Development of STT-MRAM for embedded memory applications

P. Wang, G. Jan, L. Thomas, Y. Lee, H. Liu, J. Zhu, S. Le, J. Iwata-Harms, S. Guisan, R. Tong, S. Patel, V. Sundar, D. Shen, R. He, J. Haq, J. Teng, V. Lam, Y. Wang, and T. Zhong

TDK-Headway Technologies, Inc., Milpitas, California

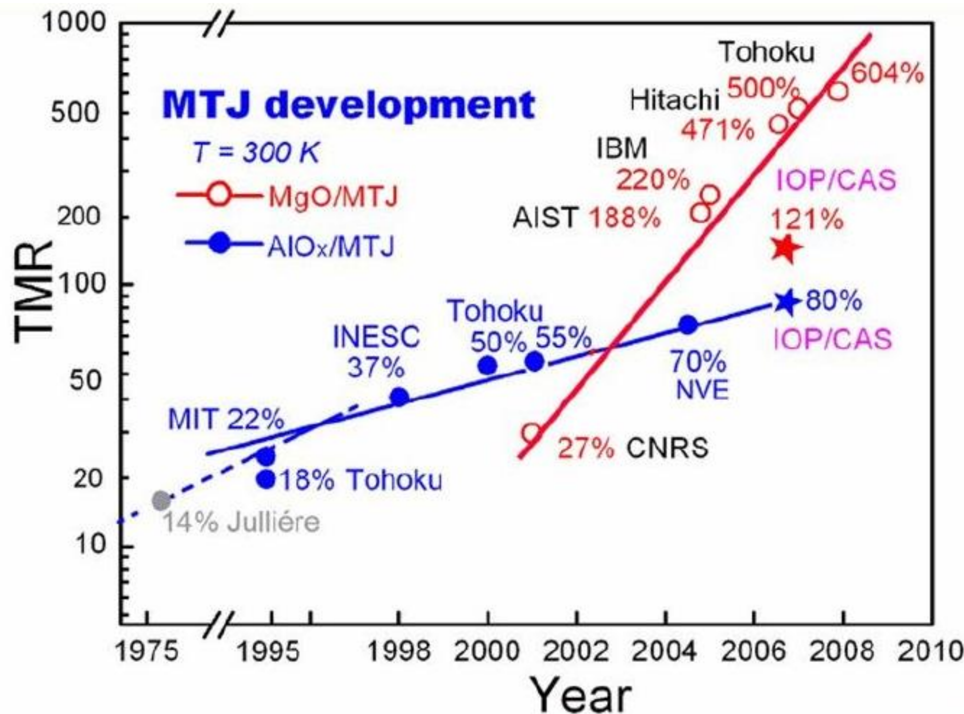
June 2017

Outline

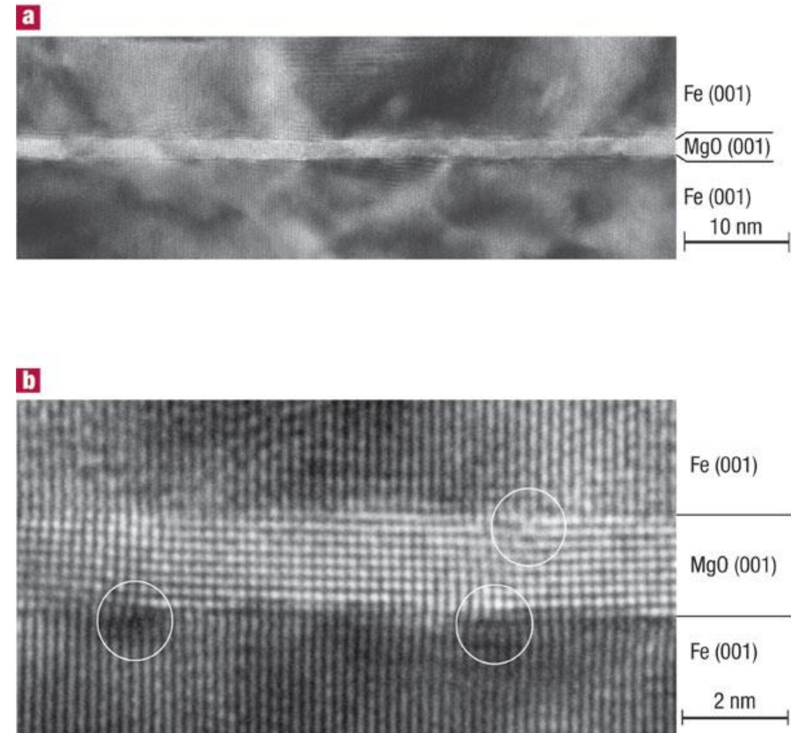
- **Basic principles of STT-MRAM**
- Embedded memory applications
- STT-MRAM integration and chip level results
- Tunnel barrier reliability at chip level

Magnetic tunnel Junction (MTJ) device

- Two ferromagnetic electrodes separated by a thin MgO tunnel barrier
- Tunnel Magnetoresistance (TMR): device resistance depends on the relative orientation of the magnetization of the two magnetic electrodes



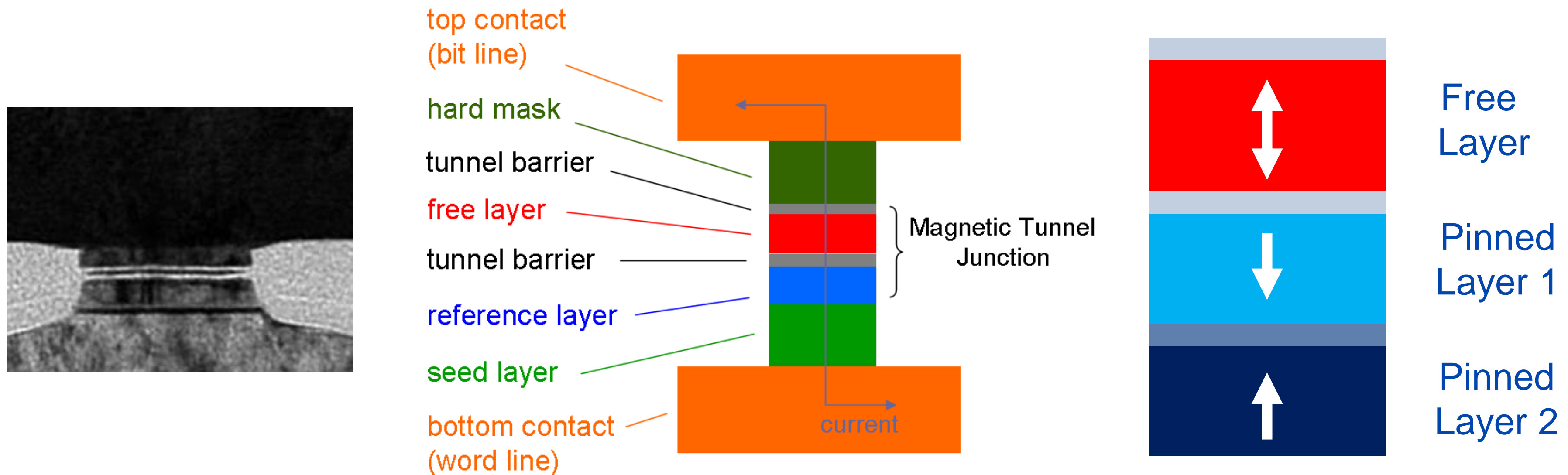
Reproduced from website of MultiDimension Technology Co.,Ltd.



Yuasa et al. (AIST) Nature Materials2004

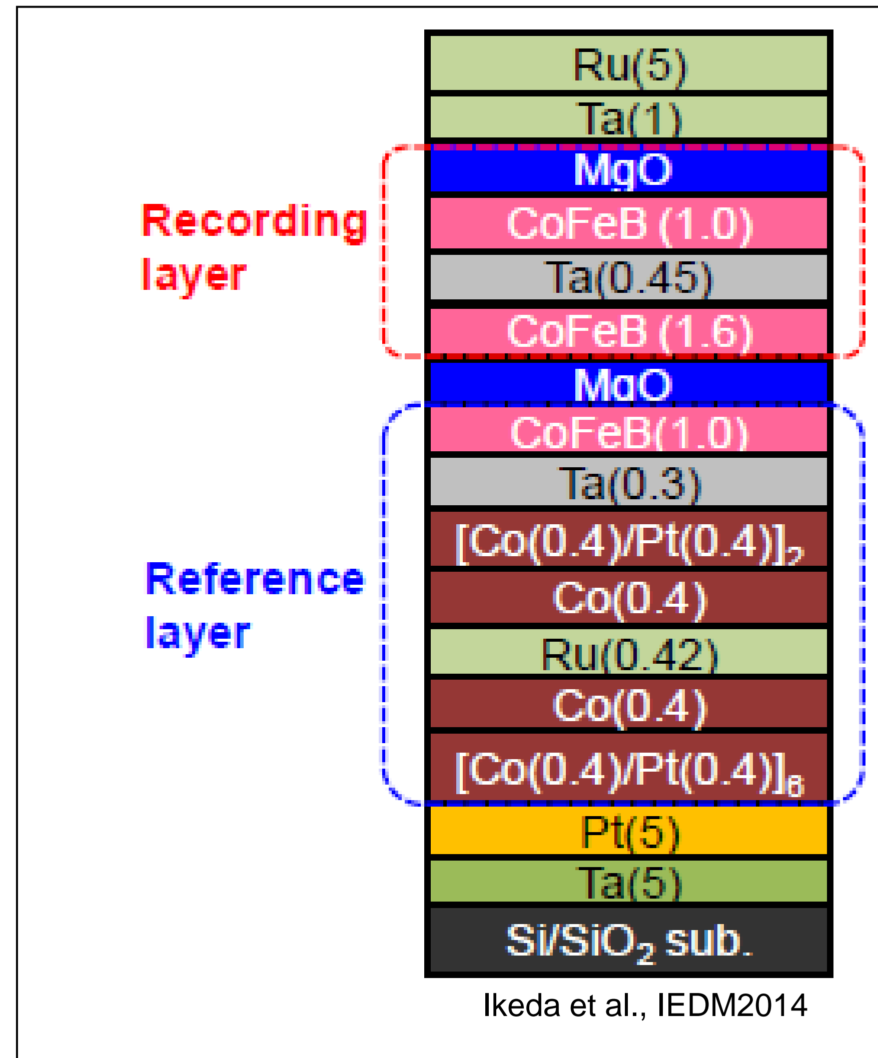
Perpendicular Magnetic Anisotropy (PMA) MTJ

- PMA is needed for data retention scaling and writing efficiency
- PMA is based on interfacial anisotropy between MgO and CoFeB
  (Ikeda et al., Nature Mat. 2011, Worledge et al., APL 2012)
- Free layer sandwiched between two MgO interfaces for enhanced anisotropy and data retention
- Dual reference layer for reducing dipolar fields and enhanced stability



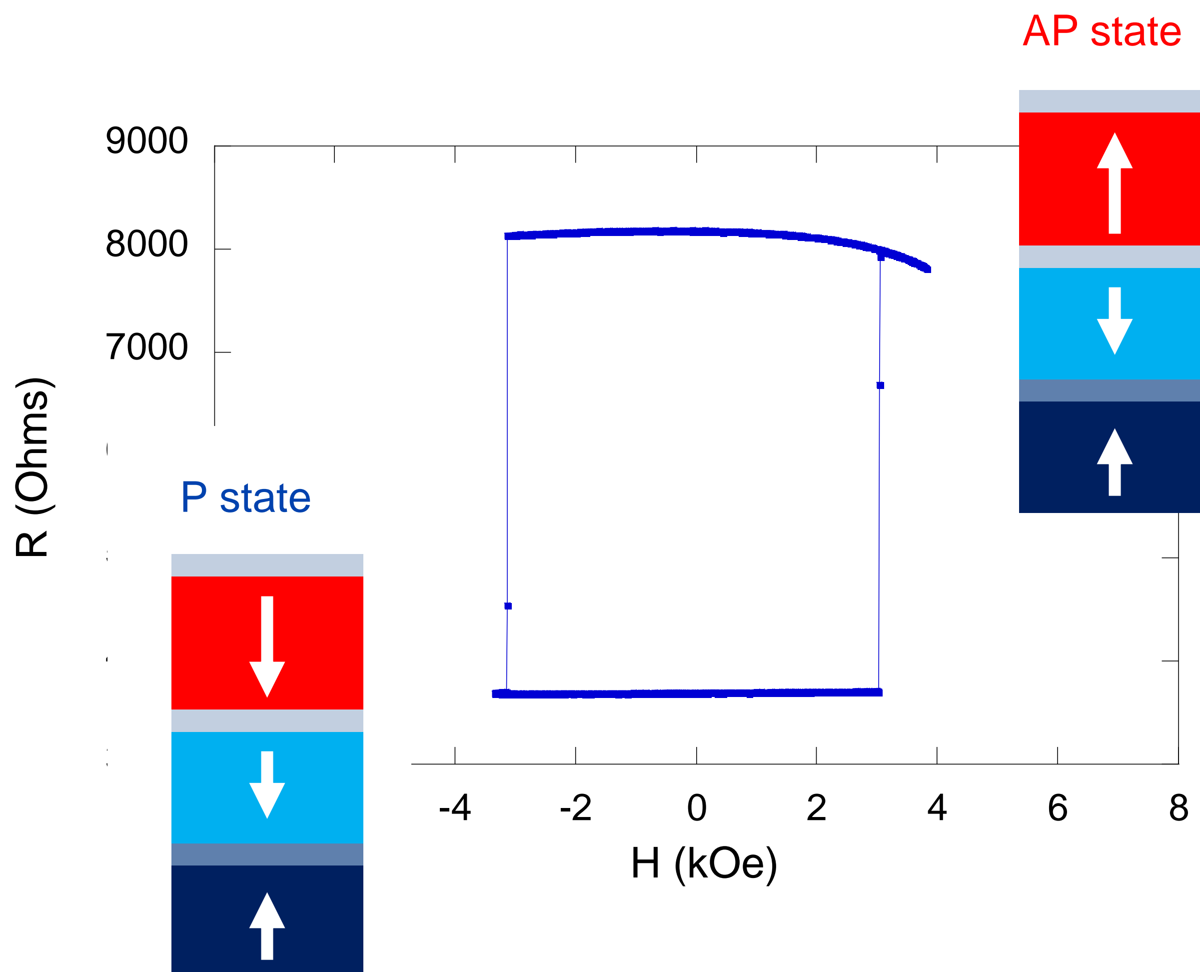
An example of perpendicular MTJ

- ~ 30 sub layers, with thickness ranging from 0.3 to 5 nm
- PMA is based on interfacial anisotropy between MgO and CoFeB
- Specialized PVD tools can achieve >20 wafers/hour throughput



Ikeda et al., IEDM2014

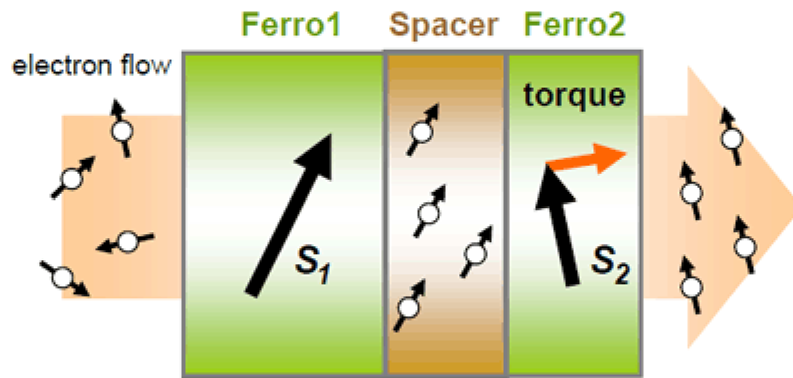
Resistance vs magnetic field hysteresis loop



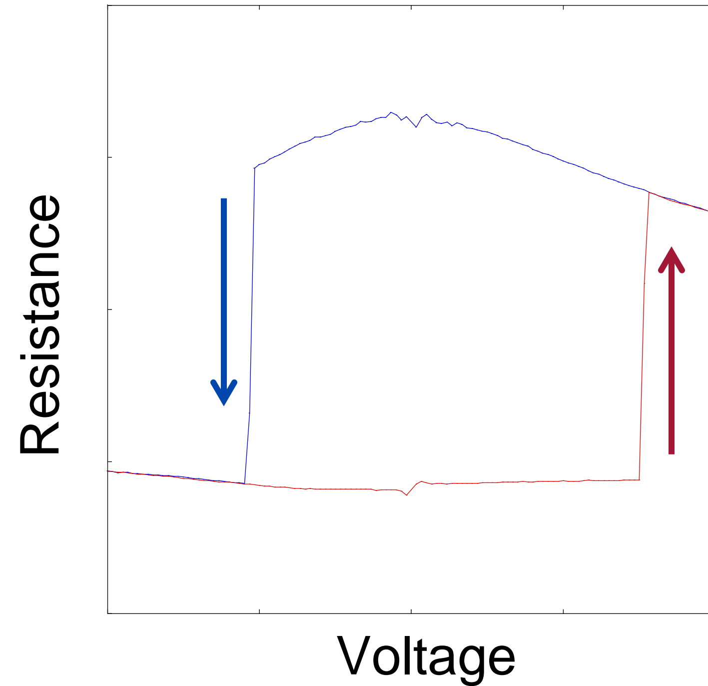
- Two well-defined resistance states depending on orientation of magnetic electrodes

Writing with Spin-Transfer Torque

- Transfer of spin-angular momentum from polarized conduction electrons to electrode magnetization

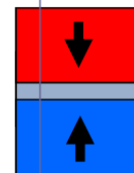


Reproduced from Quantumwise.com



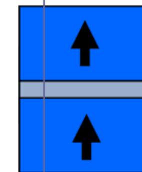
Write:
Spin Transfer Torque

Write [0]



electron flow

Write [1]



4

electron flow

Outline

- Basic principles of STT-MRAM
- **Embedded memory applications**
- STT-MRAM integration and chip level results
- Tunnel barrier reliability at chip level

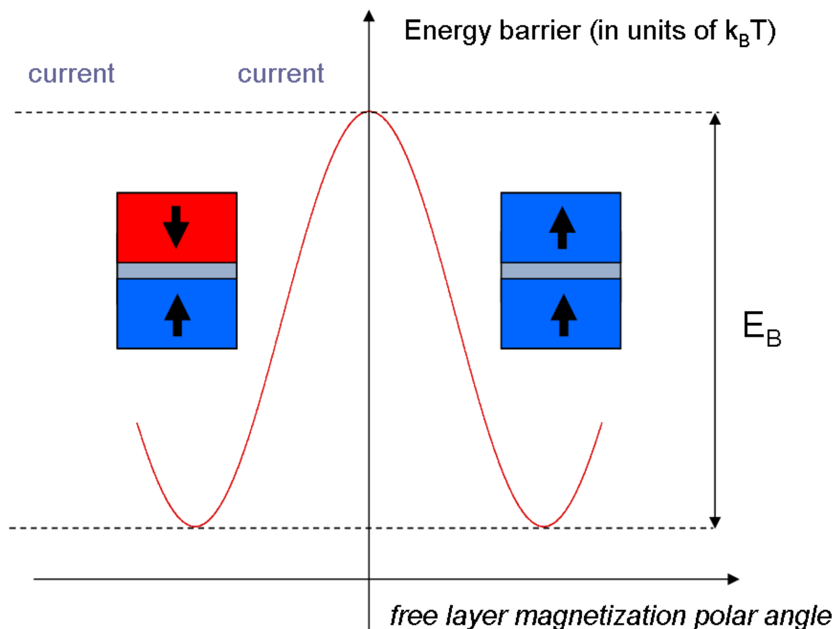
Trade-offs of STT writing

➔ Write current scale with energy barrier for data retention

Energy barrier: $E_B \sim K_u V$

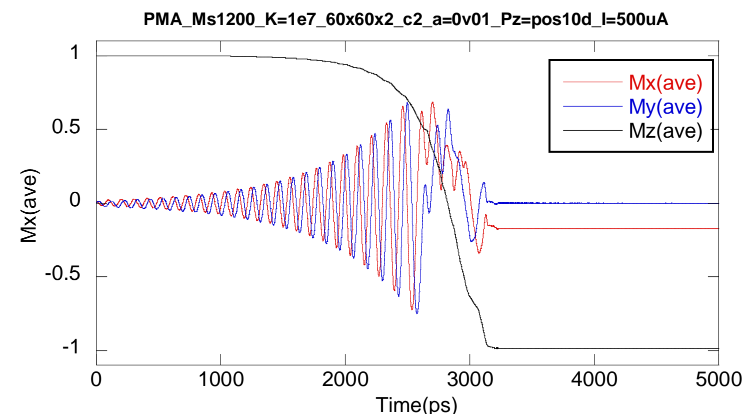
Write current: $I_{c0} = (4e/\hbar)\,(\alpha/P)\,E_B$

STT efficiency: E_B/I_{c0} ~ 1-2 in $k_BT/\mu A$



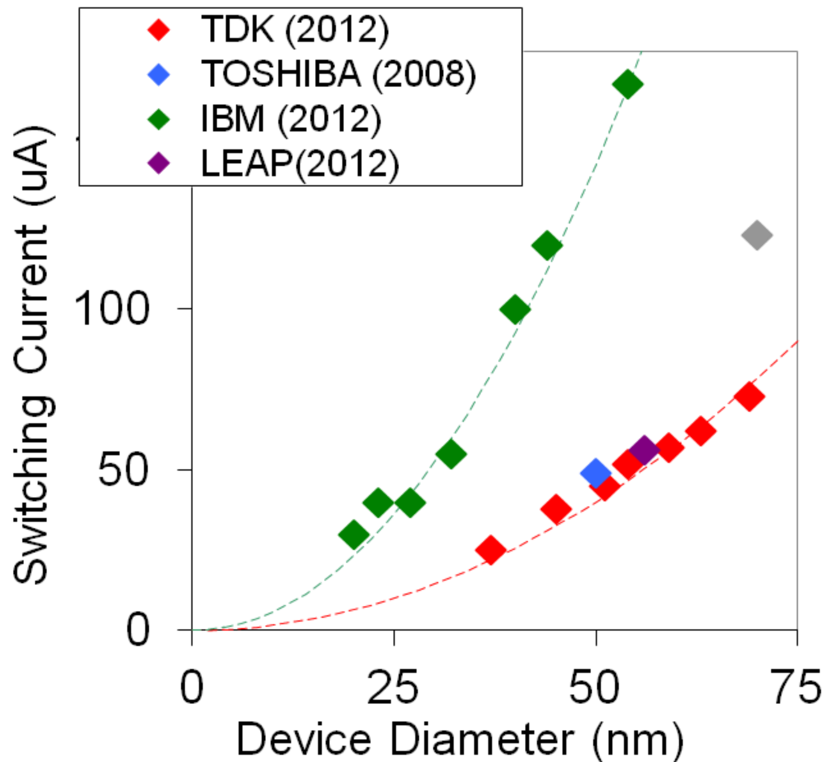
➔ Writing is probabilistic

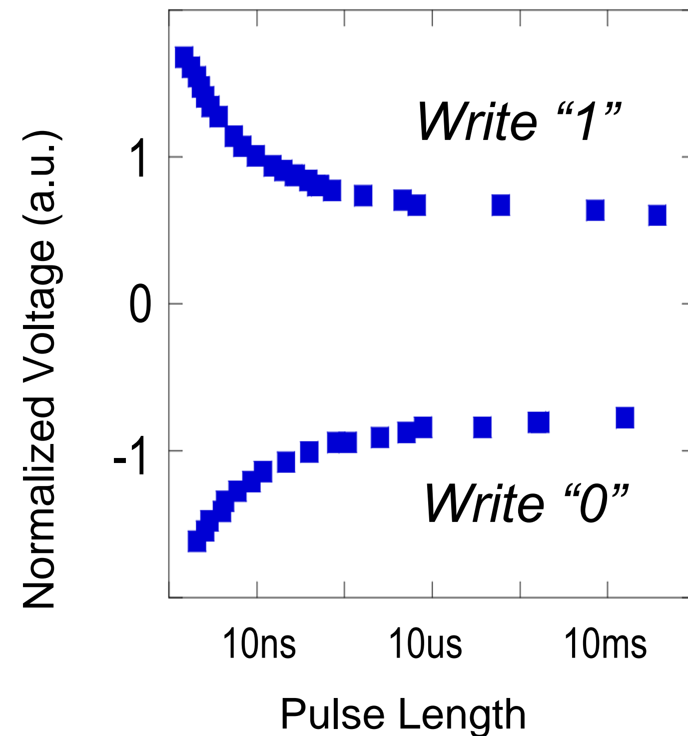
- STT vanishes for parallel alignment of PL and FL
- Switching time inversely proportional to angle between PL and FL
- Thermal fluctuations provide initial 'kick'



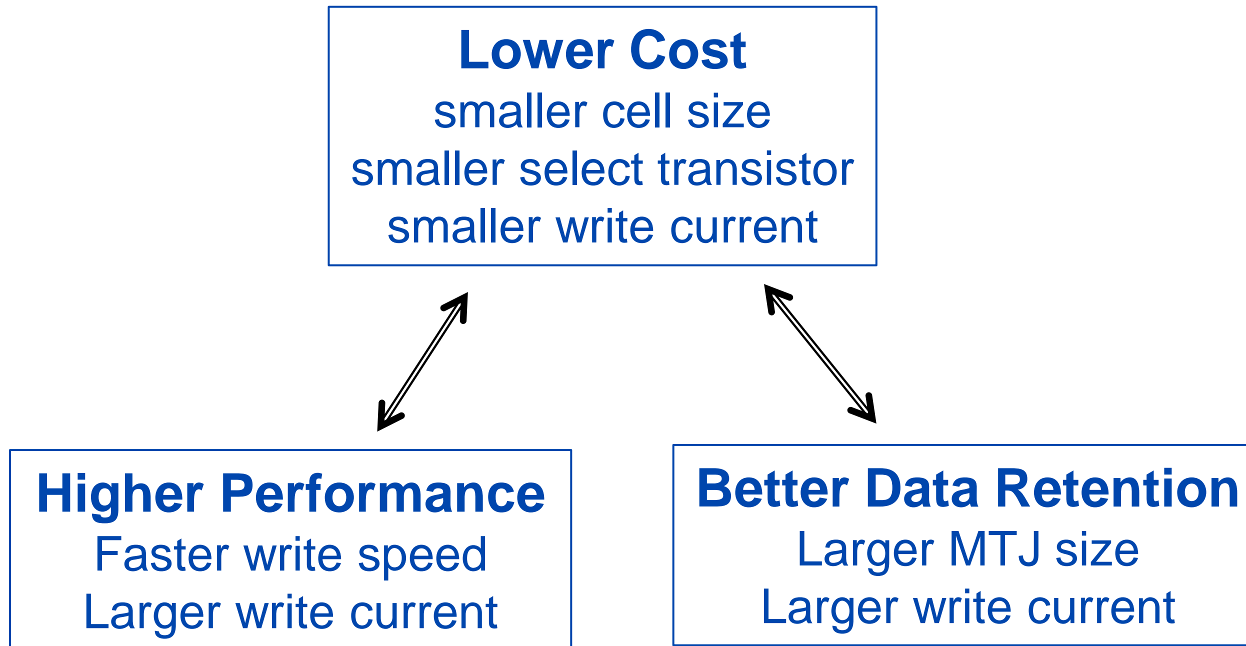
Trade-offs of STT writing (cont'd)

- Switching Current scales with MTJ area (constant current density)
  - smaller MTJ → smaller current requirement
  - smaller MTJ → worse data retention
- Current inversely proportional to pulse width at ~ ns speed
  - faster → higher current requirement





Considerations in STT-MRAM applications



- Cell size is not limited by MTJ size, but by the size of select transistor
- Generally need to prioritize the requirements between performance and data retention

Two applications for embedded STT-MRAM

	NVM	LLC
Data retention	10 years at 85-150°C	Hours to days
Write speed	20 – 200 ns	< 10 ns
Existing technology	eFlash (~ 20 masks below 28 nm node)	SRAM (over 500F^2 at 7 nm node)
MTJ size	> 50 nm	< 30 nm
Write current	> 100 μA	< 50 μA
Production	2018	?

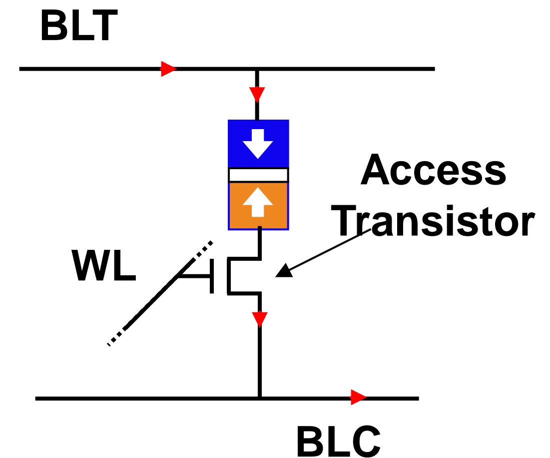
- Range of requirements within each application
  e.g. data retention through solder reflow process (at 260°C)
- Possibly a 3rd category in between NVM and LLC for mobile applications

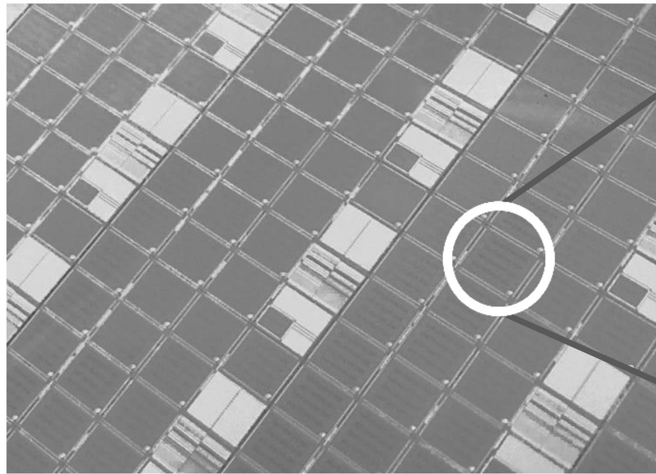
Outline

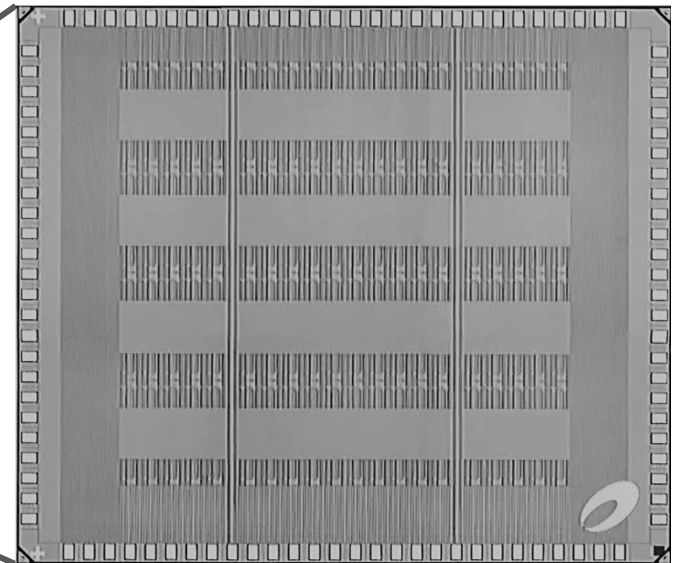
- Basic principles of STT-MRAM
- Embedded memory applications
- **STT-MRAM integration and chip level results**
- Tunnel barrier reliability at chip level

Integration of 8 Mb test chips at TDK Headway

- 8Mbits (16x512k) 1T-1MTJ
- IBM's 90nm CMOS technology
- 50F^2 cell size
- Redundancy and 2bit ECC
- FEOL in IBM foundry
- BEOL in TDK-Headway's fab

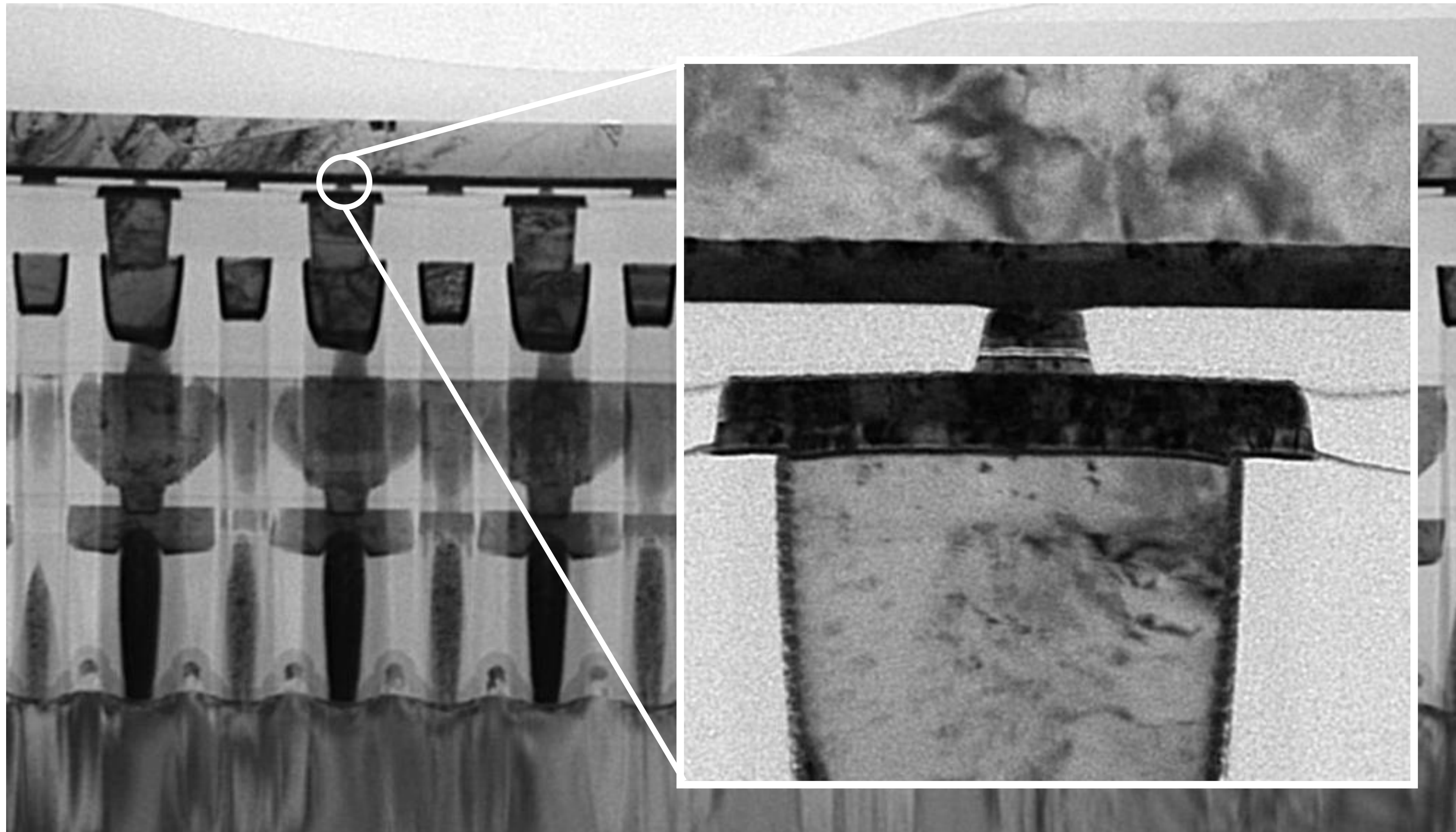






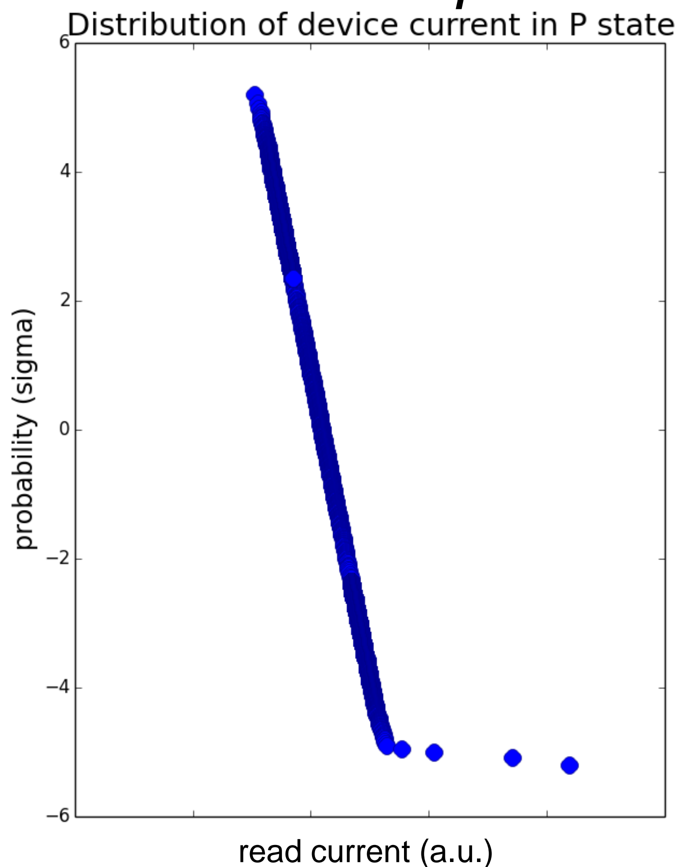
STT MRAM process integration

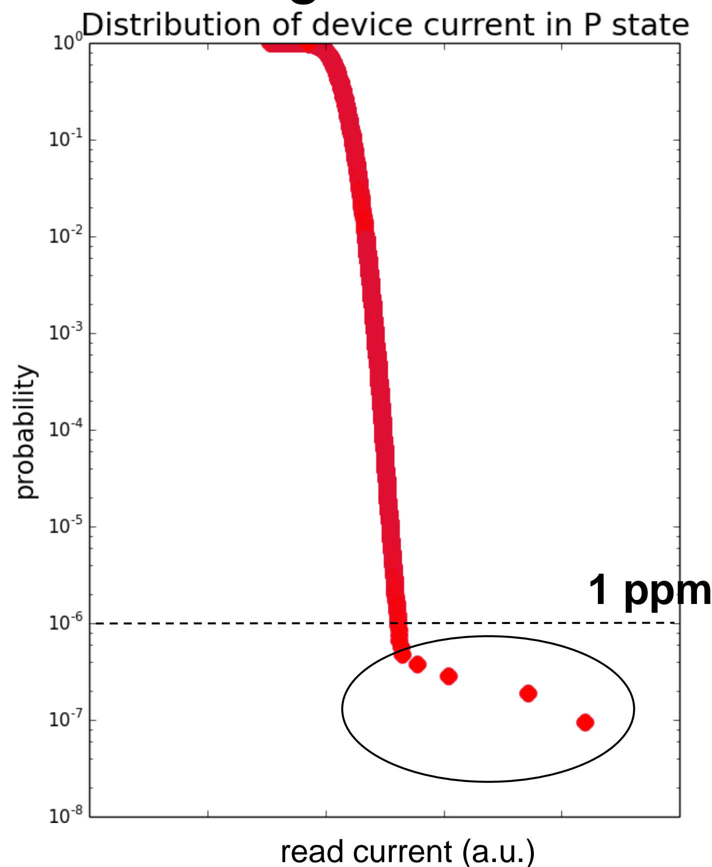
- MRAM only add two additional layers (MTJ and bottom electrodes) to standard CMOS BEOL: 3 to 4 mask adder
- MTJ stack is about 20 nm thick, can be easily integrated into CMOS backend process



Defect rate of 8 Mb chip

- Distribution of device current in the P state

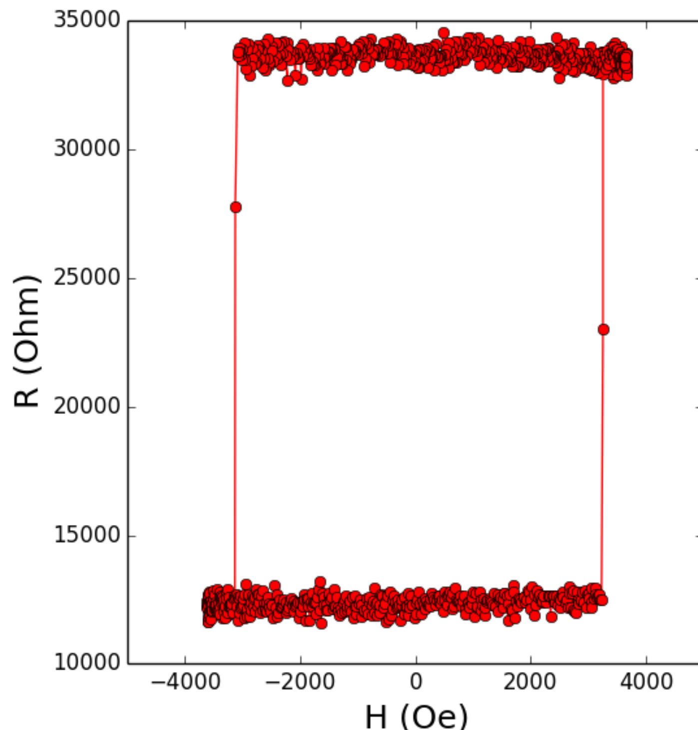




➔ less than 0.4 ppm defect rate

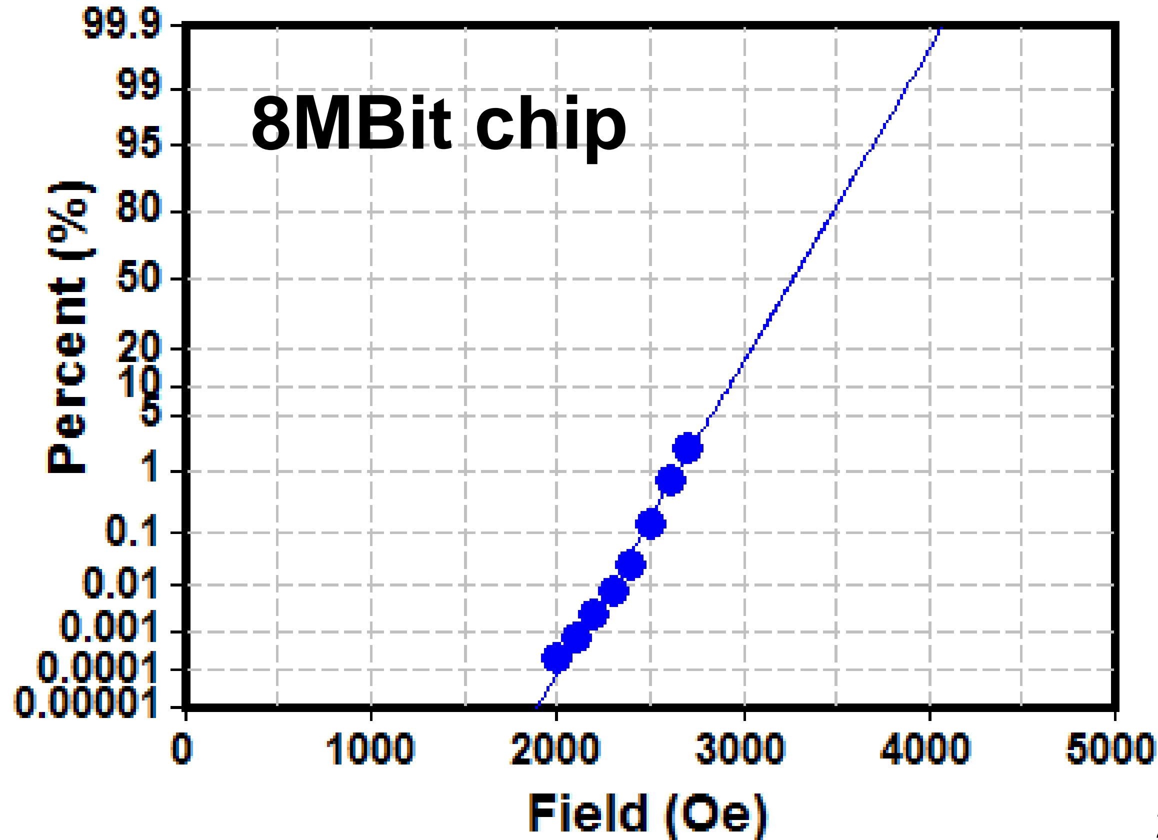
400C annealing after MTJ patterning

- ➔ 400C BEOL process can add up to several hours, depending on how many metal layers on top of MTJ
- ➔ Elemental movements and morphology changes can degrade anisotropy, exchange coupling, and defect level
  - selection of materials, diffusion barrier and interface/growth quality
  - Thorough engineering needed for electrodes, film stack, process, encapsulation



- ➔ **2.5 hours @400ºC after MTJ etching**
- ➔ **Diameter ~ 30 nm (electrical)**
- ➔ **DRR = 175%**
- ➔ **RA of 8.5 Ω-μm²**
- ➔ **H_C = 3300 Oe with no offset**

Robust against magnetic field disturbance



2016 VLSI-TSMC/TDK

- H_C mean over 3000 Oe, much higher than brown magnetic stripe card (~300 Oe) and similar to black mag-strip card (~2750 Oe)

Data retention and thermal stability factor

- Data retention determined by the thermal stability factor of energy barrier divided by $\kappa_B T$ ($\Delta = E_B / \kappa_B T$)
- From single MTJ's, different acceleration methods (magnetic field vs. current) and different switching process model (domain wall vs. macro-spin) can yield vastly different results
- Need to reply on direct retention test at the array level (with ppm failure rate), using only temperature as the acceleration parameter

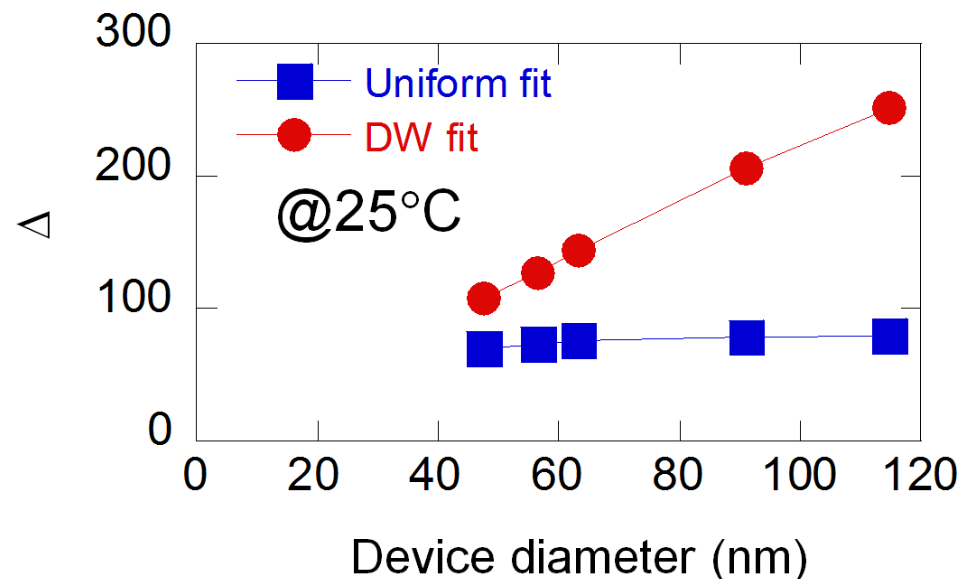
Fitting switching field distribution by a domain-wall mitigated model vs. a uniform switching model

To reach 1ppm failure rate

Δ=54 → 10 years

Δ=80 → 10^{12} years

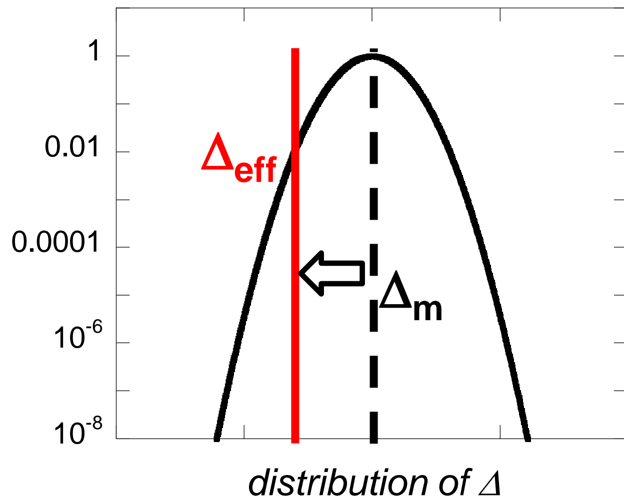
Δ=100 → 10^{20} years



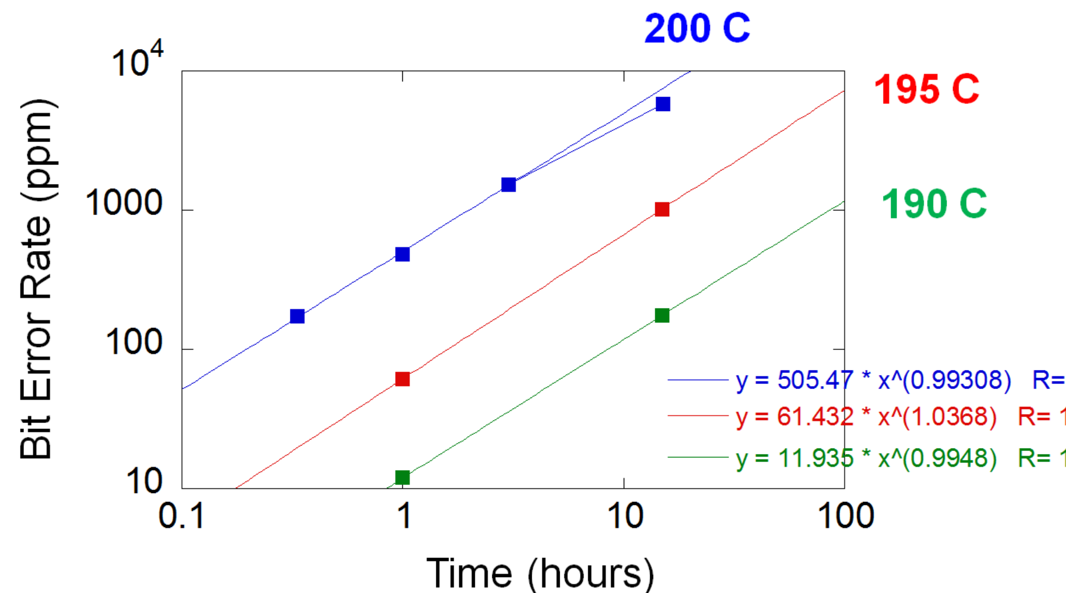
Chip level data retention (Δ_{eff} method)

- Chip level data retention is worsen by the distribution in energy barrier
- At low error rate (linear regime), effect of distribution can be described simply as an effective thermal stability factor
  $\Delta_{eff} = \Delta_m - \sigma^2/2$

$$\ln(BER) \sim \ln(t) + \ln(f_0) - \left[\Delta_m - \frac{\sigma_\Delta^2}{2}\right]$$

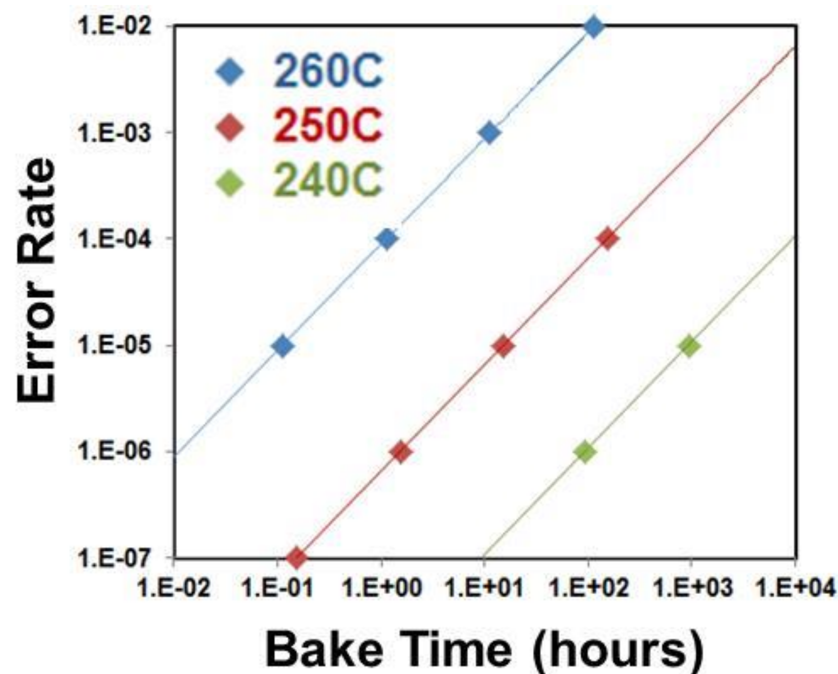


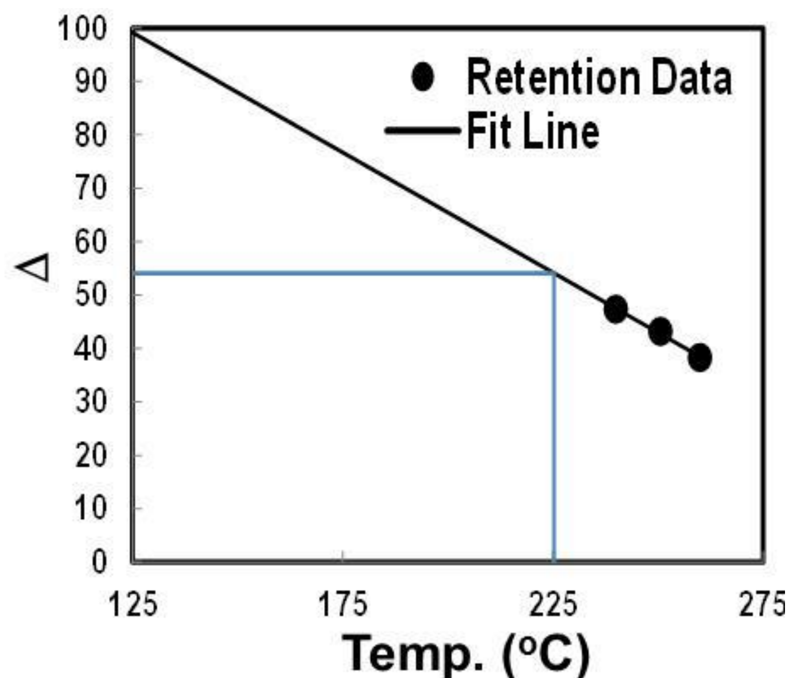
Thomas et al,. APL106, 172615 (2015)



MTJ for solder reflow compatibility

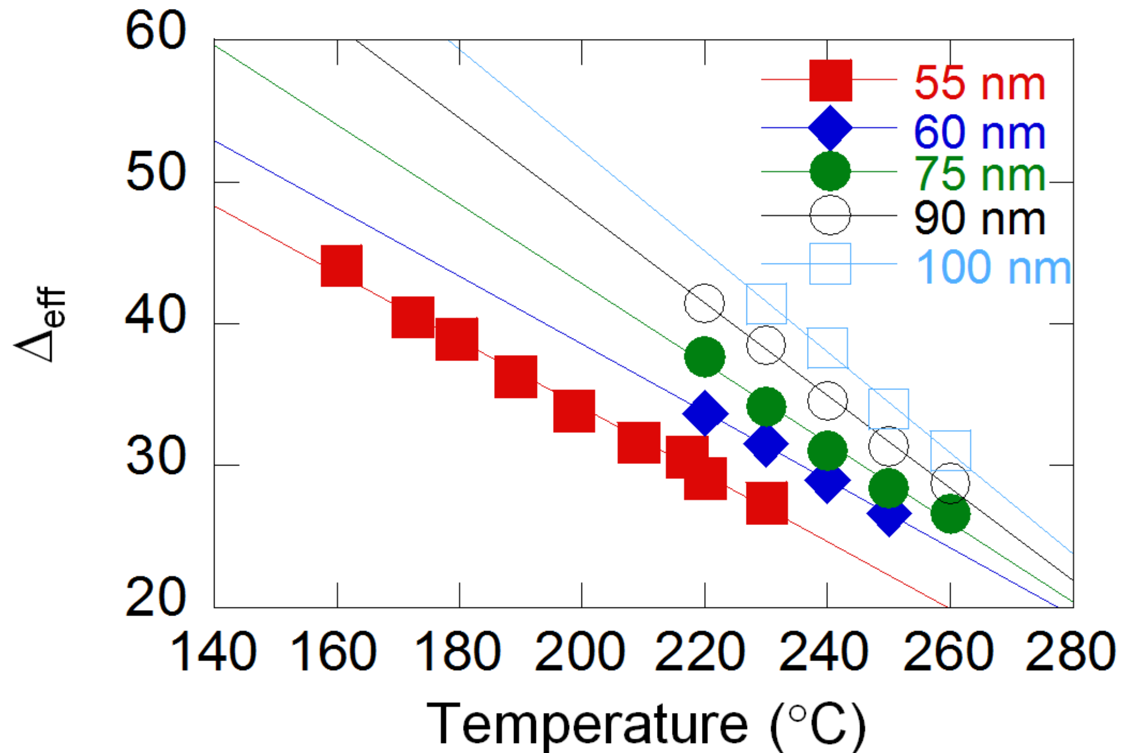
- Developed a MTJ stack of high PMA and thermal stability to satisfy solder reflow requirement of 260ºC for 90 seconds *(2016 VLSI TSMC/TDK)*
- Effective thermal stability method projects 1 ppm failure rate after 10 years at 225ºC





1ppm 10 years retention at 225ºC

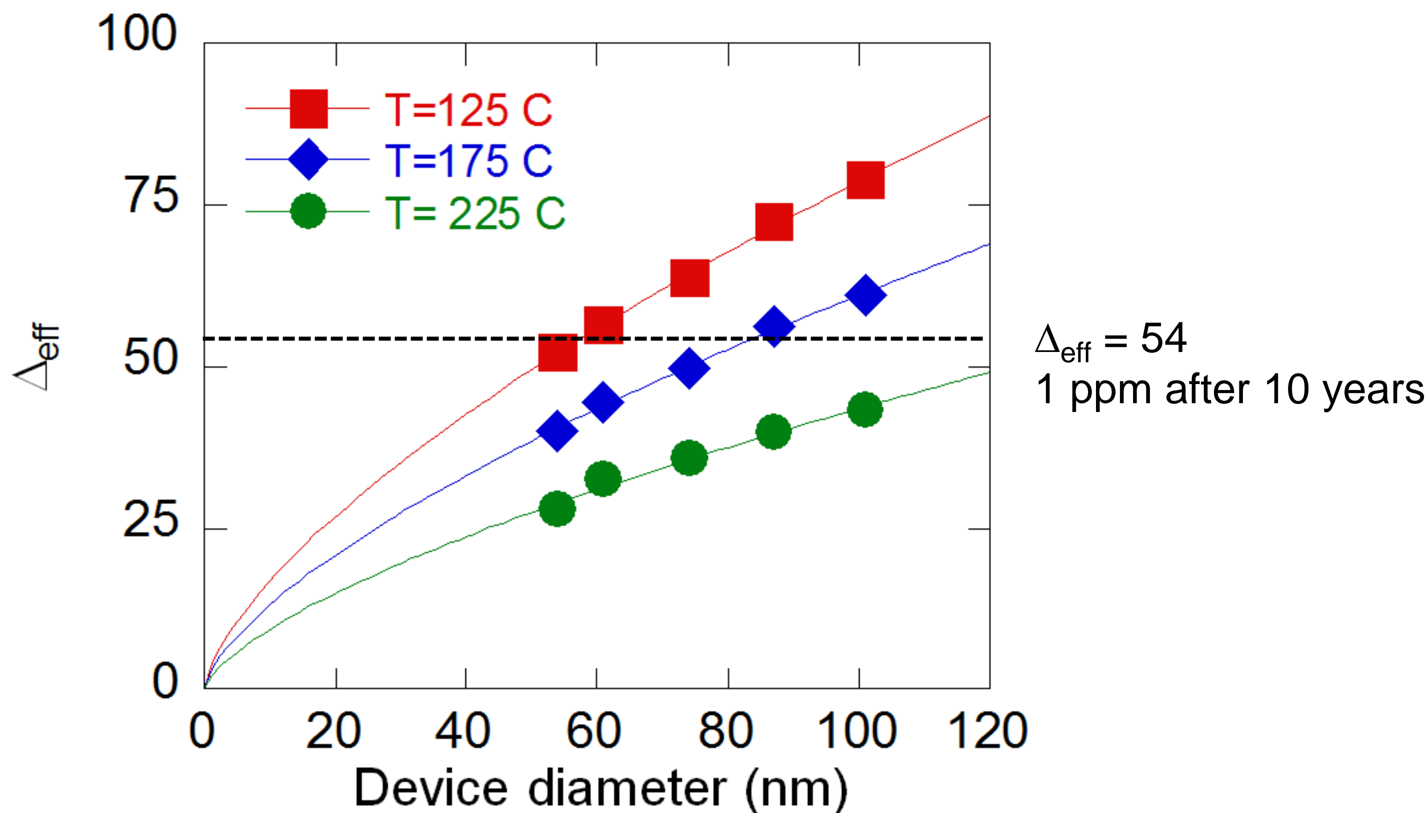
Data retention vs. size



- Thermal stability decreases with temperature because of $1/\kappa_B T$ and temperature dependence energy barrier (decrease of anisotropy and magnetic moment)
- Linear dependence on temperature in the temperature range of interest
- Data retention has significant size dependence

Data retention vs. size (cont'd)

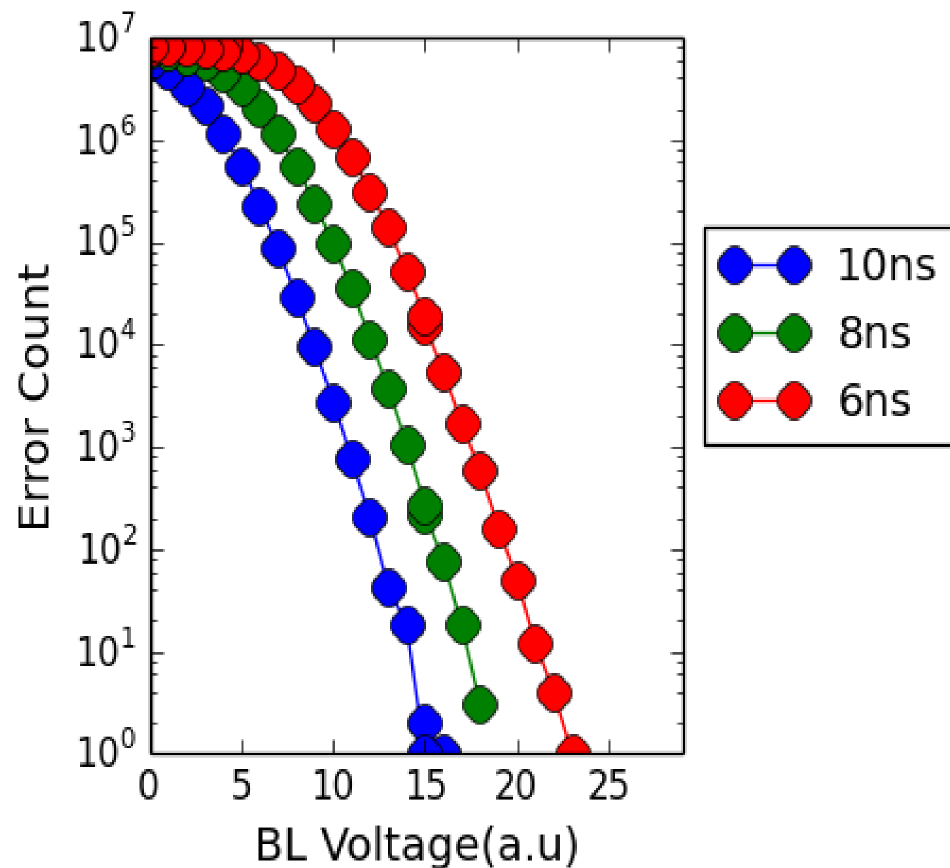
- Linear extrapolation is used to estimate Δ_{eff} down to 125C

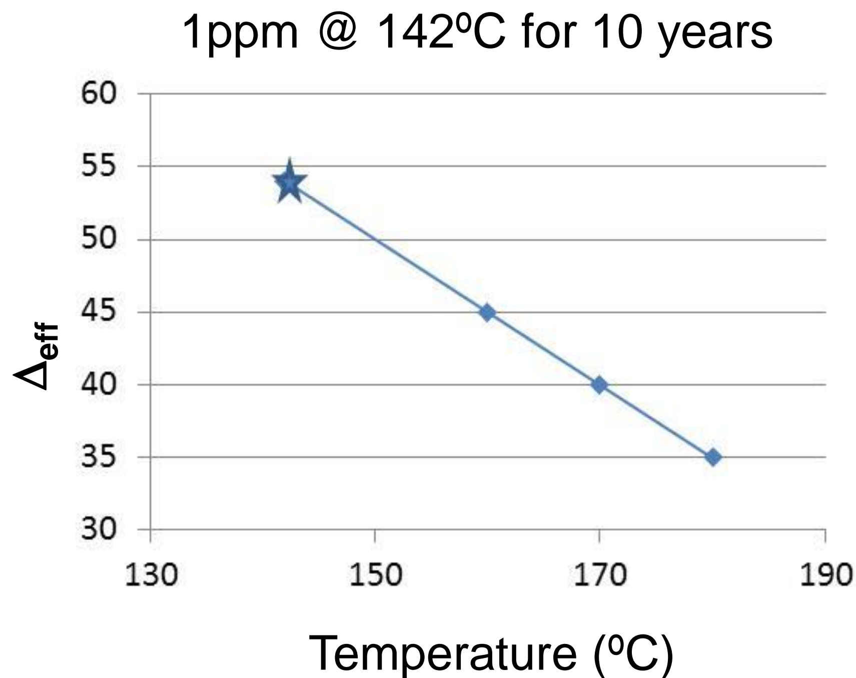


- ➢ Size dependence of energy barrier well fitted by a power law size^0.67
- ➢ Deviation from linear dependence of domain wall energy is due to energy barrier distributions

Error free writing in chip level

➔ Error free writing on 8 Mb chips without ECC
- Down to 6 ns write pulse
- While keep data retention to 142ºC for 10 years



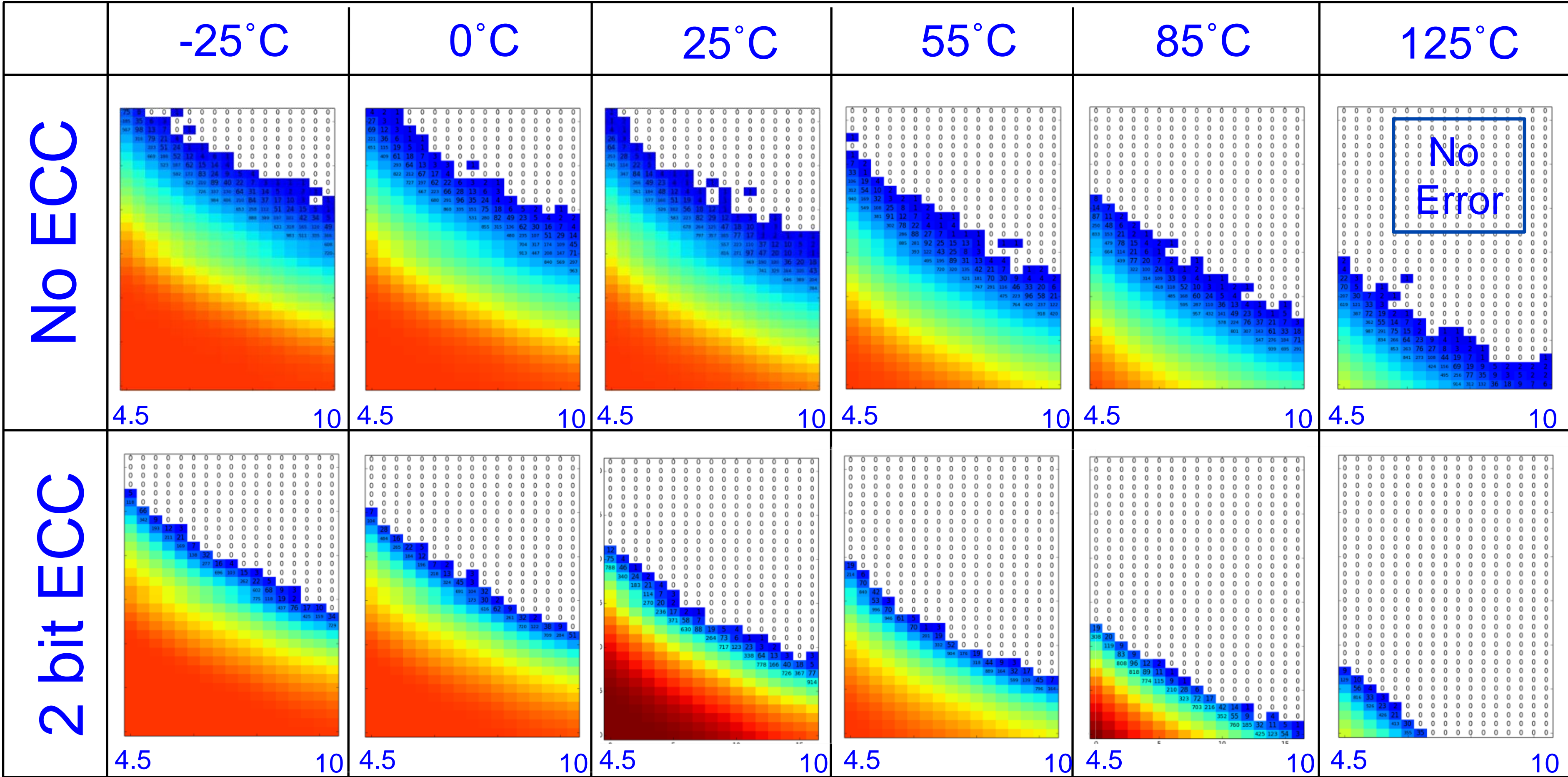


Write Schmoo vs. pulse length (without ECC)

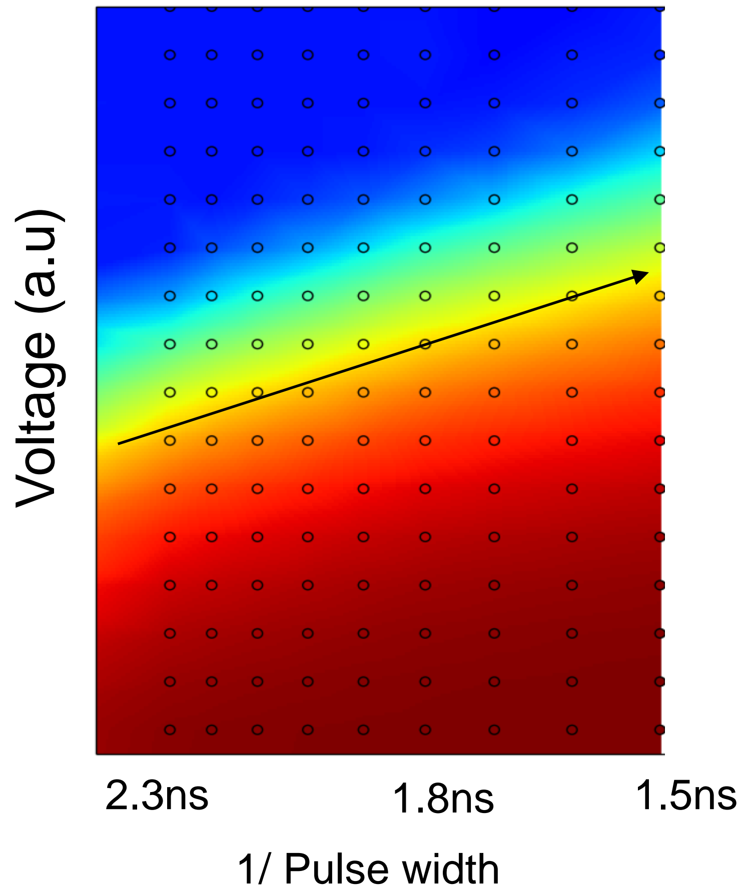
- 8 Mb chip without ECC
- Wide margin in the sub 10 ns writing regime
  - No back hopping (pinned layer issue)
  - Occasional single bit error to be corrected by ECC

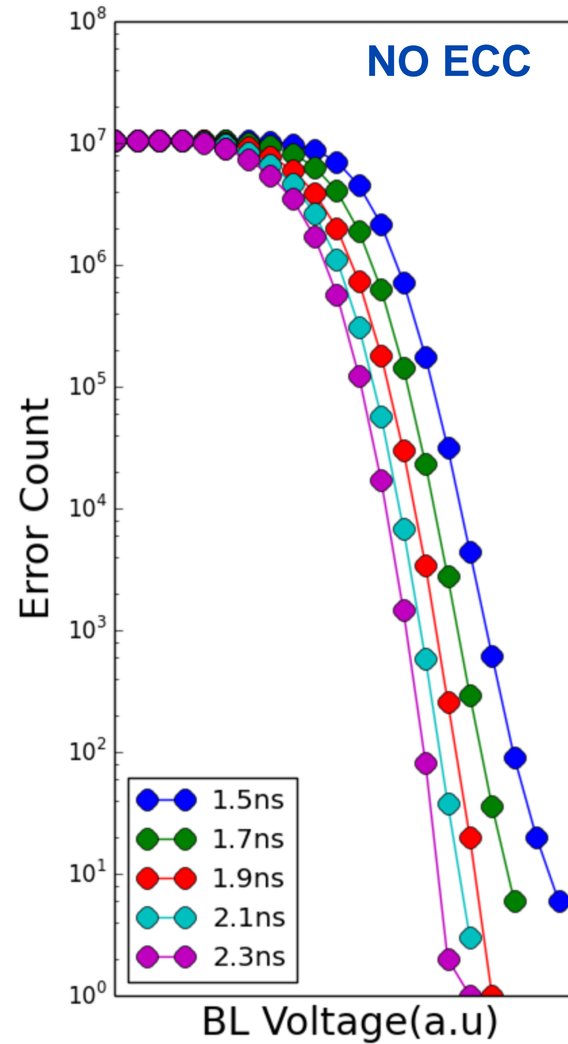
Temperature dependence

Fast operation down to 4.5 ns demonstrated over wide temperature range



Potential for even faster speed





8 Mb written without error with 1.5 ns write pulse

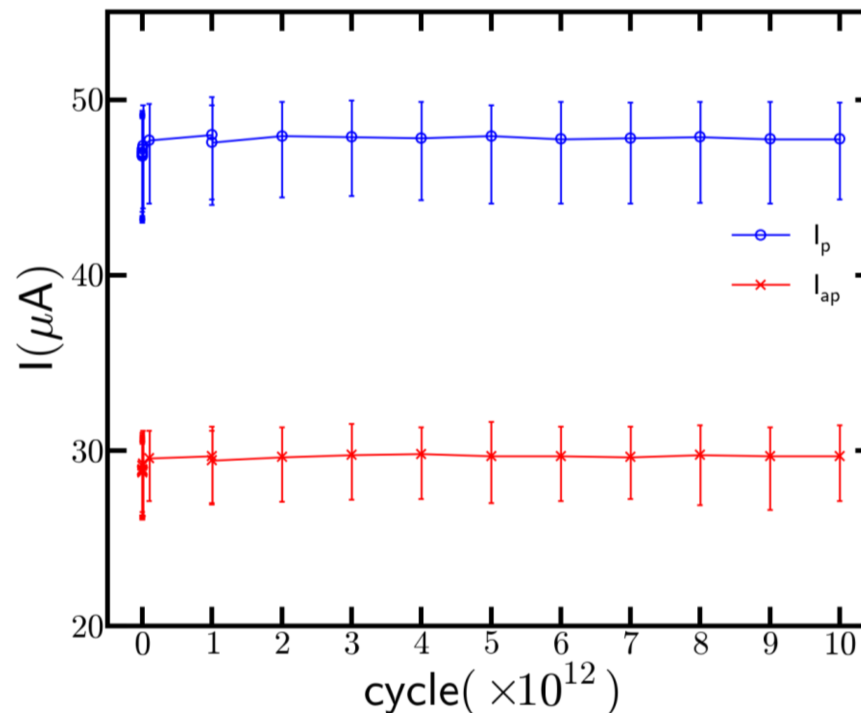
Outline

- Basic principles of STT-MRAM
- Embedded memory applications
- STT-MRAM integration and chip level results
- **Tunnel barrier reliability at chip level**

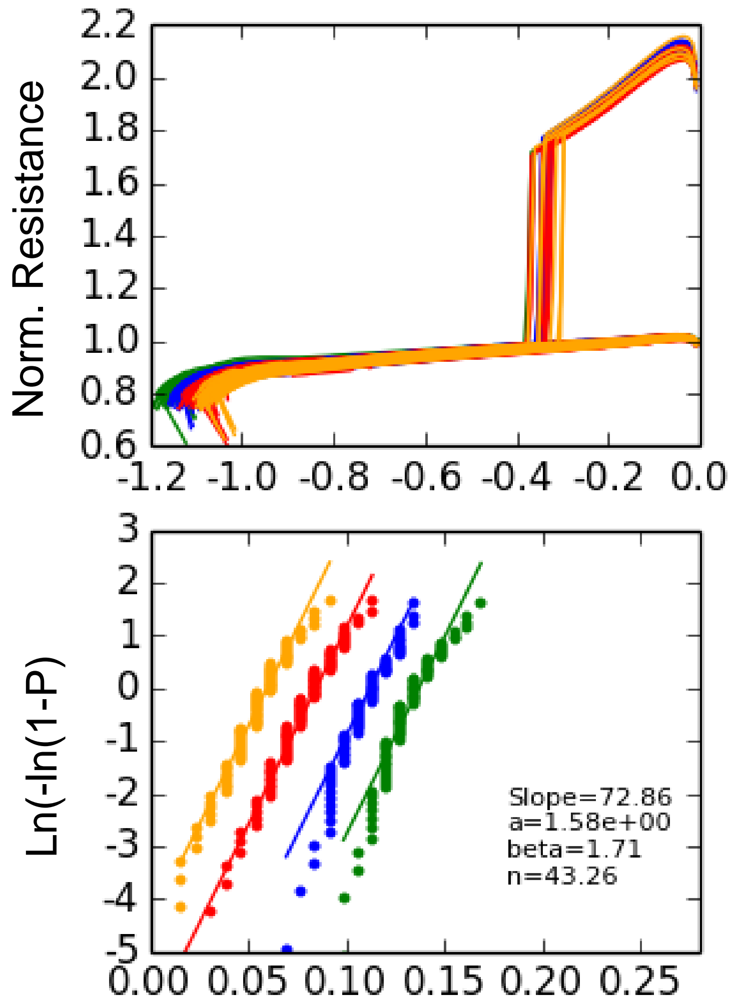
Endurance: 10^{13} cycles of 10ns write pulses

- No error found in 64 bits after 10^{13} cycles
- No drift observed in MTJ resistance throughout the 10^{13} cycles

MTJ sense current (Midian and range of 64 bits) throughout the cycling



MgO Integrity: TDDB at MTJ level



- Traditional time dependent dielectric breakdown (TDDB) measurements
- Measure on discrete devices with ramp voltage source; fitting power law

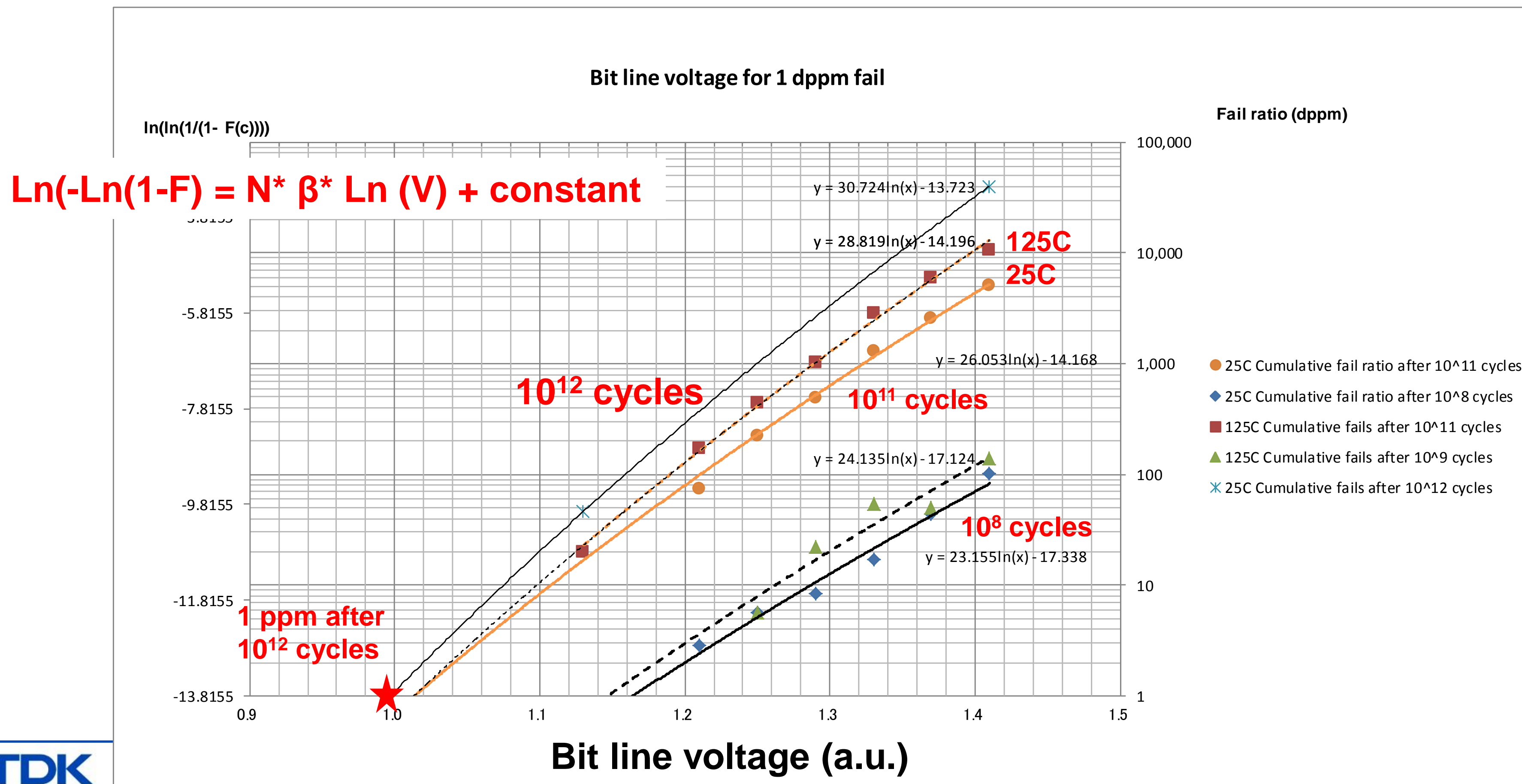
$$F_{CVS}(t,V) = 1 - \exp\left[-\left(\frac{t}{\eta(V)}\right)^{\beta}\right]$$

$$\eta(V) = a \cdot V^{-n}$$

- Clean breakdown
- Test conditions
  - 4 ramp rates (1 ms, 3 ms, 10 ms, 30 ms per step)
  - 8 mV per step (0→2V in 250 steps)
- Good fit to Weibull distribution
  - Shape parameter of 1.7
  - Can project endurance to ppm level

Endurance: chip level results

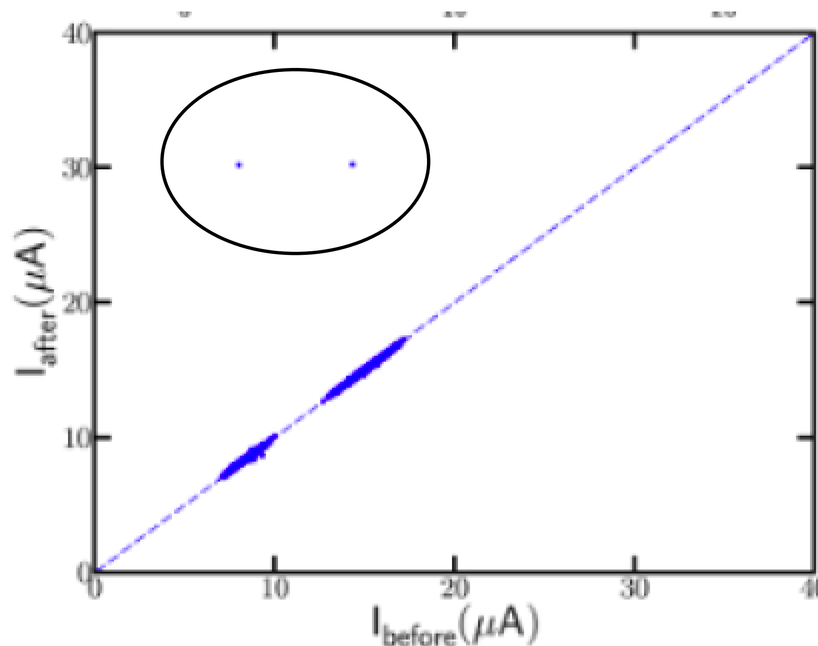
- Stress up to 10^{12} cycles
  - 5 Kb/chip up to 400 chips
  - Bit line voltage divided between MTJ's and select transistors, both with variations
- Chip level endurance resutls consistent with device level TDDB projections

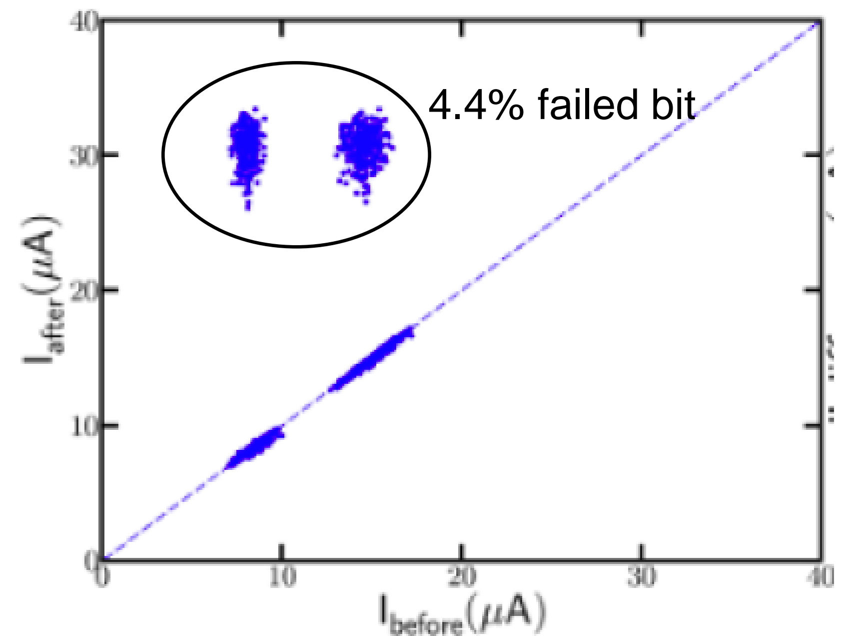


Endurance: no gradual degradation

- Survived bits show no change in electrical characteristics after cycling
  - Even after 10^{11} cycles at high stress voltage with high failure rate

5 Kb MTJ sense current before and after 10^{11} write cycling





STT-MRAM for embedded memory applications

- STT-MARM has much lower cost than eFalsh and LLC SRAM
- STT-MRAM is CMOS process compatible (400ºC thermal budget and low defect rate)
- STT-MARM is adaptable to suit varying requirements in data retention and performance
- STT-MRAM has demonstrated >10^{12} endurance at chip level