

Near-Threshold Cache Architecture for Ultra-Low Energy Computing

Tohru ISHIHARA
Nagoya University

(mostly done by Jun Shiomi and Hongjie Xu)

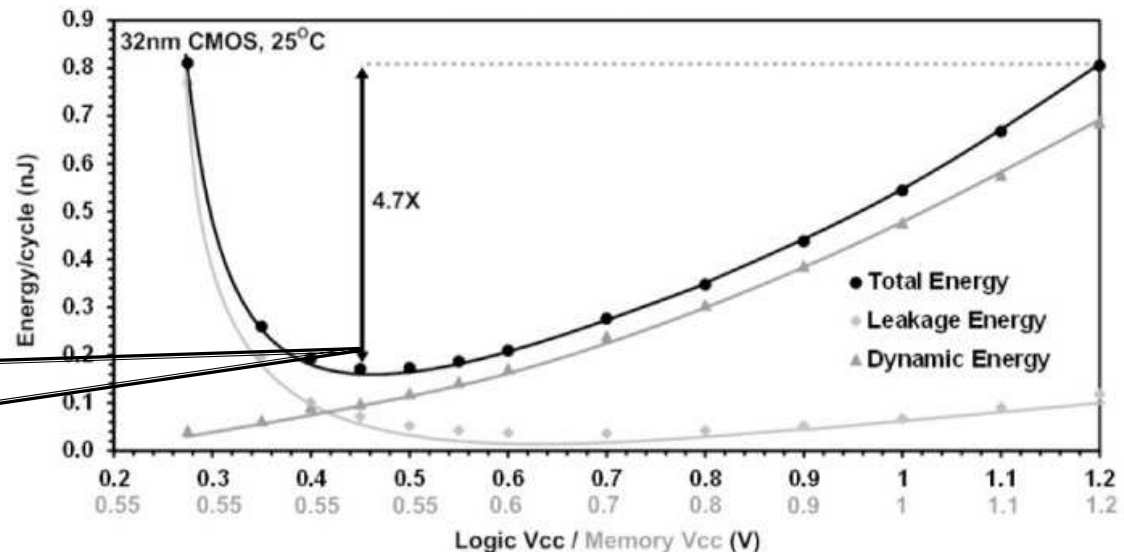
My Talk at MPSoC 2014

- = minimum energy consumption
- Highest energy efficiency can be obtained at **Near-Threshold Voltage** (NTV)

Reference: Research@Intel 2012, June 26, 2012

32nm test processor
“Claremont”

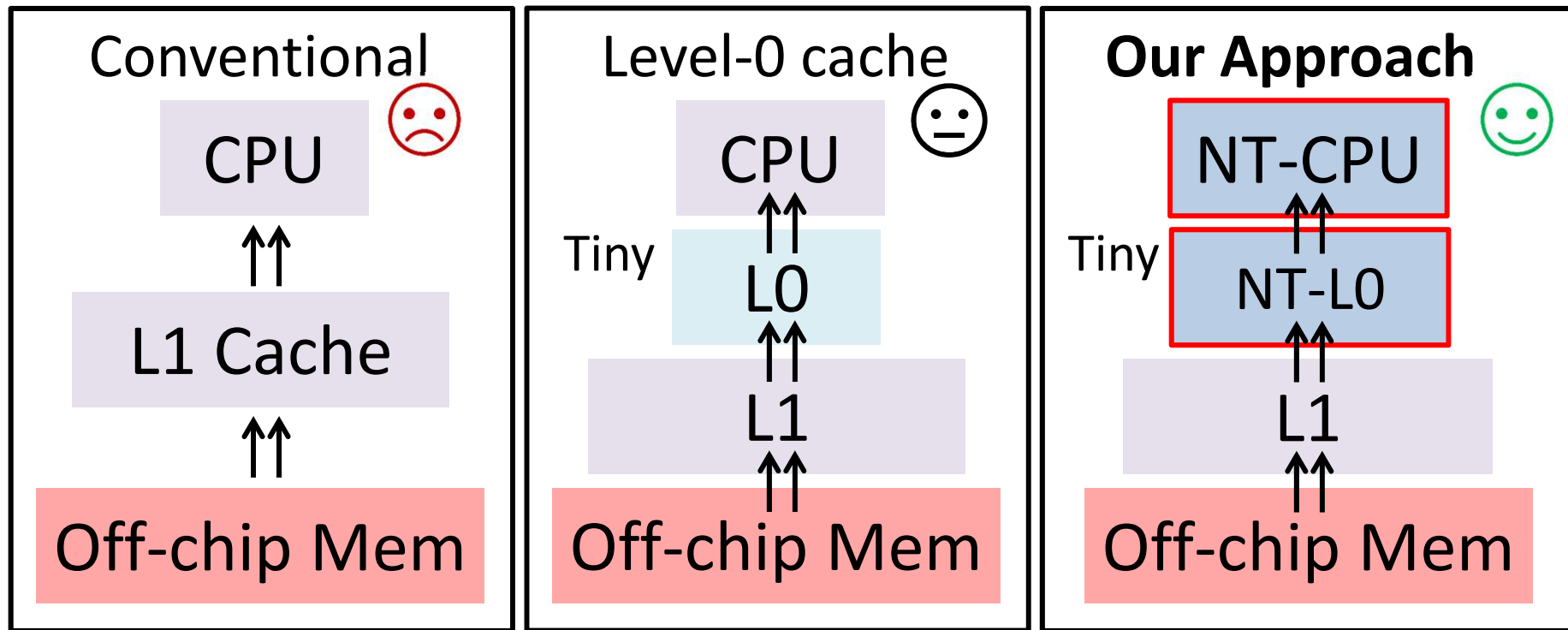
4.7x better energy
efficiency at 0.45V



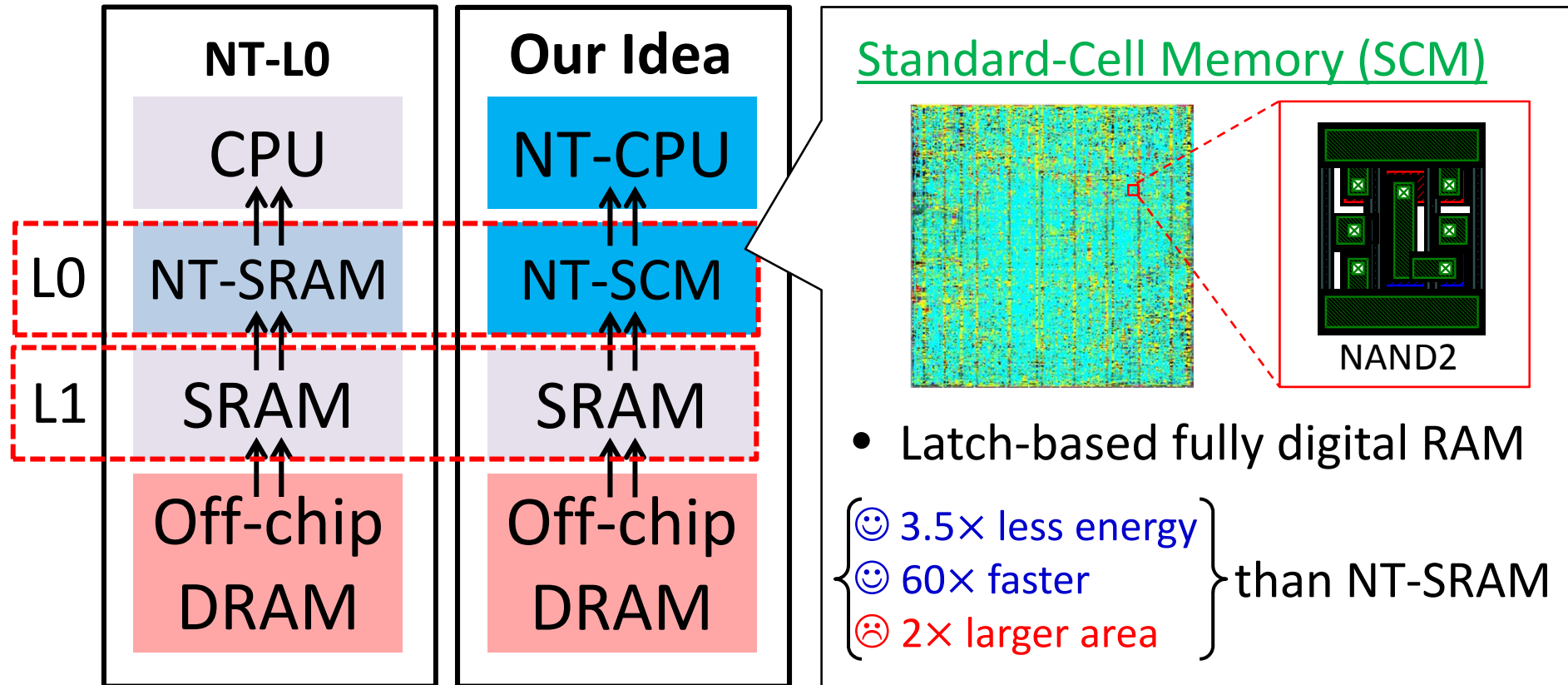
Discuss how we can enjoy E_{min} computing

Near-Threshold for Cache Memory

- L0-cache exploits fast and LP nature of tiny SRAM
- **Our idea:** Utilize NT for further energy reduction
- Targeting instruction cache only as a first step



Standard-Cell Memory for NT-L0-Cache



⇒ **Need optimizing energy-area tradeoff**

Combine energy-efficient **SCM** & area-efficient **SRAM**

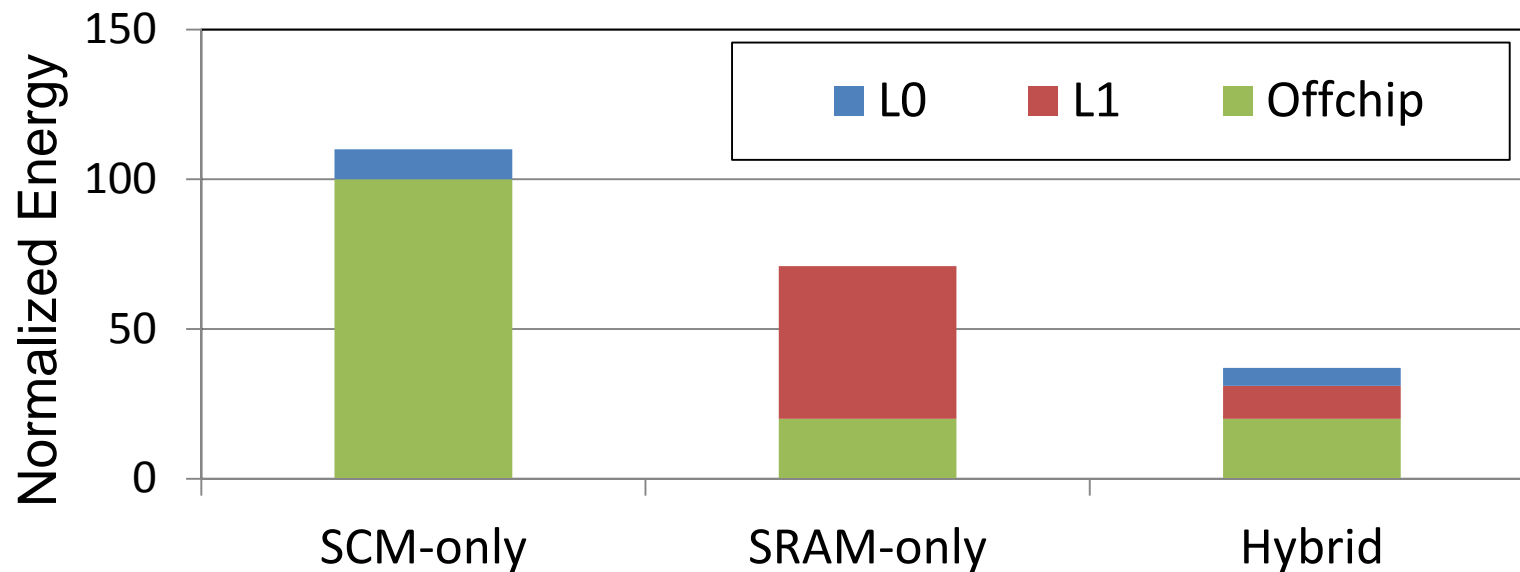
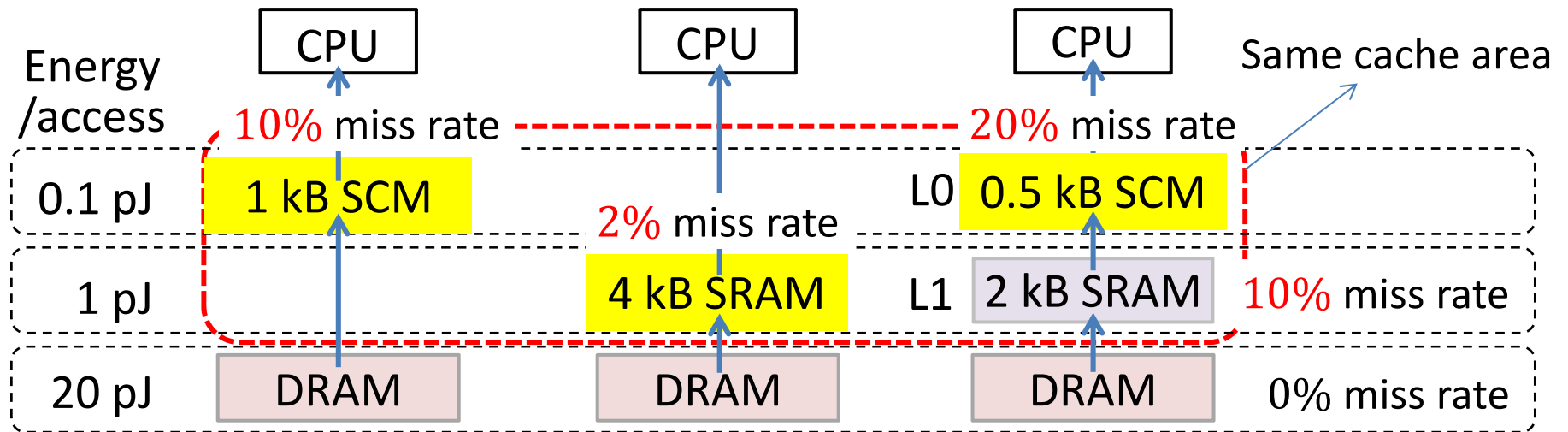
Comparison of SCM and SRAMs

	Normal SRAM [2]	NT-SRAM [3]	This work [1]
Technology	65 nm	65 nm	65 nm
Bit-cell Type	6T-SRAM	10T-SRAM	Latch-based SCM
Energy / Bit [fJ/bit]	101 fJ	39 fJ*	11 fJ
Voltage	0.8 V	0.35 V	0.4 V
Frequency	1.3 GHz	500 kHz	30.6 MHz
Area [$\mu\text{m}^2/\text{bit}$]	1.6	3.9	8.0

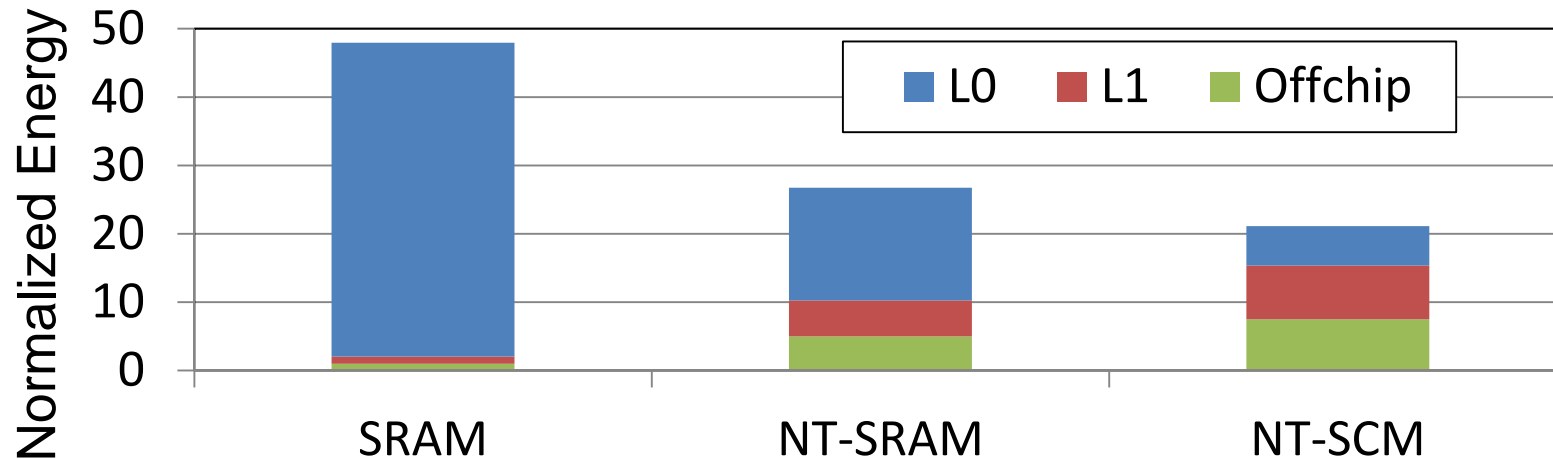
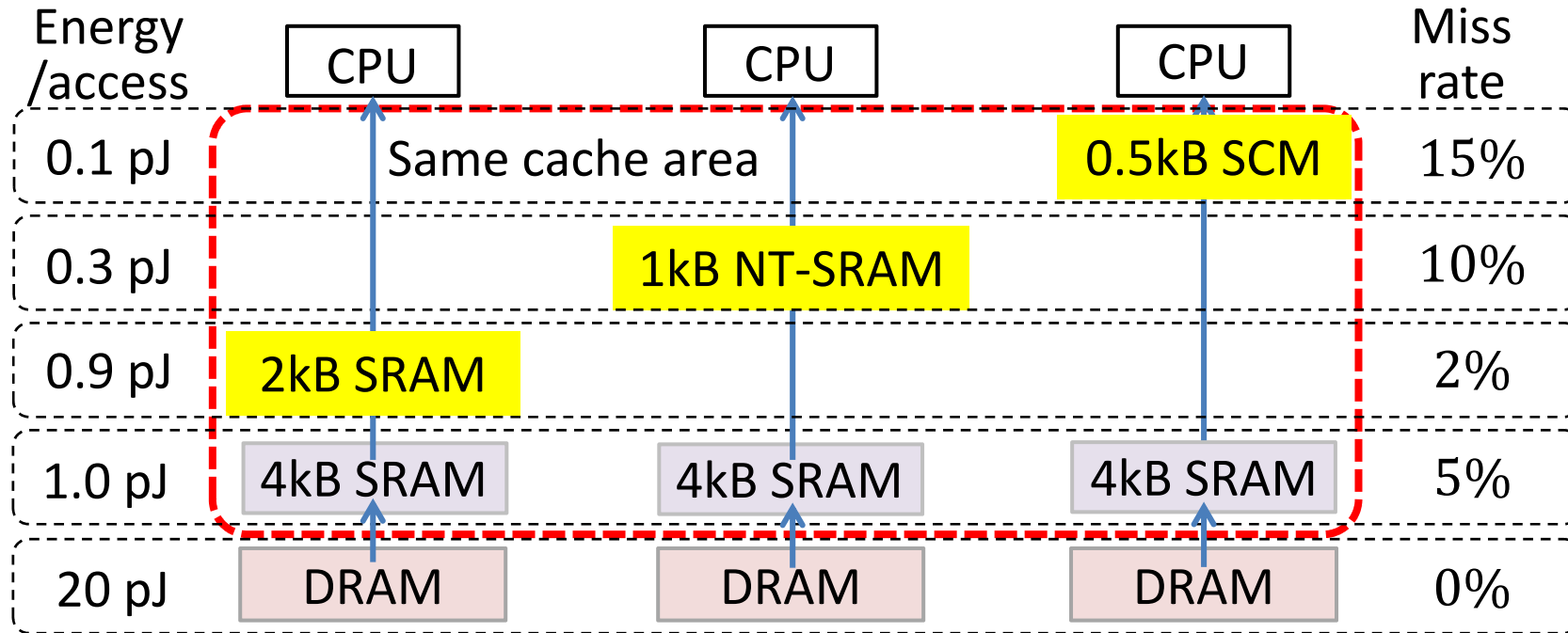
* Energy is estimated by assuming $E_{\text{energy}} \propto \log_2(\text{Capacity})$ [2]

- [1] J. Shiomi, T. Ishihara, H. Onodera, "Area-efficient fully digital memory using minimum height standard cells for near-threshold voltage computing," Elsevier VLSI Journal, 2017.
- [2] S.J.E. Wilton and N. Jouppi, "CACTI: an Enhanced Cache Access and Cycle Time Model," Journal of Solid State Circuits, vol.31, no.5, pp.677–688, May 1996.
- [3] S. Clerc, F. Abouzeid, G. Gasiot, D. Gauthier, P. Roche, "A 65 nm SRAM Achieving 250 mV Retention and 350 mV, 1 MHz, 55fJ/bit Access Energy, with Bit-Interleaved Radiation Soft Error Tolerance," in Proc. of ESSCIRC, 2012, pp.313-316.

Motivational Example

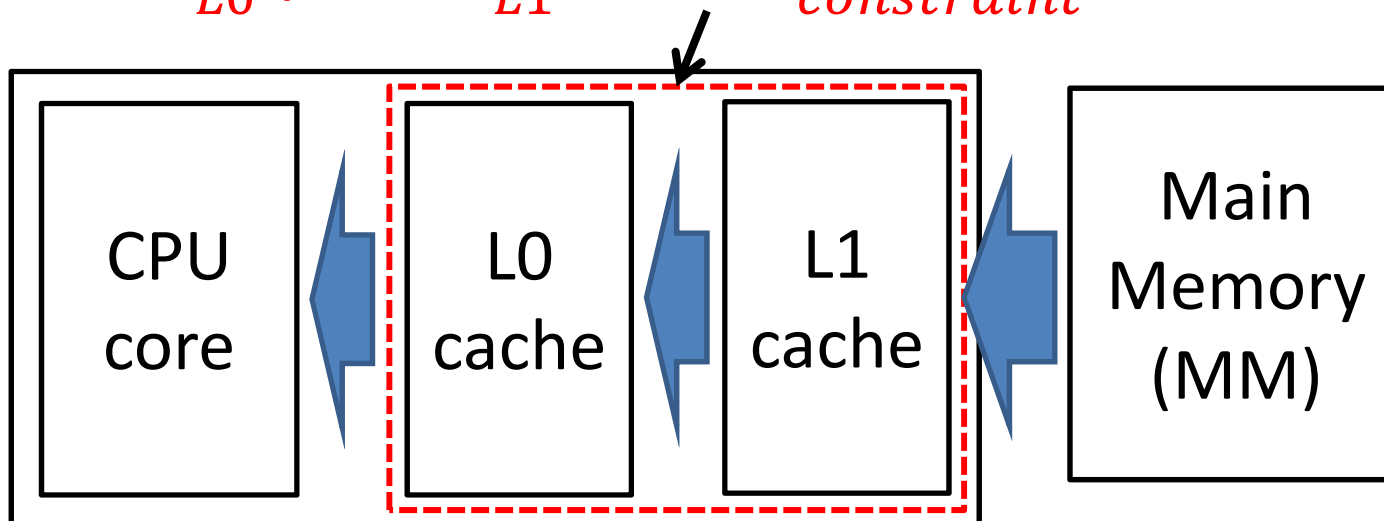


Impact of L0 Cache Type



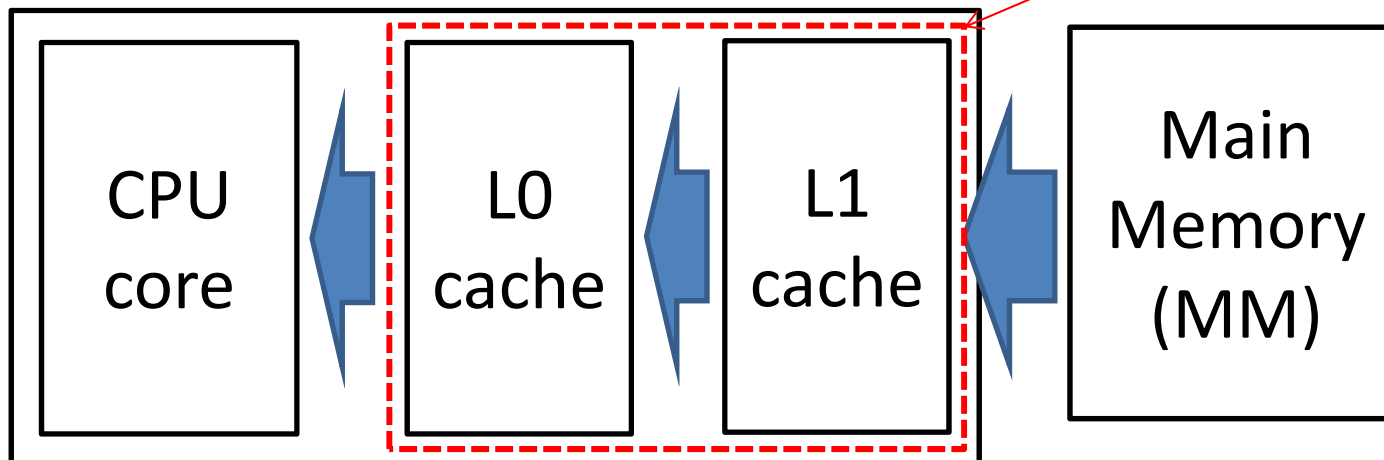
Cache Size Optimization

- Targeting instruction cache only
- Minimize total energy in memory hierarchy
 - Minimize: $Energy_{L0} + Energy_{L1} + Energy_{MM}$
- Subject to specific on-chip memory area
 - $Area_{L0} + Area_{L1} < Area_{constraint}$



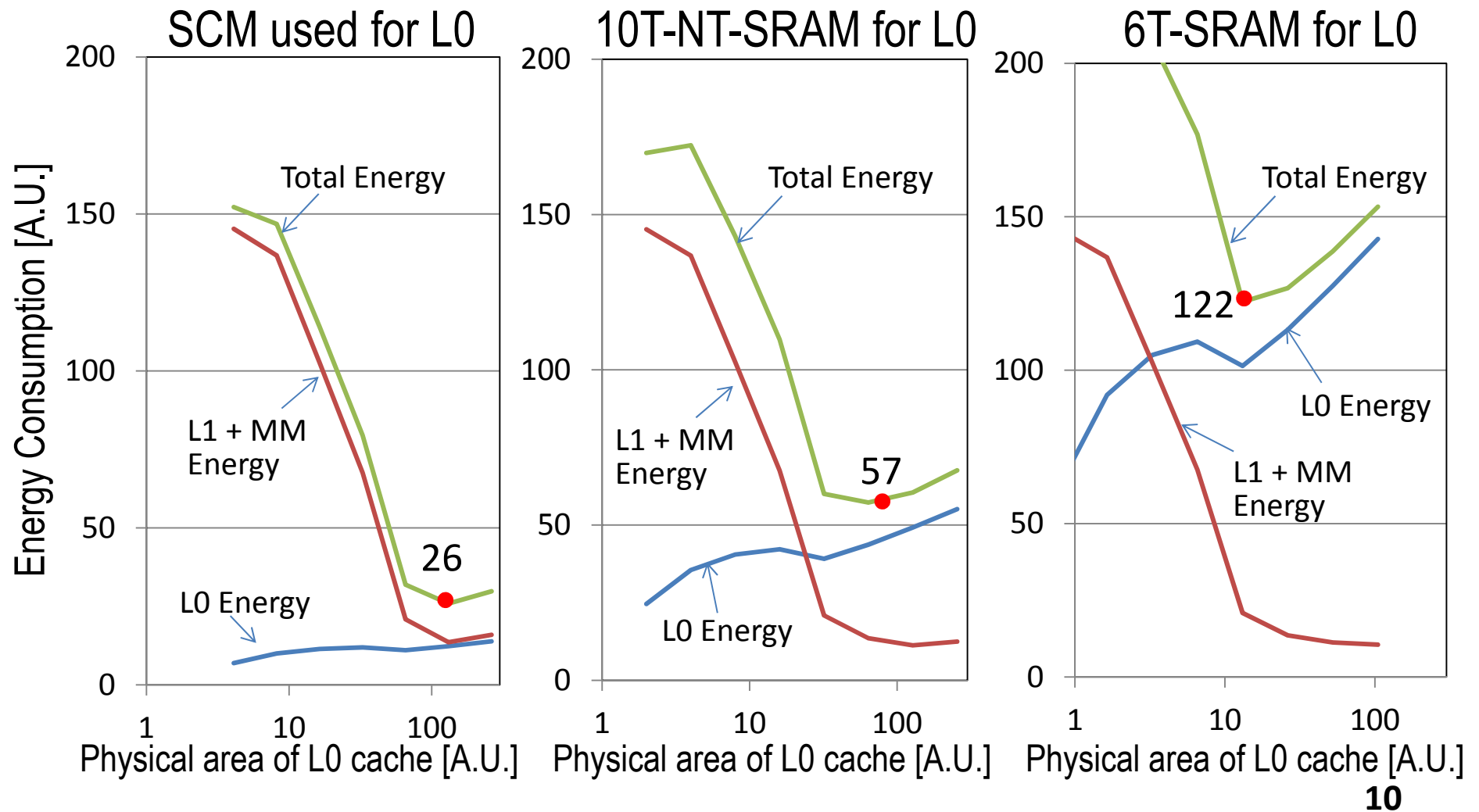
Experimental Setup

- Target processor: RISC-V
- Benchmark: AES, Conv2D, DCT, FFT, IPM, SHA
 - Synthetic program combining above 6 programs
- L0 & L1 configuration: 2-way, Line size 16B
 - Only index size is changed under area constraint



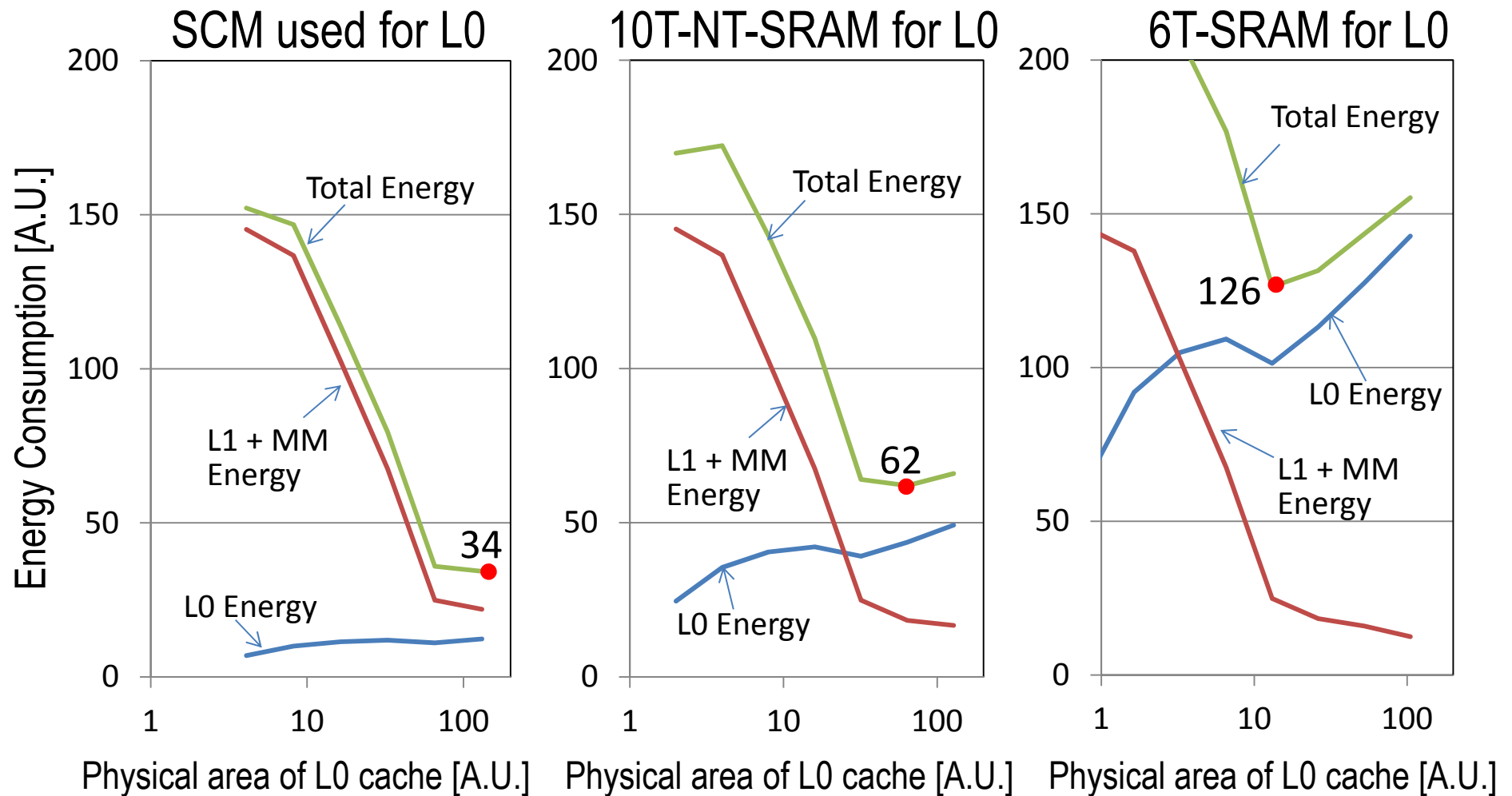
Energy Consumption Results

$Area_{constraint} \cong$ Area of **32kB** SRAM



Energy Consumption Results

$Area_{constraint} \cong$ Area of **16kB** SRAM



Conclusions

- Show energy efficiency of near-Vt SCM cache
 - Exploit energy efficient SCM and area efficient SRAM
 - 4X better than SRAM-L0 and 2X better than NT-SRAM
- Future Work
 - Extension towards data memory
 - Real chip validation

