Novel Nonvolatile Memory Hierarchies to Realize "Normally-Off Mobile Processors"

ASP-DAC 2014

Shinobu Fujita, Kumiko Nomura, Hiroki Noguchi,
Susumu Takeda , Keiko Abe

Toshiba Corporation, R&D Center
Advanced LSI technology laboratory

Acknowledgement

This work was partly supported by **Normally-off Computing PJ** (NEDO) in Japan.

OUTLINE

- Introduction:

  Normally-off (**N-off**) Processor (from ver.0 to ver.1.)

- Key Point 1: Advanced STT-MRAM

- Key Point 2: Decrease in power for short CPU standby state by applying new memory cell design

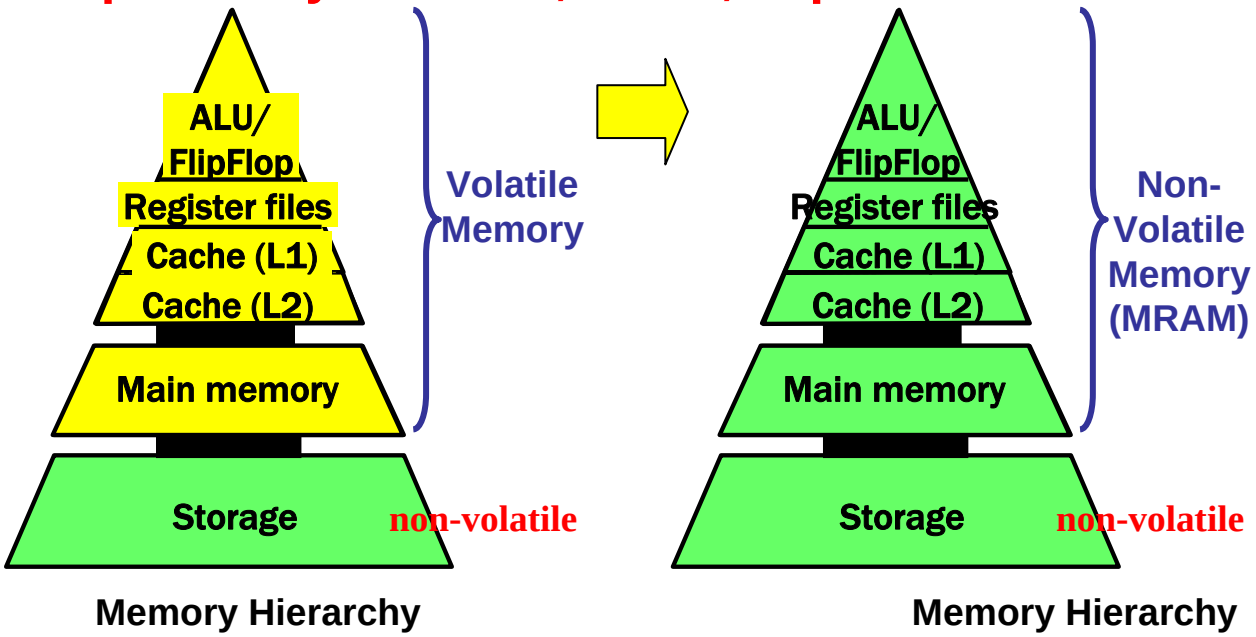
- Key Point 3: Power Decrease for long CPU standby state by Ultra-Fast- Power Gating

- Conclusions Towards N-off ver 2.

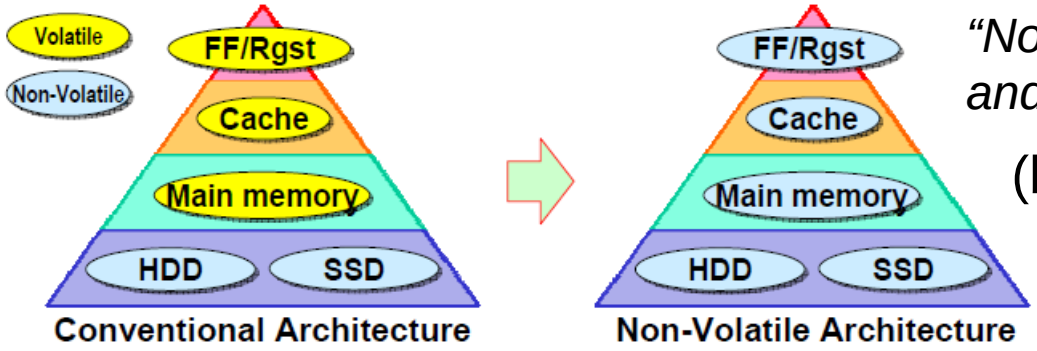
History of Concept on Normally-Off Computer

Normally-Off Computer Ver.0（2001）

Proposed by K. Ando, AIST, Japan (FED journal Japan, 2001)



The same Ver.0 concept presented by T. Kawahara, ASP-DAC 2011.

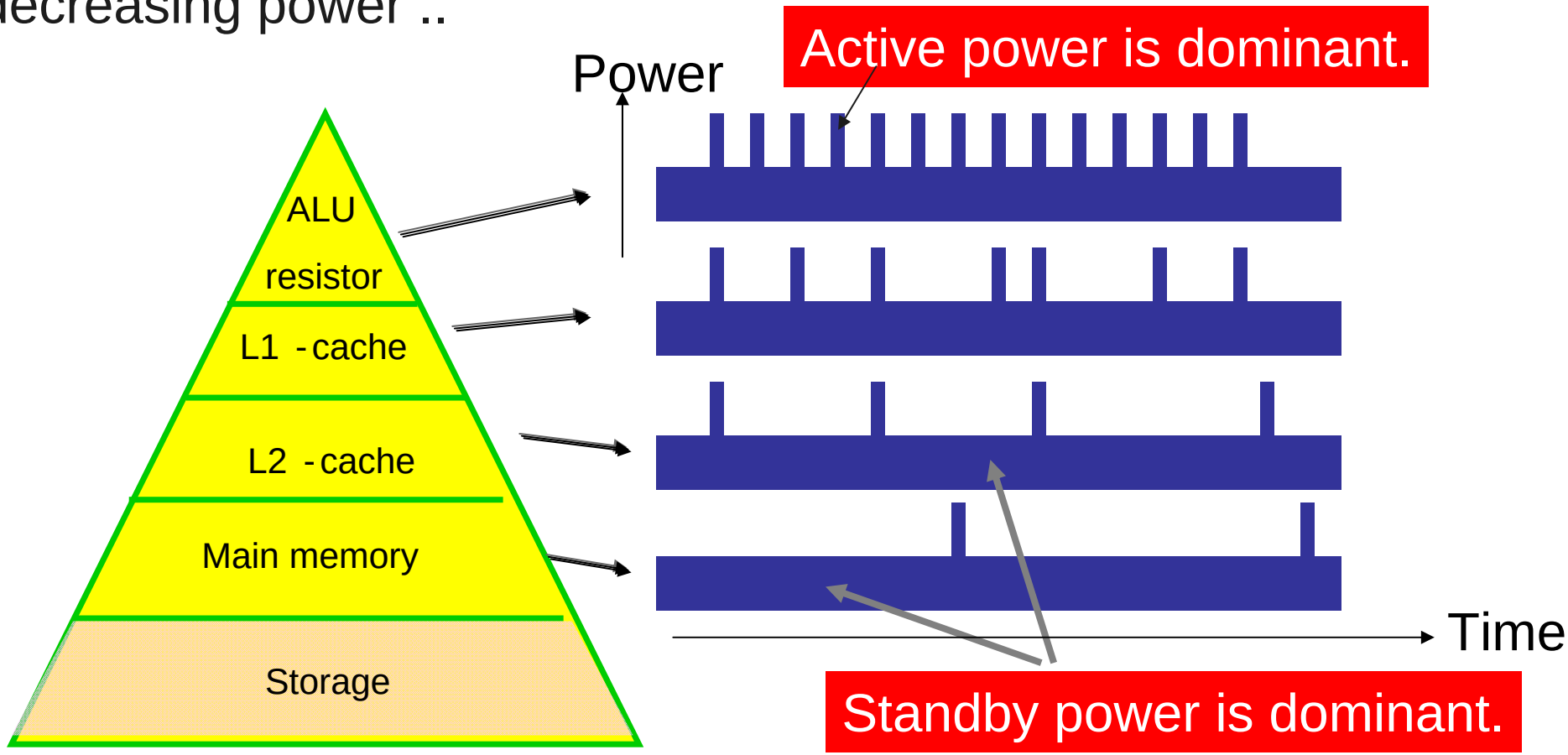


"Nonvolatile memory and normally-off computing"

(based on MRAM)

Rethink Normally-off Concept Ver.0

Ver.0 (All Nonvolatile Memory Hierarchy) is not suitable for decreasing power ..



Attention:

-Active power (write power) of nonvolatile memory is so large!

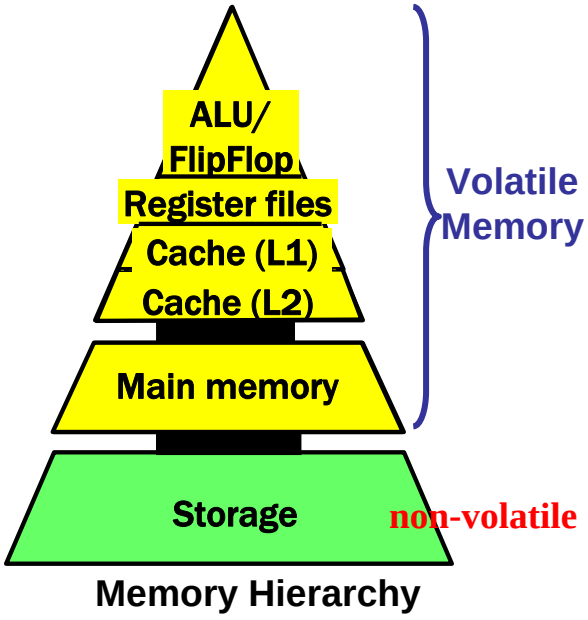
-Speed of NV-Memory is much slower than that of SRAM.

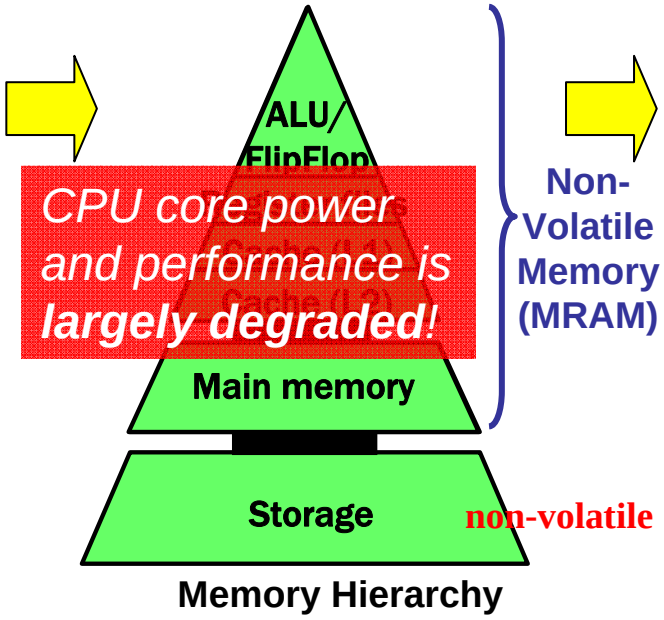
*(CPU core power and performance is **largely degraded** by Ver.0!)*

History of Concept on Normally-Off Computer (2)

Normally-Off Computer Ver.0（2001）

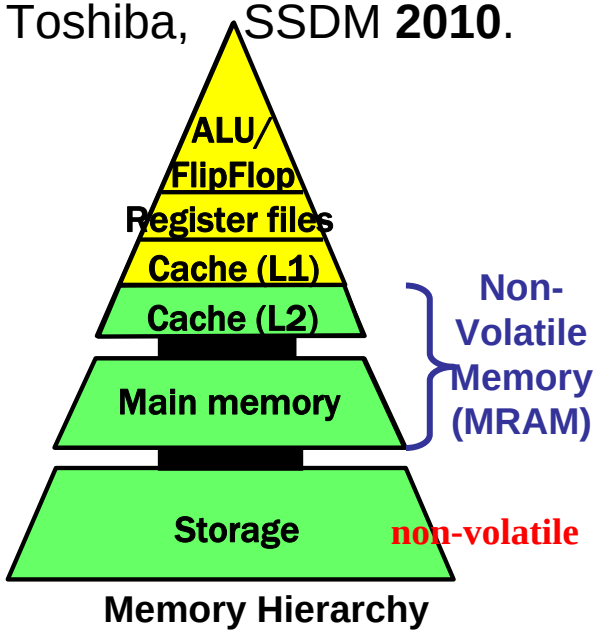
K. Ando, AIST, Japan (FED journal Japan, 2001)



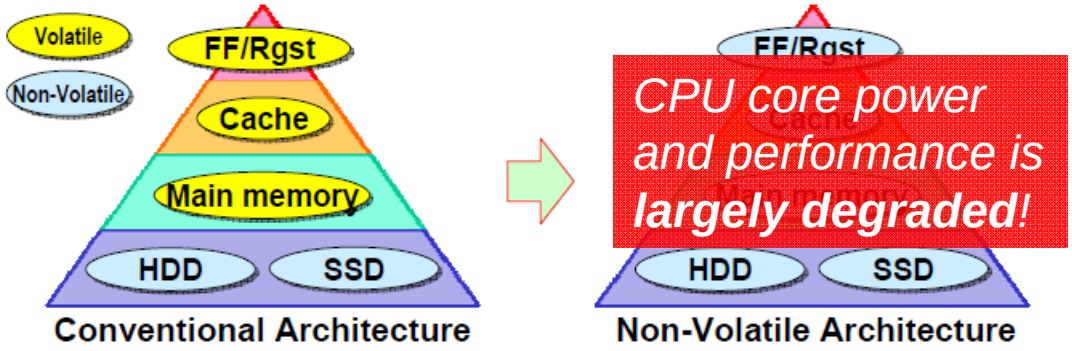


Ver.1（2010）

K. Abe, S. Fujita et al.,
Toshiba, SSDM **2010**.



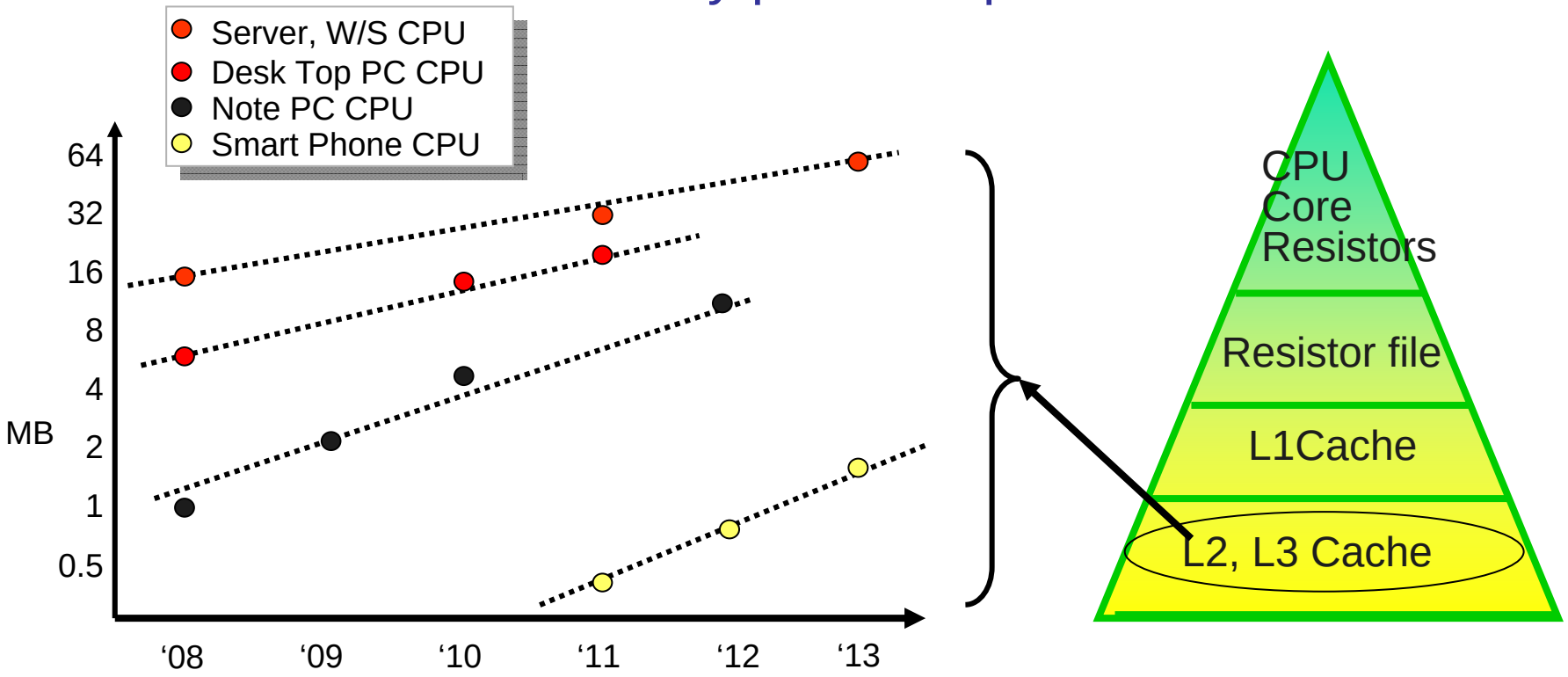
Ver.0 (2011) T. Kawahara, ASP-DAC 2011.



Ultra low power applications such as Sensor Networks etc.

Why nonvolatile L2 , L3, LL Cache?

Capacitance of Cache Memory in CPU is increasing, which increases standby power of processors!

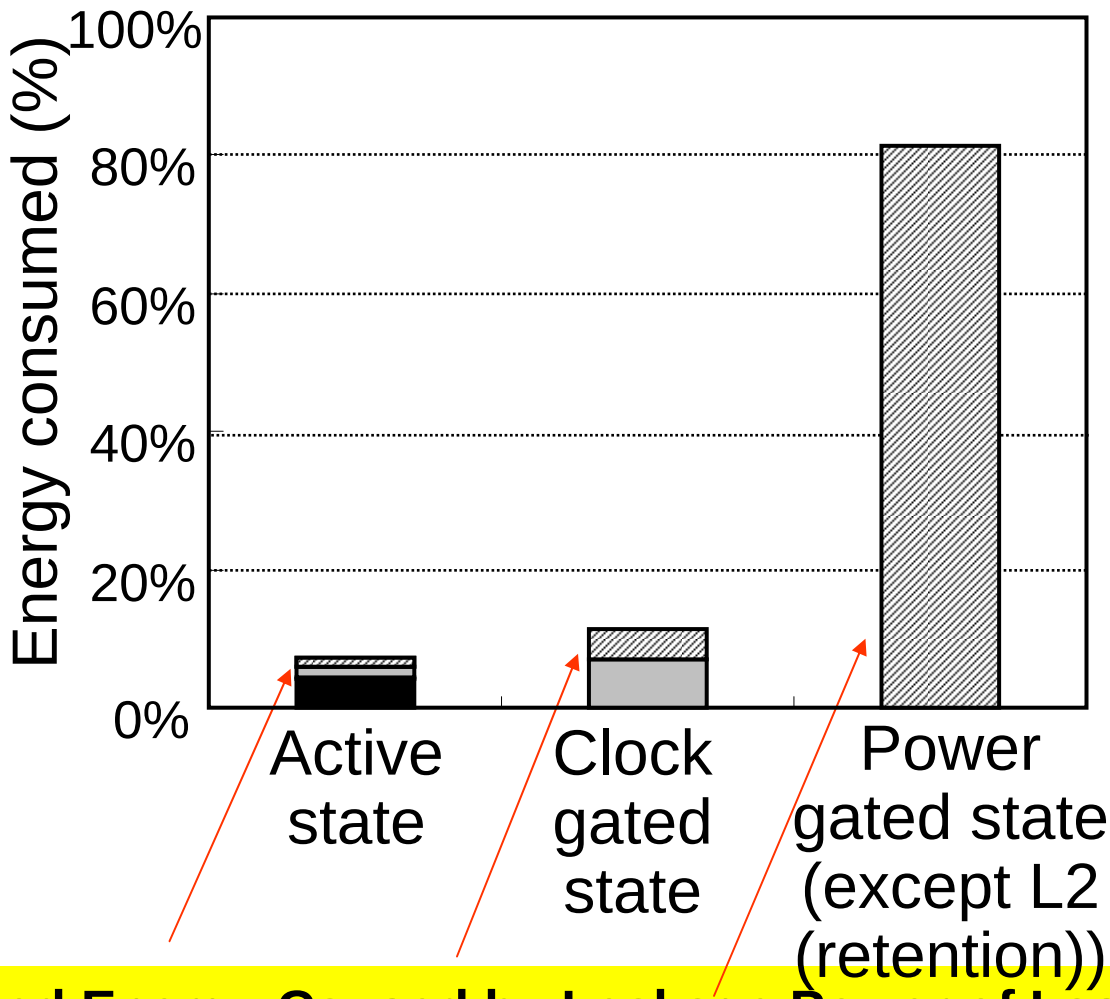


More cache, More Leakage..

<Background>

- Increase performance not by increasing clock frequency.
- Multi-core.

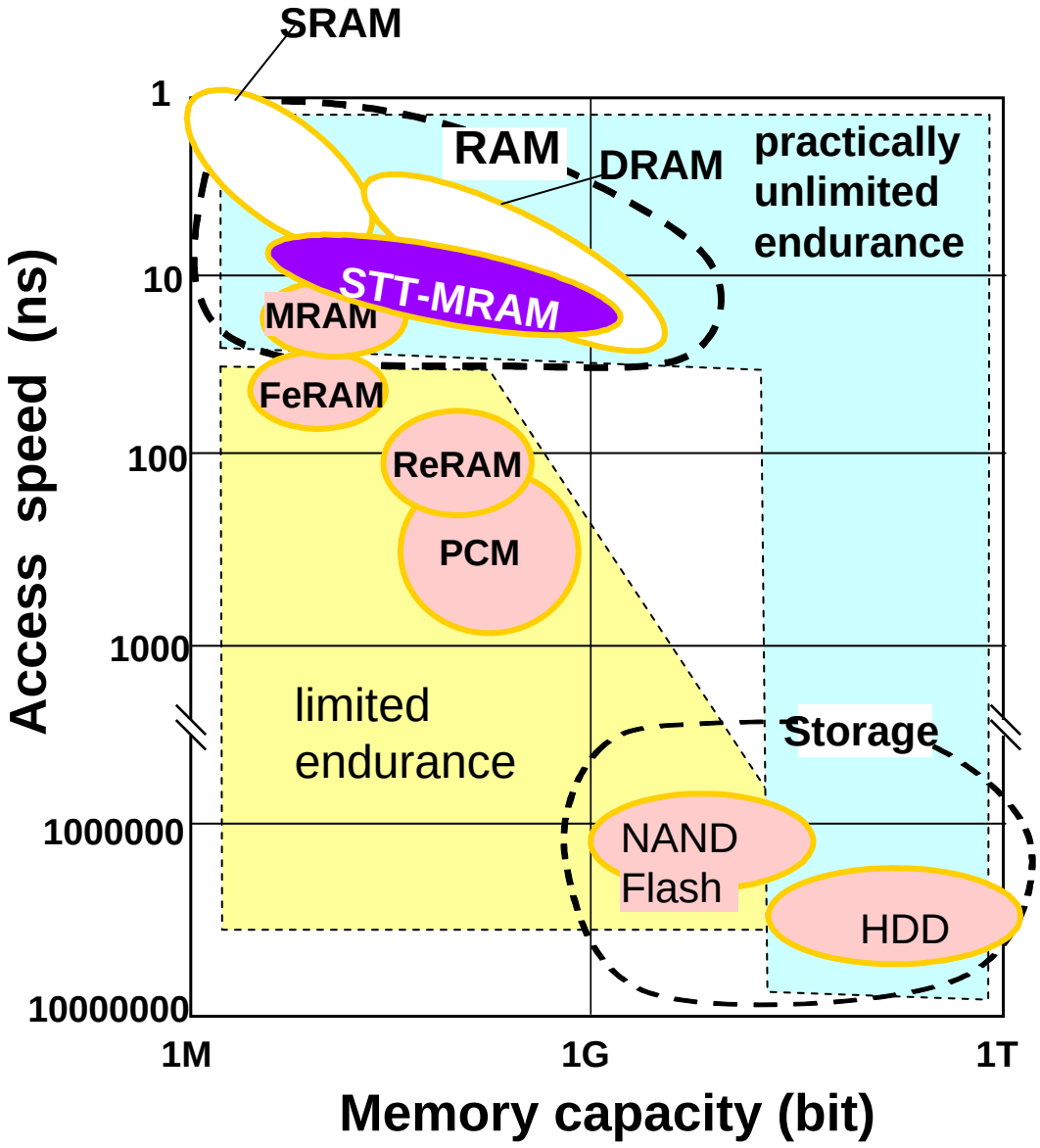
Especially for Mobile-Processor,
not Standby Power but Leakage Power is Dominant!



(Evaluation from one-day use case.)

Consumed Energy Caused by Leakage Power of Last Level Cache (L2$)

STT-MRAM is the best in NVM, but..

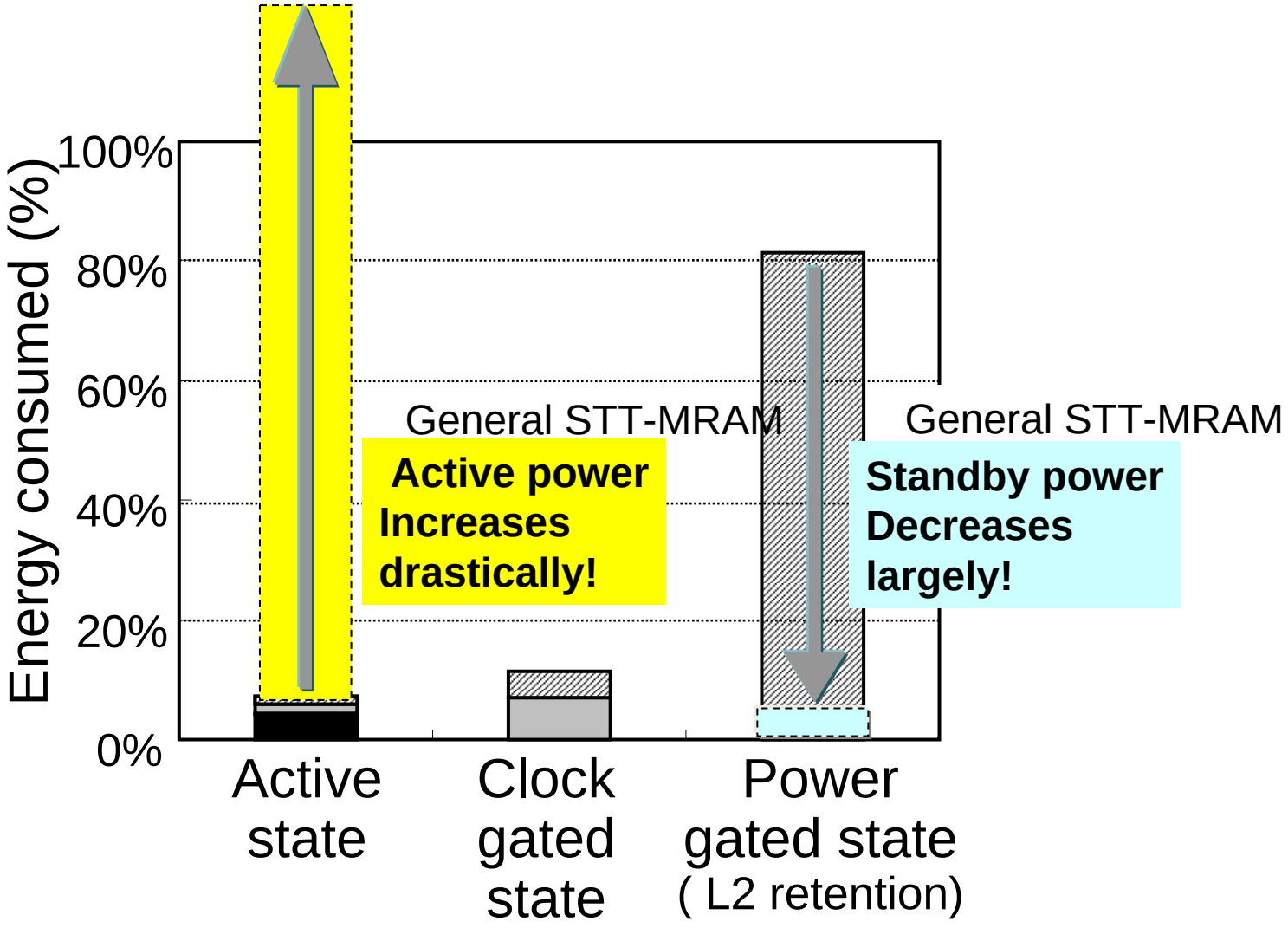


Its operation speed is slow and its power is high for cache memory.

Standby power is low,

but active energy is extremely higher than that of SRAM even using *conventional* STT-MRAM.

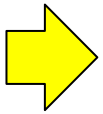
"Dilemma of Nonvolatile Memory! "



OUTLINE

- Introduction: New Design Concept

  Normally-off (**N-off**) Processor (from ver.0 to ver.1.)



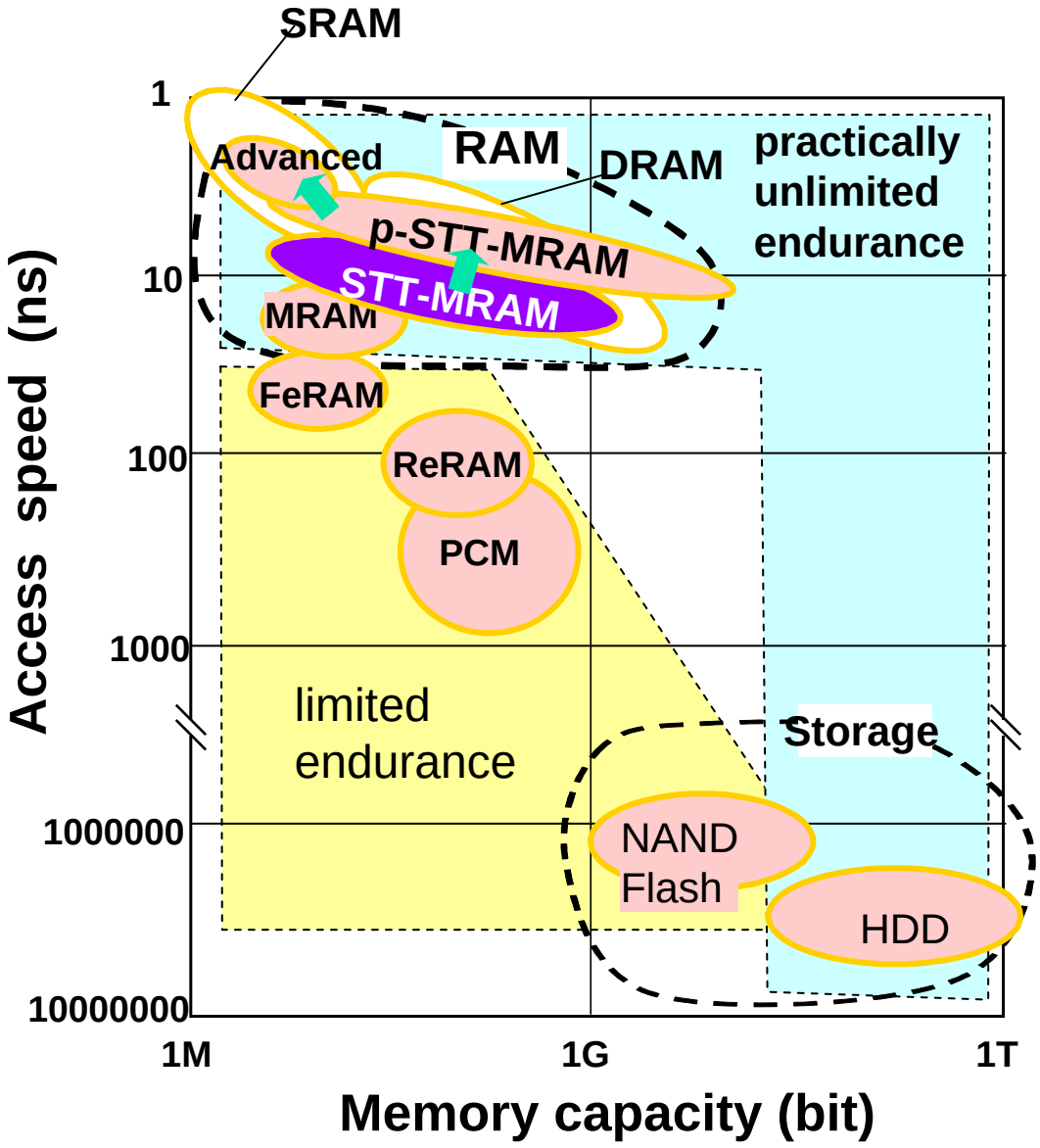
- Key Point 1: Advanced STT-MRAM

- Key Point 2: Decrease in power for **short** CPU standby state (in CPU active state) by applying **new memory cell design**

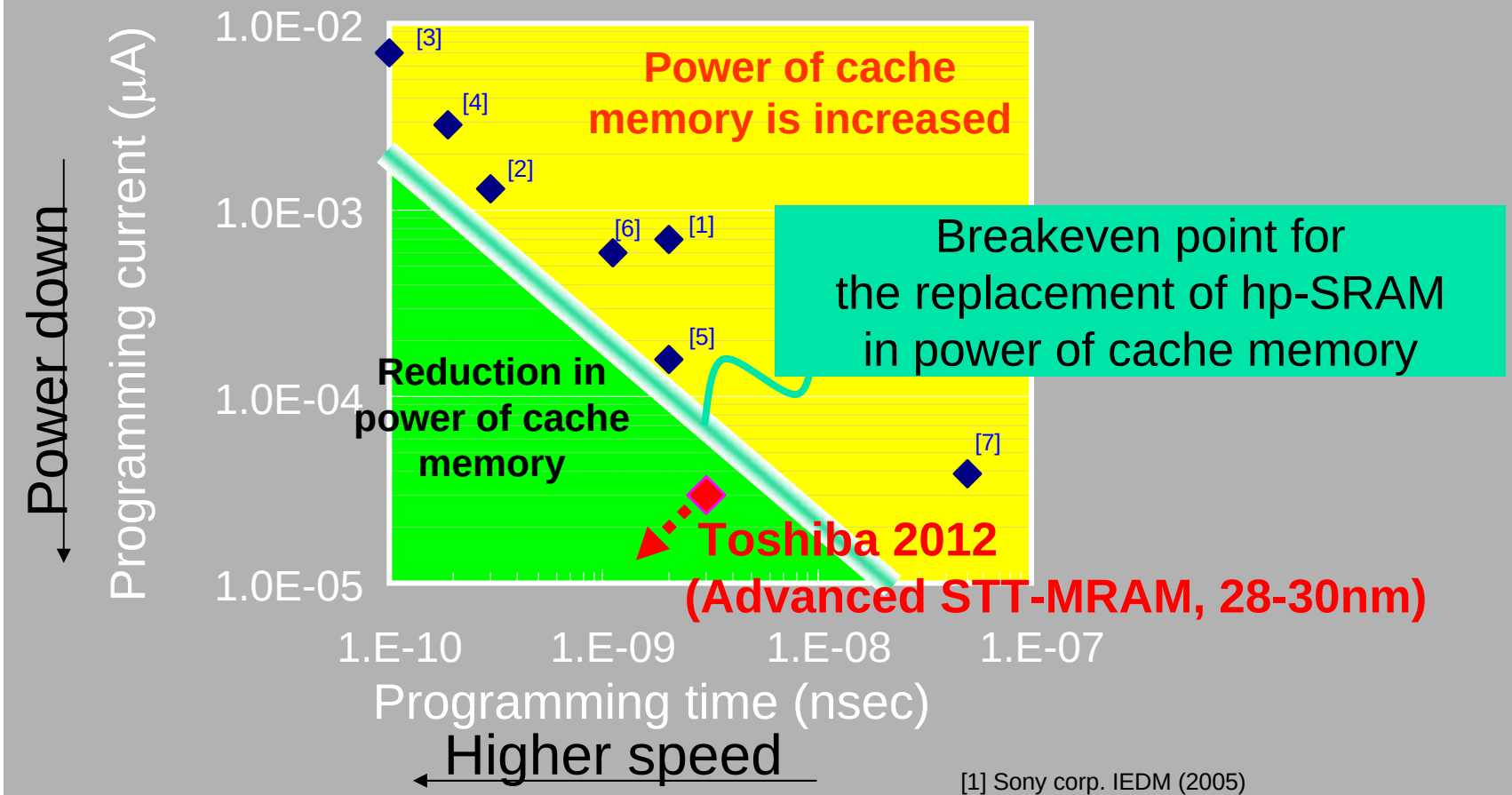
- Key Point 3: Power Decrease for **long** CPU standby state by **Ultra-Fast- Power Gating**

- Conclusions Towards **N-off ver 2**.

Advanced STT-MRAM has been developed!

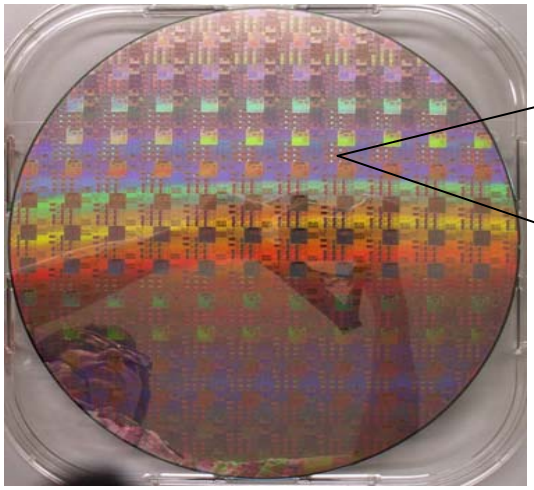


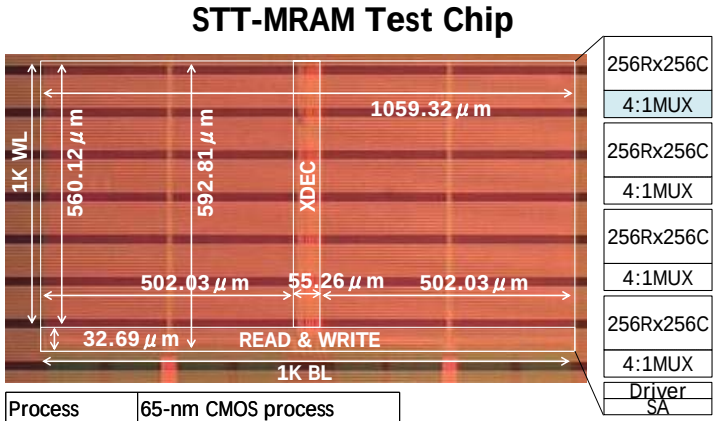
Breakthrough
by Toshiba's advanced STT-MRAM



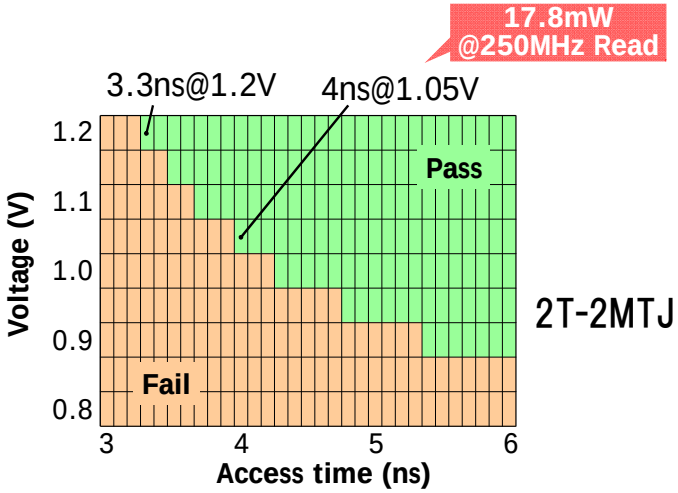
[1] Sony corp. IEDM (2005)
[2] New York univ. APPLIED PHYSICS LETTERS 97, 242510 (2010)
[3] Cornel Univ. APPLIED PHYSICS LETTERS 95, 012506 (2009)
[4] Minnesota univ. J. Phys. D: Appl. Phys. 45, 025001 (2012).
[5] NEC corp. Symposium on VLSI Circuits 7.3 (2012).
[6] IBM corp. Appl Phys Lett 98, 022501 (2011).
[7] TDK-Headway Applied Physics Express 5 093008 (2012)

Embedded Memory Integration (by Toshiba N-off PJ)





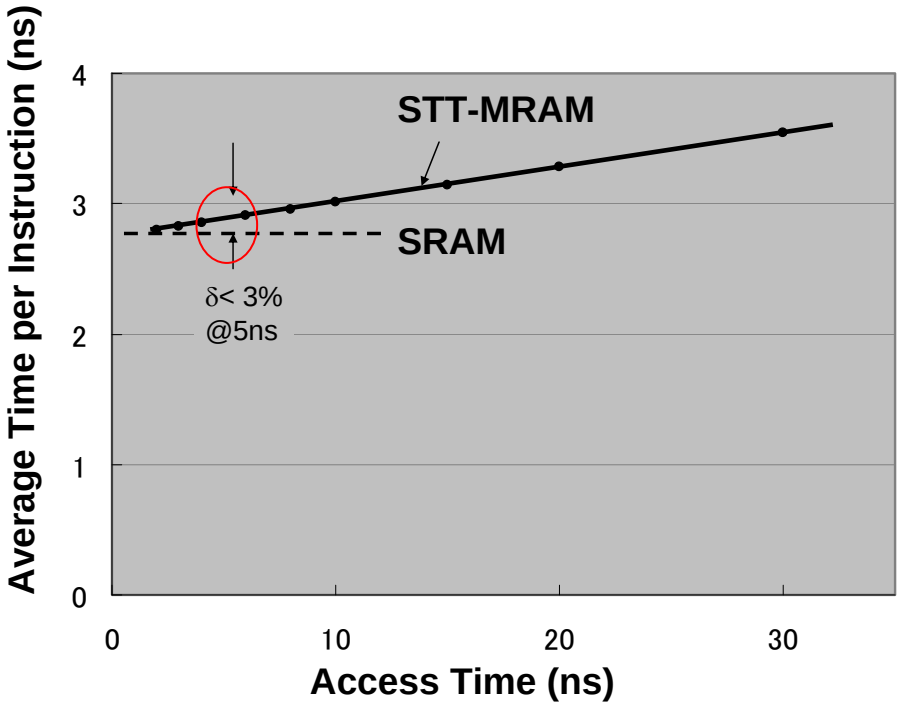
Process	65-nm CMOS process
Macro size	0.628 mm^2
Organization	4K words x 256 bits = 1 Mb

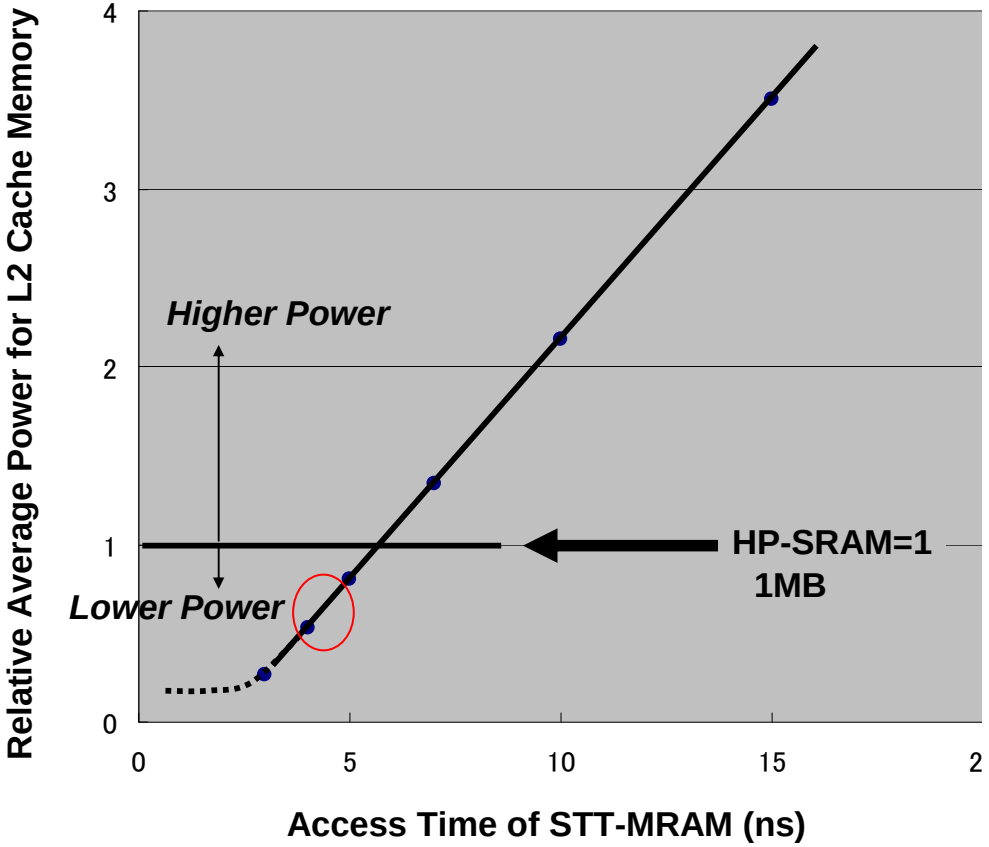


H. Noguchi et al., VLSI circuit symposium, 2013

Access time ≤ 4ns

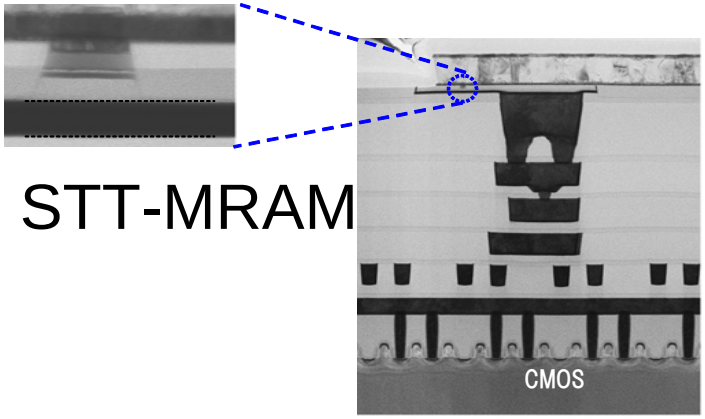
High speed STT-MRAM is NOT for high CPU performance, but for lower power CPU!





Development of “STT-MRAM-top Integration”

Specific MRAM Integration Process



Cross section image

Pool -top construction
(Marina Bay Sands Hotel)

Conventional CMOS Process
(in-house fab, foundry..)

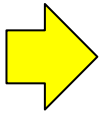
To be presented in VLSI-TSA 2014.

OUTLINE

- Introduction: New Design Concept

  Normally-off (**N-off**) Processor (from ver.0 to ver.1.)

- Key Point 1: Advanced STT-MRAM



- Key Point 2: Decrease in power for **short** CPU standby state (in CPU active state) by applying **new memory cell design (normally-off type design)**

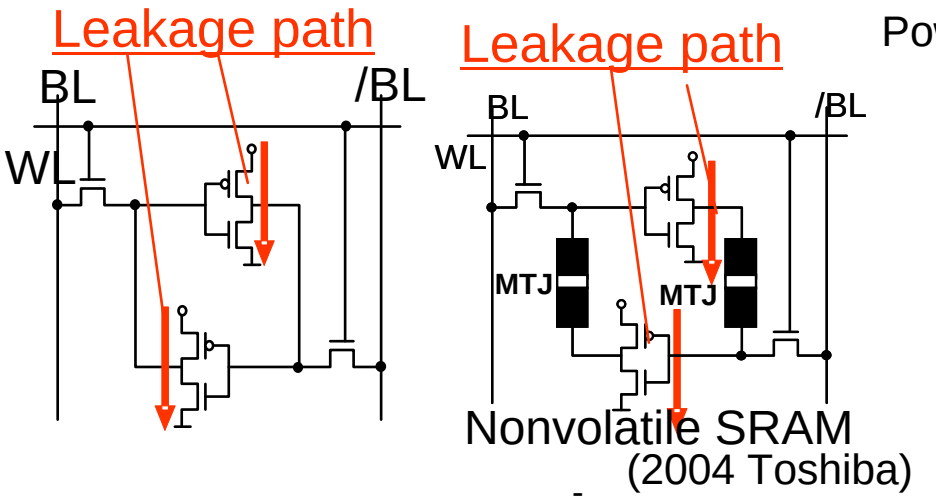
- Key Point 3: Power Decrease for **long** CPU standby state by **Ultra-Fast- Power Gating**

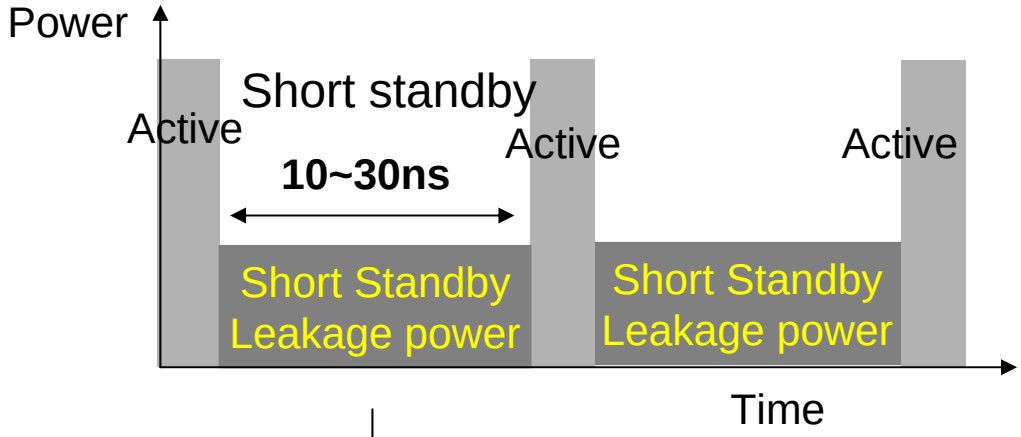
- Conclusions Towards **N-off ver 2**.

From “**Normally-On Type Memory** with Power Gating” to “**Normally-Off Type Memory** without Power Gating”

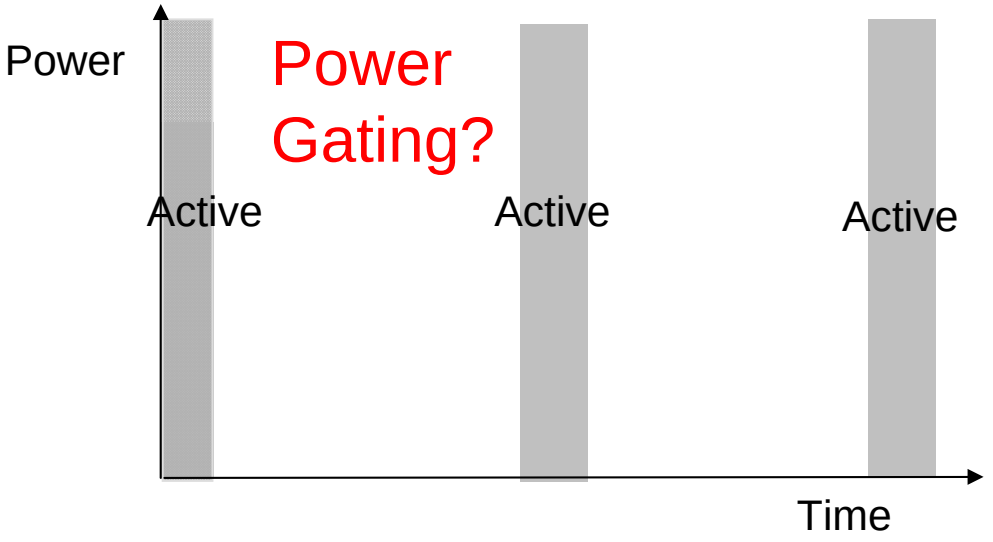
Normally-On Type

(1) SRAM and Nonvolatile SRAM without Power Gating





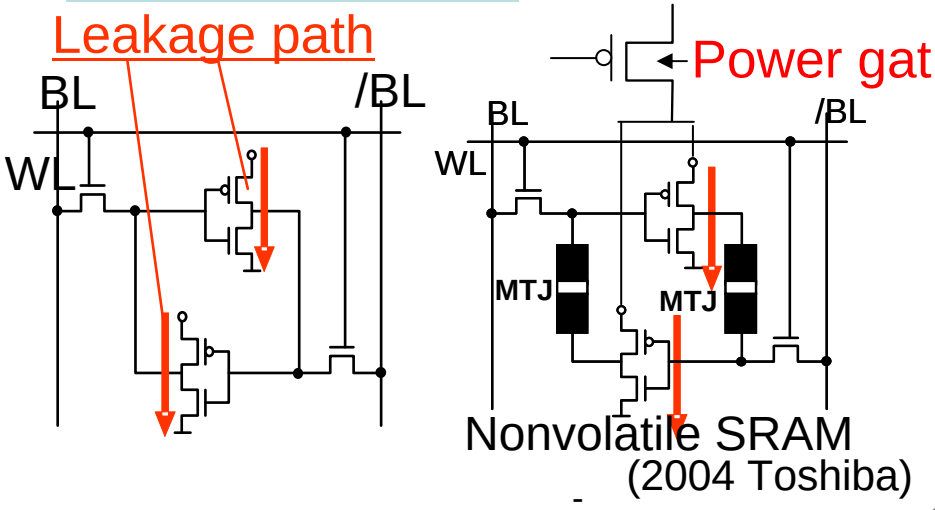
NV-SRAM for High Speed!

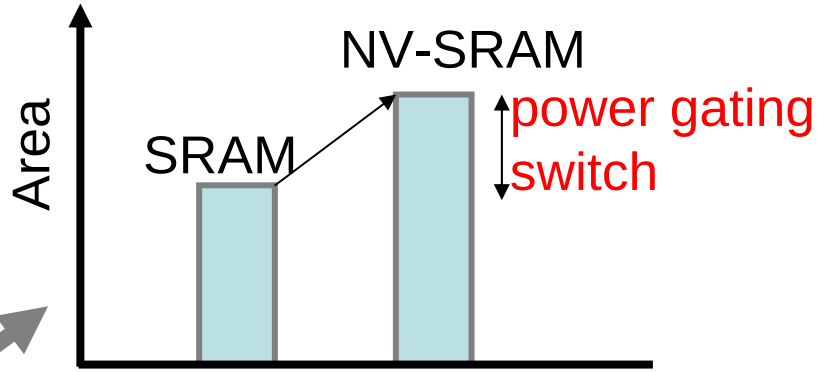


From “**Normally-On Type Memory** with Power Gating”
to “**Normally-Off Type Memory** without Power Gating”

Normally-On Type

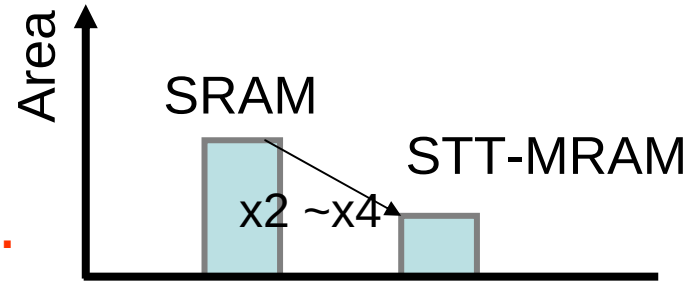
(1) SRAM and Nonvolatile SRAM with Power Gating





Overhead of power gating switch
is much large!
(Delay and Power overhead also)

(2) Normally-off Type Memory without Power Gating

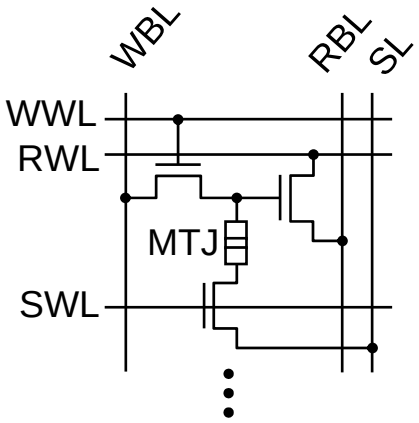


New design: “Normally-off Type”
(Next page)

No Leakage path, No power gating switch.

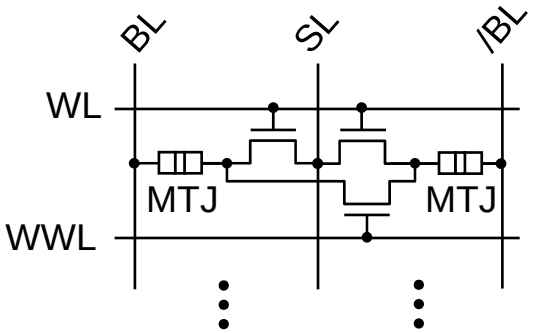
STT-MRAM cell is much smaller than SRAM cell.

Various kinds of **Normally-off Type Memory Cell designs** using advanced p-STT-MRAM presented by Toshiba.



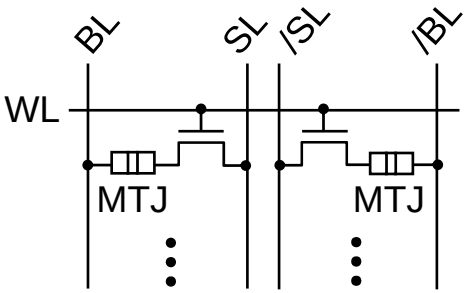
(a) D-MRAM

K. Abe et al.
IEDM2012
(Toshiba)



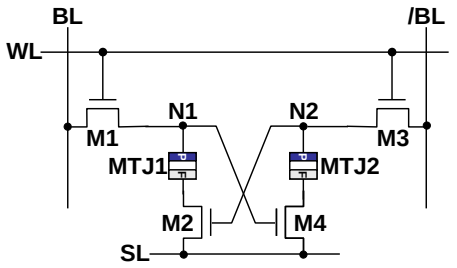
(b) 3T-2MTJ

A. Kawasumi et al.
IMW2013
(Toshiba)



(c) 2T-2MTJ

H. Noguchi et al.
VLSI Circuit 2013
(Toshiba)



(d) 4T-2MTJ

C. Tanaka et al.
SSDM 2013
(Toshiba)

As there are No Leakage paths like SRAM, no power gating switch is needed in the memory arrays.

CPU-Simulation
(ARM core, Linux-OS)

Processors	
# of cores	1
Frequency	1GHz
Issue width	1(out of order)
ISA	ARMv7

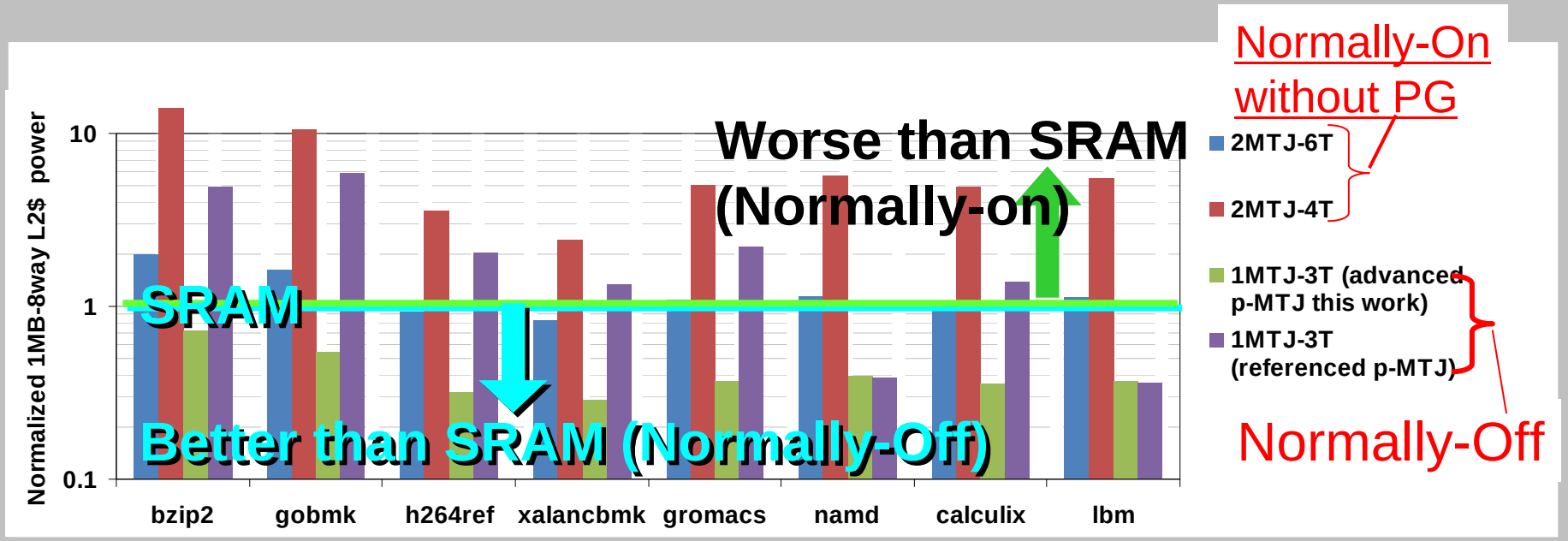
Execution	
Warm-up	1M inst.
Execution	10M inst.

Memory	
L1 cache	32+32kB, 4way, 64B line, Write-back, 1 read/write port, 1ns latency
L2 cache	1MB, 8 way, 64B line, Write-back,1 read/write port

Cell Type	MTJ device Write Time / Current
SRAM (Reference)	--
2MTJ-6T	3ns / 50uA (Advanced p-MTJ)
2MTJ-4T	25ns / 120uA (Reference p-MTJ)
D-MRAM (1MTJ-3T, This work)	3ns / 50uA (Advanced p-MTJ)

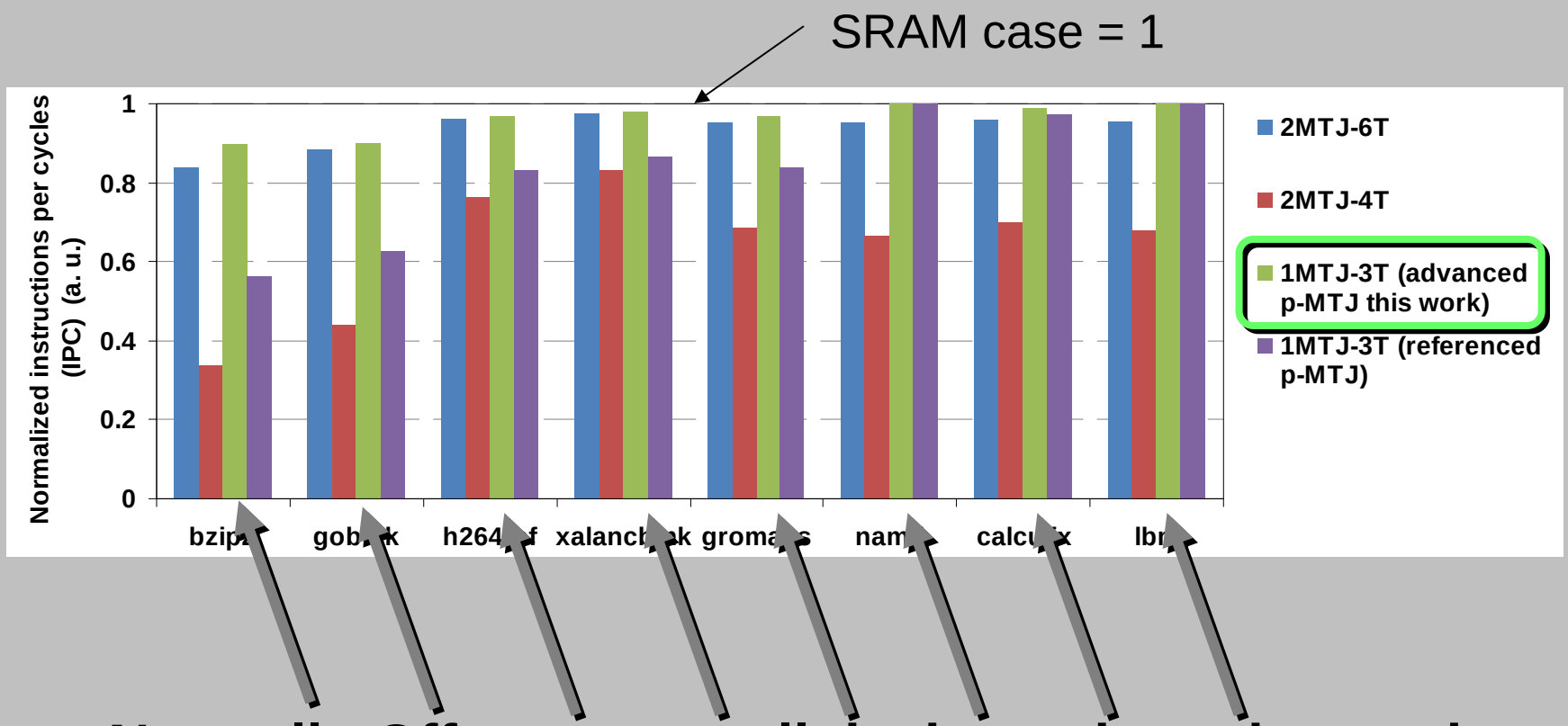
Processor benchmark sets: SPEC2006

Results of Power of Cache Memory (Short standby state) (case study: (a) D-MRAM)



Normally-Off memory using low-power and advanced p-STT-MRAM can reduce the cache power the most effectively.

Processor performance (Short standby state) (case study: Normally-off STT-MRAM)



Normally-Off memory cell design using advanced p-MTJ(STT-MRAM) has the best performance comparable to that of SRAM.

OUTLINE

- Introduction: New Design Concept

  Normally-off (**N-off**) Processor (from ver.0 to ver.1.)

- Key Point 1: Advanced STT-MRAM

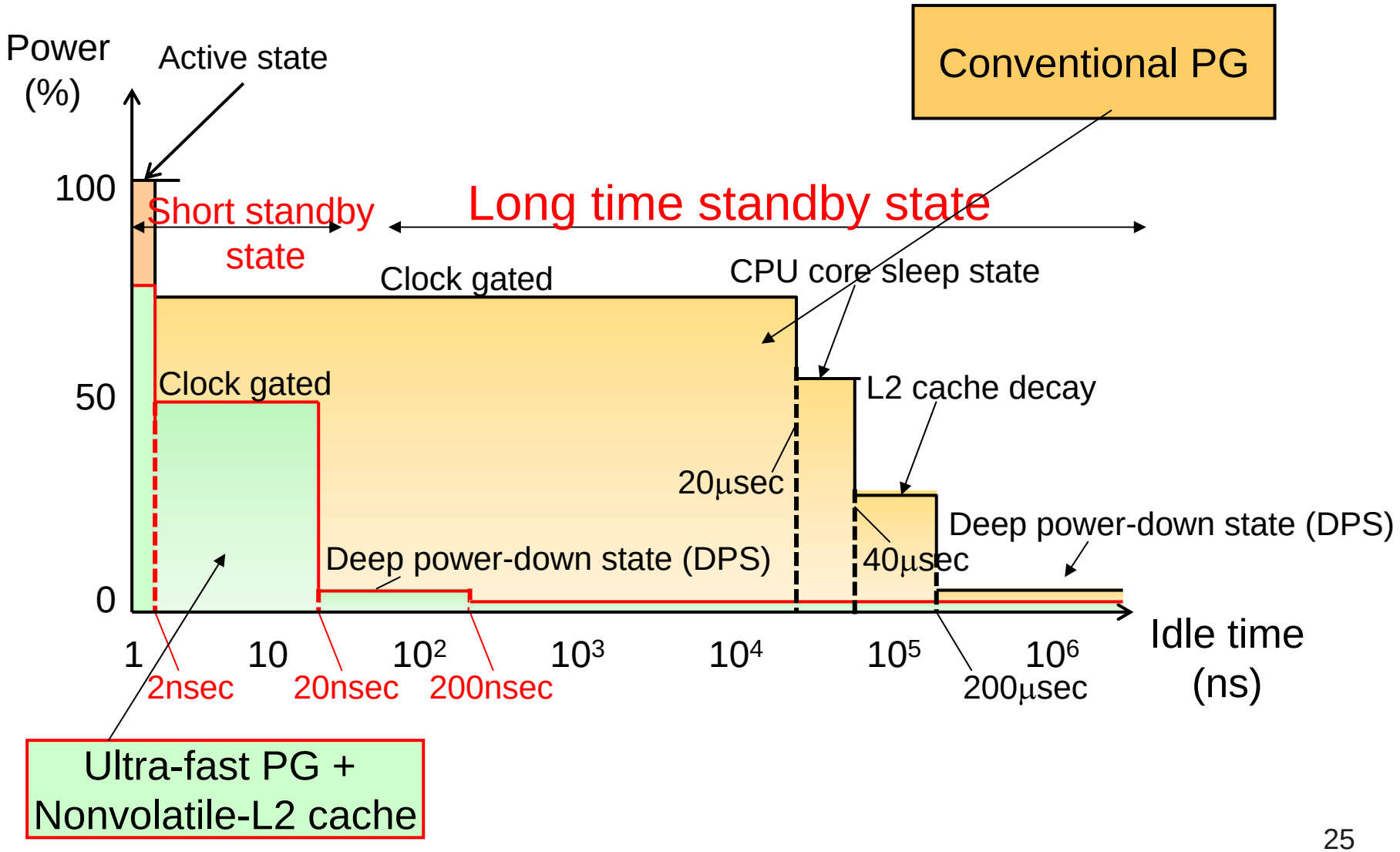
- Key Point 2: Decrease in power for **short** CPU standby state (in CPU active state) by applying **new memory cell design**

- Key Point 3: Power Decrease for **long** CPU standby state by **Ultra-Fast- Power Gating**

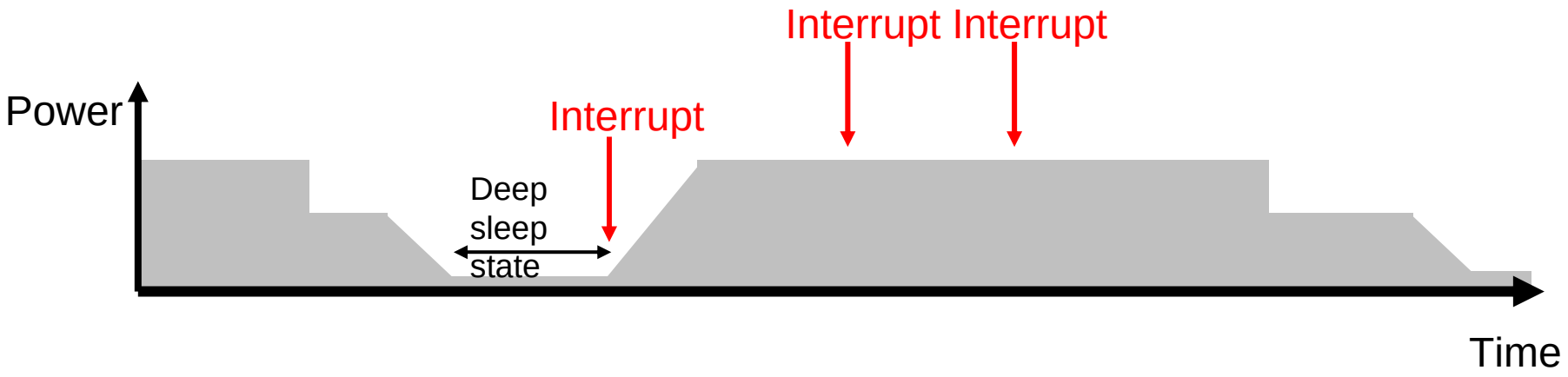
- Conclusions Towards **N-off ver 2**.

Conventional Power Gating	Clock	L1-$	L2-$	Others	Recovery time to active-state
Active state					
Clock gated state	OFF				~1ns
CPU core sleep (L2-Cache retention)	OFF	OFF			10μs
CPU core Sleep (L2-Cache decay)	OFF	OFF	decay		20μs
Deep power-down state (DPS)	OFF	OFF	OFF	OFF except State SRAM	100μs
High-speed PG with NV-cache	**Clock**	**L1-$**	**L2-logic +prepheral**	**L2-memory**	**Recovery time to active-state**
Active state				Normally OFF	
Clock gated state	OFF			Normally OFF	~1ns
CPU core sleep; Deep power-down state (DPS)	OFF		OFF	Normally OFF	~10ns
Deeper power-down state	OFF	OFF	OFF	Normally OFF	~100ns

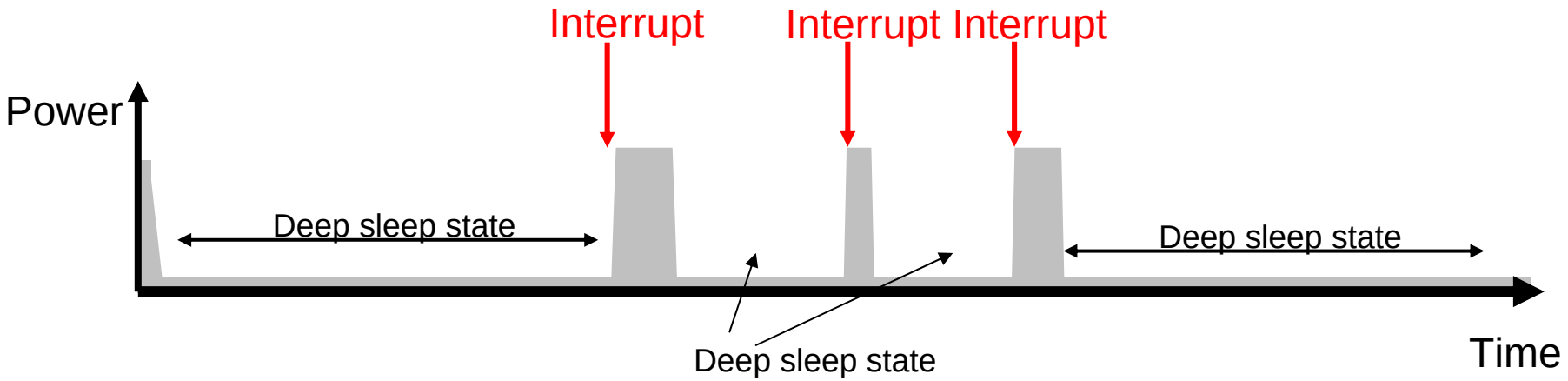
State transition policy for long time standby state: Conventional power gating (PG) vs. Ultra-fast PG with NV-L2-cache



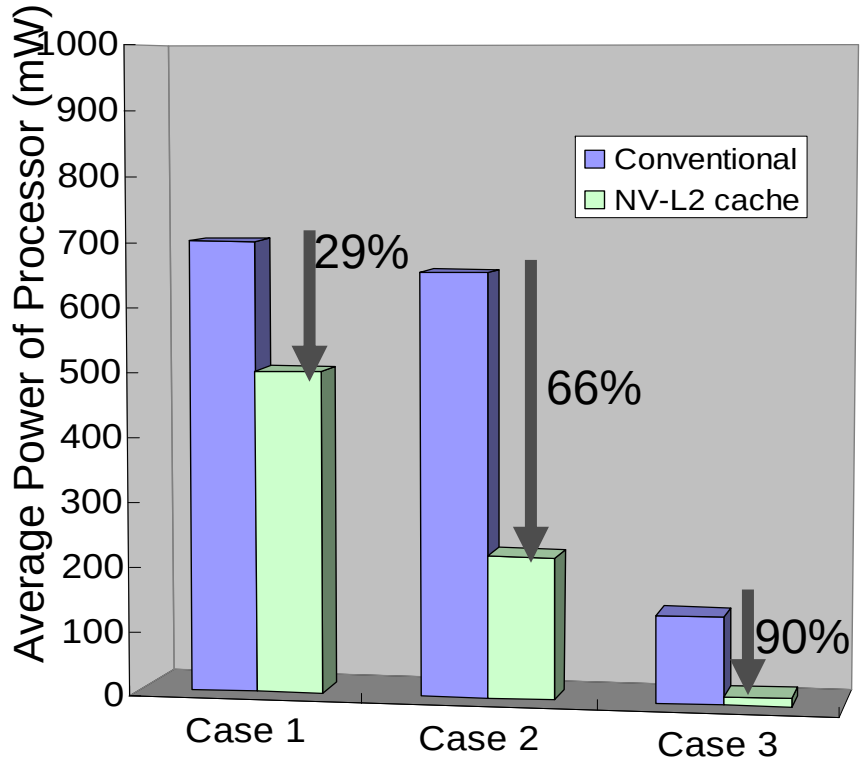
a) Conventional PG (Power Gating)



b) Ultra-fast PG+NV-cache



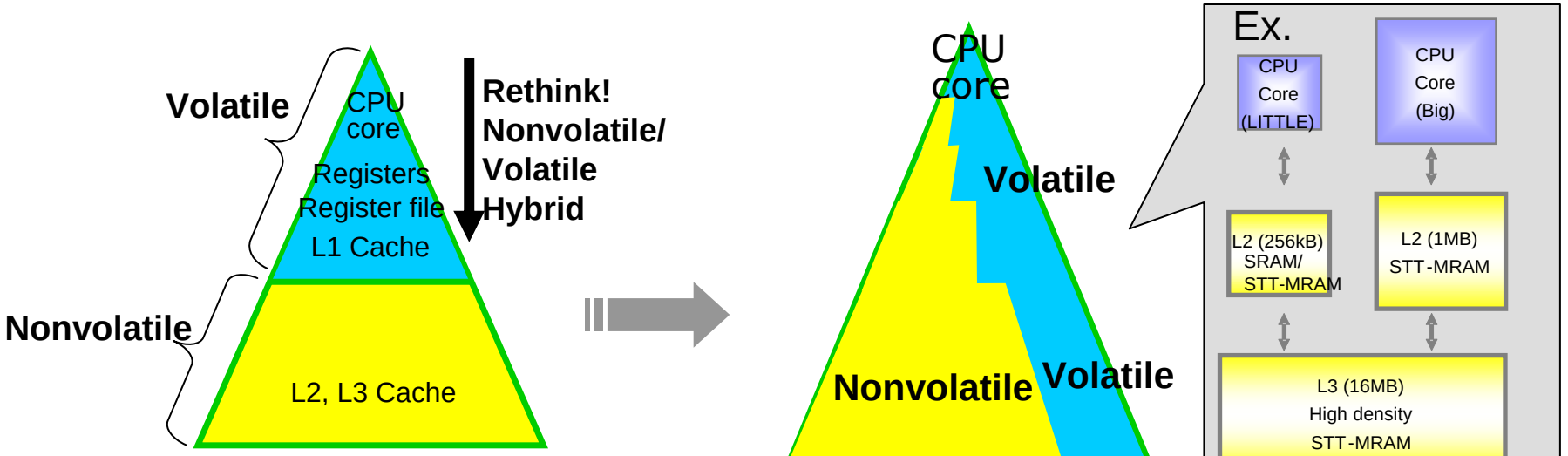
- Case studies: Decrease in average power of processor by ultra-fast PG with nonvolatile-L2 cache.



Case1:CPU active state dominant,
Case2: Moderate,
Case3: CPU idle state dominant.

Conclusion

- For HP-mobile processors, we proposed N-off processor ver.1; volatile L1-cache/ nonvolatile L2,LLC.
- To realize N-off processor ver.1, advanced STT-MRAM, normally-off type memory cell design, ultra-fast power gating are three key points.
- By applying new memory cell designs without leakage paths, not only CPU standby power but CPU active power has been effectively reduced.
- Average power reduction by 29 to 90% can be expected with little degradation of CPU performance.
- N-off processor concept shifting Ver.1 to Ver.2 has been in progress.



Normally-off Processor (**Concept Ver.1**) 2010

Normally-off Processor (**Concept Ver.2**) from 2012

Thank you!