Optimizing Memory Hierarchy within an Internet Router for High-Throughput and Energy-Efficient Packet Processing

Kyosuke Tanaka

(Supervisors: Hayato Yamaki, Shinobu Miwa, Hiroki Honda)

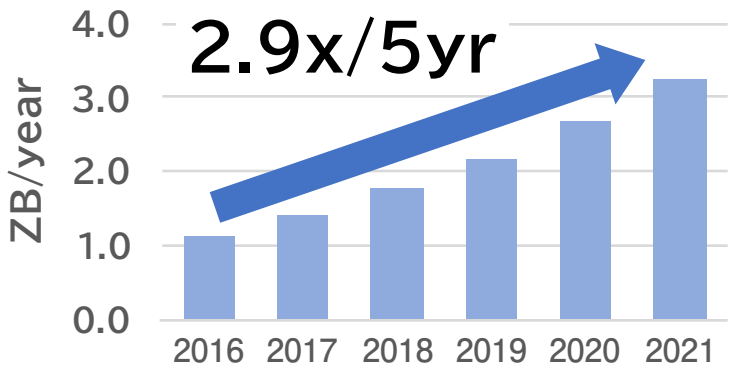
The University of Electro-Communications, Japan

Background

Dramatic increase in network traffic*

- Current standard of network is 100Gbps
- Next standard will be 400Gbps (IEEE802.3bs)

4x

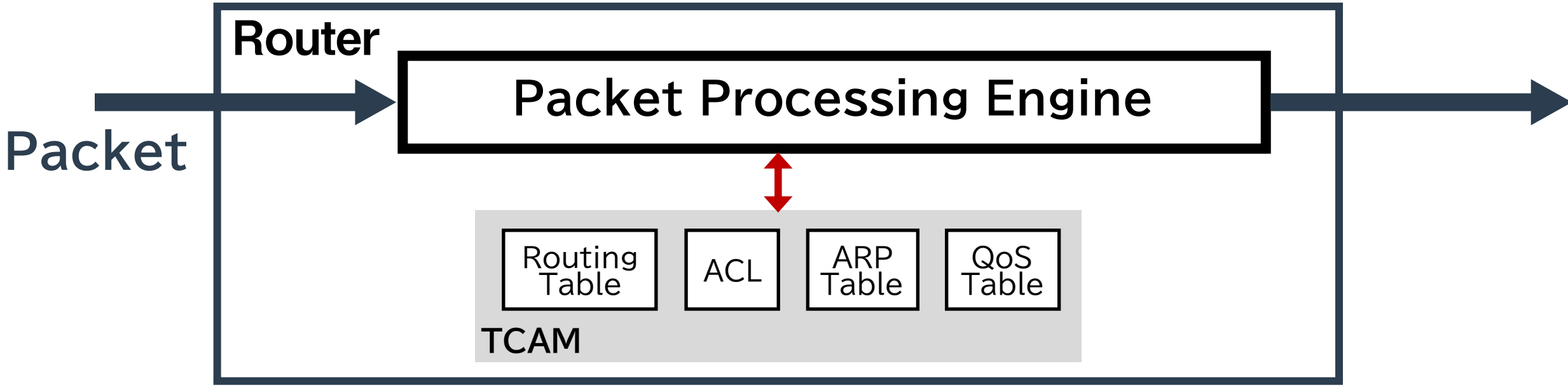


A main challenge towards 400Gbps:

- **Table lookups** are known as bottleneck of internet routers
- Realize **high-speed table lookup** with small power

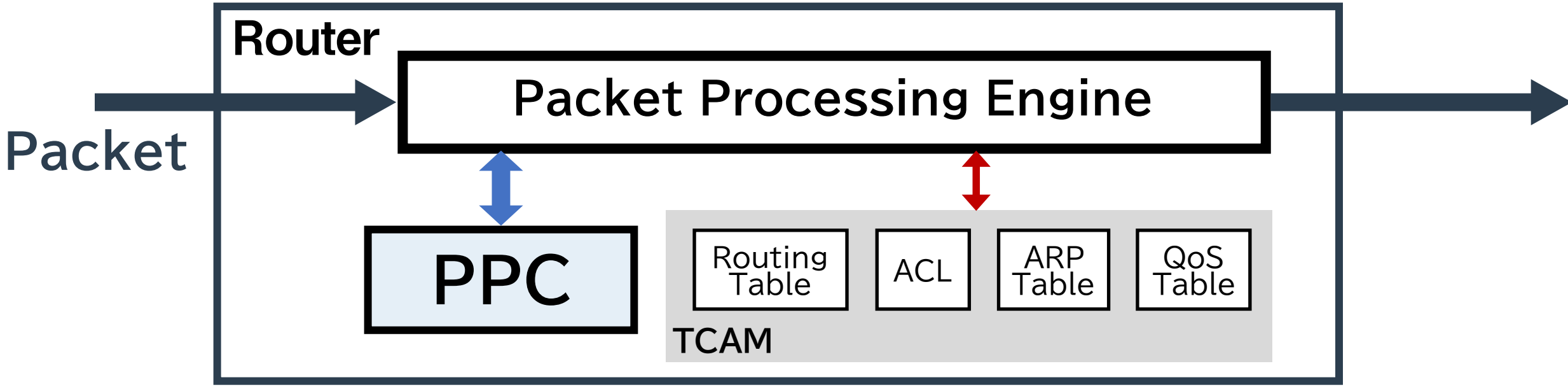
* https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

Router Architecture



- Accesses multiple tables for packet processing
- **TCAM** is used for high-speed search, **but**
  - **Difficult to achieve 100Gbps**
  - **Consume large power**

PPC: Packet Processing Cache (1/2)



- Reuse TCAM lookups results
- Improve performance and energy efficiency (e.g., **278.9 Gbps** / 206.3 mW)

PPC: Packet Processing Cache (2/2)

- **400Gbps has not yet been achieved**
  - PPC miss rate is **still high** (~18%)
  - **Large gap** between cache and TCAM

	PPC (Conv.)		TCAM
Capacity (entries)	28KB (1K)	90x →	2.5MB (0.5M)
Access latency	0.6ns	16x →	10ns
Dynamic Energy	0.02nJ	500x →	10.5nJ

Introducing multi-level caches

Experimental Methodology

Evaluate various combinations of L1/L2 PPC

- Use **in-house PPC simulator** + CACTI 6.5

  Available at GitHub (kyontan/cache_simulator)

- Use **10 real network traces** from WIDE MAWI and NLANR PMA

CACTI Parameter	Value
Process technology	32nm
Ports	1 (read) + 1(write)
Transistor model	L1: ITRS-HP, L2: ITRS-LSTP

* WIDE MAWI Working Group - http://mawi.wide.ad.jp/
** RIPE Labs – Data Repository – Data Sets – NLANR AMP Data - https://labs.ripe.net/datarepository/data-sets/nlanr-amp-data

PPC Configuration

Flow information (cache tags) | TCAM lookup results

Src IP	Dst IP	Src Port	Dst Port	Protocol	Output IF	Action	…
10.9.2.3	1.2.3.4	443	65382	TCP	2	Permit	…
10.2.1.4	192.30.0.8	60001	11467	UDP	3	Deny	

13bytes | 15bytes

- Indexed with hashed flow information
- 4-way set-associative
- LRU
- Write-through
- Inclusive

Throughput Model

$$Th_{PPC} = L / \max\left(t_{L1}, t_{L2} \cdot m_{L1}, t_{TCAM} \cdot m_{L1} \cdot m_{L2}\right)$$

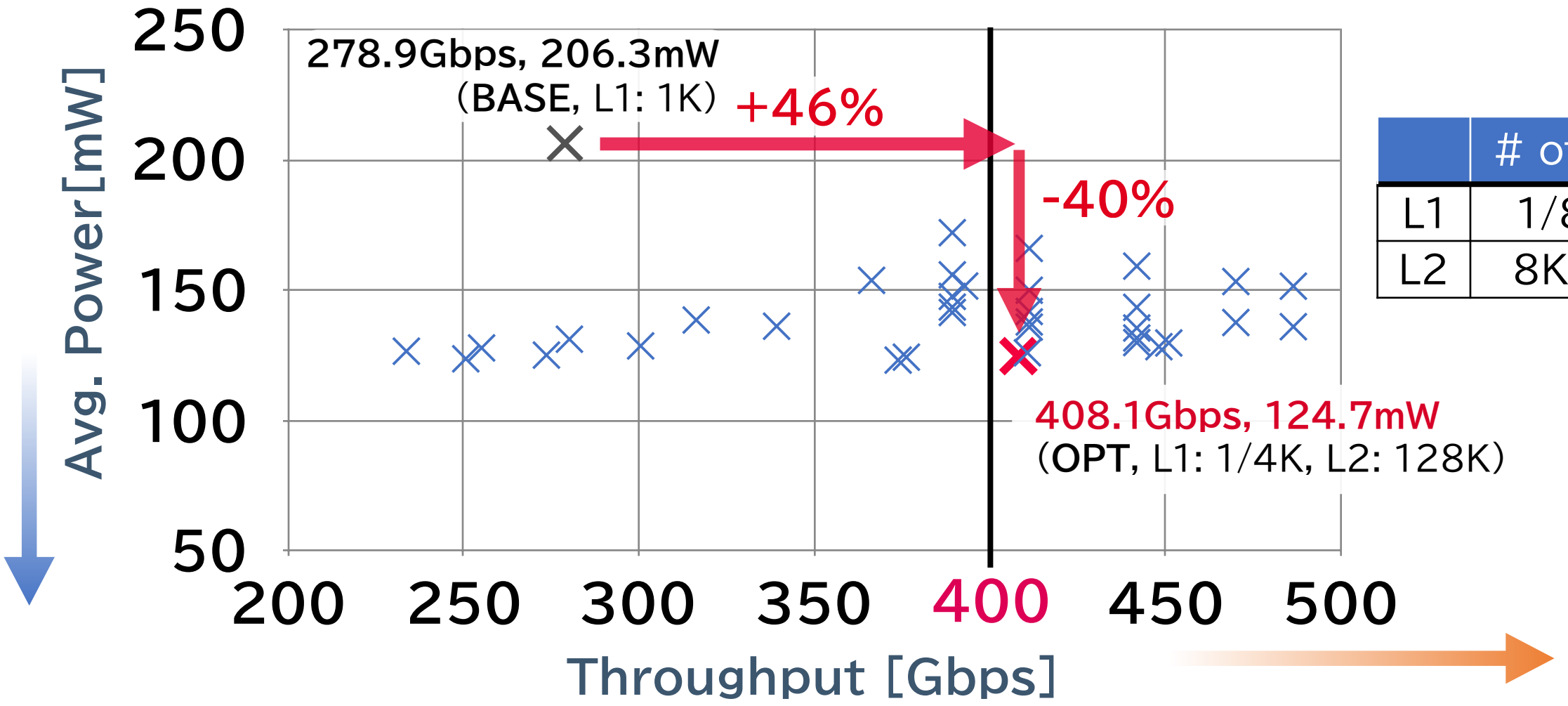
Variables	Description
Th_{PPC} [bps]	Throughput of PPC
L [bit]	Shortest packet length ($= 512bit$)
t_{L1}, t_{L2}, t_{TCAM} [s]	Average lookup latency of L1, L2 caches, TCAM and PPC
m_{L1}, m_{L2}, m_{TCAM}	Miss rate of L1 and L2 caches

Power Model

$$P_{PPC} = \underbrace{\left(DE_{L1} + DE_{L2} \cdot m_{L1} + 4DE_{TCAM} \cdot m_{L1} \cdot m_{L2}\right) \cdot n}_{\text{Dynamic Energy x Accesses / sec}} + \underbrace{SP_{L1} + SP_{L2} + SP_{TCAM}}_{\text{Static Power}}$$

Variables	Description
P_{PPC} [W]	Average power consumption of PPC systems
$DE_{L1}, DE_{L2}, DE_{TCAM}$ [J]	Dynamic energy of L1, L2 caches and TCAM per access
n [packets/s]	Number of packets that arrive at the router per second
$SP_{L1}, SP_{L2}, SP_{TCAM}$ [W]	Static power of L1, L2 caches and TCAM

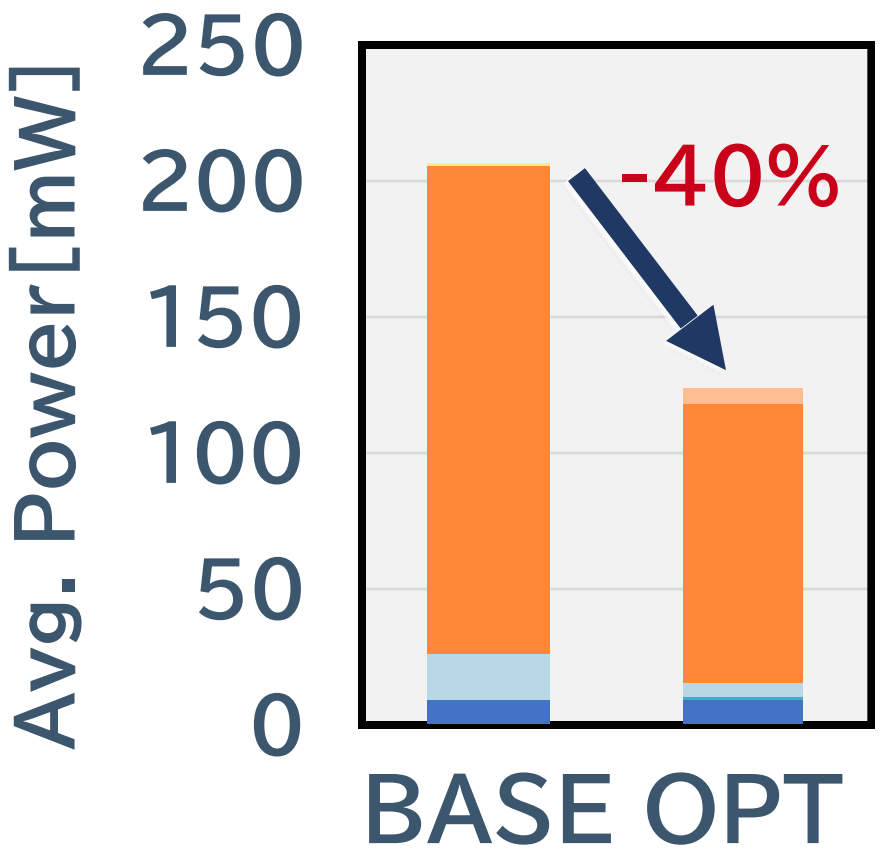
Experimental Result (1/3)



	# of Entries
L1	1/8K - 2K
L2	8K - 512K

Experimental Result (2/3)

Power consumption

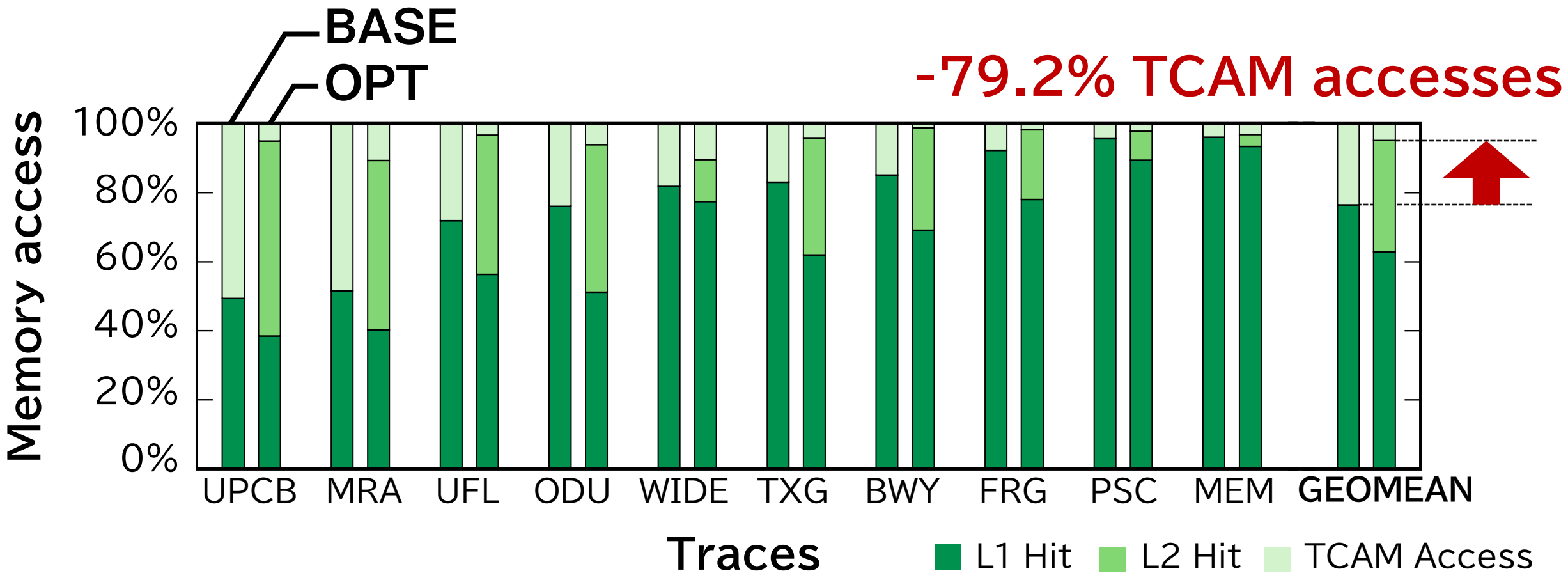


- L2 PPC mainly contributes to **reducing dynamic power of TCAM**
- L2 PPC has little impact on overall power consumption

DP(L1) DP(L2) DP(TCAM)
SP(L1) SP(L2) SP(TCAM)

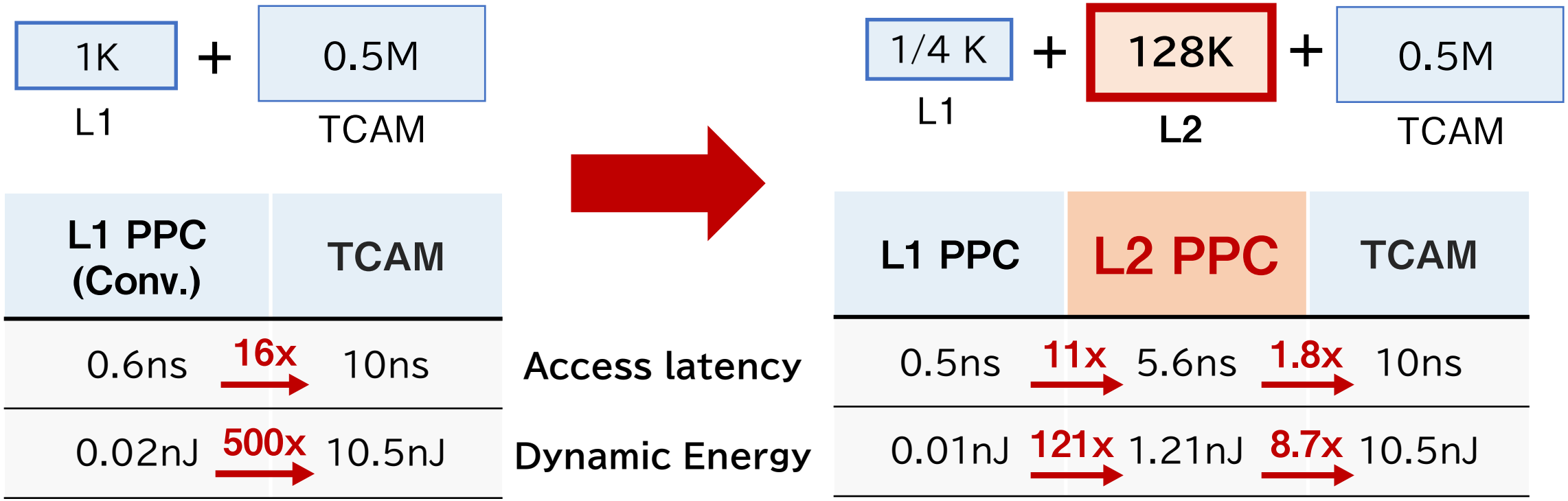
Experimental Result (3/3)

Breakdown of memory accesses



Discussion

- **Reduce the gap** among the caches and TCAM



Conclusion / Future Work

Introducing L2 PPC enables 400Gbps packet processing

Future Work

- Evaluate the impact of L3+ PPCs
- Develop a new insertion policy that uses large caches effectively

Analysis of PPC misses

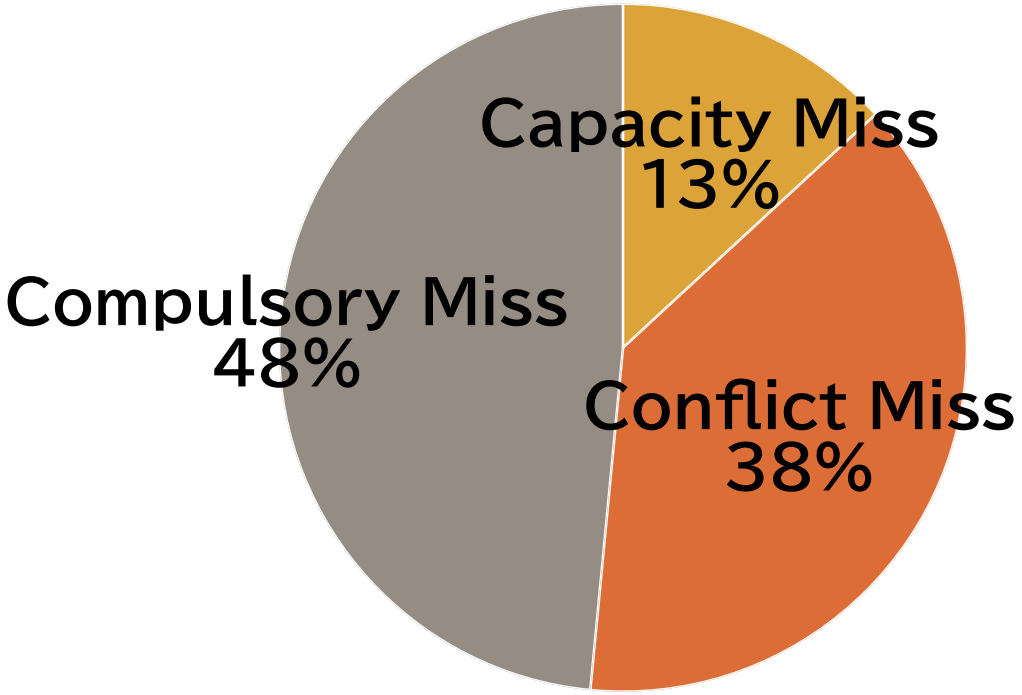
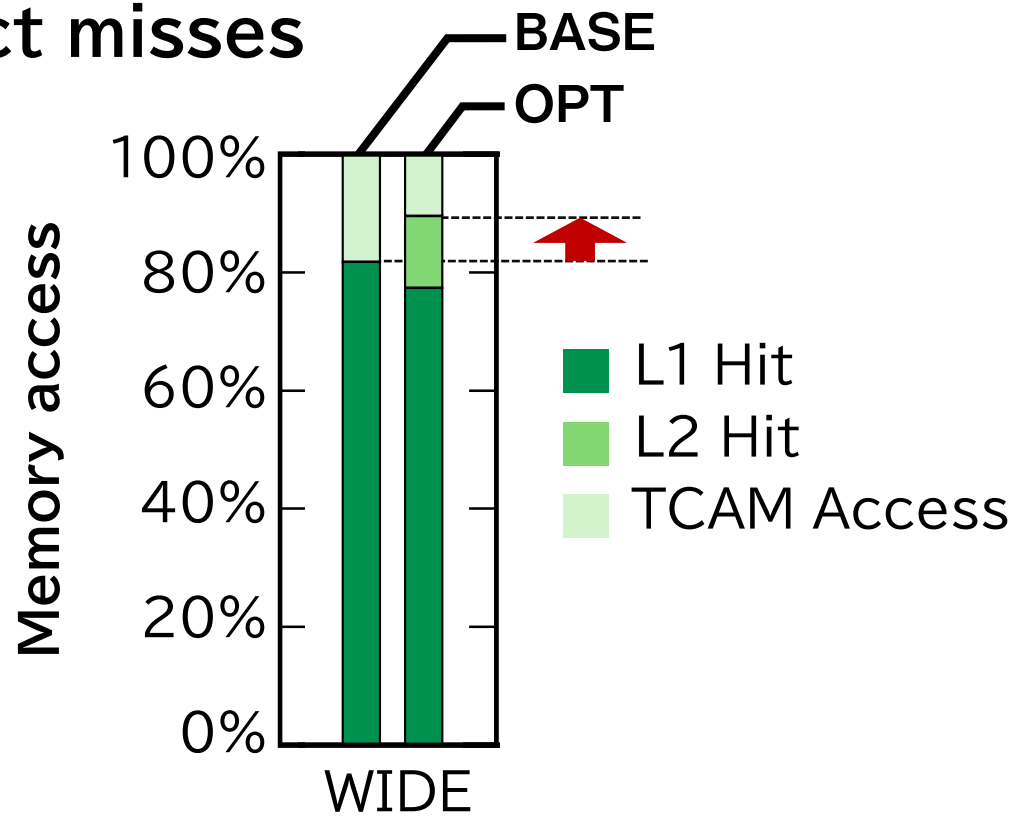


Fig: *Miss causes of 1-level PPC (1K-entry) on WIDE network trace

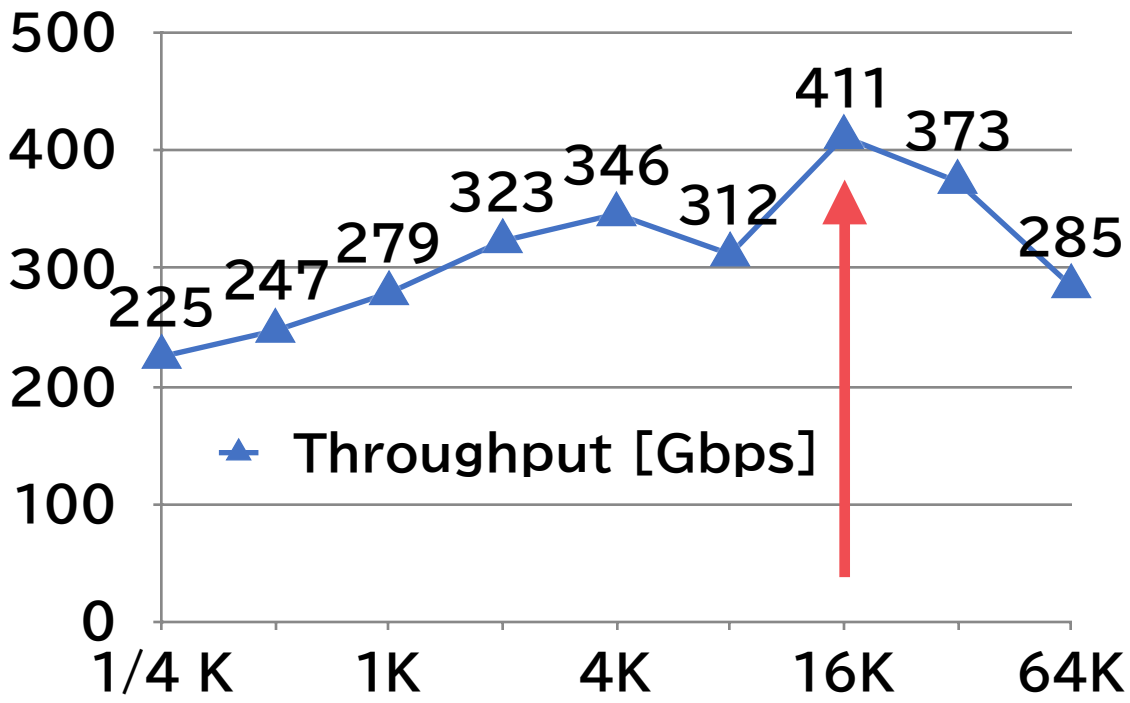
- L2 PPC reduces PPC misses by **42.7%** **(in WIDE network trace)**
- Cut most of **capacity misses** and **conflict misses**

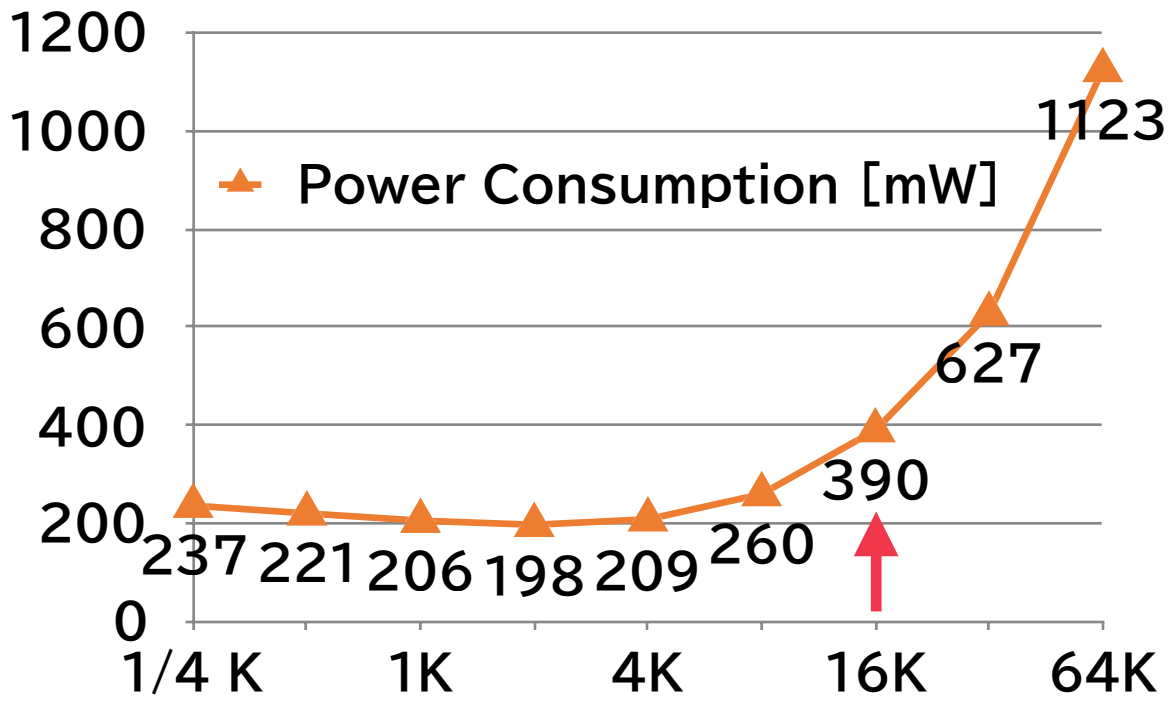


* H. Yamaki, et al.,Data prediction for response flows in packet processing cache, DAC, pp.1-6, 2018.

Experimental Result (Increasing # of L1 entries)

- 16K-entry L1 PPC shows 411.2Gbps with 390.4mW
  - +47.4% throughput with 89.2% more power consumption

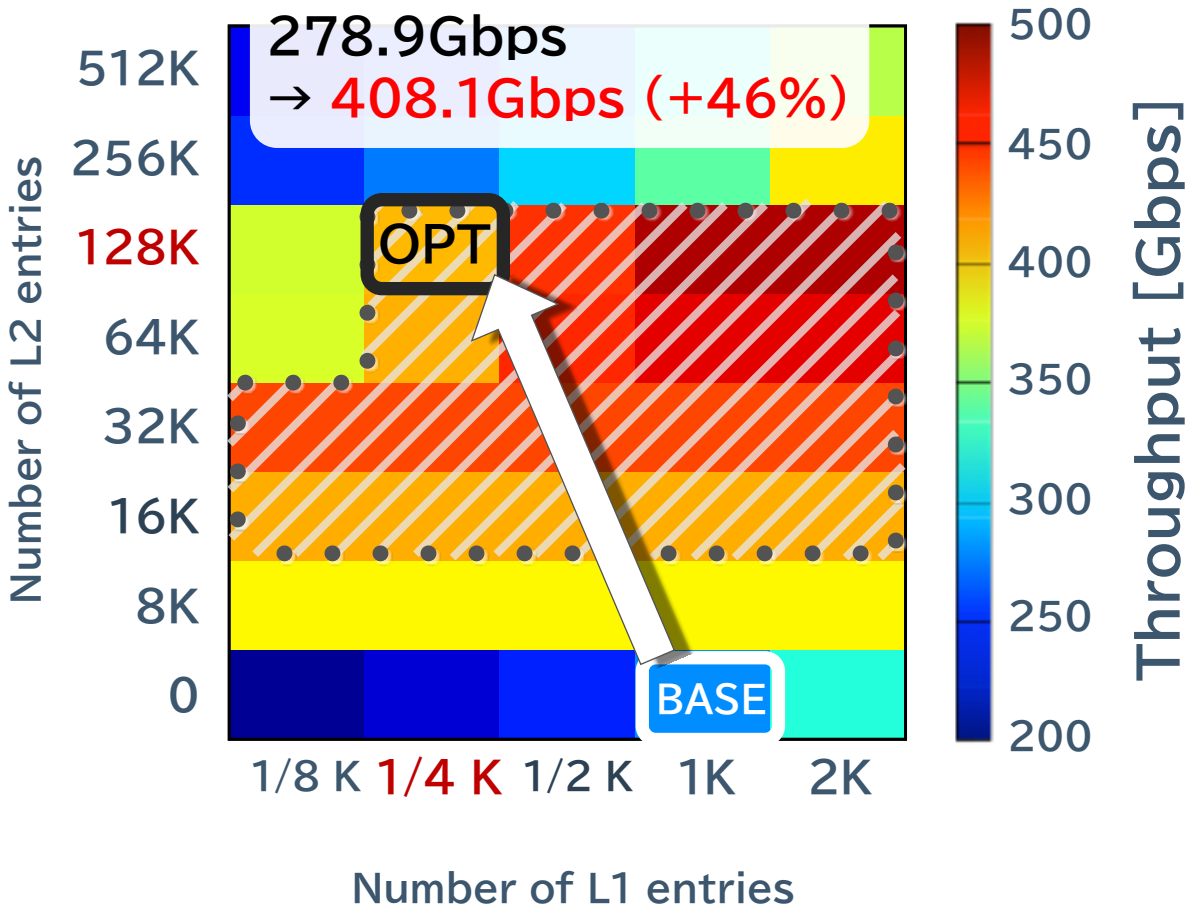




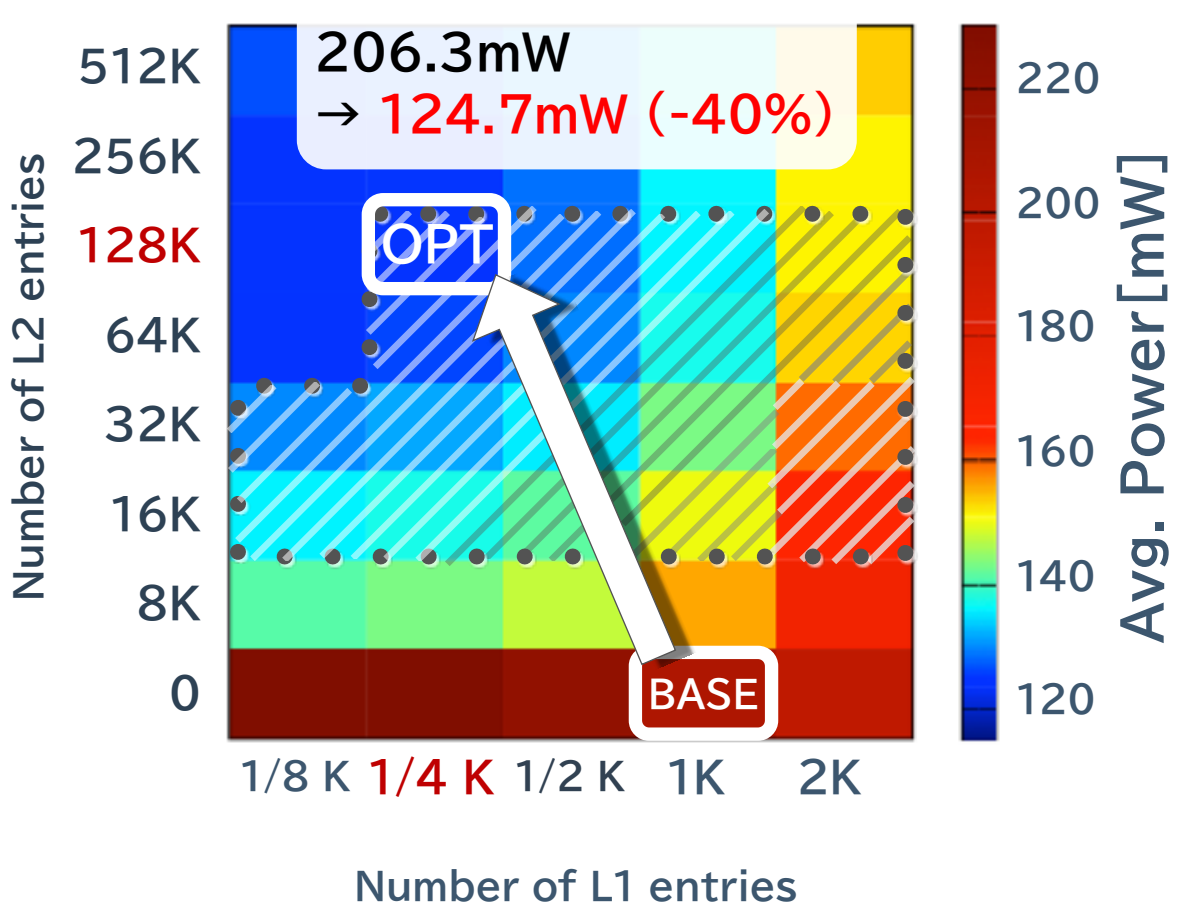
Experimental Result (Heatmaps for WIDE)

Throughput

: 400+ Gbps

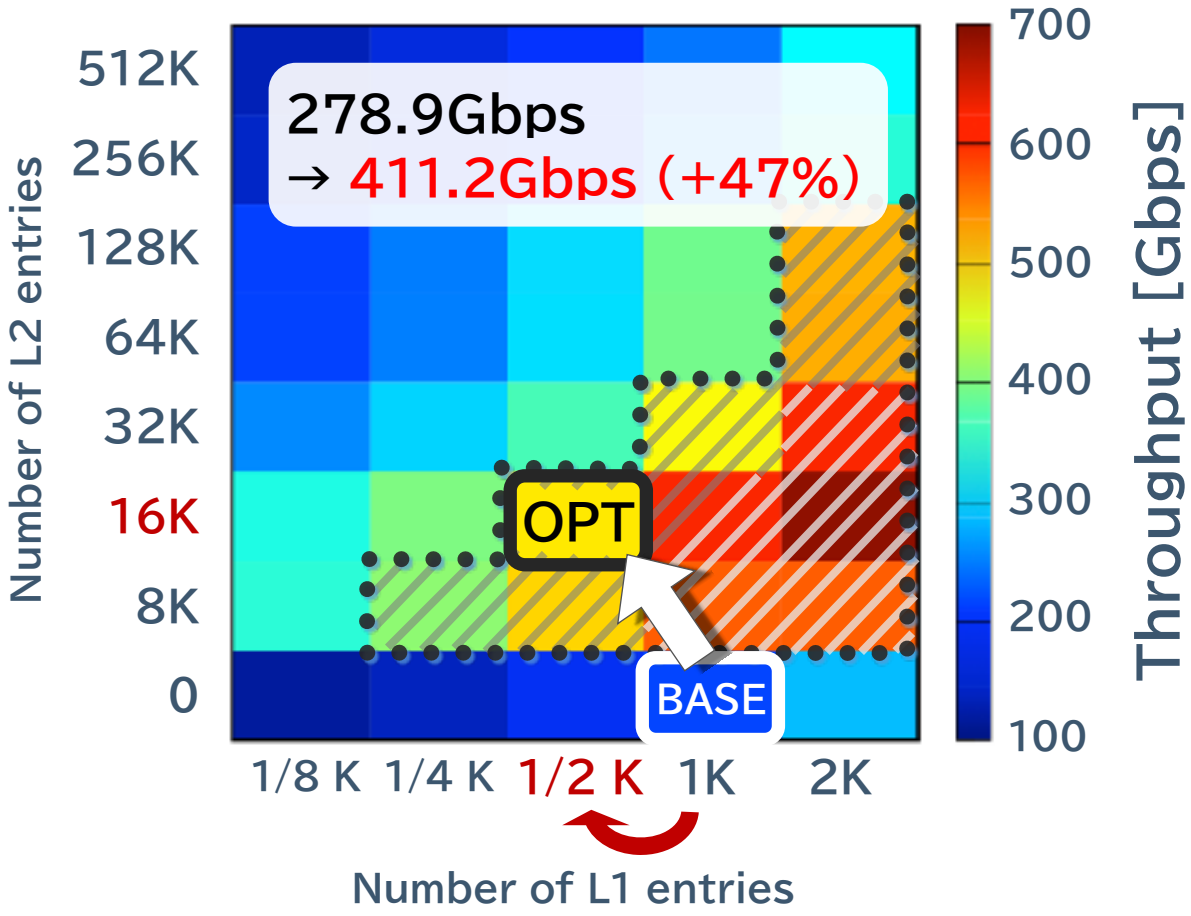


Power consumption

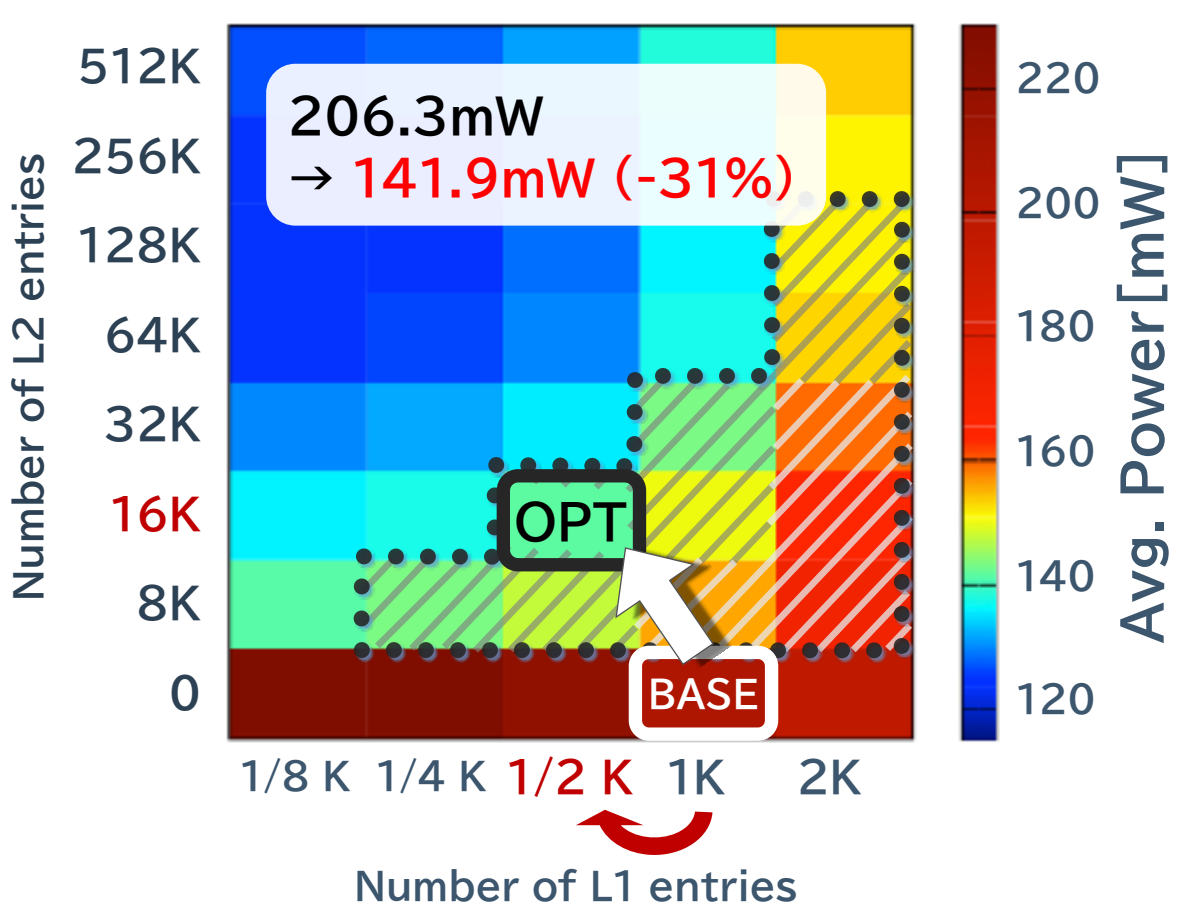


(Based on all network traces)

Throughput : 400+ Gbps



Power consumption



Experimental Result (4/3)

Area

	BASE	OPT
Total [mm^2] (# of entries)	90.22	109.2 (+21%)
L1 (BASE: 1K, OPT: 1/4 K)	0.22	0.14
L2 (128K)	N/A	19.0
TCAM (0.5M)	90	90

- Overhead of L2 cache is small