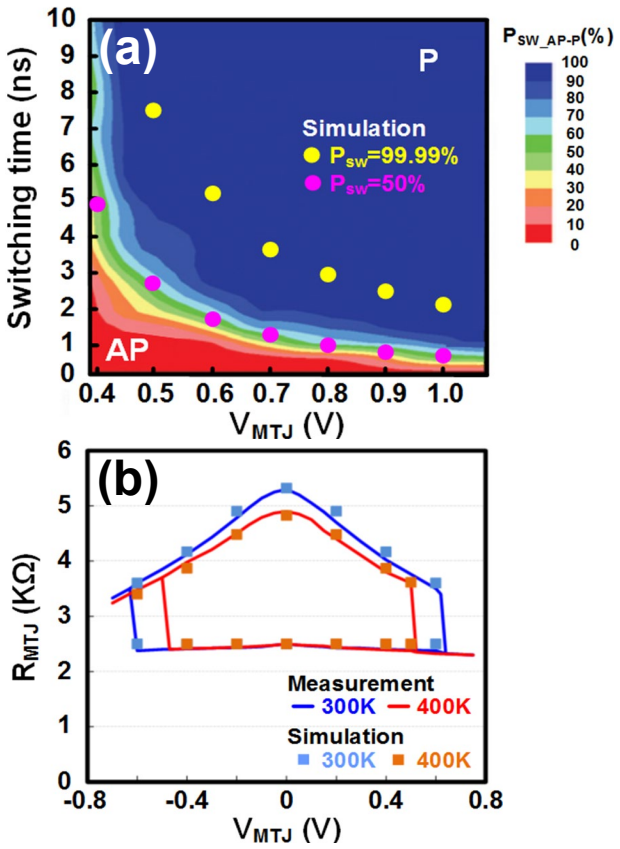


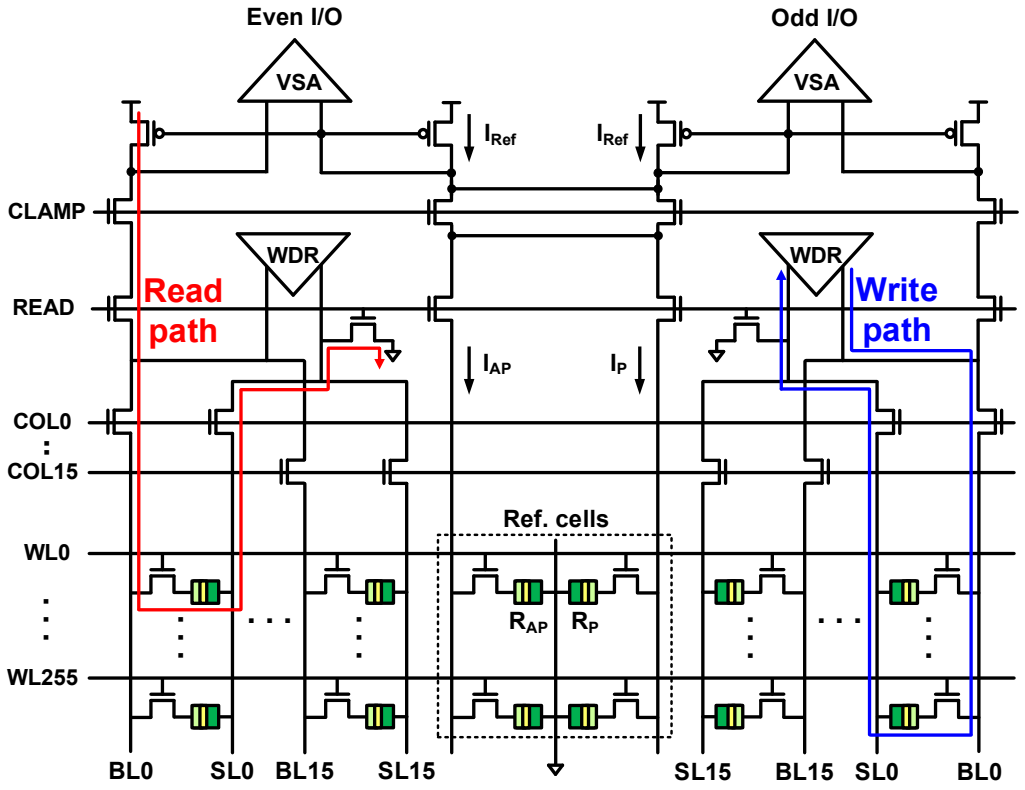
HSPICE script (y-coordinate):

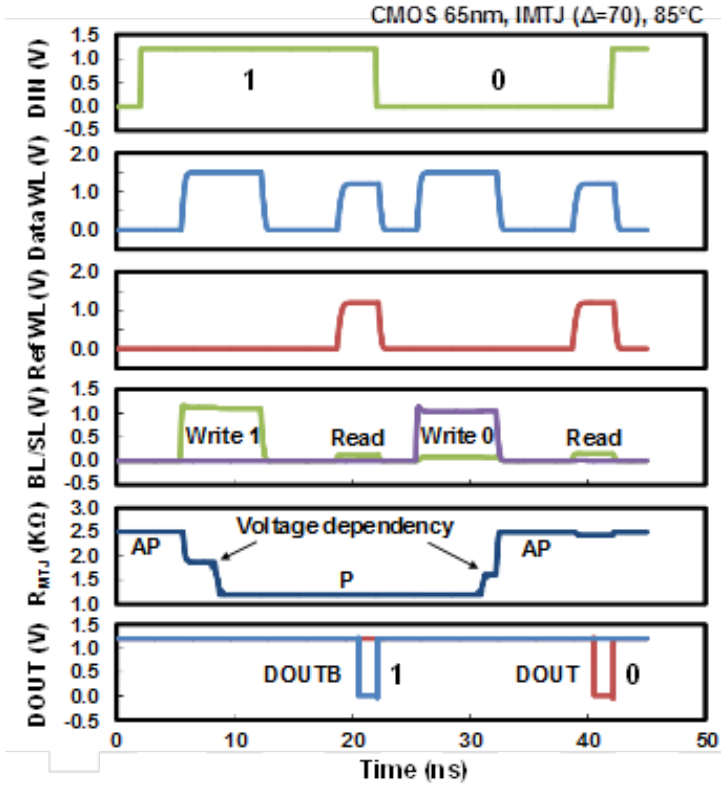
```
C_My      My  0   '(1+α²)/γ'
G_dMy_prc 0   My  cur='-(v(Mz)·v(HKefx)-v(HKefz)·v(Mx))'
G_dMy_dmp 0   My  cur='-α·(v(Mz)·(v(My0)·v(HKefz)-v(HKefy)·v(Mz))-(v(Mx)·v(HKefy)-v(HKefx)·v(My0))·v(Mx))'
G_dMy_stt 0   My  cur='v(Astt)·(v(Mz)·(v(My0)·Mpz-Mpy·v(Mz))-(v(Mx)·Mpy-Mpx·v(My0))·v(Mx))'
E_My0     My0 0   vol='v(My)'  max='cos(v(θc))' min='-cos(v(θc))' → thresholding
```

J. Kim, et. al., CICC, 2015.

MTJ SPICE Model Verification



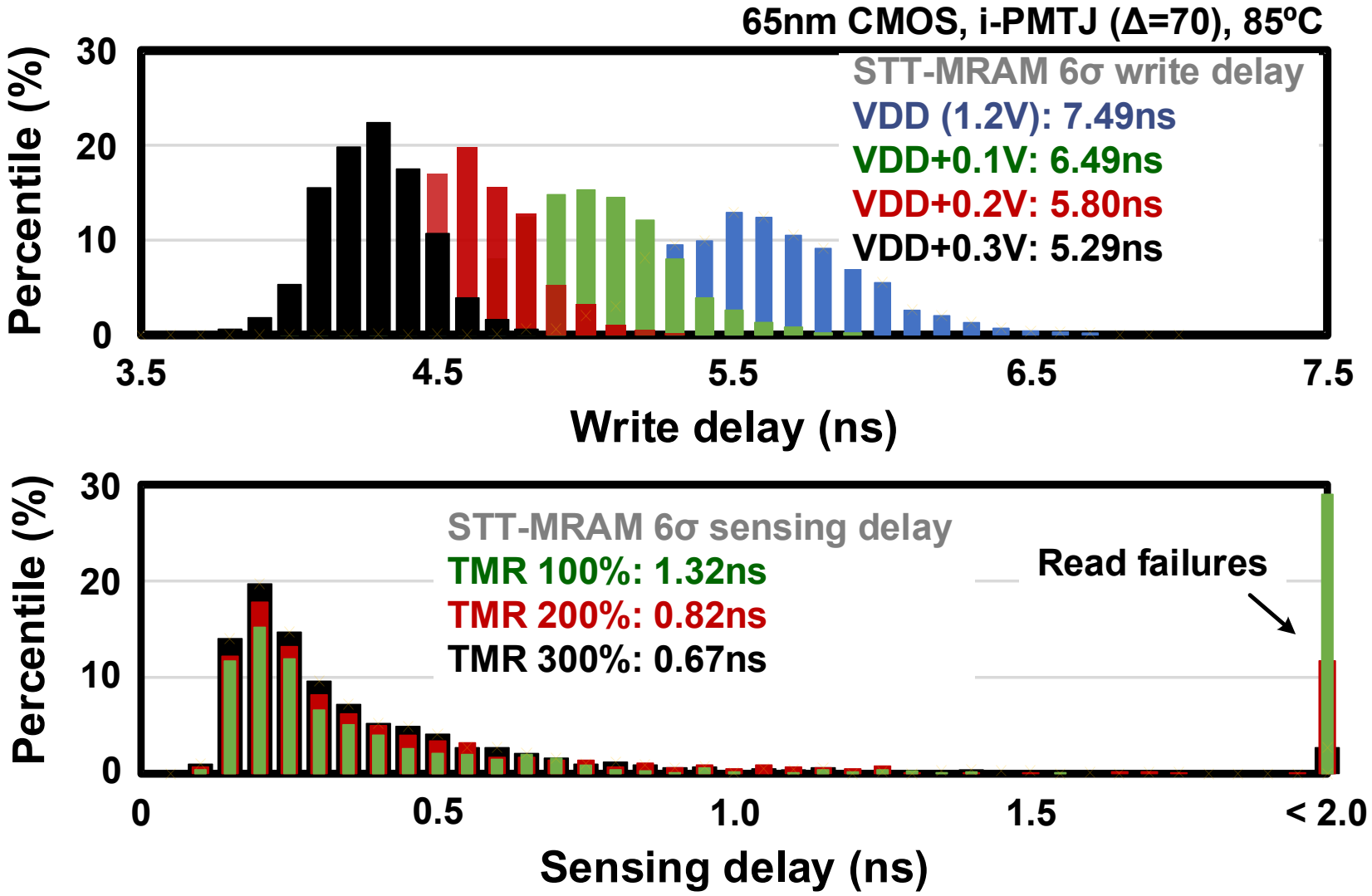




J. Kim, et. al., CICC, 2015.

- 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

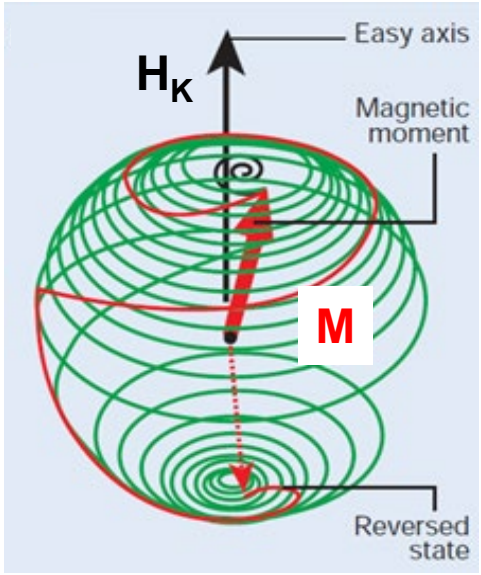
Simulated Write and Read Delay Distributions



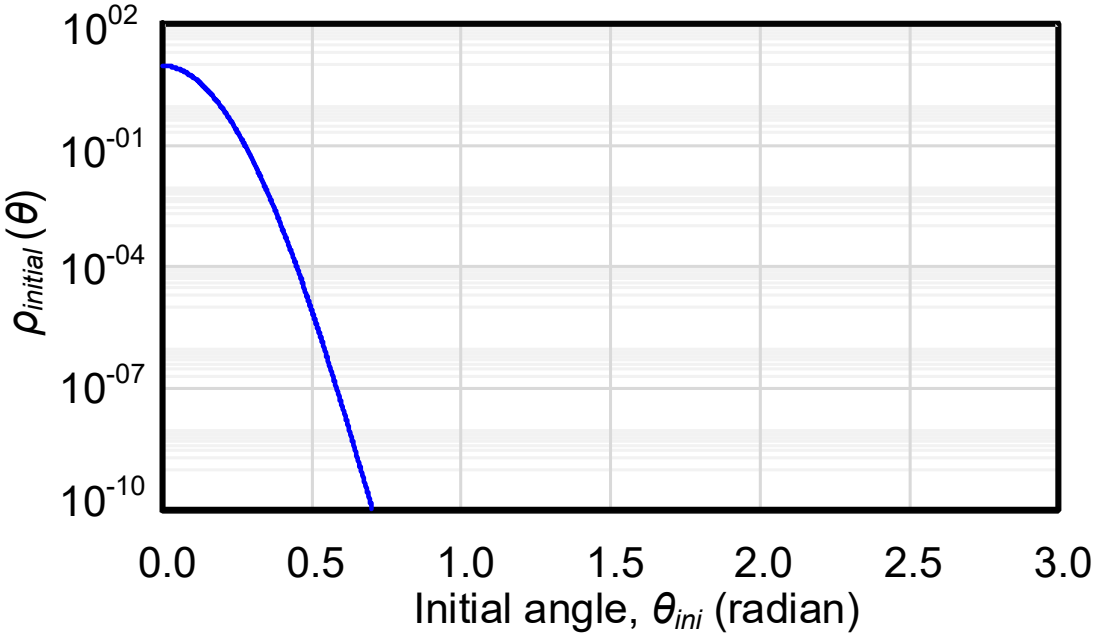
- Includes realistic variation for both MTJ (i.e. W, L, t_F, RA) and CMOS (i.e. transistor W, L, Vth, tox)
- Top: Write delay distribution is narrower for a higher write voltage due to faster switching.
- Bottom: TMR ratio ($=(R_{AP}-R_P)/R_P$) impact on read sensing margin.

J. Kim, et. al., CICC 2015

Incorporating Stochastic Switching Behavior



J. Sun, Nature 2003 (IBM)



- 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

$$\rho_{initial}(\theta) = \frac{exp(-\Delta\, sin^2\,\theta)}{\int_0^{\frac{\pi}{2}} sin\,\theta\, exp(-\Delta\, sin^2\,\theta) d\theta}$$

Δ = Thermal stability factor,

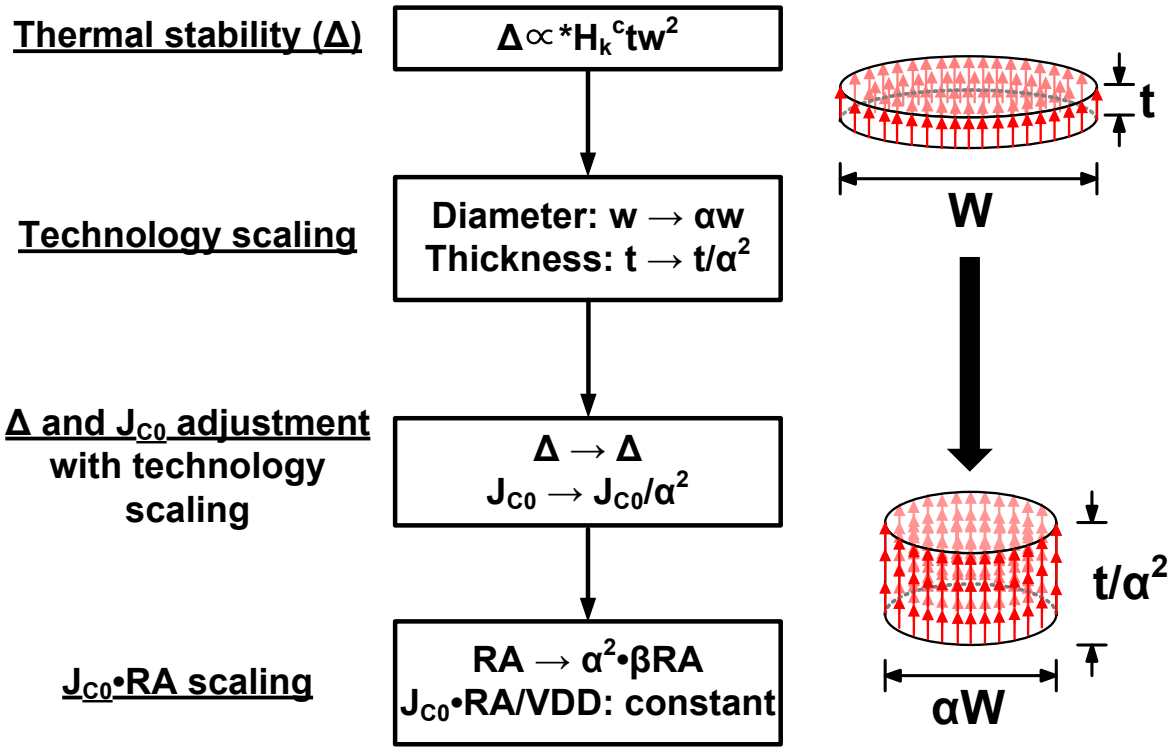
θ = Initial angle between magnetization and Z-axis

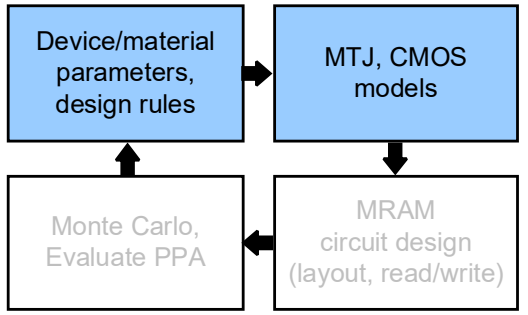
I. Ahmed, et al., JxCDC 2017

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



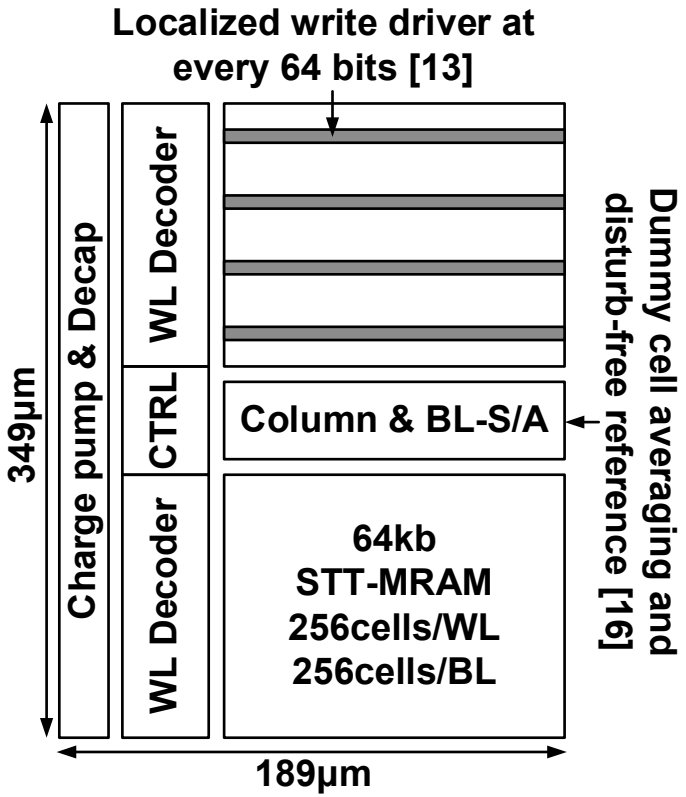


K. Chun, et al., JSSC 2013
*based on MTJ macromodels, before MTJ SPICE models became available

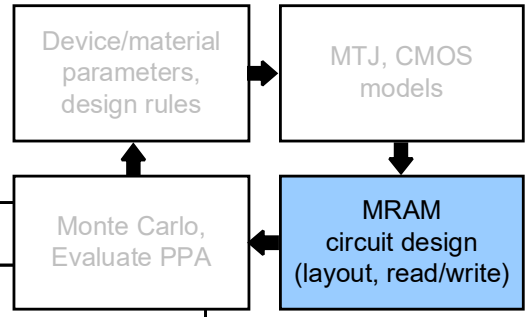
α: dimension scaling factor
β: voltage scaling factor
J_{c0}: Critical switching current density
RA: Resistance area product

- First order scaling analysis based on MTJ macromodels
- Δ 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}•RA/V_{DD} scaling scenario

“Dennard” Style Scaling Study for STT-MRAM Caches



K. Chun, et al., JSSC 2013

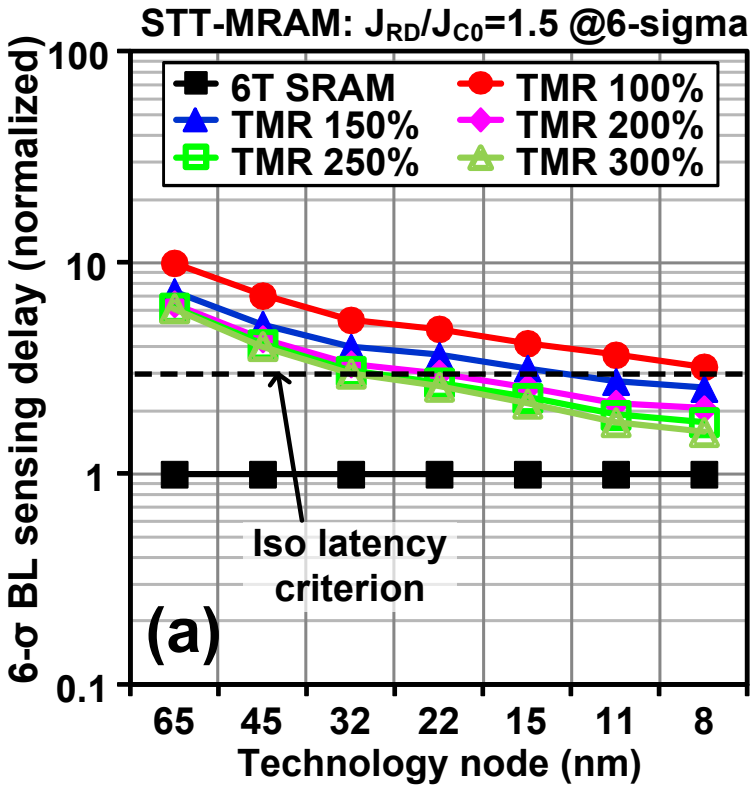


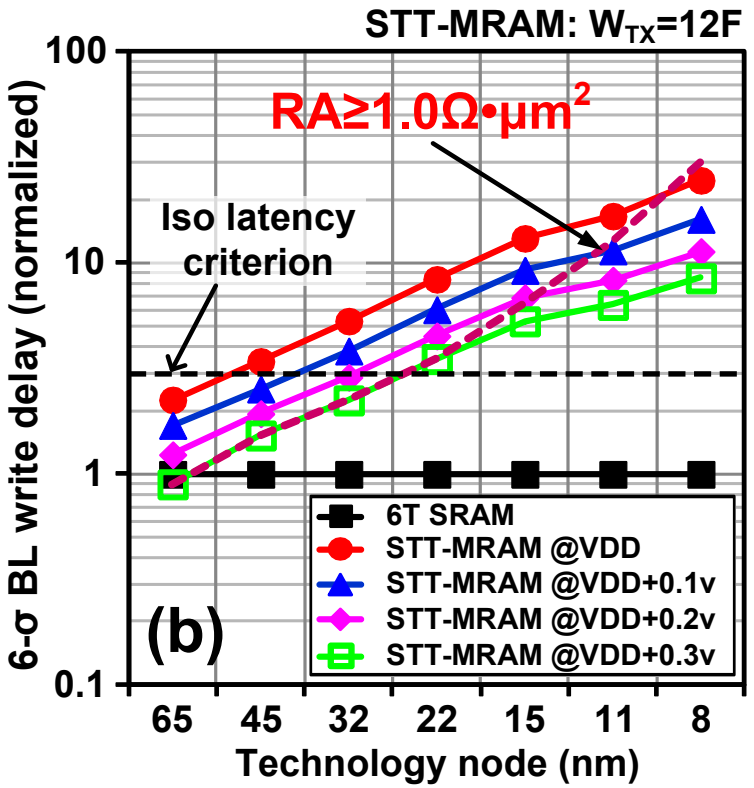
Power supply noise	-10% to account for supply noise
Bit-cell	Device mismatches
Parasitic capacitance (C_{BL})	σ/µ=5%: each µ are calculated based on sub-array size
Resistance area product	σ/µ=5%
Sense Amplifier (S/A)	I-applying and V-sensing method (AP direction read) + Voltage S/A : I_{REF} σ/µ=2.5%, S/A pair mismatches
Reference cell	Reference cell averaging scheme with MTJ replica cells
Write threshold current	σ/µ=5%

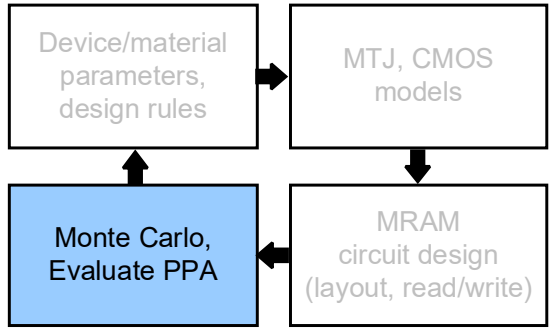
* Mismatches are based on inverse square root relationship of devices' areas.
* Based on historic data, we assume σ_{Vt}/F is constant with tech. scaling
* µ(C_{BL}) is assumed to be scaled proportional to scaling factor.

- 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

6σ BL Sensing and Write Delay Trends







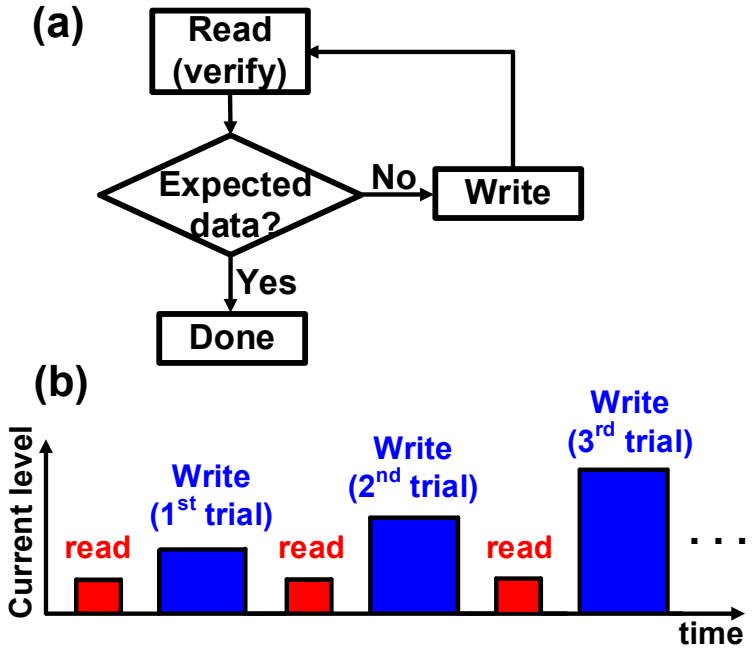
- Sensing delay decreases with scaling, and with a higher TMR ratio
- Write delay becomes worse due to the lower drive current (in planar CMOS).
- The trends follow basic circuit intuition (i.e. read and write always have a conflict)

K. Chun, et al., JSSC, 2013.

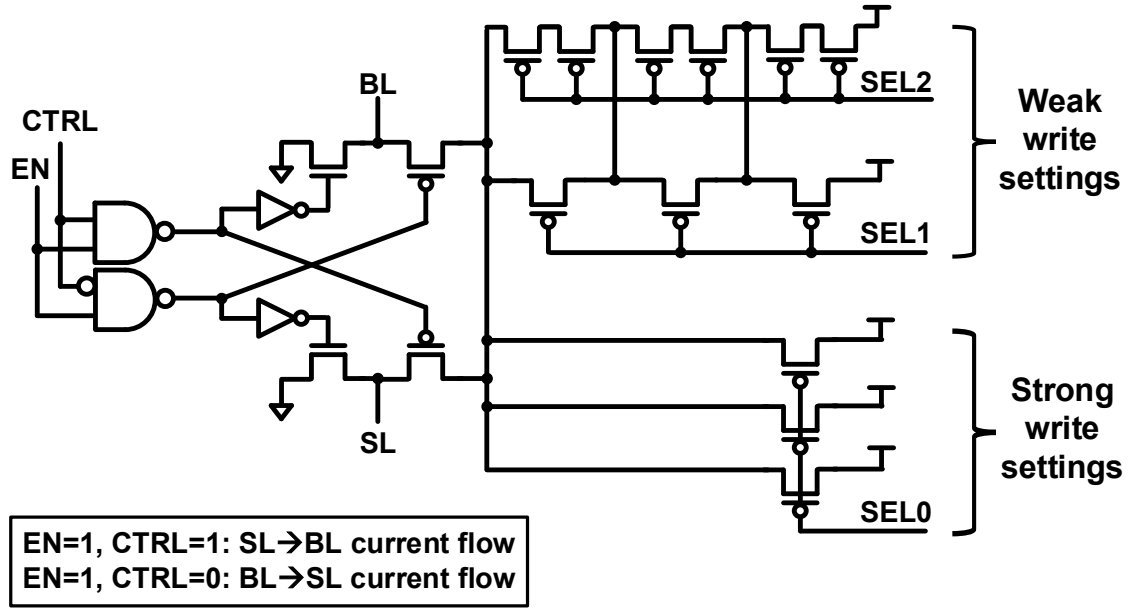
Outline

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

Write Verify Scheme



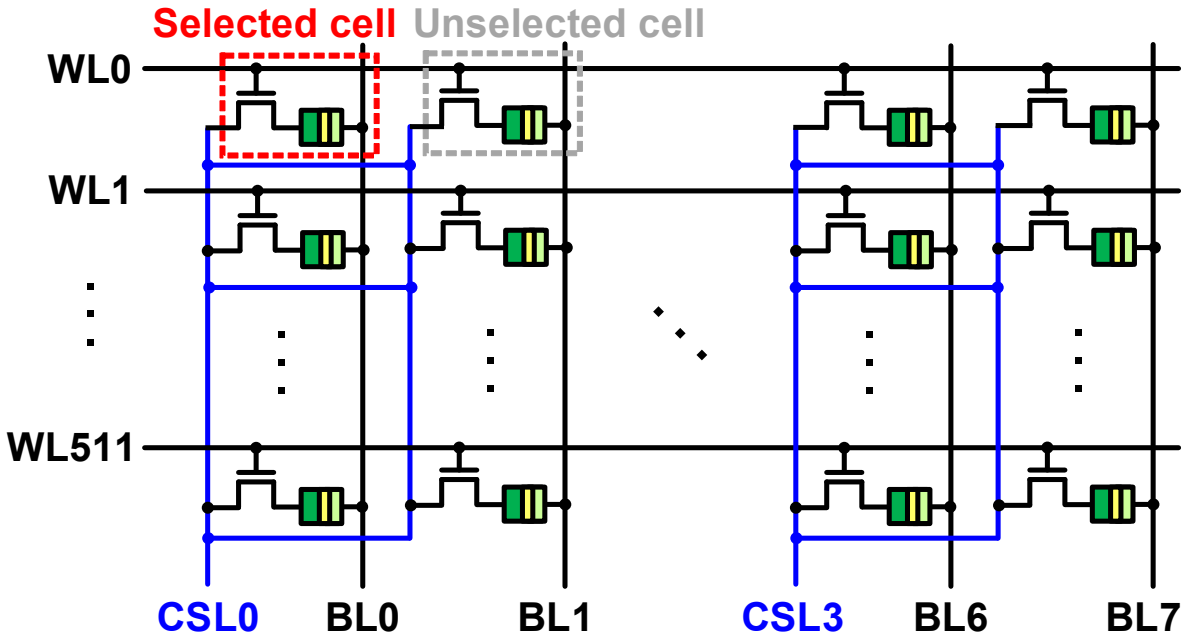
H. Noguchi, et al., ISSCC 2016 (Toshiba 65nm)



L. Wei, et al., ISSCC 2019 (Intel 22nm)

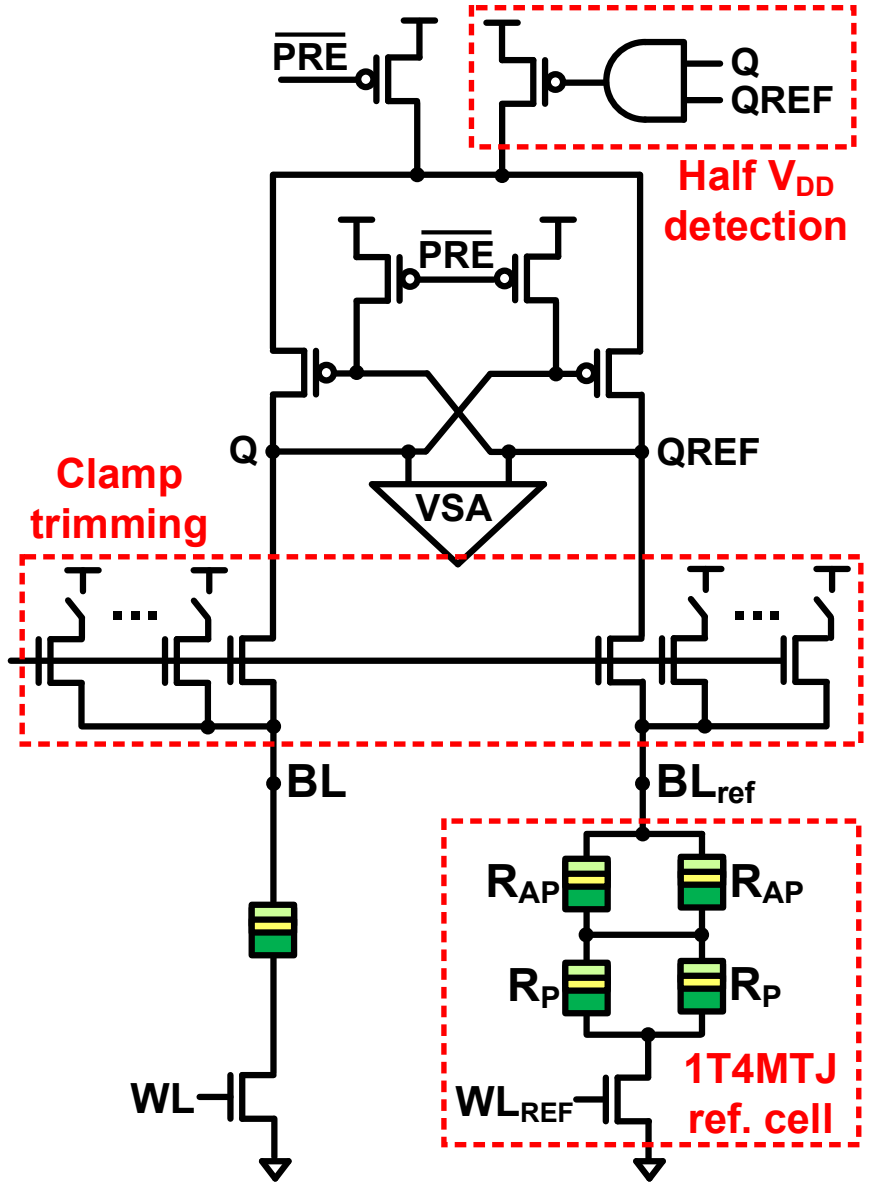
- MRAM cell is repetitively written until correct value is verified → significant write error rate reduction at the expense of high write energy and long write time
- More suitable for NVM applications than cache memory
- Write driver has programmable write strength

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)



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

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

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.