Spin Transfer Torque Devices as an Embedded non-volatile Memory

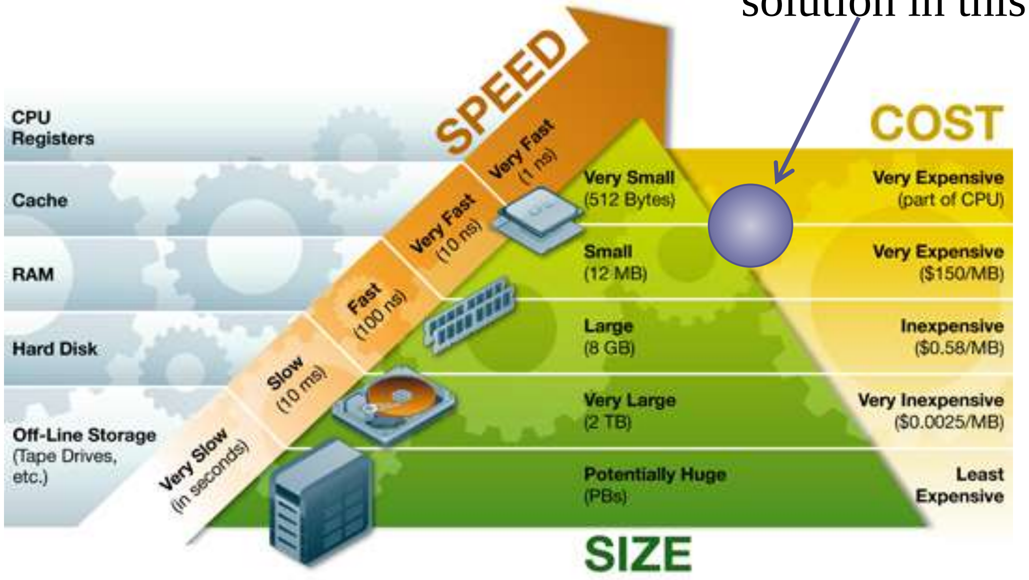
Sayeef Salahuddin

Electrical Engineering and Computer Sciences, UC Berkeley

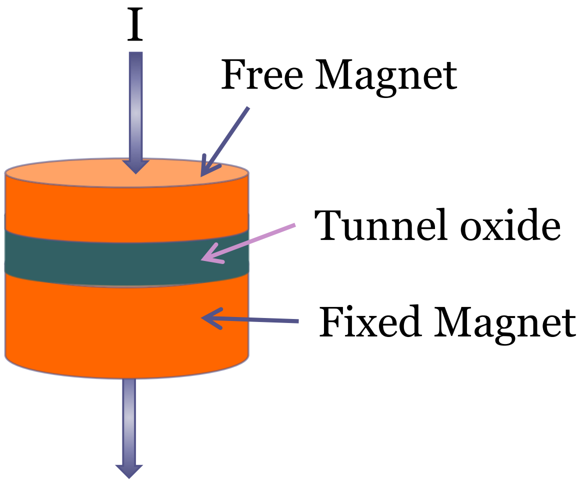
Research group: LEED (Laboratory for Emerging and Exploratory Devices)
http://leed.eecs.berkeley.edu

Memory Technologies and MRAM

Plausible non-volatile solution in this space



source: http://www.ts.avnet.com/

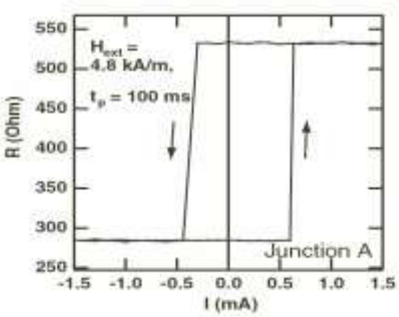


I

Free Magnet

Tunnel oxide

Fixed Magnet



Kubota et. al., JJAP, 44, 40, 1237,2005

64 MB DRAM like STT MRAM
Available from Everspin

0.5 V, non-volatile, high endurance, write speed ~ few ns

A recent announcement

GLOBALFOUNDRIES / News & Events / Press Releases / GLOBALFOUNDRIES Announces Availability of Embedded MRAM on Leading 22FDX® FD-SOI Platform

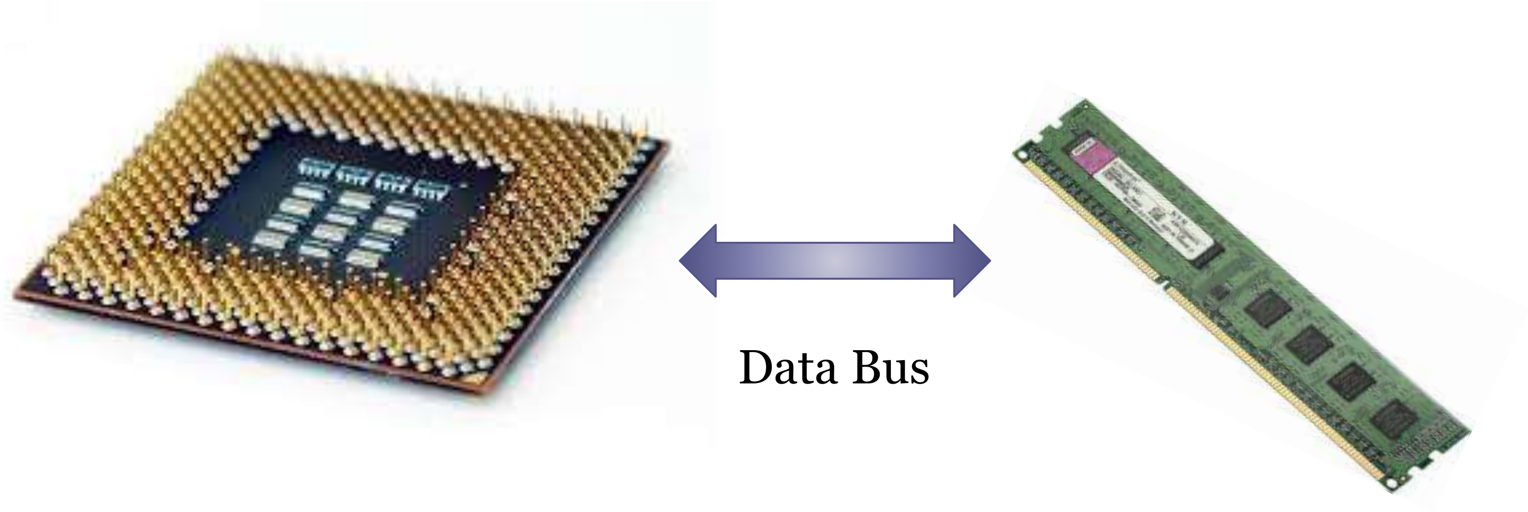
GLOBALFOUNDRIES Announces Availability of Embedded MRAM on Leading 22FDX® FD-SOI Platform

Sep 20, 2017

Advanced embedded non-volatile memory solution delivers 'connected intelligence' by expanding SoC capabilities on the 22nm process node

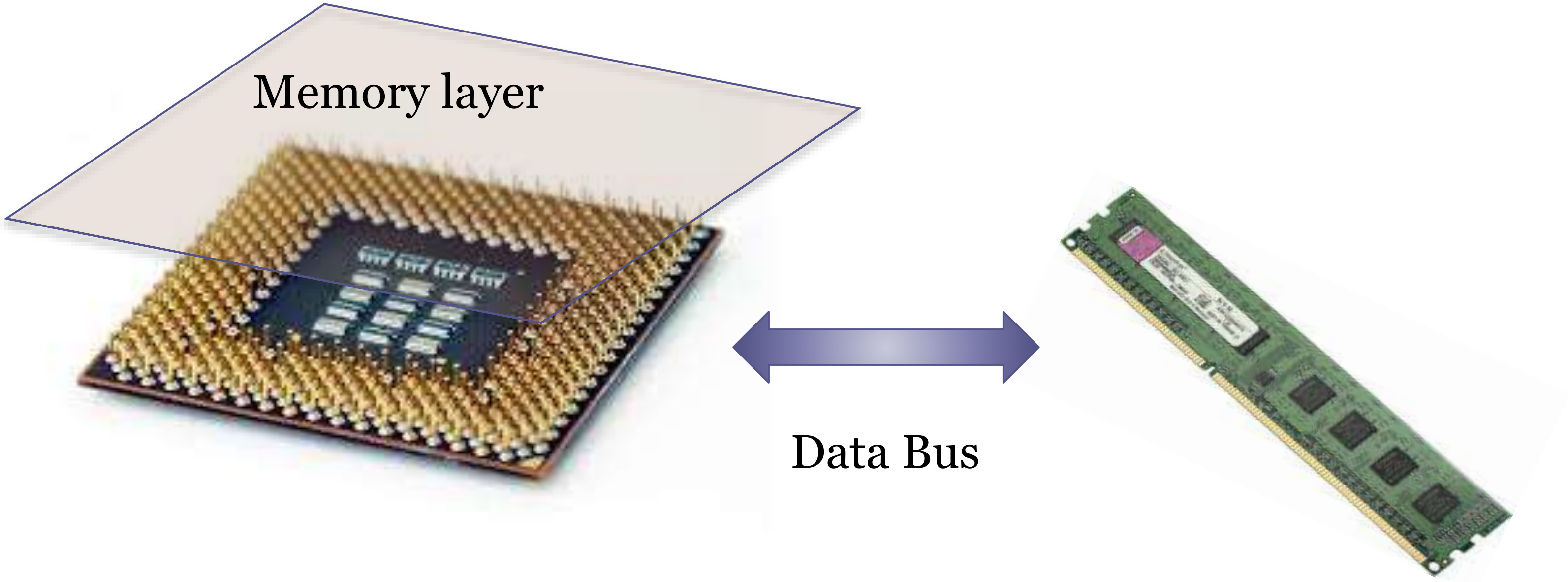
Santa Clara, Calif., September 20, 2017 -- GLOBALFOUNDRIES today announced the availability of its scalable, embedded magnetoresistive non-volatile memory (eMRAM) technology on the company's 22nm FD-SOI (22FDX®) platform. As the industry's most advanced embedded memory solution, GF's 22FDX eMRAM provides high performance and superior reliability for broad applications in consumer and industrial controllers, data centers, Internet of Things (IoT), and automotive.

Embedded Non-Volatile Memory



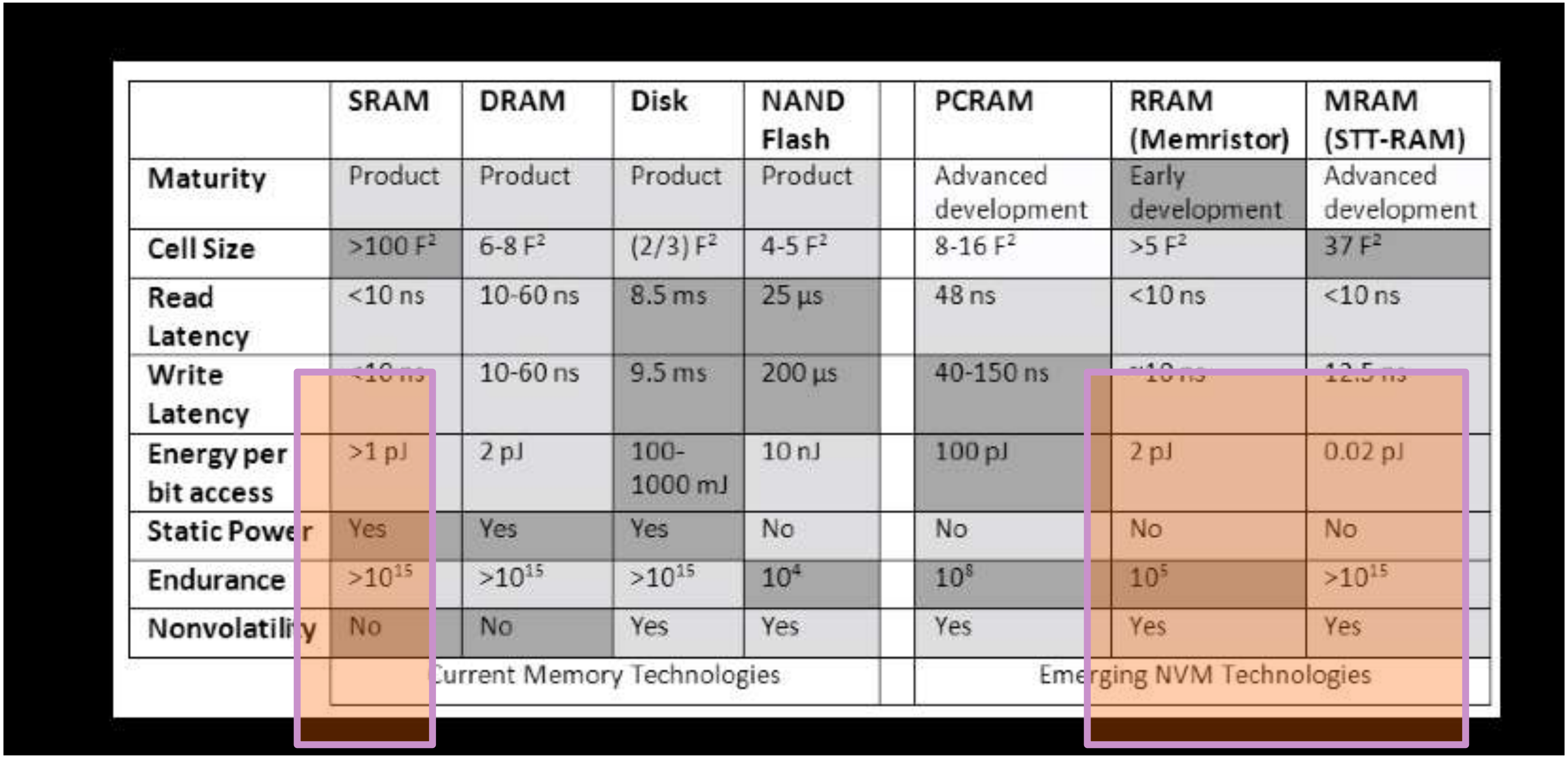
Data Bus

Embedded Non-Volatile Memory



- Significant reduction in energy → mobile computing
- Non volatility is a prime requirement for machine learning applications

Competition

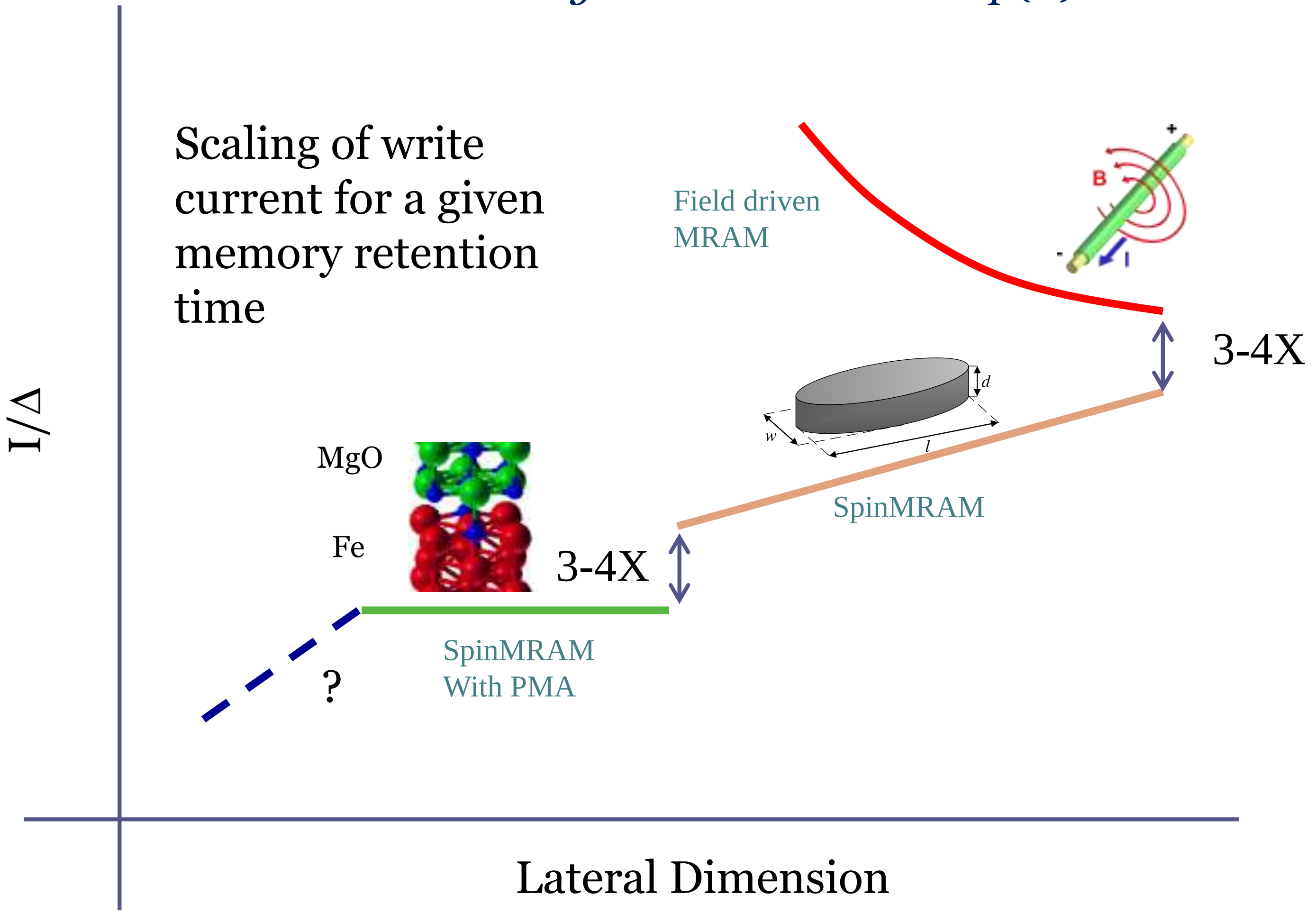


	SRAM	DRAM	Disk	NAND Flash	PCRAM	RRAM (Memristor)	MRAM (STT-RAM)
Maturity	Product	Product	Product	Product	Advanced development	Early development	Advanced development
Cell Size	>100 F^2	6-8 F^2	(2/3) F^2	4-5 F^2	8-16 F^2	>5 F^2	37 F^2
Read Latency	<10 ns	10-60 ns	8.5 ms	25 µs	48 ns	<10 ns	<10 ns
Write Latency	<10 ns	10-60 ns	9.5 ms	200 µs	40-150 ns	<10 ns	12.5 ns
Energy per bit access	>1 pJ	2 pJ	100-1000 mJ	10 nJ	100 pJ	2 pJ	0.02 pJ
Static Power	Yes	Yes	Yes	No	No	No	No
Endurance	$>10^{15}$	$>10^{15}$	$>10^{15}$	10^4	10^8	10^5	$>10^{15}$
Nonvolatility	No	No	Yes	Yes	Yes	Yes	Yes
	Current Memory Technologies				Emerging NVM Technologies		

- Slightly slower than SRAM
- Smaller than SRAM
- Same endurance

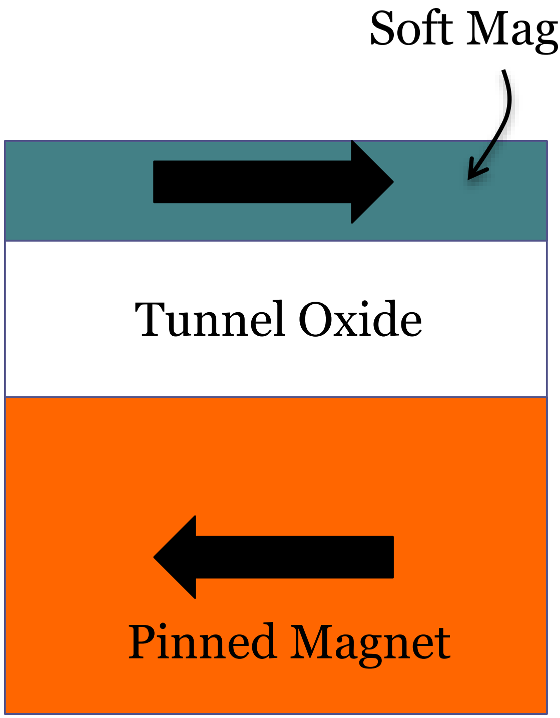
MRAM scaling Trends

Memory retention time ~ exp(Δ)

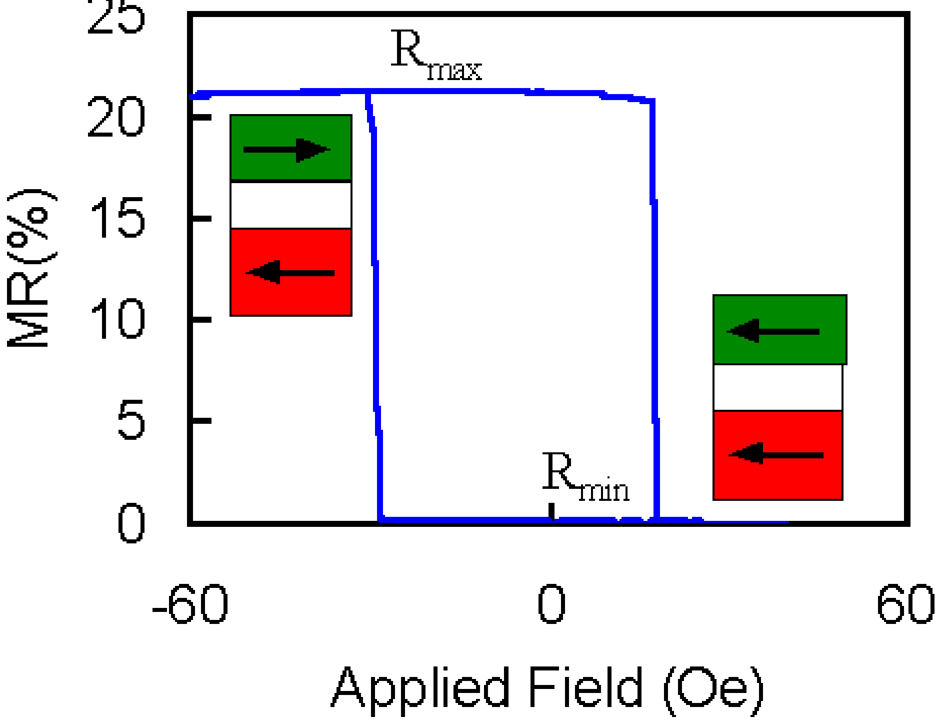


Scaling of write current for a given memory retention time

The Traditional MRAM



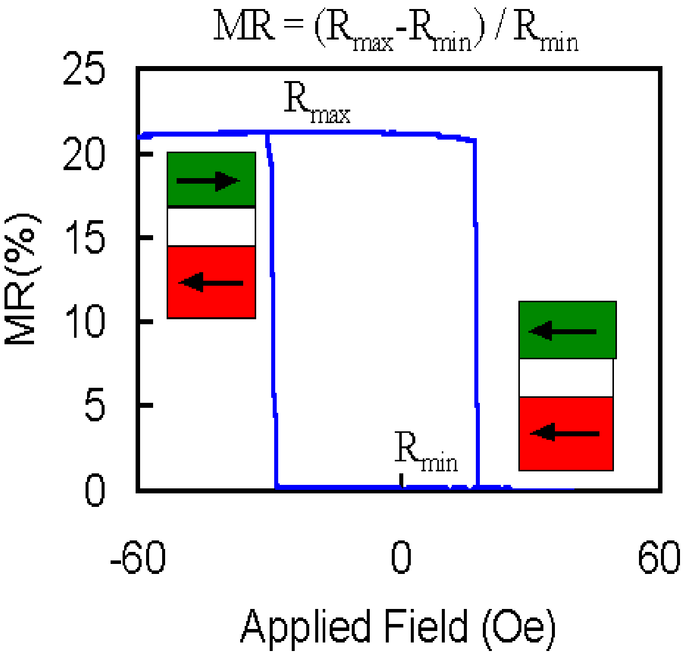
$$MR = (R_{max} - R_{min}) / R_{min}$$

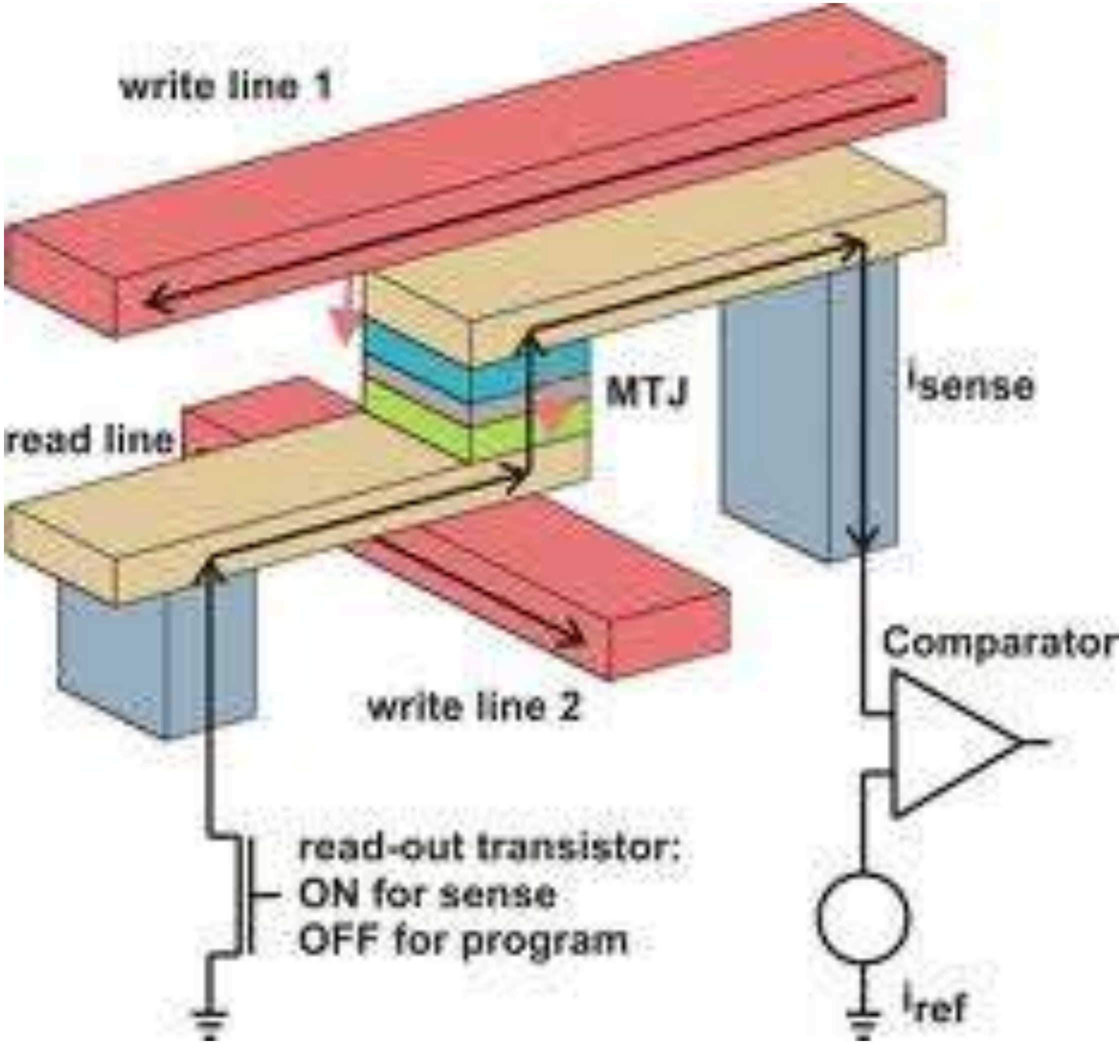


Flowing electrons are predominantly composed of the type of spins based on the magnetization direction of the magnet that injects them.

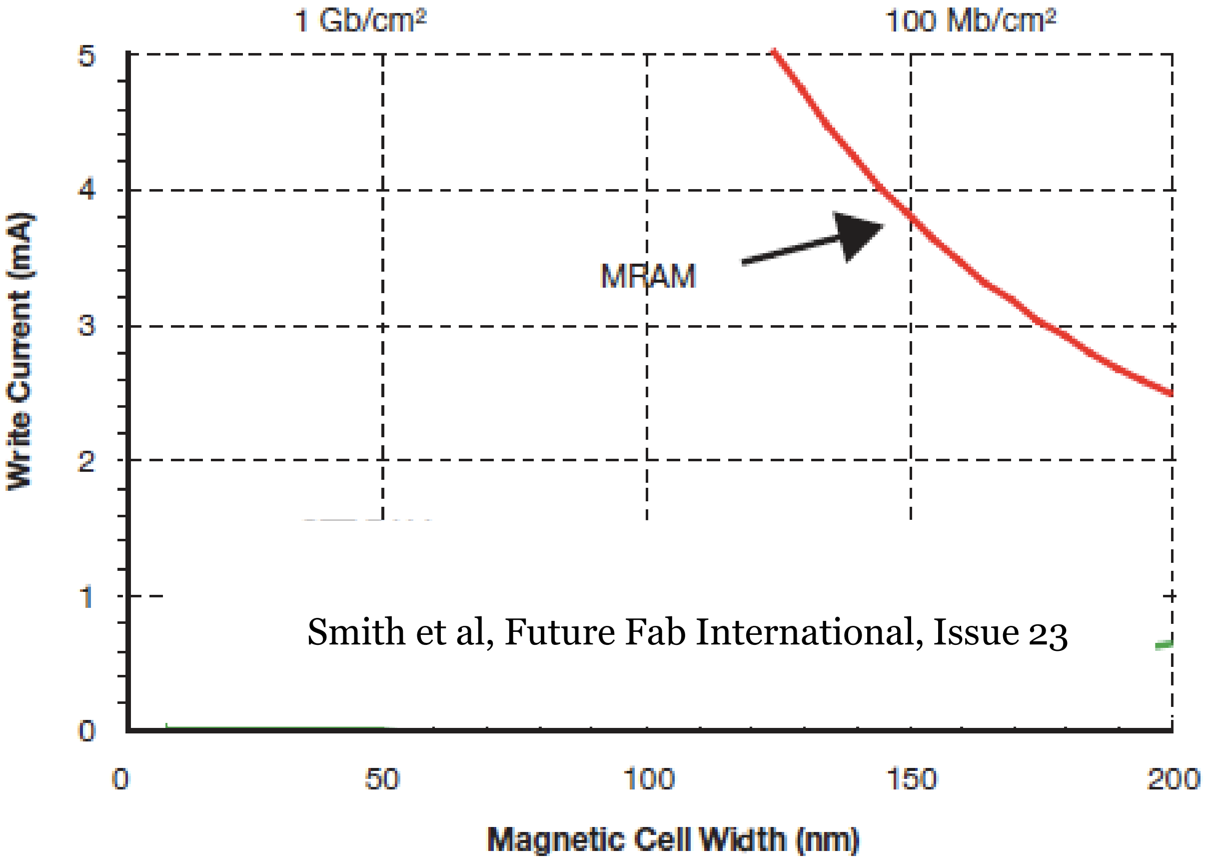
- Non-volatility > 10 years (typical)
- Unlimited Endurance

MRAM Cell


MR = (R_max-R_min) / R_min
R_max
R_min
MR(%)
25
20
15
10
5
0
-60
0
60
Applied Field (Oe)


write line 1
read line
MTJ
I_sense
Comparator
write line 2
read-out transistor:
ON for sense
OFF for program
I_ref

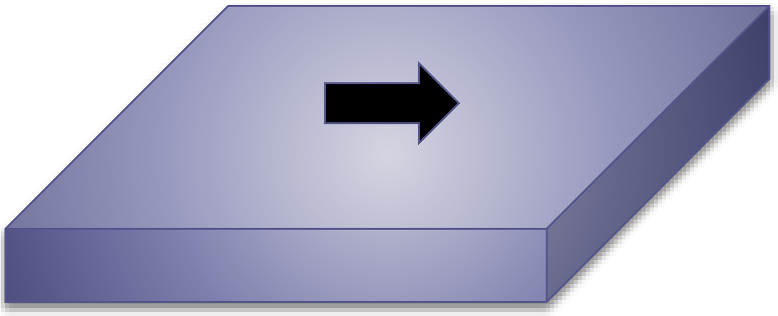
But..MRAM does not scale



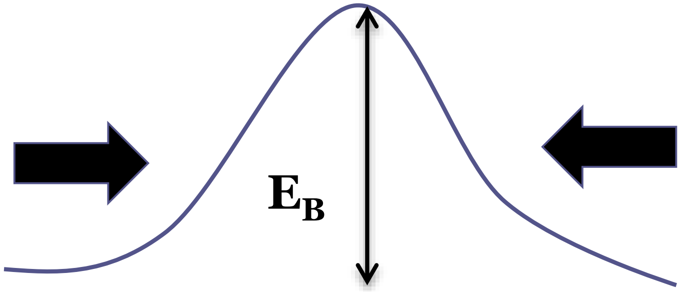
Smith et al, Future Fab International, Issue 23

$$I \propto 1/L$$

Magnet and thermal stability, Δ



Thin film magnet



N=density of spins #/cm³
V=volume of the magnet

$N\mu_B V H_k = E_B$
$M_s = N\mu_B$

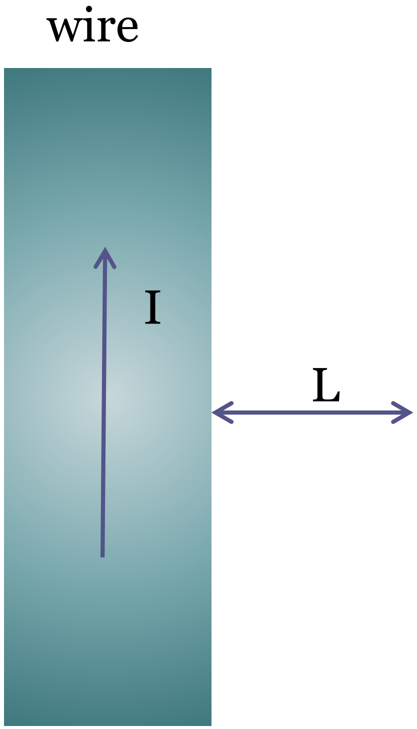
Time necessary to spontaneously go from one state to the other

$$t = t_0 e^{E_B/kT}$$

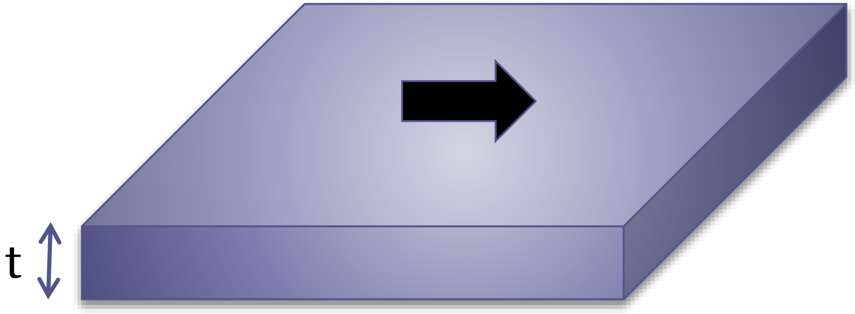
- Usually E_B is parameterized by $\Delta = E_B/kT$

Field needed to switch the magnet, $H_k = \dfrac{E_B}{NV\mu_B}$

MRAM scaling challenge



$$H = \frac{I}{2\pi L}$$

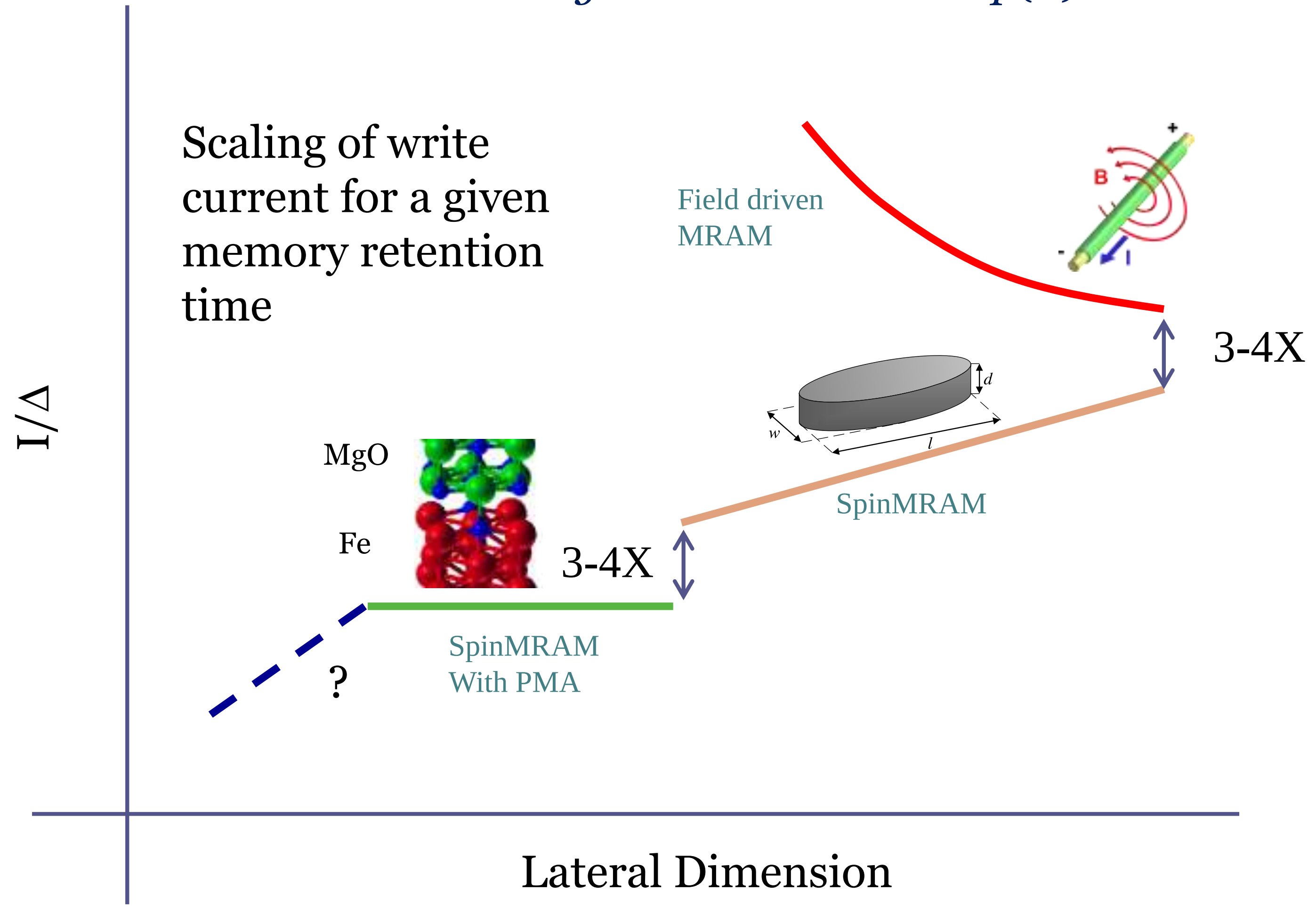


$$H_k = \frac{E_B}{N V \mu_B}$$

To achieve a specific H_k, $I = 2\pi H_k L = 2\pi L \times \frac{E_B}{N \mu_B t L^2} \propto \frac{1}{L}$

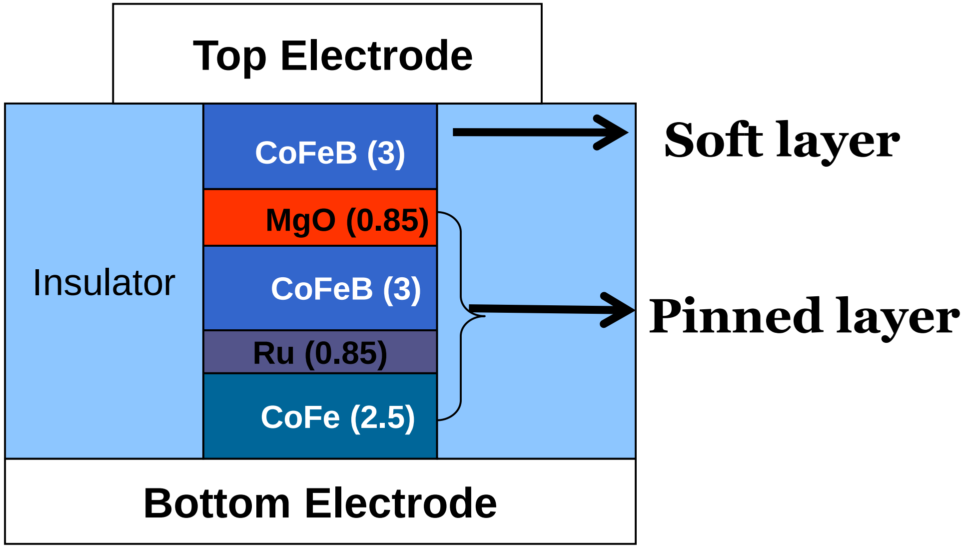
MRAM scaling Trends

Memory retention time ~ exp(Δ)



Scaling of write current for a given memory retention time

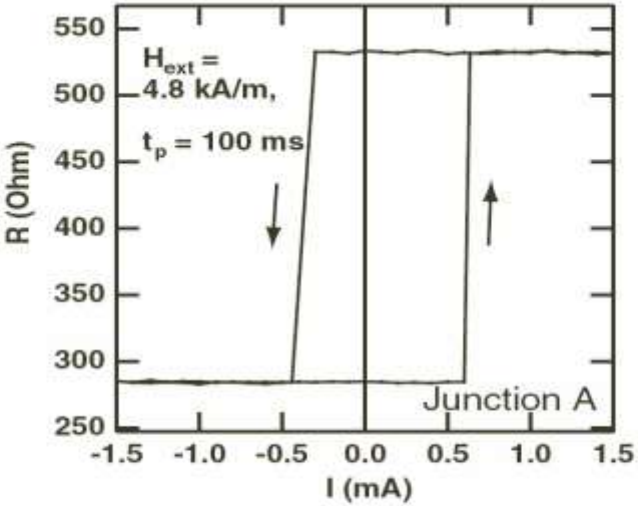
Spin Transfer Torque Devices



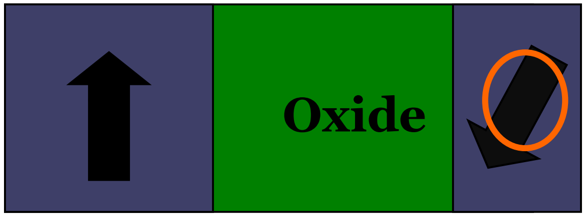
Soft layer

Pinned layer

Slonczewski:
JMMM 1996, 2002,2007
PRB 1989, 2005
First experiment:
Phys. Rev. Lett. 84, 3149 (2000)

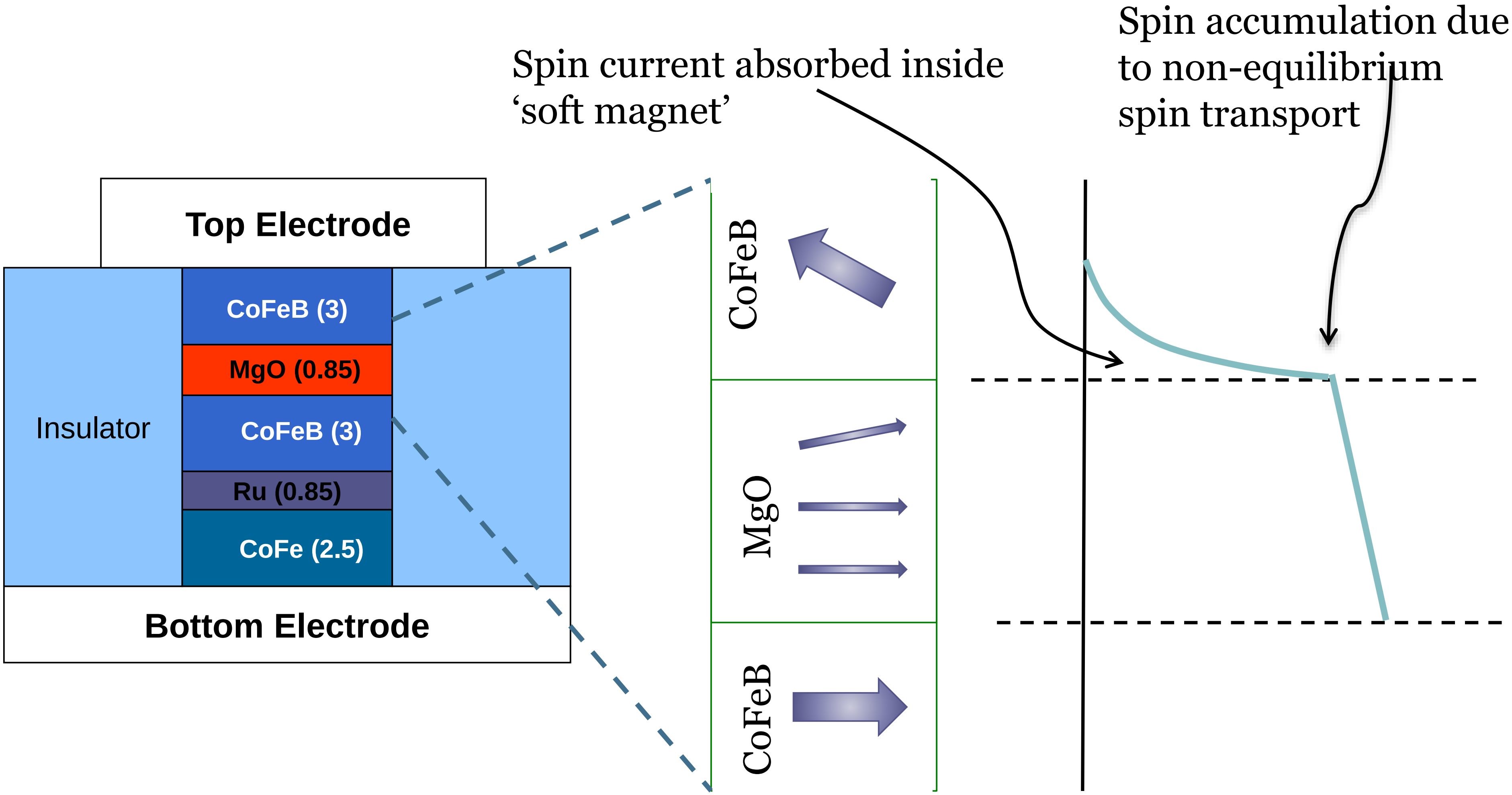


Kubota et. al., JJAP, 44, 40, 1237,2005



Pinned layer **Soft layer**

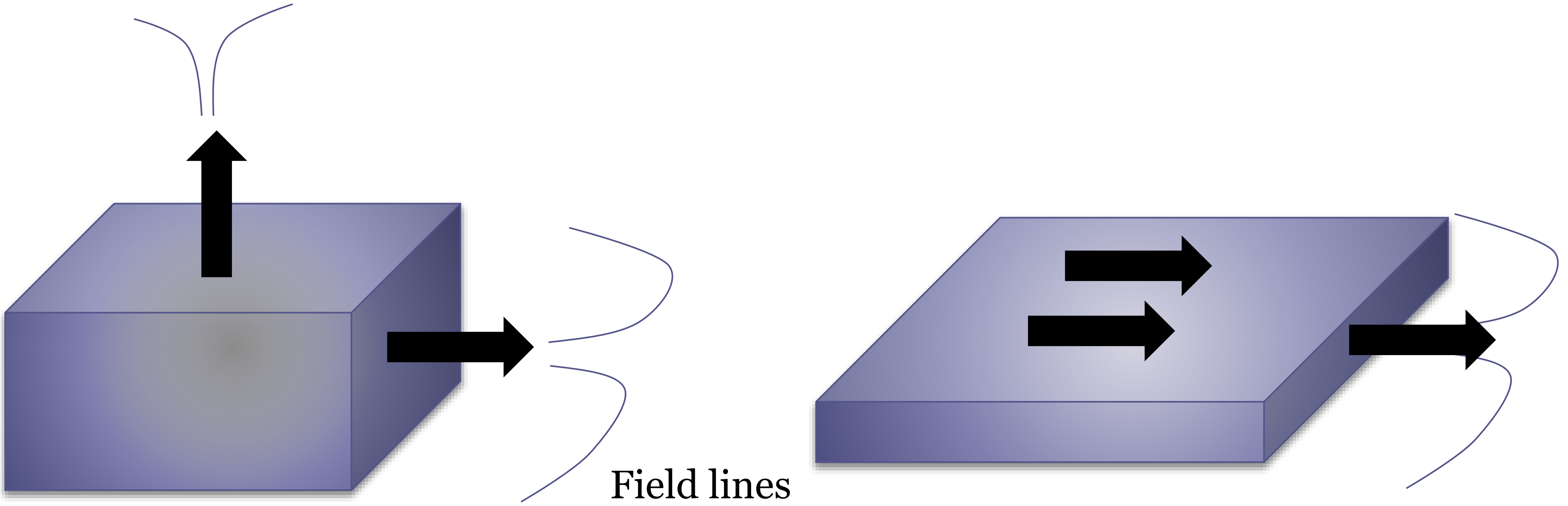
Simple Physics of Spin Torque



Spin current absorbed inside 'soft magnet'

Spin accumulation due to non-equilibrium spin transport

Physics of STT Devices: Thin film magnets



Field lines

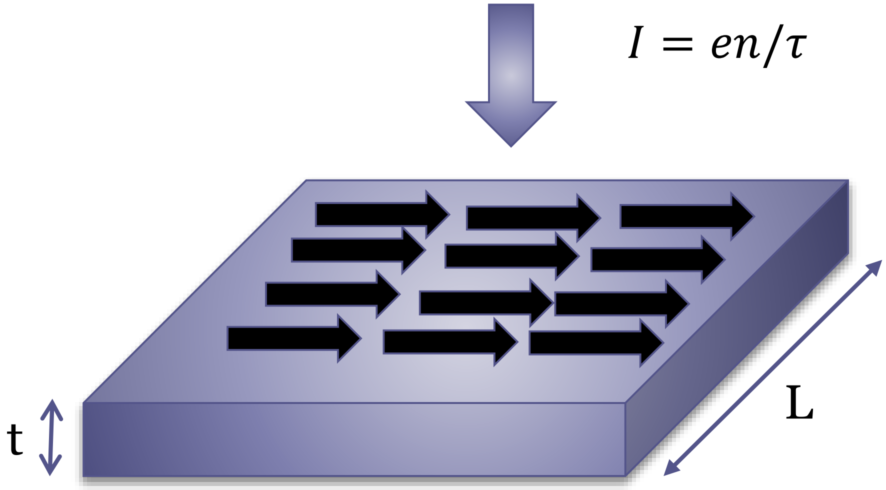
Field lines

Magnetostatics legislates that a thin film magnet be polarized in-plane

Simple Physics of Spin Torque Transfer

Spin angular momentum has to be conserved.

If we are flipping the magnet from right to left, one must provide a (NV) amount of left polarized spins



$I = en/\tau$

Required number of electrons, n = $\frac{NV}{\eta}$

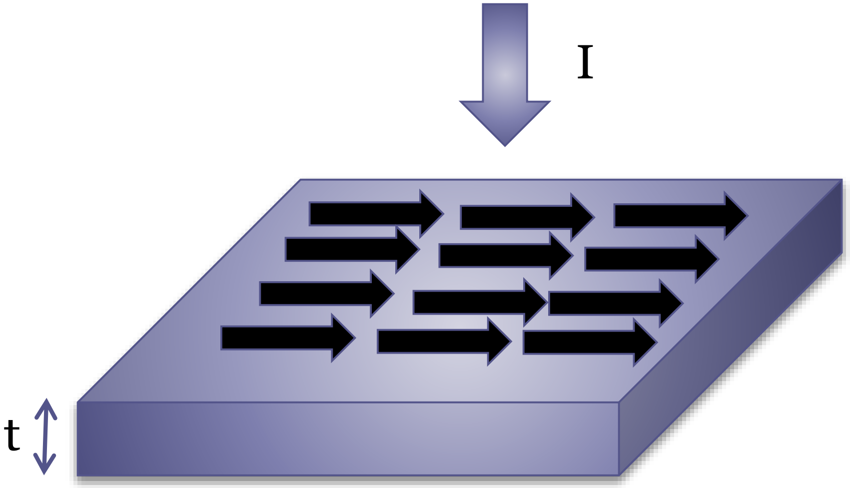
η is the spin polarization efficiency

$V = L^2 t$

Current needed to switch the magnet: $I = \frac{en}{\tau} = e\frac{N}{\eta}\left(\frac{1}{\tau}\right)(t)L^2$

Switching current *scales* with footprint area and thickness

Switching of the magnet by current



$$\frac{d\bar{m}}{dt} = -\gamma\bar{m} \times \overline{H_{eff}}$$

γ=gyromagnetic ratio=$g\mu_B/\hbar=2\mu_B/\hbar$

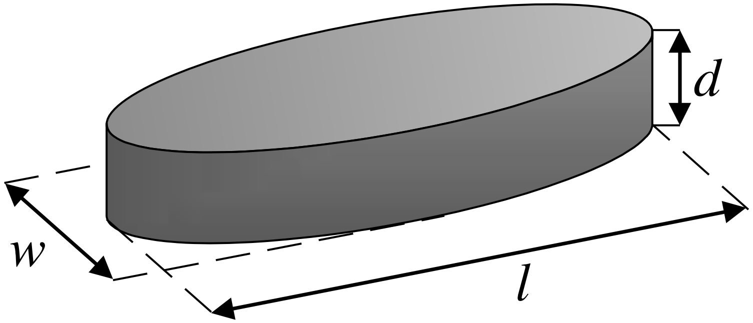
$H_{eff}=H_{anisotrpy}+2\pi M_s+H_{external}$

Characteristic time, τ, to switch a magnet:

$$\frac{1}{\tau} = \alpha\left(\gamma H_{eff}\right) \qquad ;\alpha=\text{damping}$$

$$= \alpha\left(g\mu_B / \hbar\right)\left(H_a + 2\pi M_s\right)$$

Scaling of Spin RAM



$$I = \frac{2e}{\hbar}\alpha\left(\frac{M_s}{\eta}\right)(H_a + 2\pi M_s)(t)LW$$

$$H_a \equiv H_k$$

Thermal stability : $\Delta = M_s H_k (t) LW$ Anisotropy field: $H_k \propto (t)\left(\frac{1}{W} - \frac{1}{L}\right)$

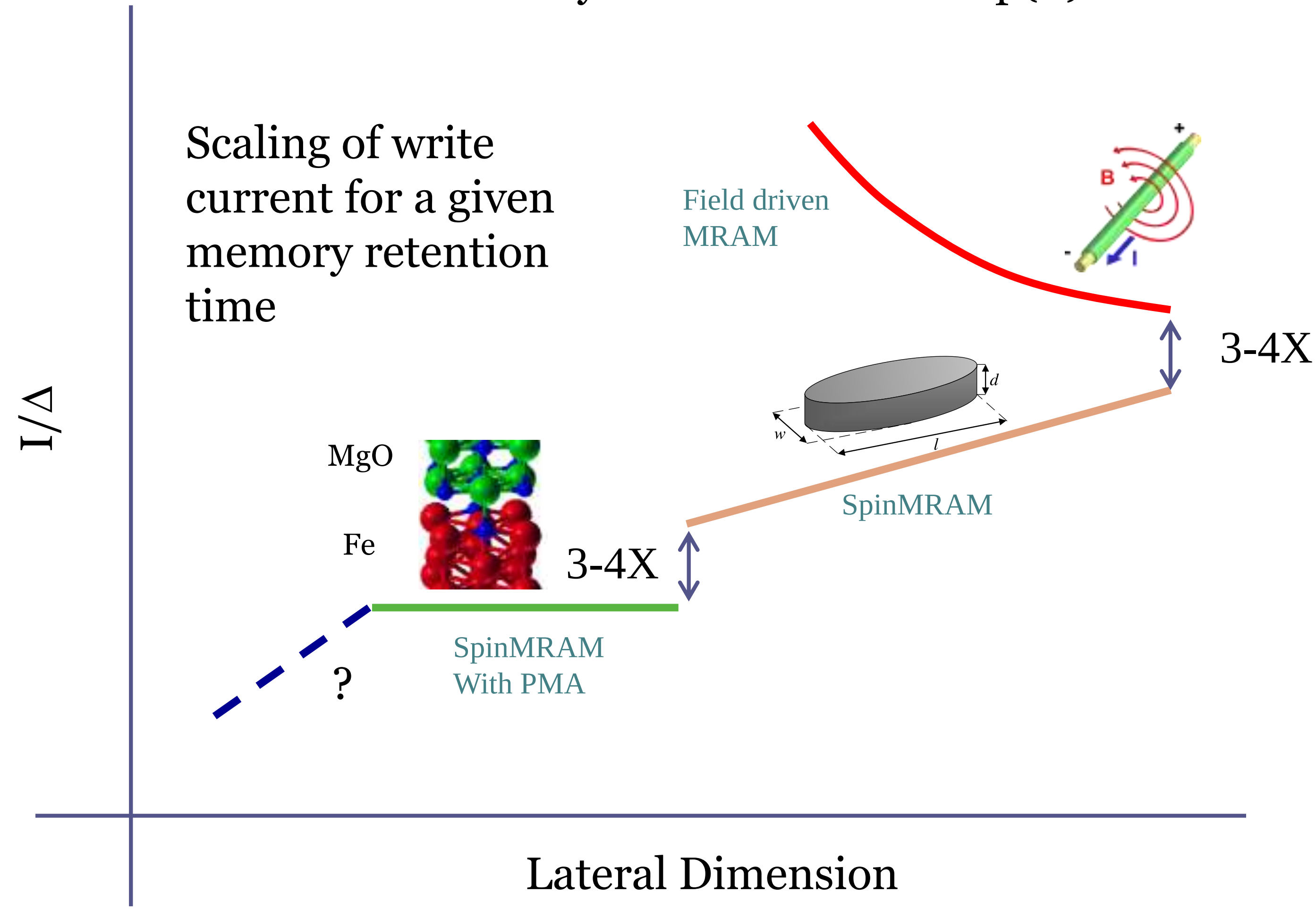
If L and W are both scaled by λ

$$\frac{I}{\Delta} \approx M_s \frac{\alpha}{\eta}\frac{WL}{(L-W)t}\left(1 + \frac{2\pi M_s}{H_k}\right) \sim \lambda$$

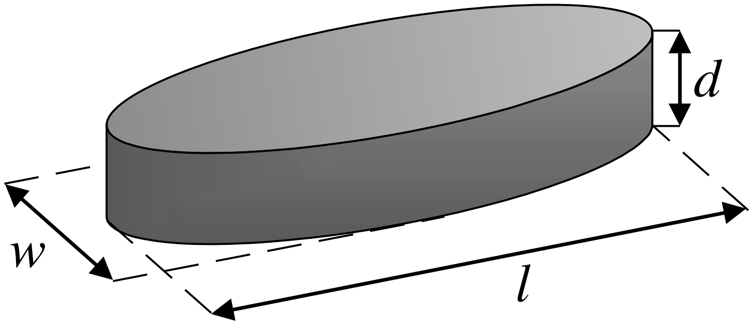
In-plane MRAM scales

MRAM scaling Trends

Memory retention time ~ exp(Δ)



Issues with in-plane MRAM



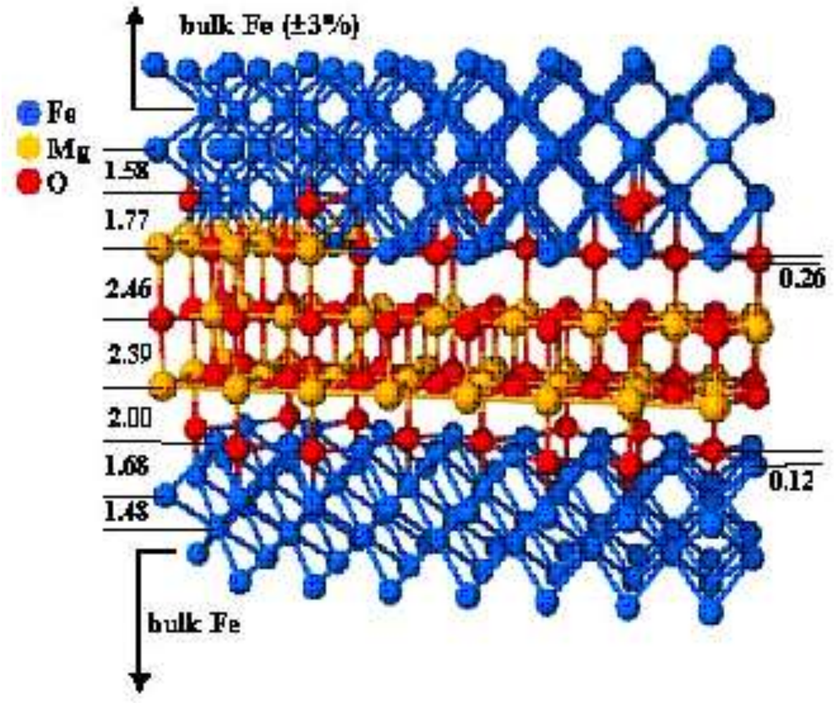
$$\frac{I}{\Delta} \approx M_s \frac{\alpha}{\eta} \frac{WL}{(L-W)t}\left(1+\frac{2\pi M_s}{H_k}\right)$$

H_k~100-200 Oe
M_s ~600-1200 Oe

$\left(1+\frac{2\pi M_s}{H_k}\right)$~20-50

- Difficult to keep reducing dimensions keeping intact the aspect ratio
- The area is always much larger compared to what is possible with minimum feature size.

Perpendicular STT MRAM



Objective is to reduce: $\left(1+\frac{2\pi M_s}{H_k}\right)$

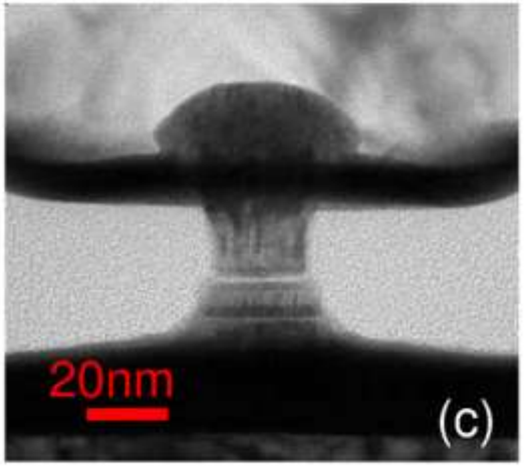
$If\ a\ crystalline\ anisotropy\ can\ be\ created$

$$H_{net} = 2\pi \mathrm{M}_s - H_{PMA} \sim 0$$

the switching current can be 10X lower

- In a Fe-MgO interface Fe-O bonds can provide the required crystalline anisotropy
- There is no reason to have shape anisotropy any more, the magnets can be circular

Current drops but scaling goes away

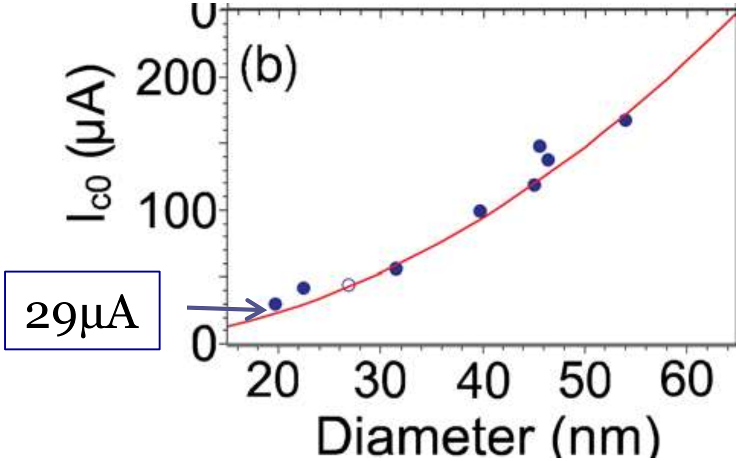


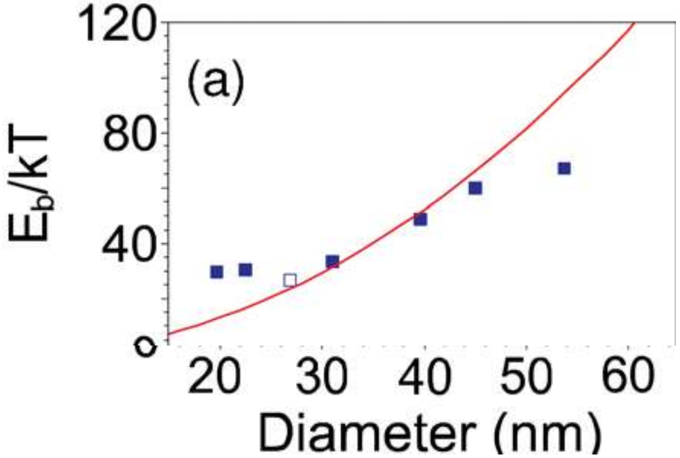
$$I = \frac{2e}{\hbar}\alpha\left(\frac{M_s}{\eta}\right)(H_a + 2\pi M_s)(t)LW \textit{ now becomes}$$

$$I = \frac{2e}{\hbar}\alpha\left(\frac{M_s}{\eta}\right)(2\pi M_s - H_{PMA})(t)LW$$

$$\Delta = M_s(2\pi M_s - H_{PMA})(t)LW$$

$$\frac{I}{\Delta} \approx \frac{\alpha}{\eta}$$ Scaling dependence is gone





M Gajek et. al, APL, 100, 132408 (2012)

Scaling Challenges

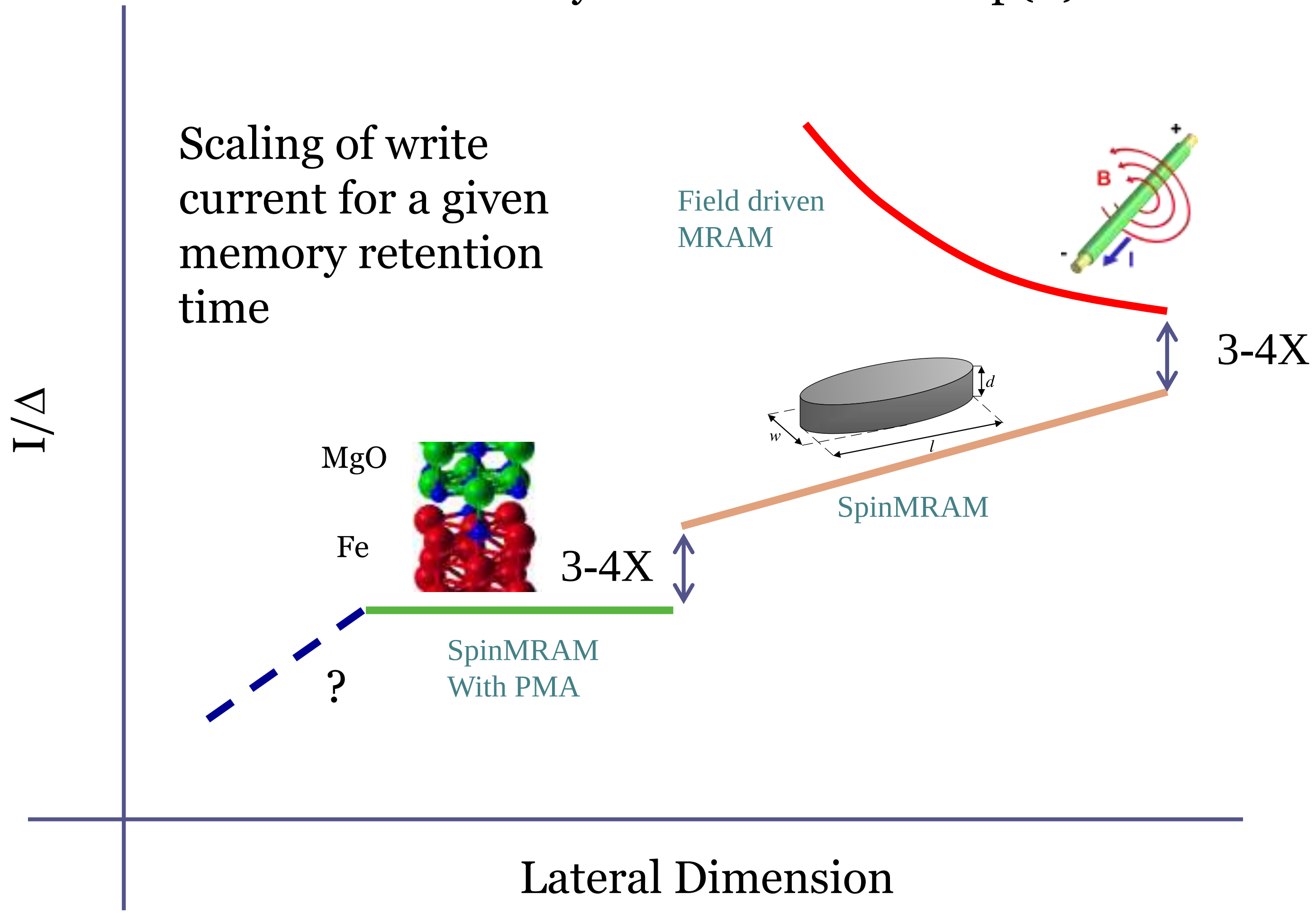
$$\frac{I}{\Delta} \approx \frac{\alpha}{\eta} \qquad \Delta = M_s(2\pi M_s - H_{PMA})(t)LW$$

- To decrease current with scaling
  - decrease α
  - increase η

- To keep the memory
  - increase $(2\pi M_s - H_{PMA})$

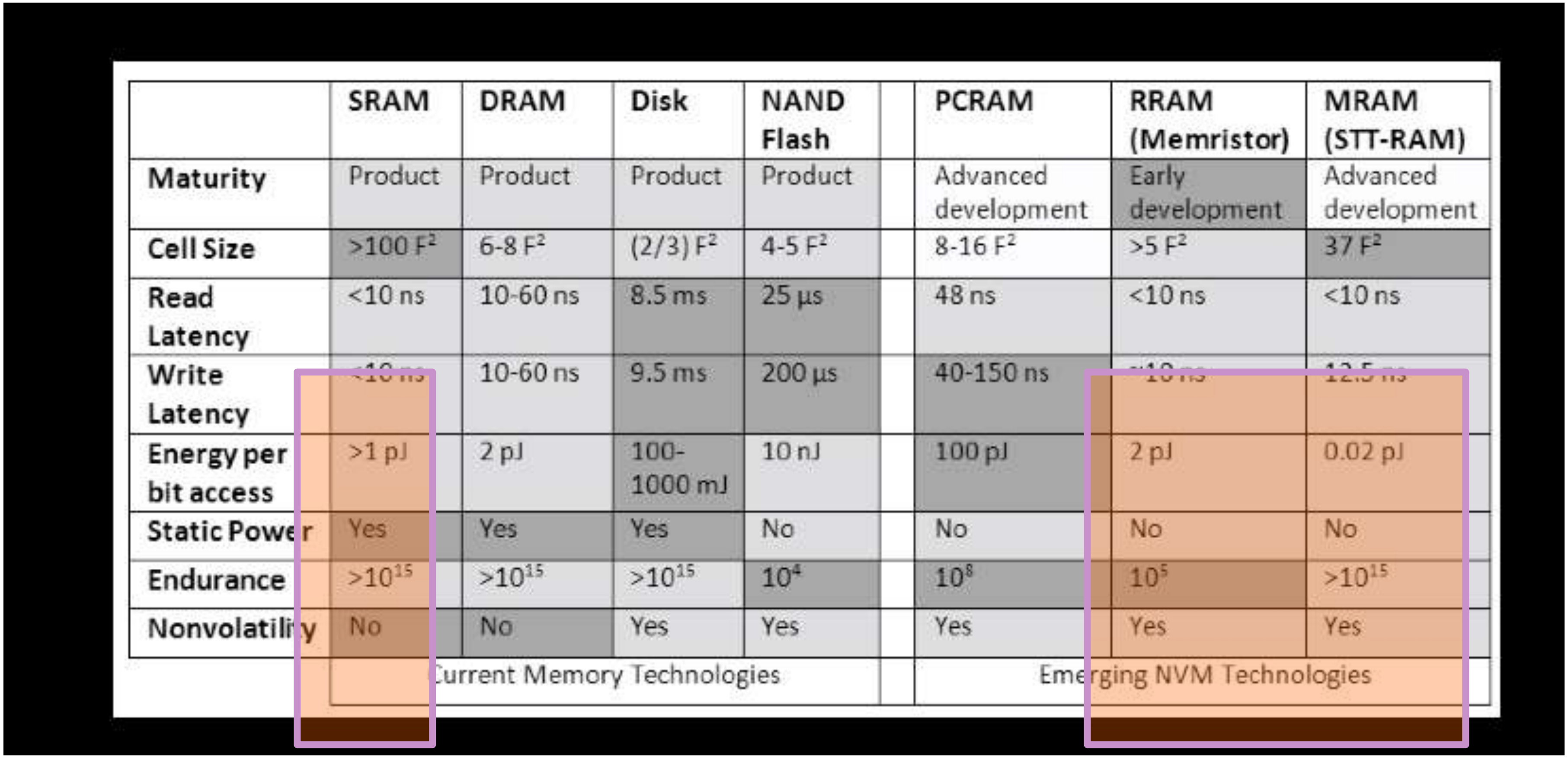
MRAM scaling Trends

Memory retention time ~ exp(Δ)



Scaling of write current for a given memory retention time

Competition

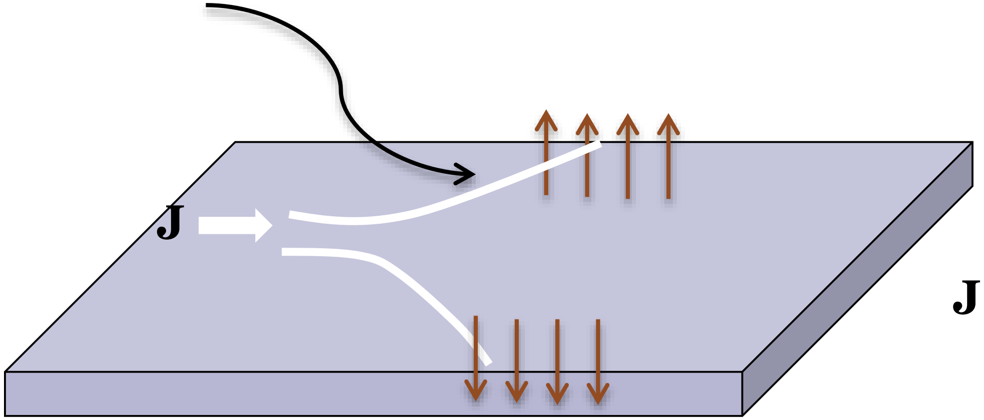


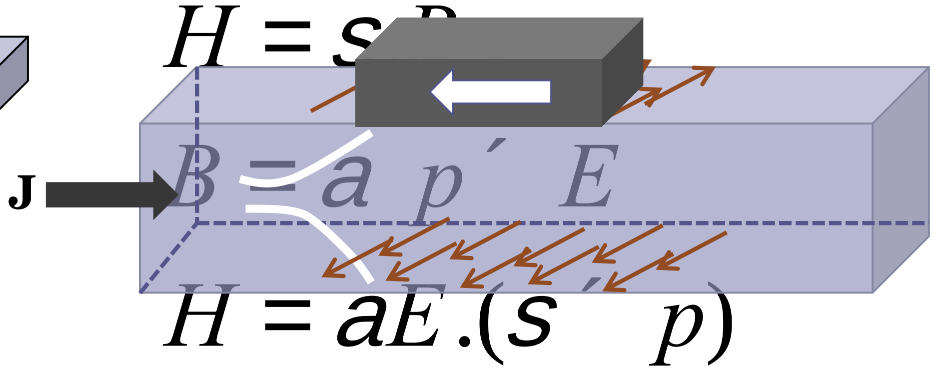
	SRAM	DRAM	Disk	NAND Flash	PCRAM	RRAM (Memristor)	MRAM (STT-RAM)
Maturity	Product	Product	Product	Product	Advanced development	Early development	Advanced development
Cell Size	>100 F^2	6-8 F^2	(2/3) F^2	4-5 F^2	8-16 F^2	>5 F^2	37 F^2
Read Latency	<10 ns	10-60 ns	8.5 ms	25 µs	48 ns	<10 ns	<10 ns
Write Latency	<10 ns	10-60 ns	9.5 ms	200 µs	40-150 ns	<10 ns	12.5 ns
Energy per bit access	>1 pJ	2 pJ	100-1000 mJ	10 nJ	100 pJ	2 pJ	0.02 pJ
Static Power	Yes	Yes	Yes	No	No	No	No
Endurance	$>10^{15}$	$>10^{15}$	$>10^{15}$	10^4	10^8	10^5	$>10^{15}$
Nonvolatility	No	No	Yes	Yes	Yes	Yes	Yes
	Current Memory Technologies				Emerging NVM Technologies		

- Slightly slower than SRAM
- **Much** smaller than SRAM
- Same endurance

Spin orbit coupling to generate spin current

Spin current flowing towards the edge





$H = s.B$

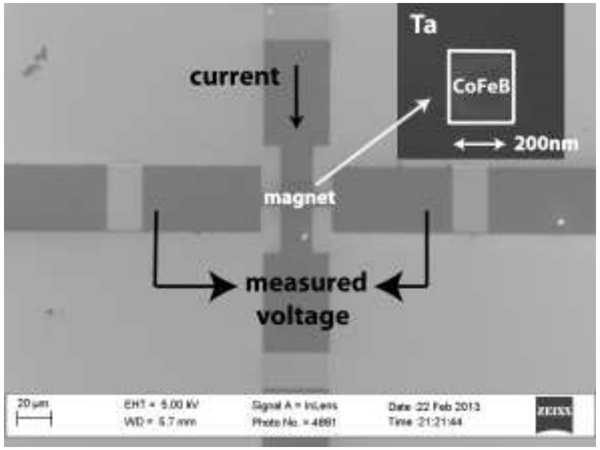
$B = a\, p \times E$

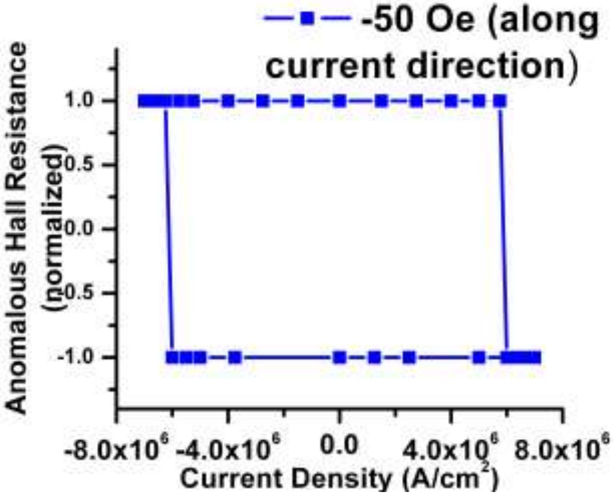
$H = aE.(s \times p)$

<u>Prediction:</u> Dyakonov and Perrel (1971)
<u>Experimental demonstration</u>: D Awschalom group in 2D electron gas of GaAs (2004)

Miron et. al. Nature Materials, 2010, Suzuki et.al., APL, 98, 142505, 2011,Ryu et. al., Nat Nano, 2012, Liu et. al., Science, 2012, Bhowmik et. al., IEDM, 2012

Spin orbit generated spin current

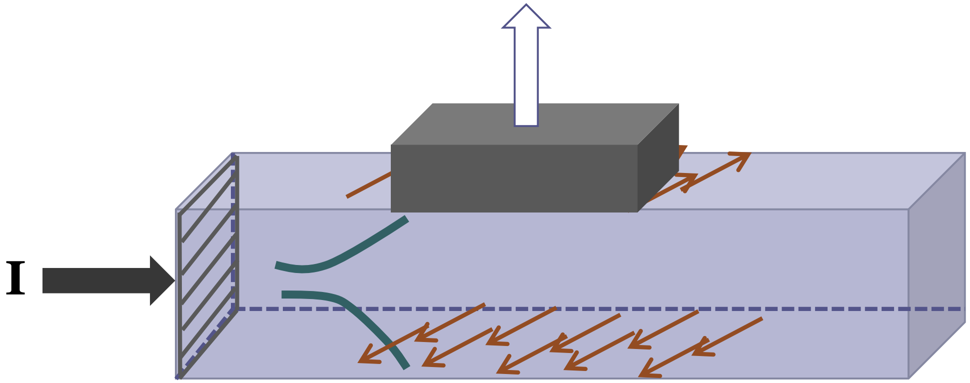




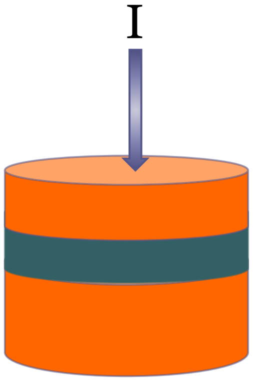
Bhowmik, You and Salahuddin, Nature Nanotechnology,9,59 (2014)

Potential reduction in WRITE current

Spin Orbit Torque Device



Spin Transfer Torque Device



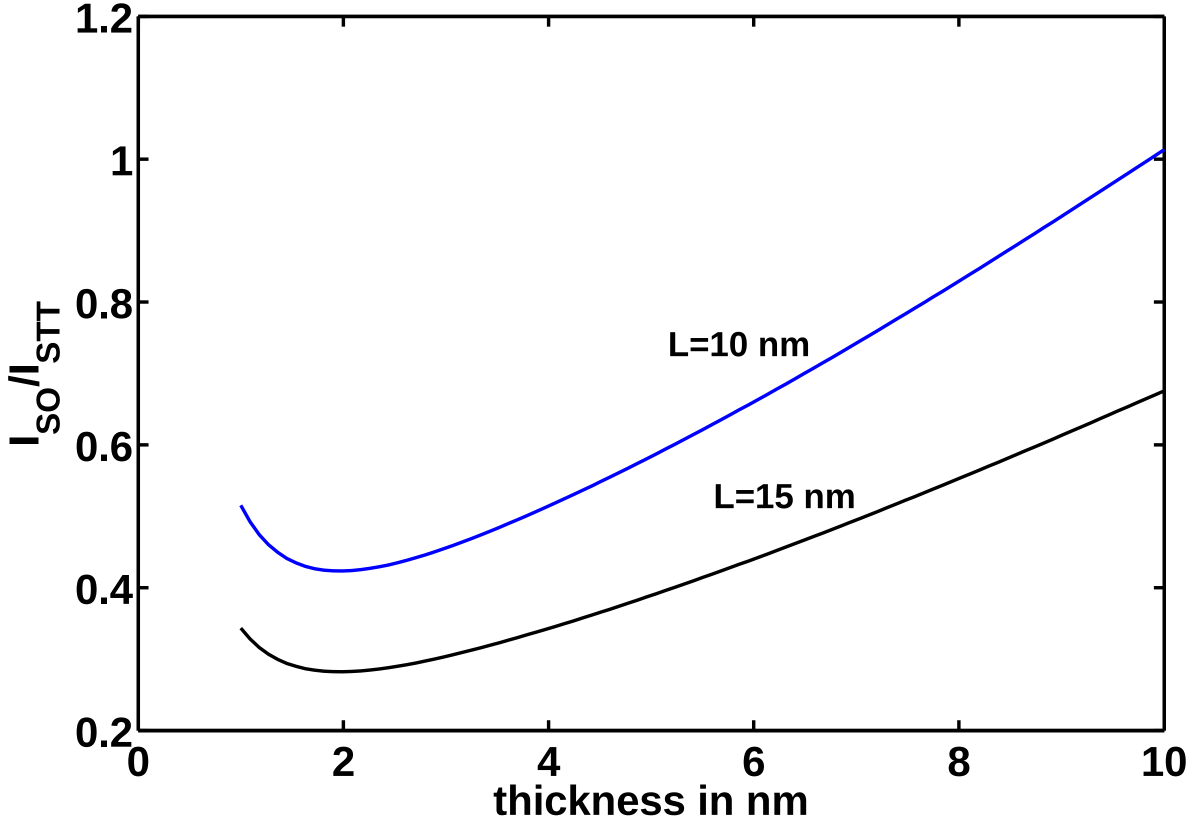
$I \propto$ Width of the magnet x length of the magnet

$I \propto$ Width of the magnet x ***thickness of the wire***

So a potential reduction of (L/t) is possible

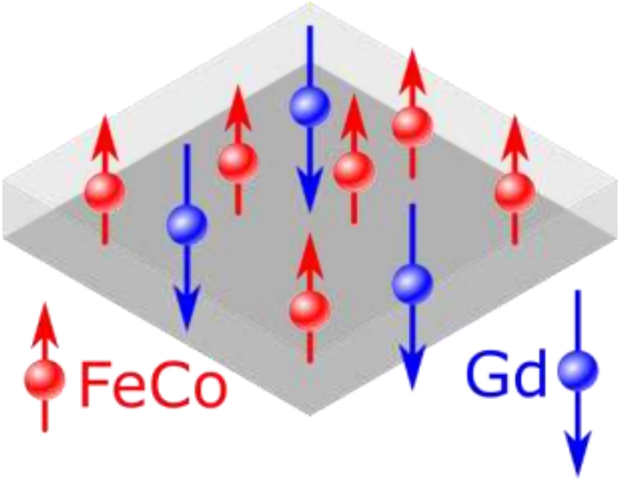
Potential Reduction in Current

$$\frac{I_{SO}}{I_{STT}} = \frac{t}{L}\,\frac{h}{q_{so}}\left[1+\frac{\cos ech(t/l)}{\tanh(t/2l)}\right] \qquad \lambda\sim 2\text{nm}; \qquad \frac{h}{q_{so}}\sim 1$$



3-4X decrease in current is possible

Thickness dependence of SOT



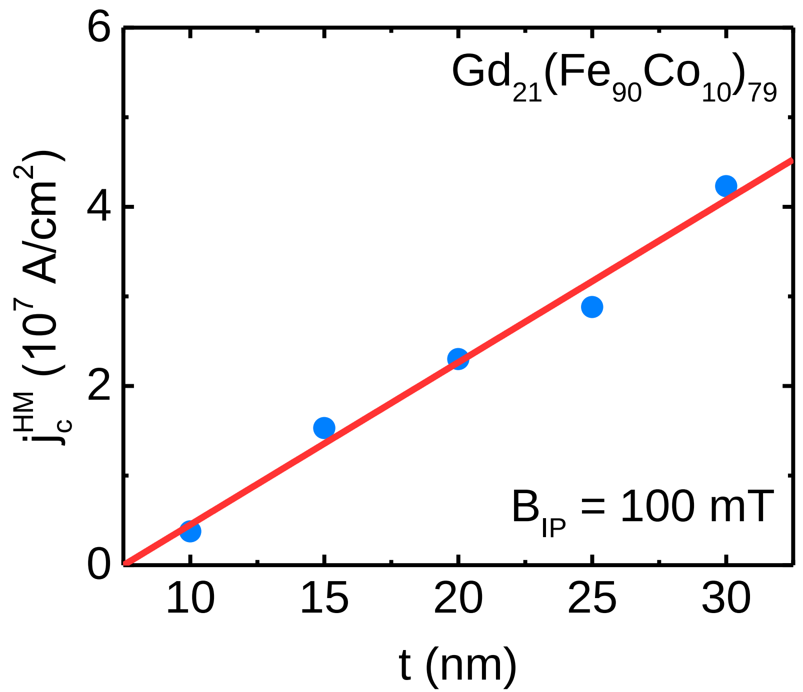
- GFC is a Bulk PMA material

- Thermal stability can be retained by increasing thickness, unlike interfacial PMA, when the areal footprint is scaled → can be very important for ultra scaled memory technologies

- Combined with lower current for SOT this could help resolve the scaling issue

But can we switch a large thickness GFC with SOT?

Scaling Trends Summary



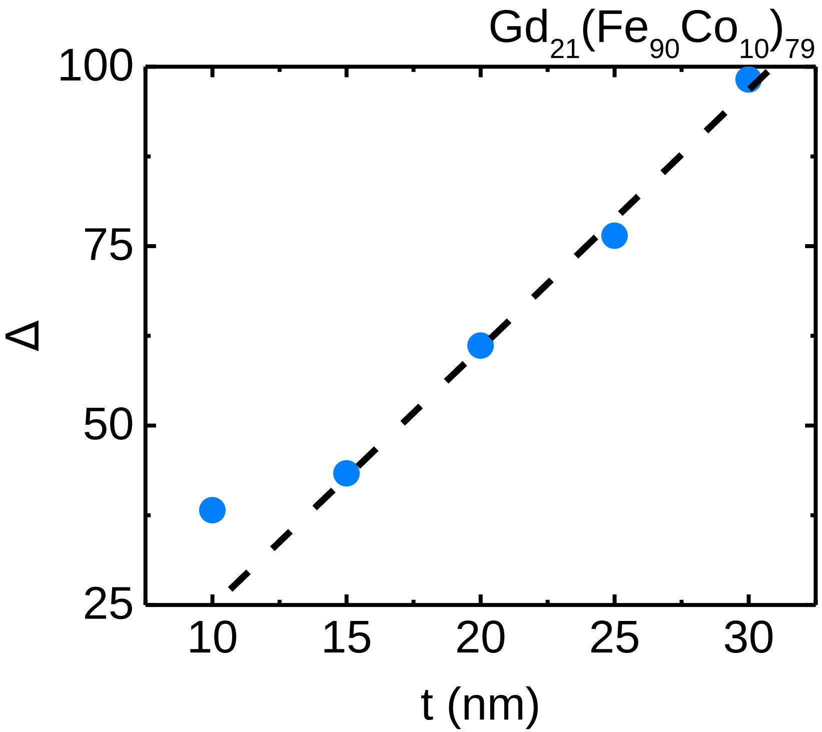
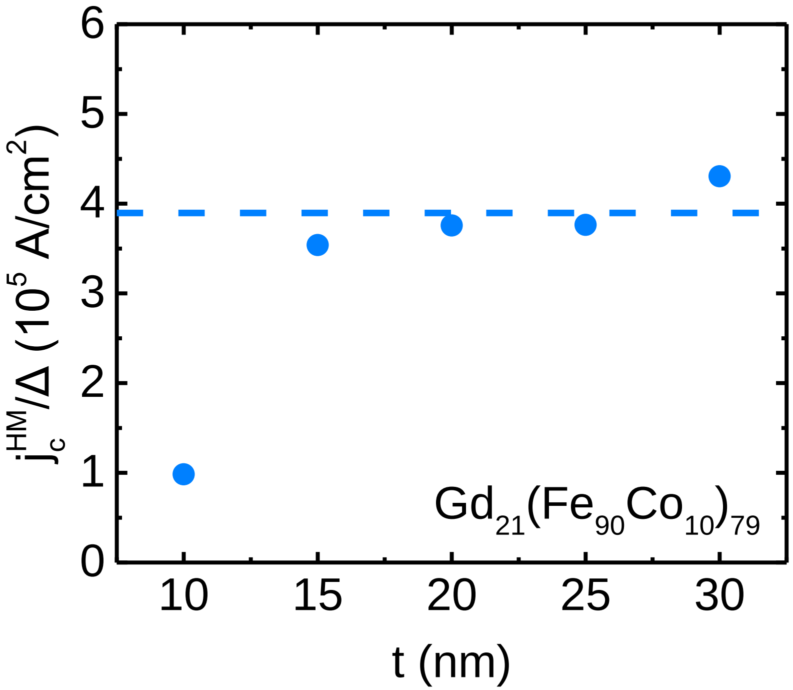
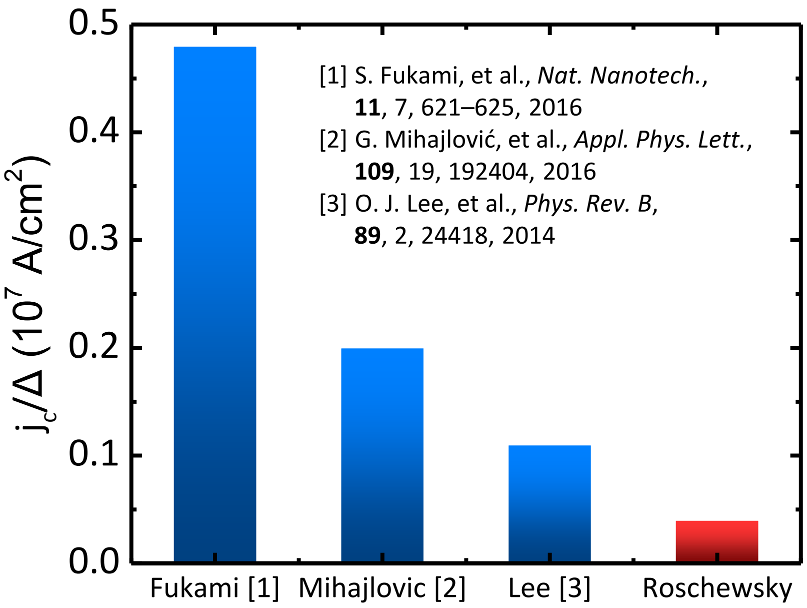


Figure of Merit of SOT Switching





[1] S. Fukami, et al., *Nat. Nanotech.*, **11**, 7, 621–625, 2016

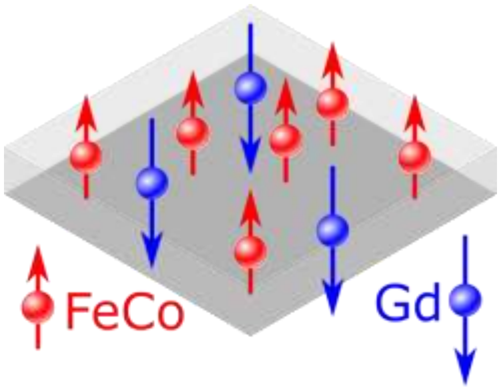
[2] G. Mihajlović, et al., *Appl. Phys. Lett.*, **109**, 19, 192404, 2016

[3] O. J. Lee, et al., *Phys. Rev. B*, **89**, 2, 24418, 2014

Very high switching efficiency in ferrimagnetic GFC

(Rochewsky. C-H, Lambert et al, PRB, 2017)

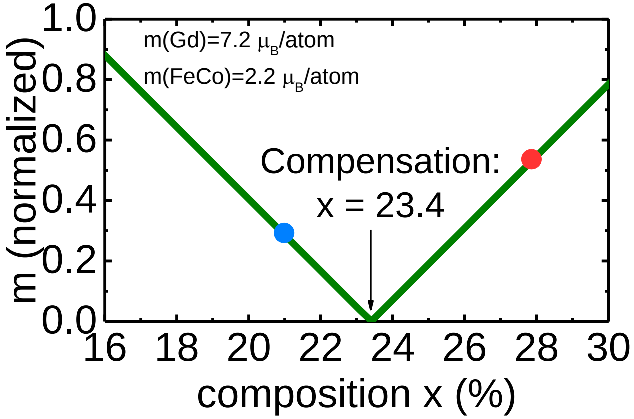
Spin Orbit Torque in Ferrimagnetic $Gd_x(Fe_{90}Co_{10})_{1-x}$

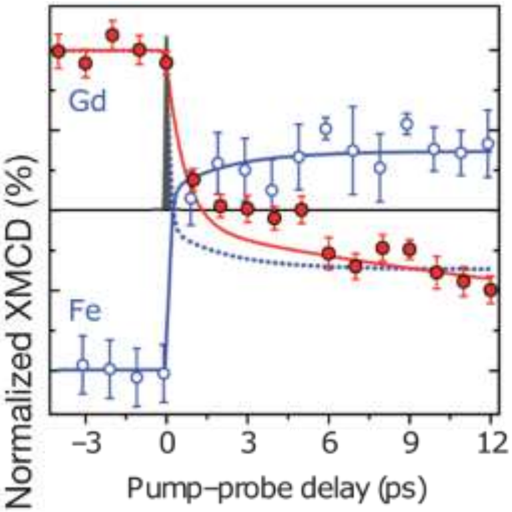


$Gd_x(Fe_{90}Co_{10})_{1-x}$

J Bokor group

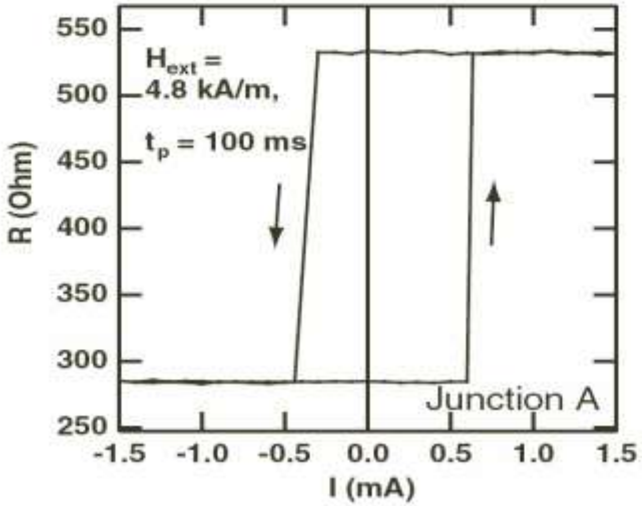
Gorchon et al, arxiv:1702.08492
Yang et al. arxiv: 1609.06392
Wilson et al. arxiv: 1609.05155



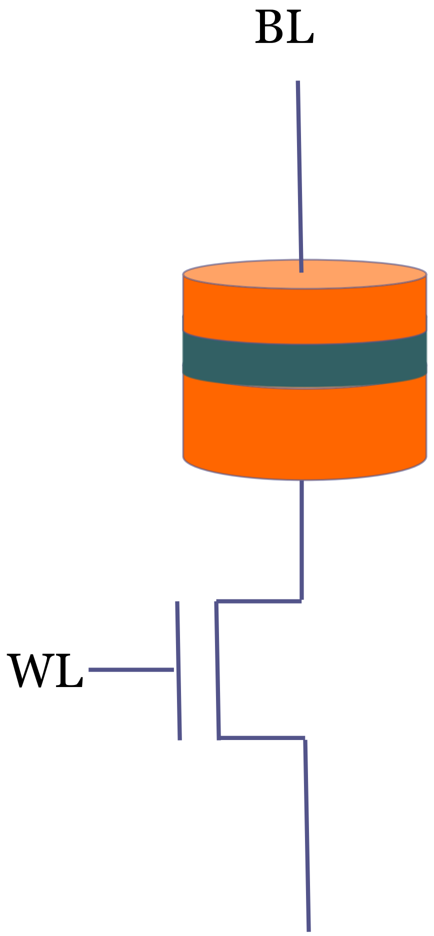


Radu et al., *Nature* 2011.

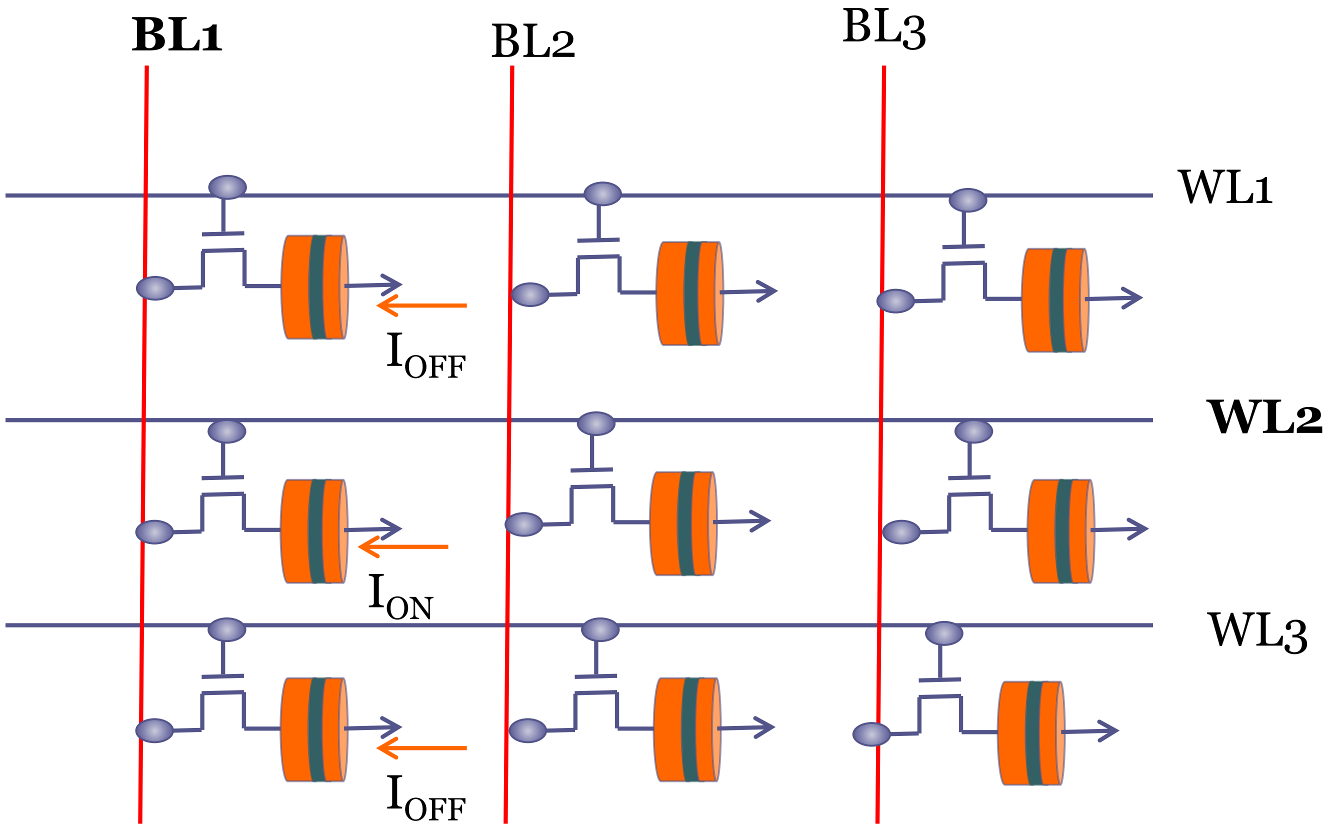
Array Issues due to low MR: Competition



Kubota et. al., JJAP, 44, 40, 1237,2005



Array Issues: Competition



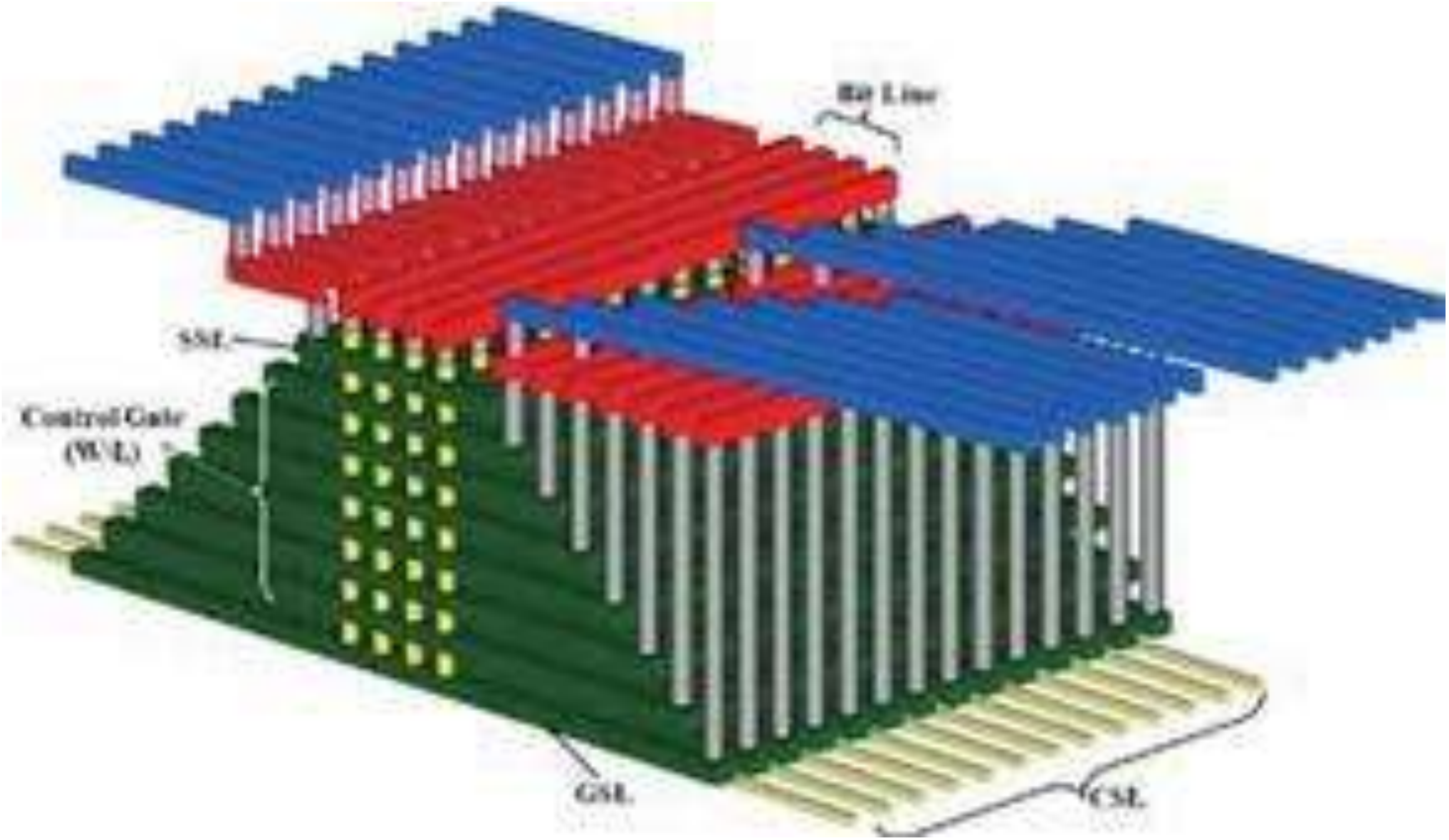
Total current flowing in BL1= ION+n*IOFF; n is number of bits in the BL

So, if ION/IOFF~2, one cannot have more than 3 elements without a transistor

Due to the need of a transistor, currently STTRAM cannot be integrated in 3D

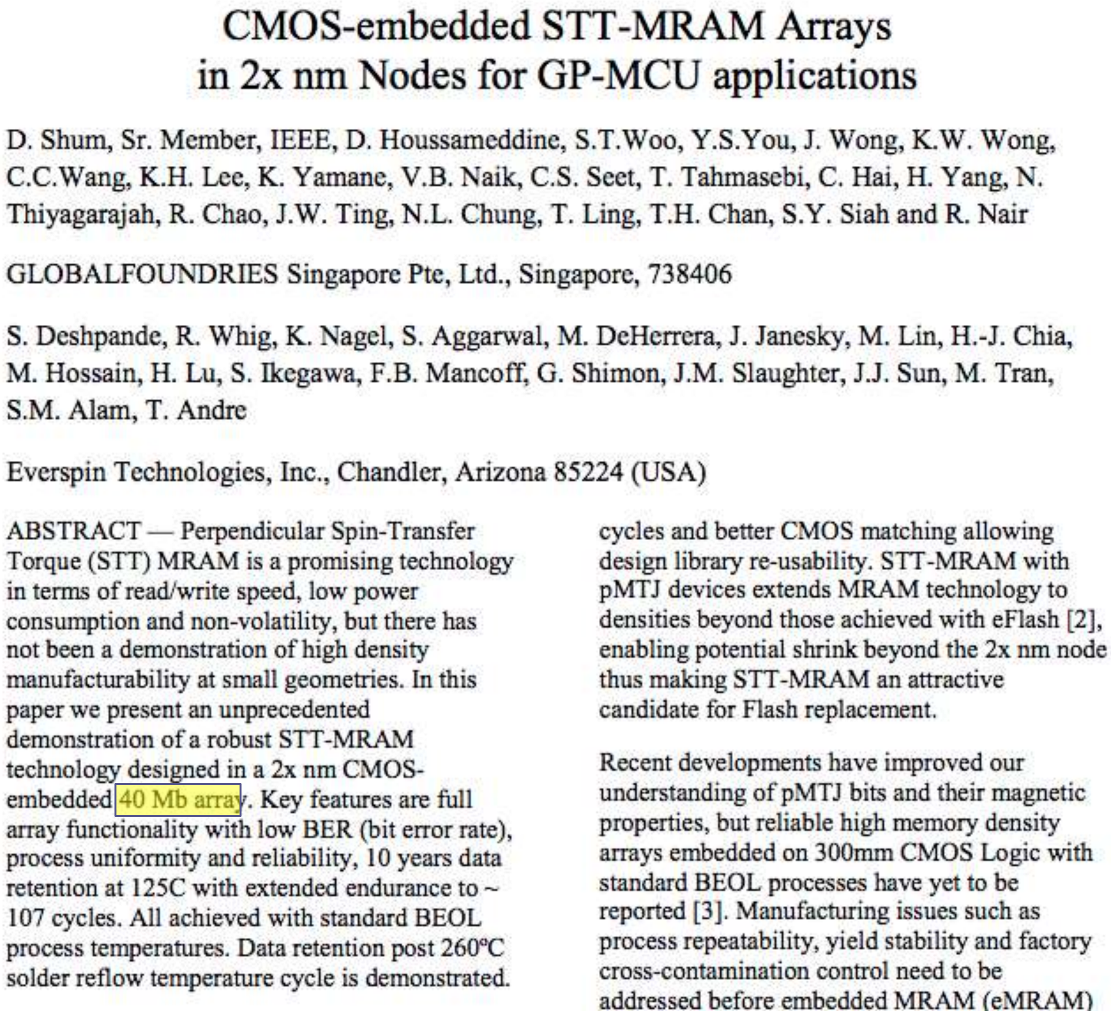
Competition

3D NAND FLASH



STT is not competitive in stand alone high density data storage

Challenges for STT MRAM as an embedded memory



CMOS-embedded STT-MRAM Arrays in 2x nm Nodes for GP-MCU applications

D. Shum, Sr. Member, IEEE, D. Houssameddine, S.T.Woo, Y.S.You, J. Wong, K.W. Wong, C.C.Wang, K.H. Lee, K. Yamane, V.B. Naik, C.S. Seet, T. Tahmasebi, C. Hai, H. Yang, N. Thiyagarajah, R. Chao, J.W. Ting, N.L. Chung, T. Ling, T.H. Chan, S.Y. Siah and R. Nair

GLOBALFOUNDRIES Singapore Pte, Ltd., Singapore, 738406

S. Deshpande, R. Whig, K. Nagel, S. Aggarwal, M. DeHerrera, J. Janesky, M. Lin, H.-J. Chia, M. Hossain, H. Lu, S. Ikegawa, F.B. Mancoff, G. Shimon, J.M. Slaughter, J.J. Sun, M. Tran, S.M. Alam, T. Andre

Everspin Technologies, Inc., Chandler, Arizona 85224 (USA)

ABSTRACT — Perpendicular Spin-Transfer Torque (STT) MRAM is a promising technology in terms of read/write speed, low power consumption and non-volatility, but there has not been a demonstration of high density manufacturability at small geometries. In this paper we present an unprecedented demonstration of a robust STT-MRAM technology designed in a 2x nm CMOS-embedded 40 Mb array. Key features are full array functionality with low BER (bit error rate), process uniformity and reliability, 10 years data retention at 125C with extended endurance to ~ 107 cycles. All achieved with standard BEOL process temperatures. Data retention post 260°C solder reflow temperature cycle is demonstrated.

cycles and better CMOS matching allowing design library re-usability. STT-MRAM with pMTJ devices extends MRAM technology to densities beyond those achieved with eFlash [2], enabling potential shrink beyond the 2x nm node thus making STT-MRAM an attractive candidate for Flash replacement.

Recent developments have improved our understanding of pMTJ bits and their magnetic properties, but reliable high memory density arrays embedded on 300mm CMOS Logic with standard BEOL processes have yet to be reported [3]. Manufacturing issues such as process repeatability, yield stability and factory cross-contamination control need to be addressed before embedded MRAM (eMRAM)

Conclusion

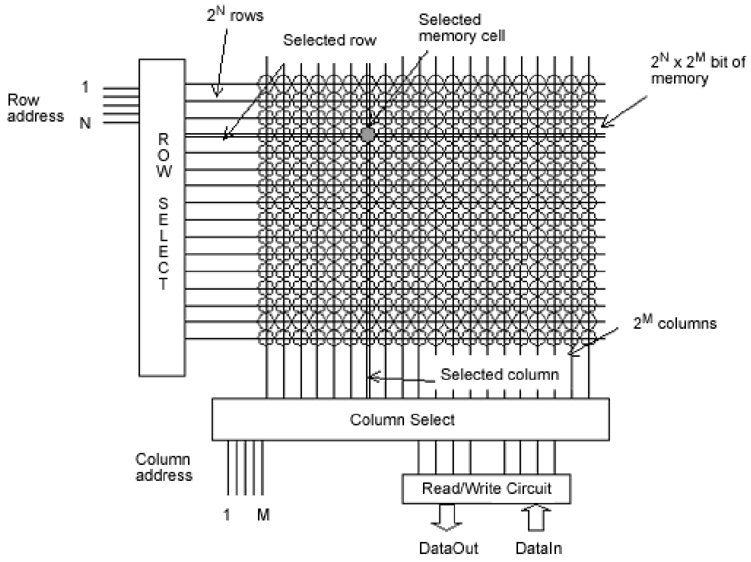
- STT RAM allows achieving magnetic storage on-chip by enabling operation without a magnetic field
- The combination of high speed switching and high endurance is unique among known non-volatile technologies → thereby an enabler for applications that need those properties.
- Scaling below 20 nm currently faces significant challenges.

Array issues

The probability that an error has occurred after a time t can be written as:

$$p(t)=1-e^{-t/t} \quad \text{where} \quad t=t_0e^{D}$$

Application designers will set a allowed rate of errors after a given time, t=t$_{s.}$ Say, it is decided that if the number of bits in the array is $\boldsymbol{N_B}$, only $\boldsymbol{m}$ number of bits are allowed to be erroneous after $\boldsymbol{t_s}$. Then



$$p(t_s)=1-e^{-t_s/t}=\frac{m}{N_B}$$

$$\Rightarrow e^{-t_s/t_0e^{D}}=1-\frac{m}{N_B}$$

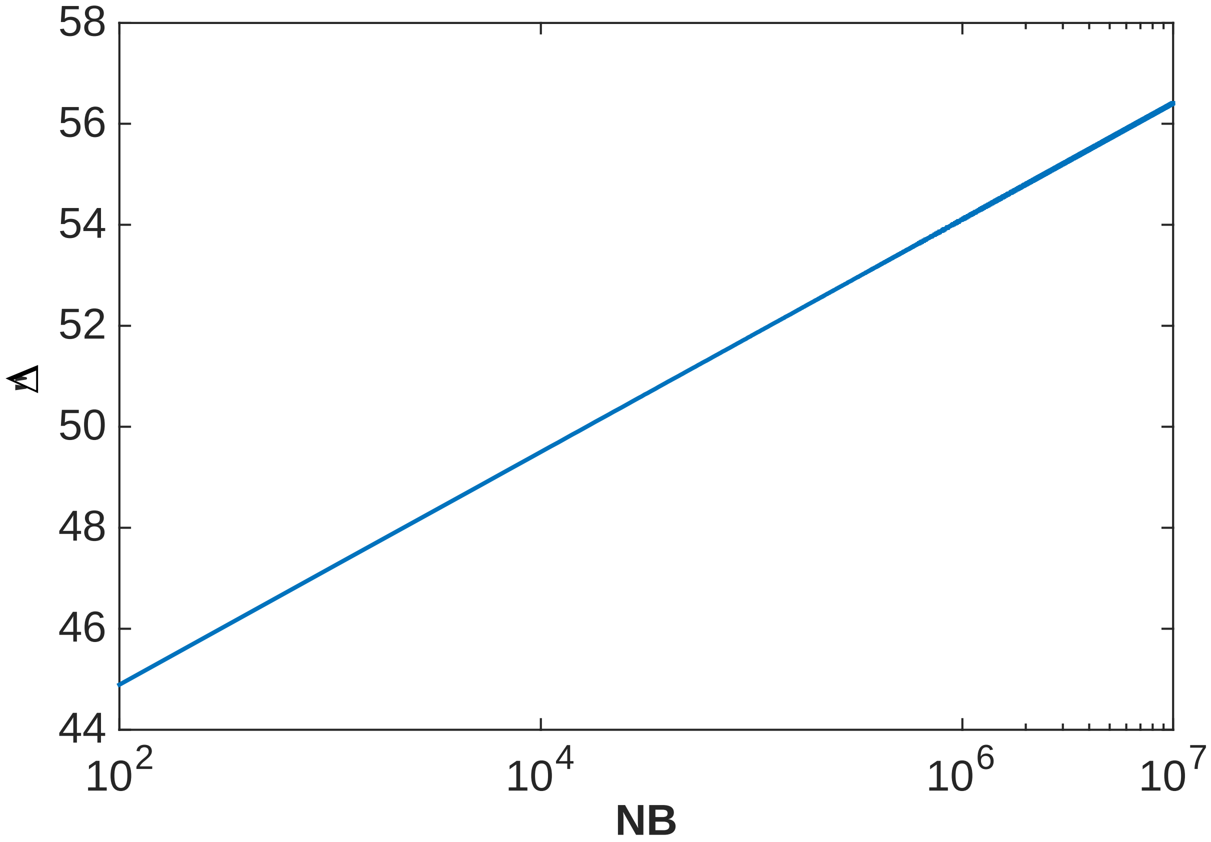
$$\Rightarrow e^{-D}t_s/t_0=-\log\left(1-\frac{m}{N_B}\right)$$

$$\boxed{D=-\log\left[-\frac{t_0}{t_s}\log\left(1-\frac{FIT}{N_B}\right)\right]}$$

$$m\,^{\circ}\ FIT(\text{failure in time})$$

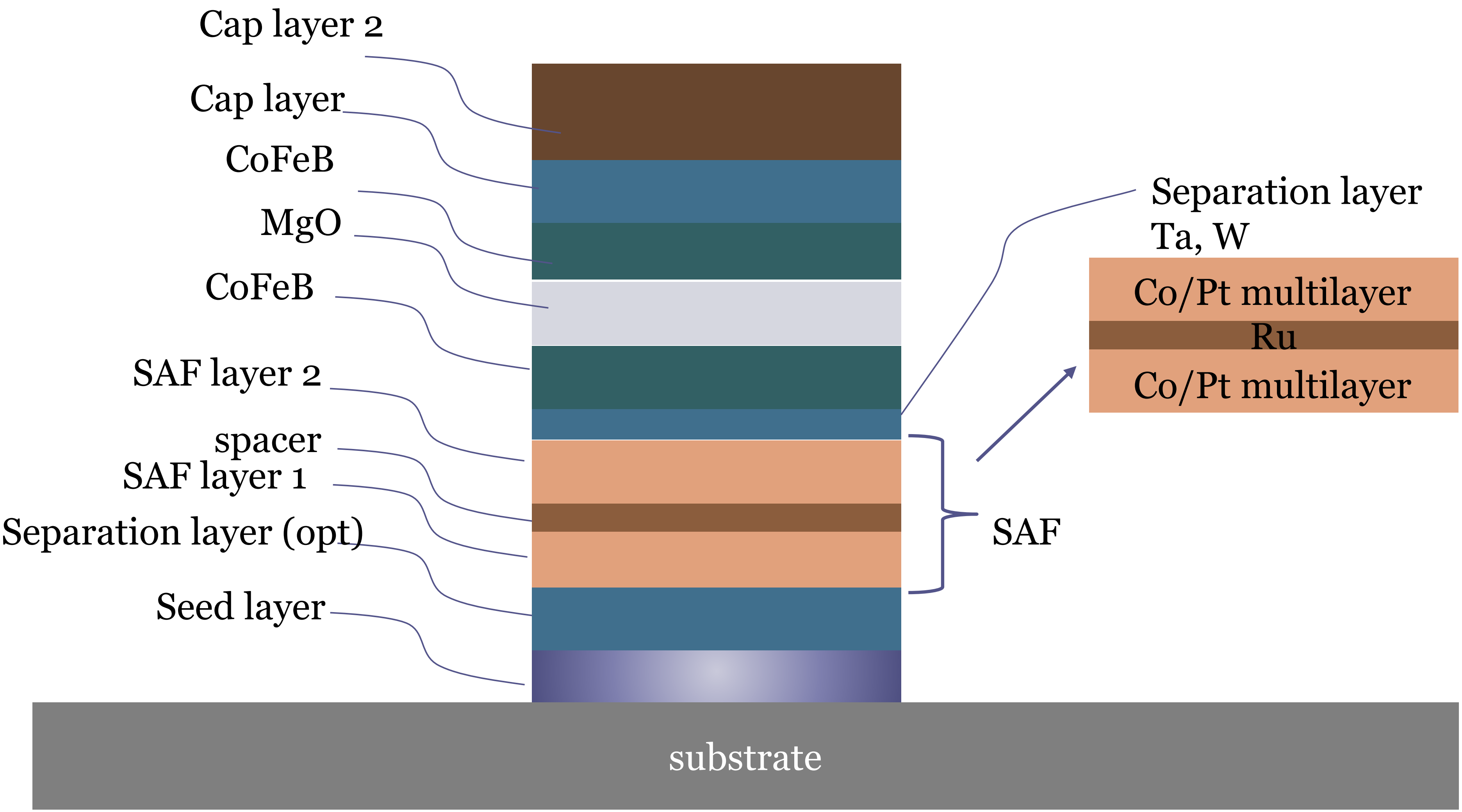
Array issues

$$D = -\log\left[-\frac{t_0}{t_s}\log\left(1 - \frac{FIT}{N_B}\right)\right]$$



Δ=70 is a good number

A typical MTJ stack



Etch issues for high density MRAM

Physical etching creates sidewalls