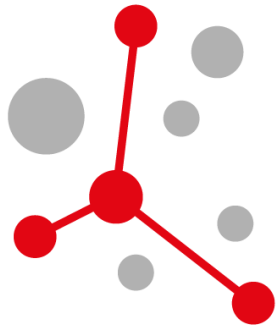
CONNECTING THE DOTS



ISC High Performance

JUNE 10 – 13, 2025 | HAMBURG, GERMANY

Analysis of Application Power Characteristics Using Performance Counters on A64FX

Ryoma Ohara[1] Keiji Yamamoto[2]
Toshihiro Hanawa[1]

1 The University of Tokyo, Japan
2 R-CCS, RIKEN

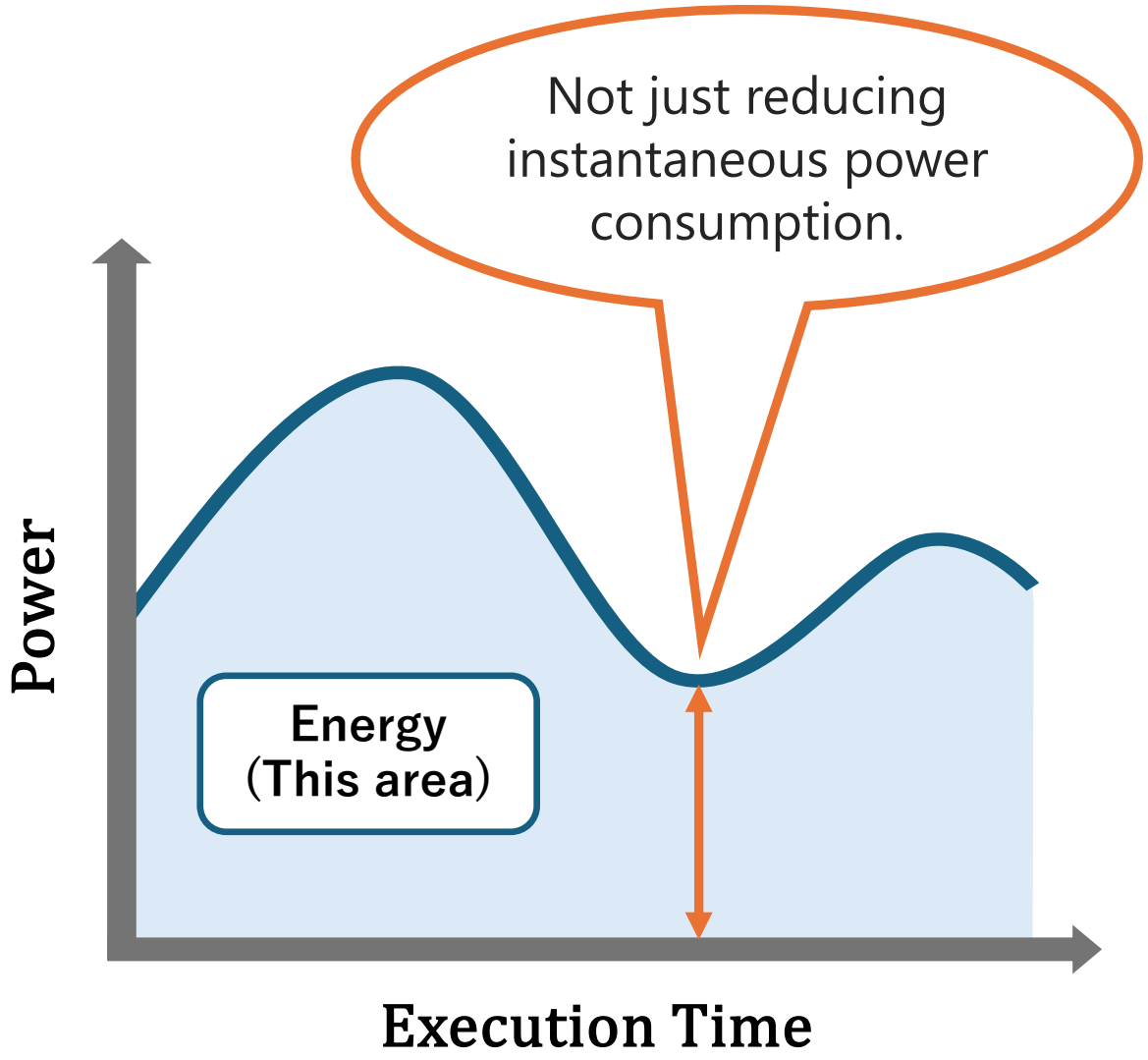
Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Introduction

!! Rising Electricity Costs at Centers

- further increases are not acceptable😥
- Few-percent energy savings
  → Saves hundreds of thousands of dollars in electricity costs

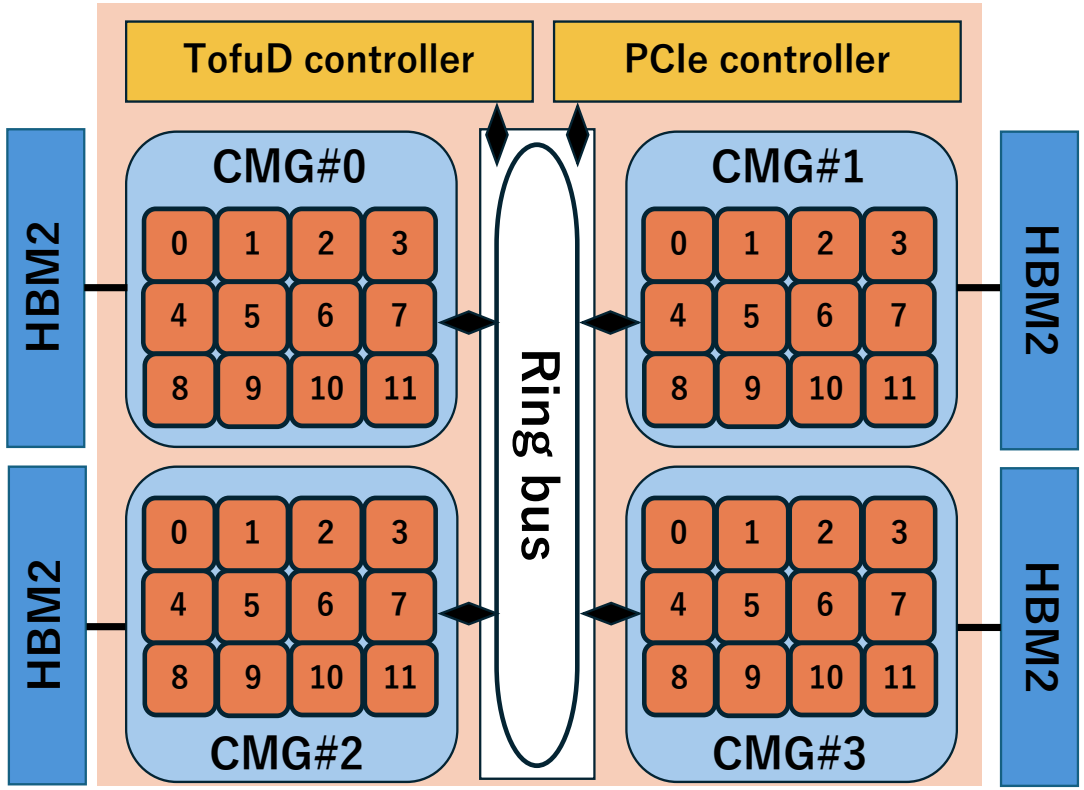


Fugaku

- Total: 158,976 nodes(1 × A64FX per node)
- TDP：180W
- Experimental Setup: 1–2 nodes

Specification of the A64FX[1]

Item	Description
Architecture	Armv8.2-A SVE 512 bit
Number of compute cores	48 cores
CPU frequency	2.0 / 2.2 GHz
Theoretical Performance	Double Precision (2.2GHz) : 3.3792 TFLOPS
Memory	HBM2 32 GiB, 1024 GB/s
Interconnect	Tofu Interconnect D (28 Gbps × 2 lanes × 10 ports)
I/O	PCIe Gen3 × 16
Technology	7nm FinFET



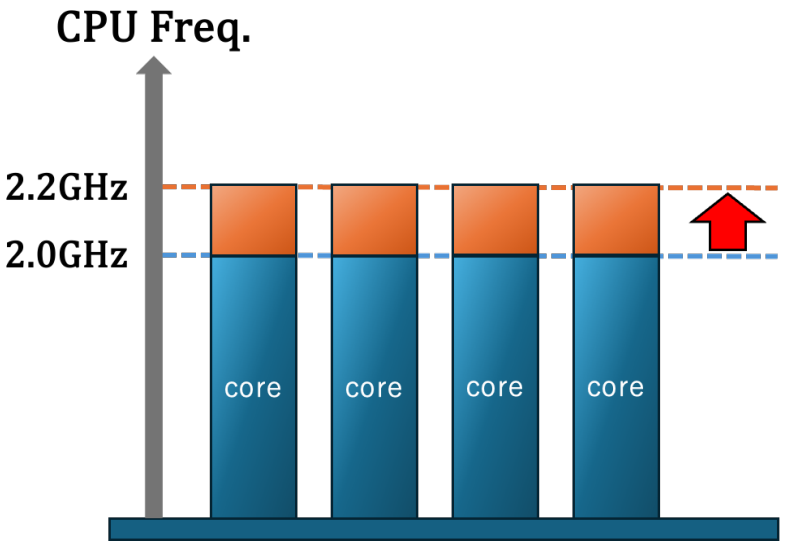
Block diagram of the A64FX CPU

Supercomputer “Fugaku” (RIKEN R-CCS)

Power Knob

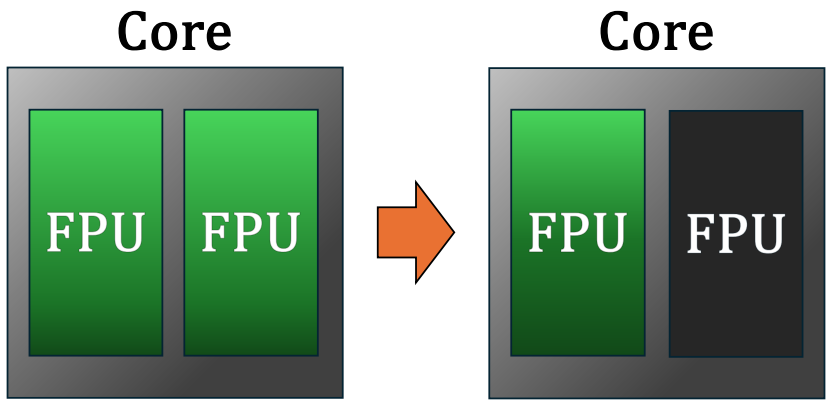
- Mechanisms to improve energy efficiency on Fugaku

Boost



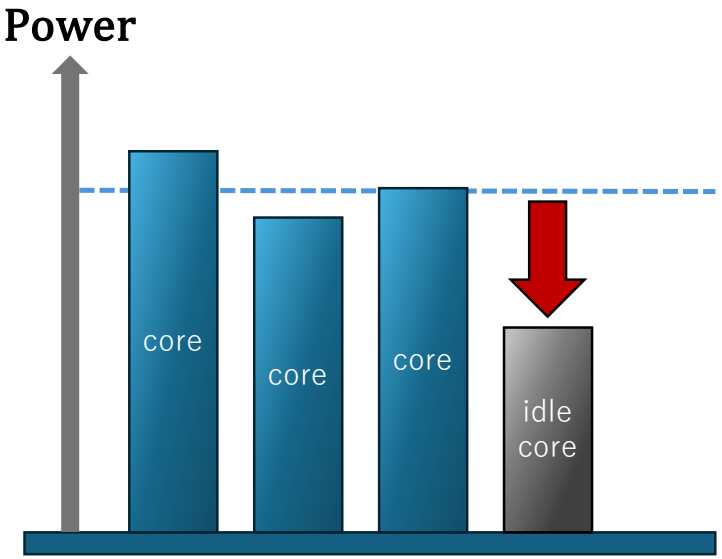
Power Consumption ▲☹

Eco



Power Consumption ▼👍

Retention



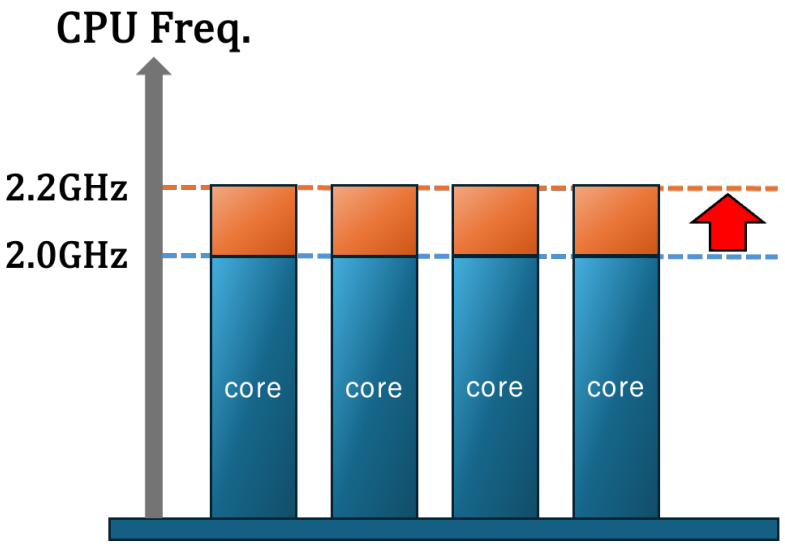
Power Consumption ▼👍

[1] Fugaku – Specifications. https://www.fujitsu.com/global/about/innovation/fugaku/specifications/, Online; accessed 25 May 2025

Power Knob

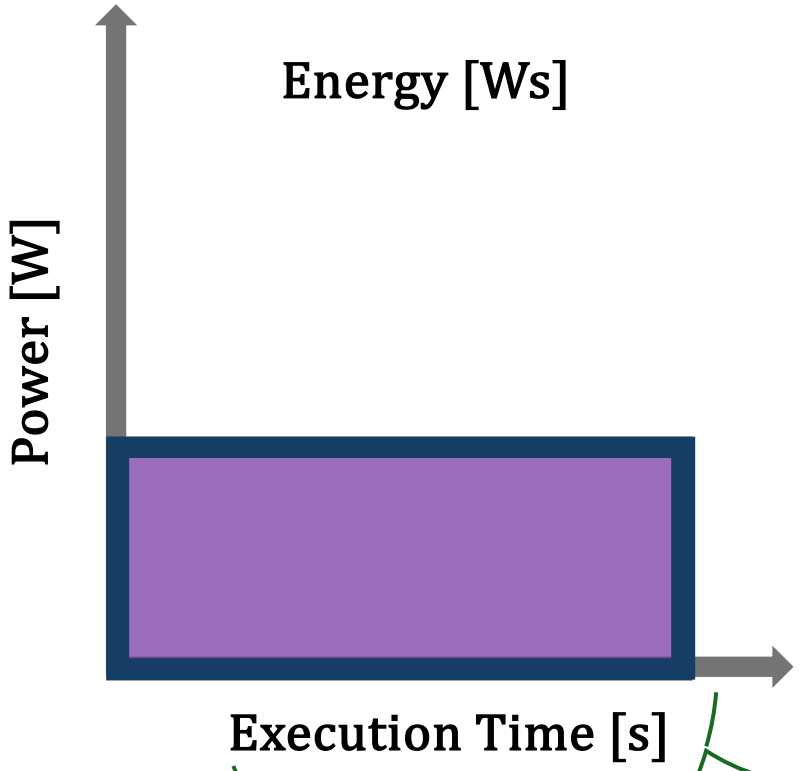
- Mechanisms to improve energy e

Boost



Power Consumption ▲☹

However, it's possible to redece "energy" consumption

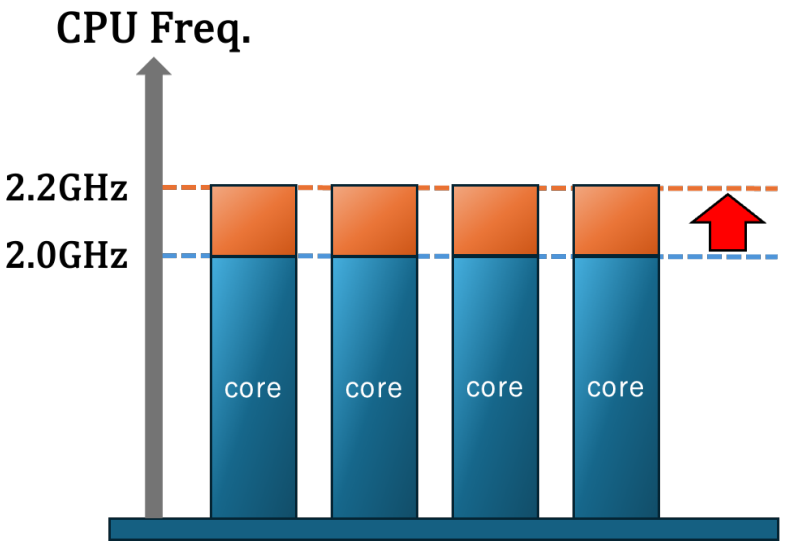


[1] Fugaku – Specifications. https://www.fujitsu.com/global/about/innovation/fugaku/specifications/, Online; accessed 25 May 2025

Power Knob

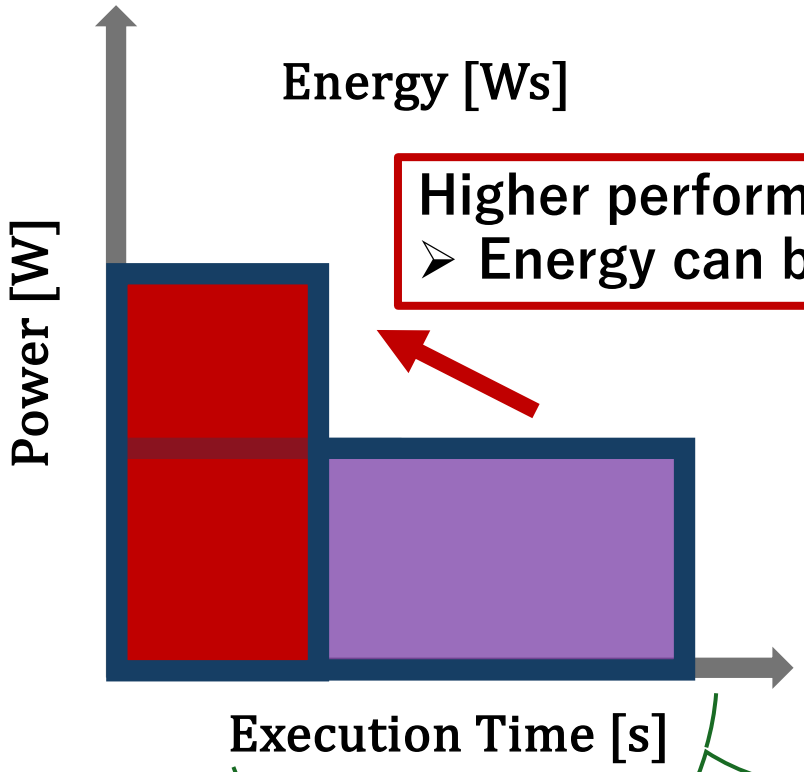
- Mechanisms to improve energy e

Boost



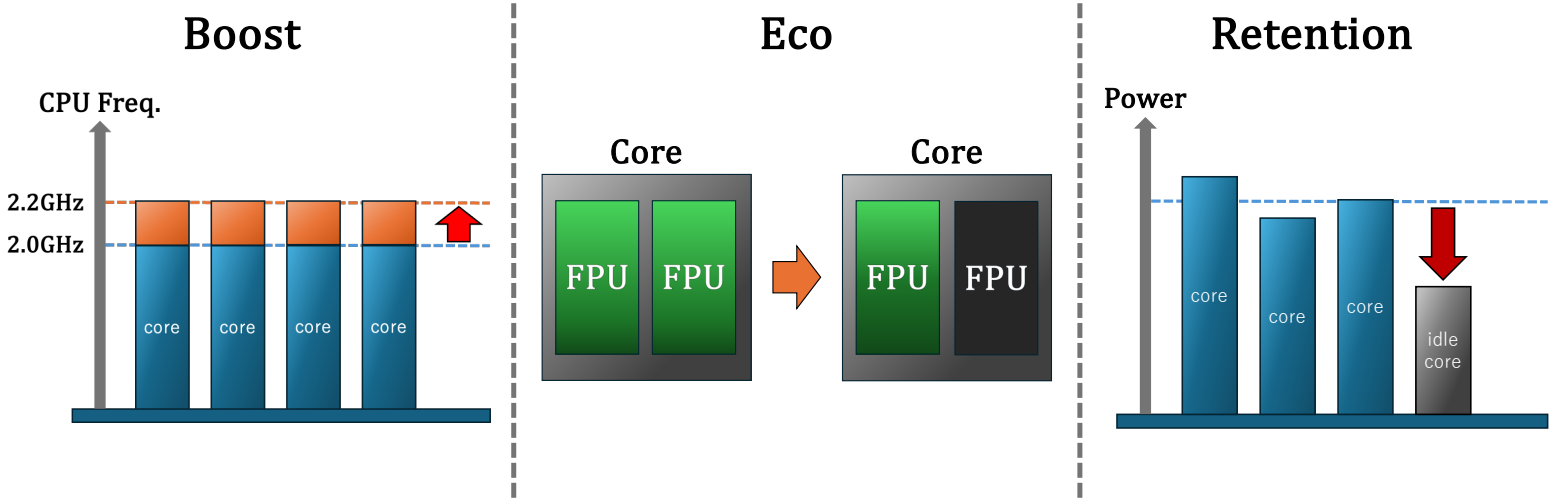
Power Consumption ▲☹

However, it's possible to redece "energy" consumption



Higher performance
- **Energy can be reduced!**

[1] Fugaku – Specifications. https://www.fujitsu.com/global/about/innovation/fugaku/specifications/, Online; accessed 25 May 2025



All eight combinations can be specified.

Power Knob	Frequency [GHz]	# of FPU pipelines	retention mode
normal	2.0	2	off
boost	**2.2**	2	off
eco	2.0	**1**	off
retention	2.0	2	**on**
boost + eco	**2.2**	**1**	off
boost + retention	**2.2**	2	**on**
eco + retention	2.0	**1**	**on**
boost+eco+retention	**2.2**	**1**	**on**

Energy Reduction of Fugaku

Method in this presentation:

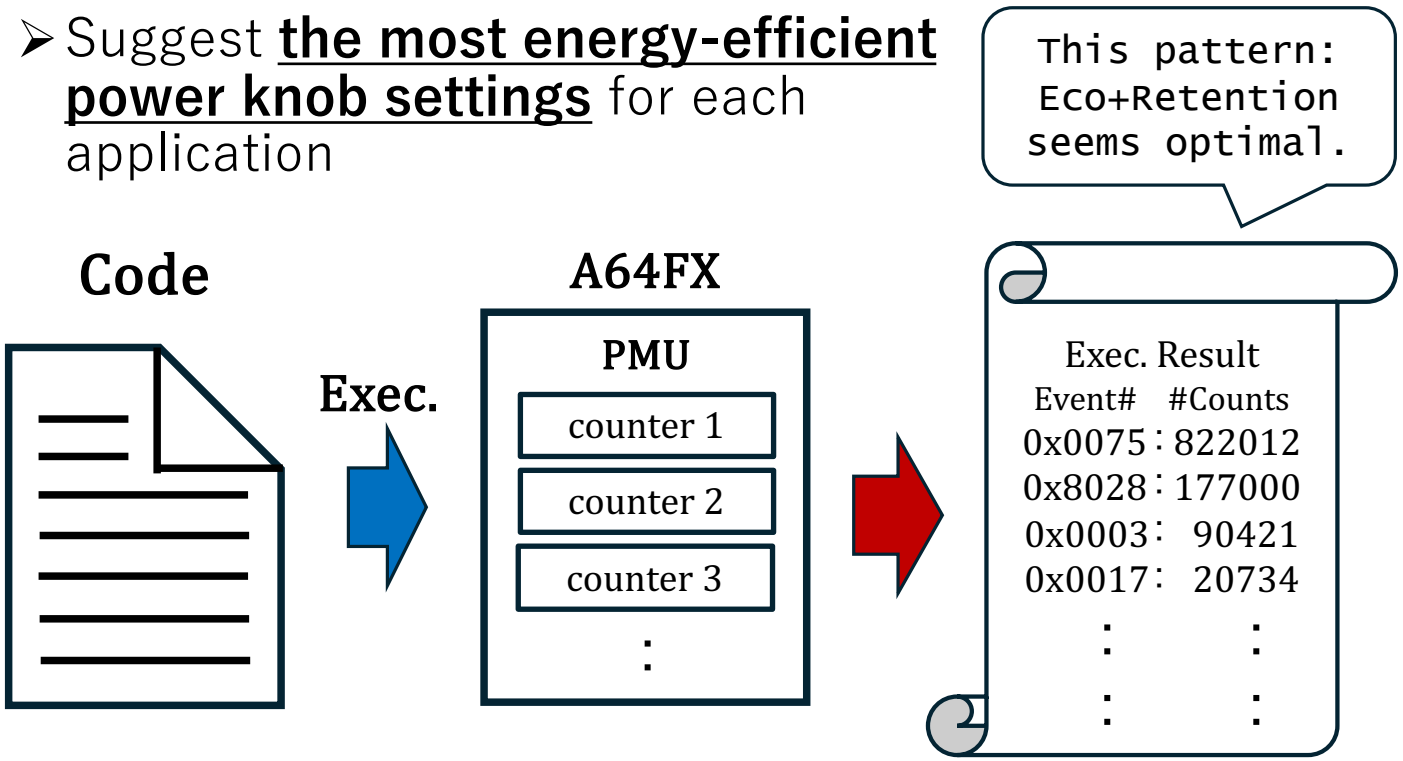
develop

Future work:

[Analysis]

Identify the power characteristics of apps from the **Performance Monitoring Unit (PMU) counter values**

- Suggest **the most energy-efficient power knob settings** for each application



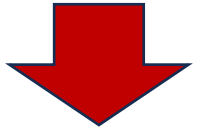
[Machine learning]

Estimation of the optimal power knob based on PMU counter values

Real-time dynamic power knob switching

Presentation Outline

1．Evaluation of microbenchmarks



2．Measurement of PMU counter values for each microbenchmark

3．Estimation of optimal power knob

Investigate the power characteristics for each benchmark.

Investigate the correlation between PMU counter values and power knobs.

Determine the optimal power knob based on metrics calculated from PMU counter values.

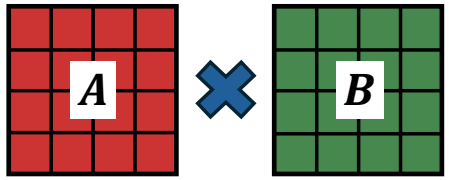
Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Microbenchmarks

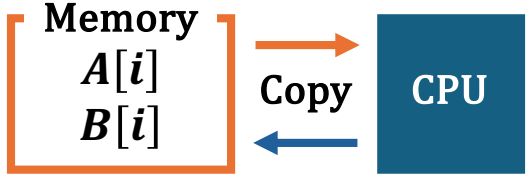
dgemm

Matrix multiplication



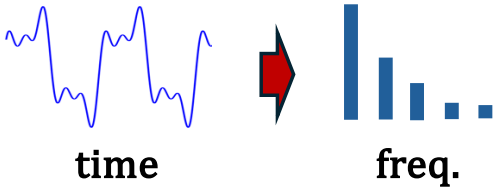
stream

Memory bandwidth test



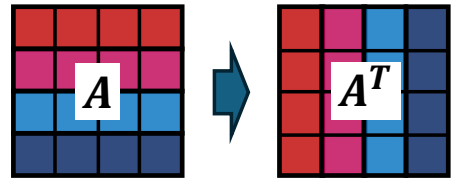
fft

FFT



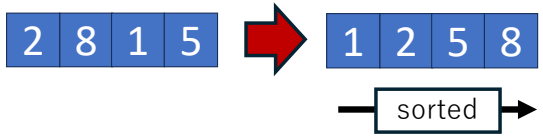
ptrans

Matrix transpose



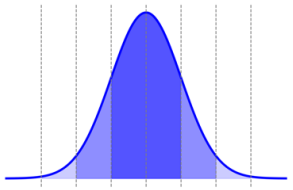
IS

Integer sort



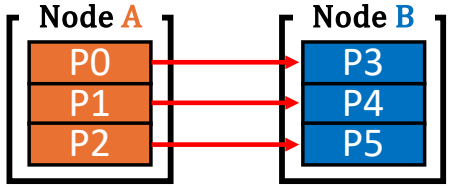
EP

Embarrassingly parallel
(Gaussian distribution calc.)



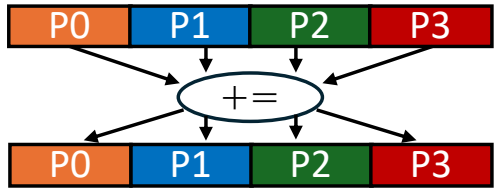
osu_mbw_mr

Inter-node bandwidth



osu_allreduce

Allreduce bet. nodes



※ dgemm, stream, fft, ptrans are selected from [2], IS and EP from [3], and osu benchmarks from [4].

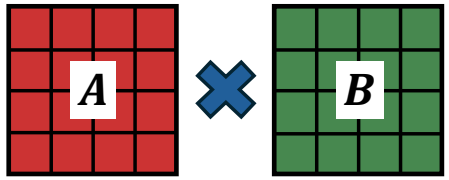
[2] HPC Challenge Benchmark. https://hpcchallenge.org/hpcc/, Online; accessed 19 December 2024.
[3] NAS Parallel Benchmarks. https://www.nas.nasa.gov/software/npb.html, Online; accessed 19 December 2024.
[4] osu-micro-benchmarks, https://github.com/forresti/osu-micro-benchmarks, Online; accessed 28 February 2025

Microbenchmarks

Single node

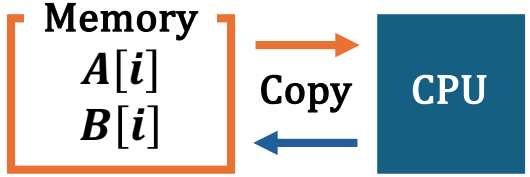
dgemm

Matrix multiplication



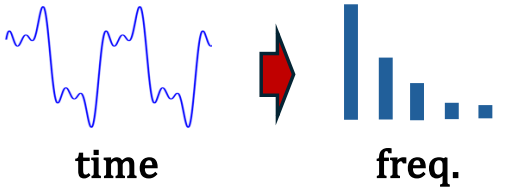
stream

Memory bandwidth test



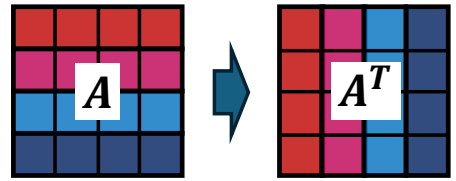
fft

FFT



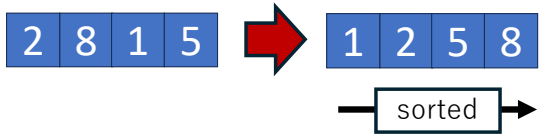
ptrans

Matrix transpose



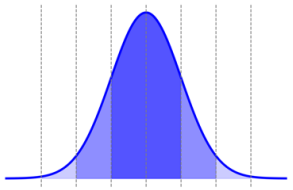
IS

Integer sort



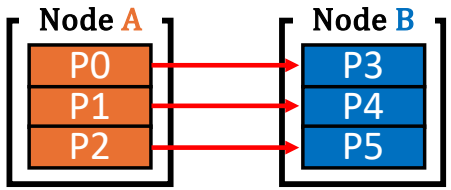
EP

Embarrassingly parallel
(Gaussian distribution calc.)



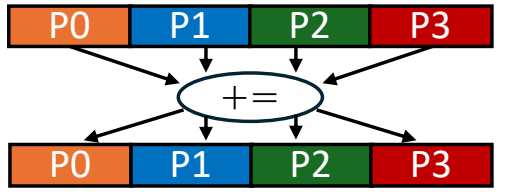
osu_mbw_mr

Inter-node bandwidth



osu_allreduce

Allreduce bet. nodes



Two adjacent nodes

[2] HPC Challenge Benchmark. https://hpcchallenge.org/hpcc/, Online; accessed 19 December 2024.
[3] NAS Parallel Benchmarks. https://www.nas.nasa.gov/software/npb.html, Online; accessed 19 December 2024.
[4] osu-micro-benchmarks, https://github.com/forresti/osu-micro-benchmarks, Online; accessed 28 February 2025

Results – non-MPI benc

- Use **a single node** and change the number of threads
- Use Power API [5] to measure actual power consump

Boost types:

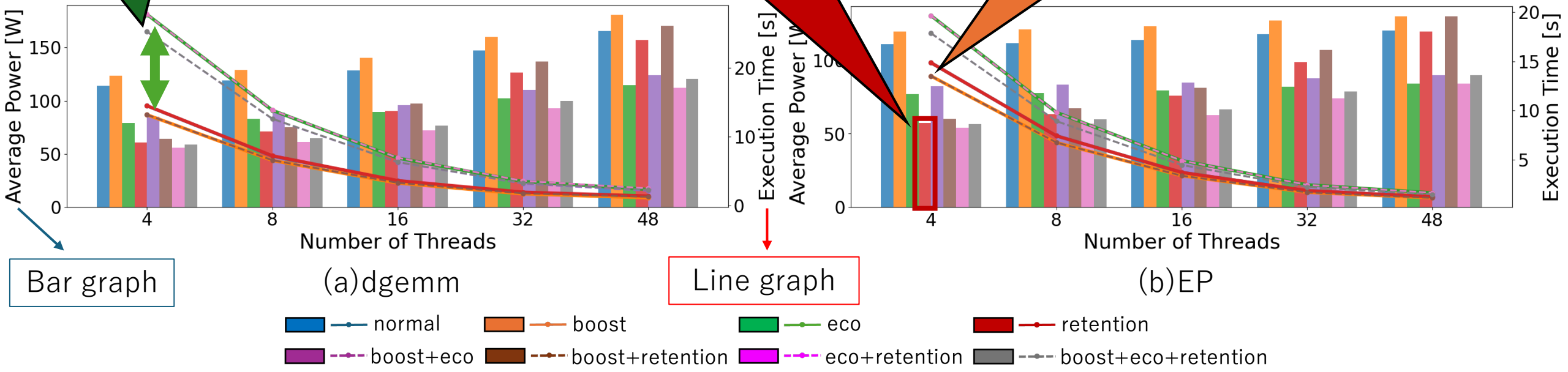
Energy = Power↑ × Exec. Time↓↓

If boost significantly reduces execution time, energy efficiency↑

Eco types: perf. ↓ ↓😩

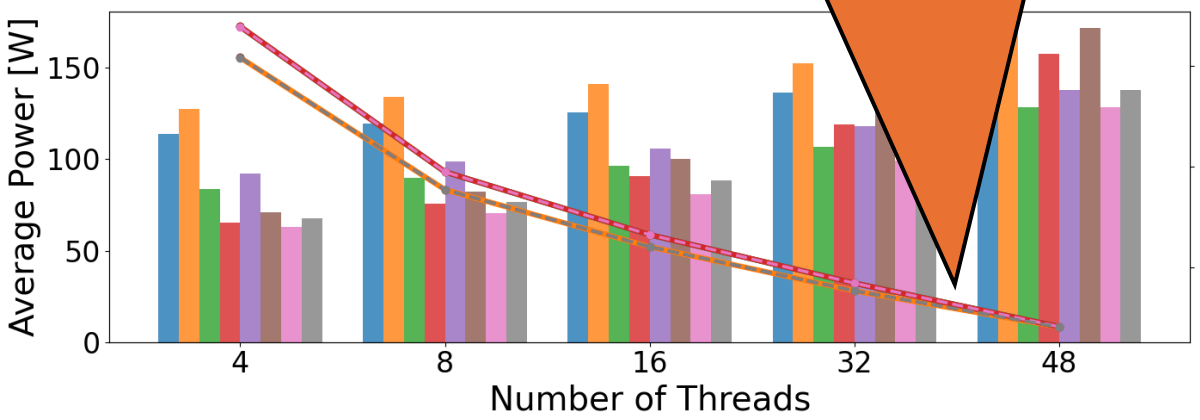
Retention types: with fewer threads, power ↓ ↓👍

Boost: Enhanced perf. over normal 👍



[5] Sandia power api. http://powerapi.sandia.gov/, Online; accessed 19 December 2024.

Boost types: minimal perf. gain😩
(under high-thread conditions)

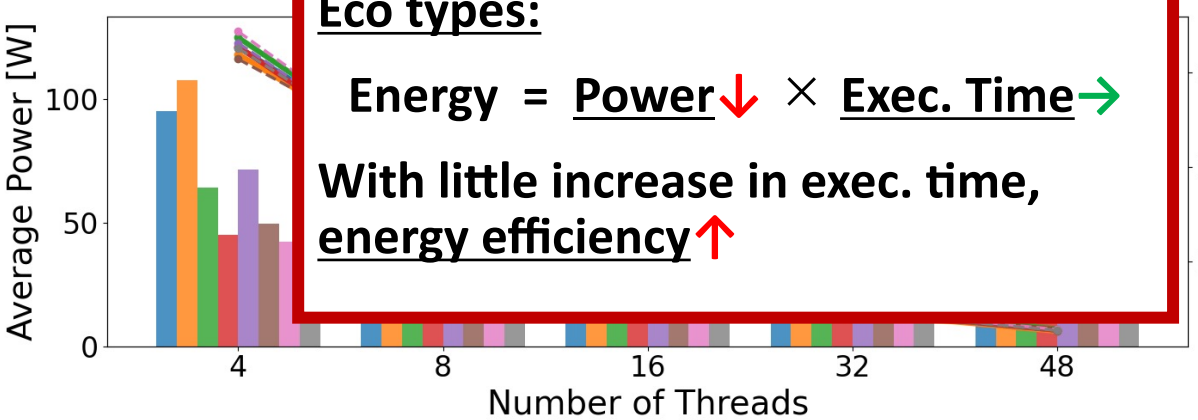


(c)stream

Eco types:

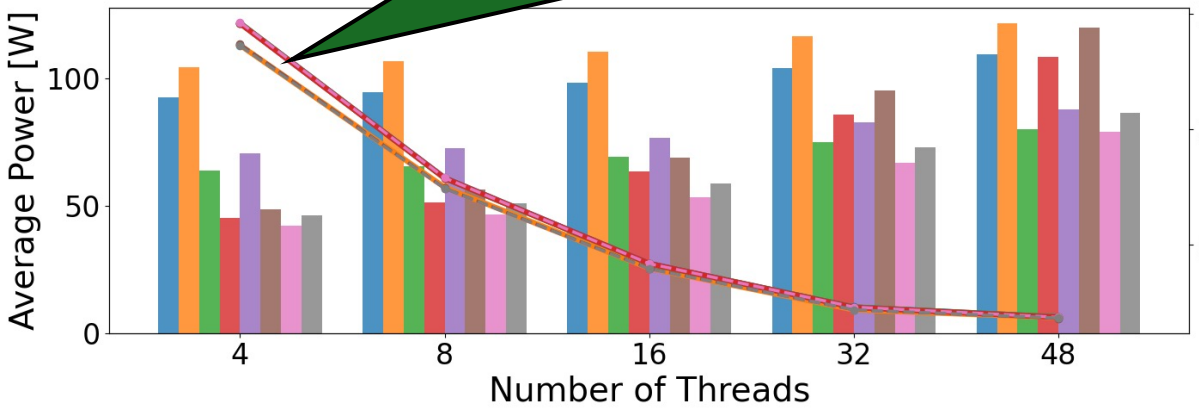
Energy = Power↓ × Exec. Time→

With little increase in exec. time, energy efficiency↑

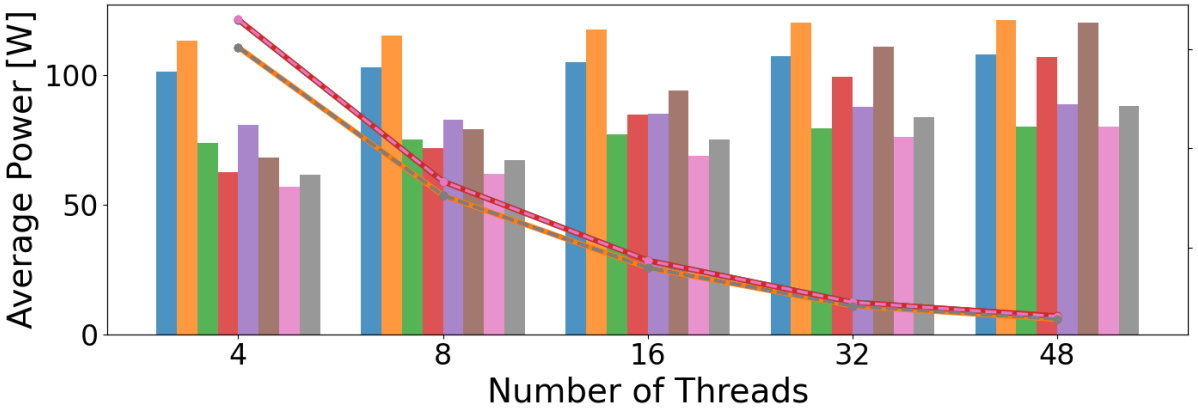


(d)fft

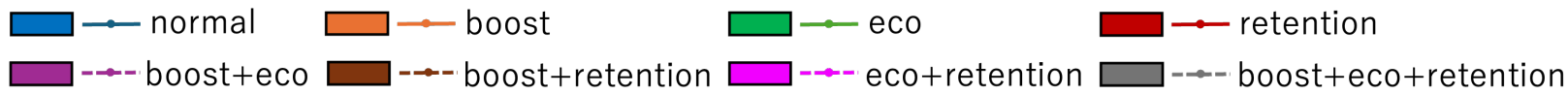
Eco types: no perf. degradation👍



(e)ptrans



(f)IS



Results – MPI benchma

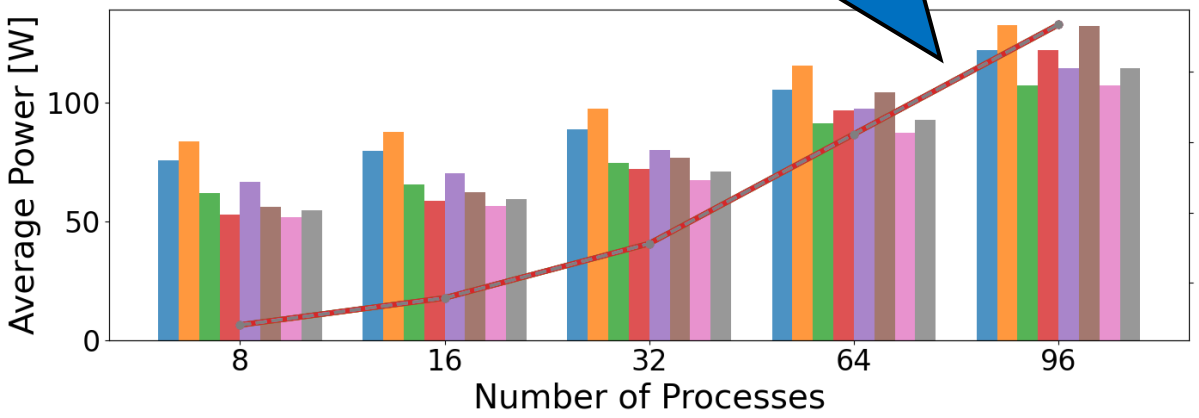
- MPI communication between **adjacent 2 nodes**
- 4, 8, 16, 32, 48 processes per node

Eco types:

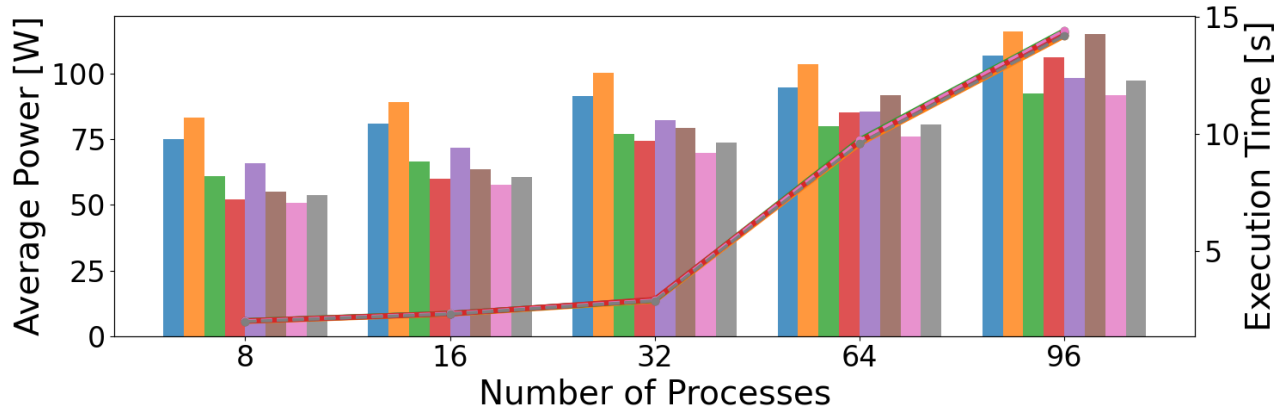
Energy = Power↓ × Exec. Time→

With little increase in exec. time, energy efficiency↑

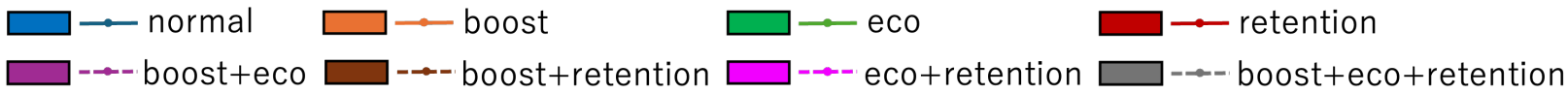
Almost no perf. change



(g)osu_mbw_mr



(h)osu_allreduce



Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Measurement of PMU counter

- For each benchmark, PMU counter values were measured using the fapp command.
  - For example, PMU event numbers were specified in fapp as shown below.
  - fapp -C -d ./pmu -Hevent_raw=0x001b,0x8010,0x8028,0x8034,0x8038,0x0105,0x8043,0x0108 ./dgemm

e.g.

	0x001b	0x8010	0x8028	0x8034	0x8038	0x0105	0x8043	0x0108
dgemm	3.20E+11	1.81E+11	1.81E+11	0	524	15165304	233419	209015

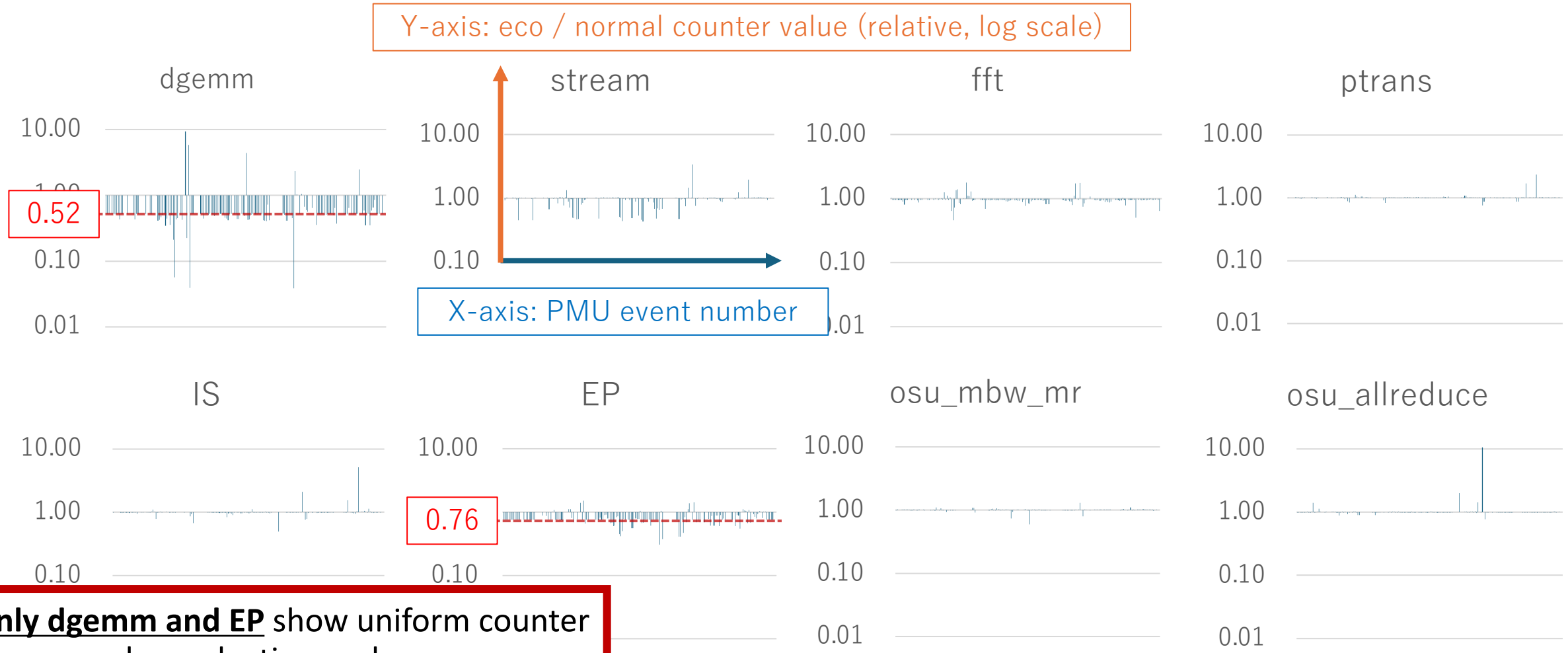
- Counter values were collected for all 183 events listed in A64FX PMU Events [6].
  - Measured with 8, 16, 32, 48 threads and 16, 32, 64, 96 processes.
- Each benchmark was repeated as needed to ensure 15-20 seconds of execution time per measurement.

[6] A64FX PMU Events, https://github.com/fujitsu/A64FX/blob/master/doc/A64FX_PMU_Events_v1.3.pdf, Online; accessed 19 December 2024.

show **how the behavior changes** when switching from normal mode to eco mode.

Results

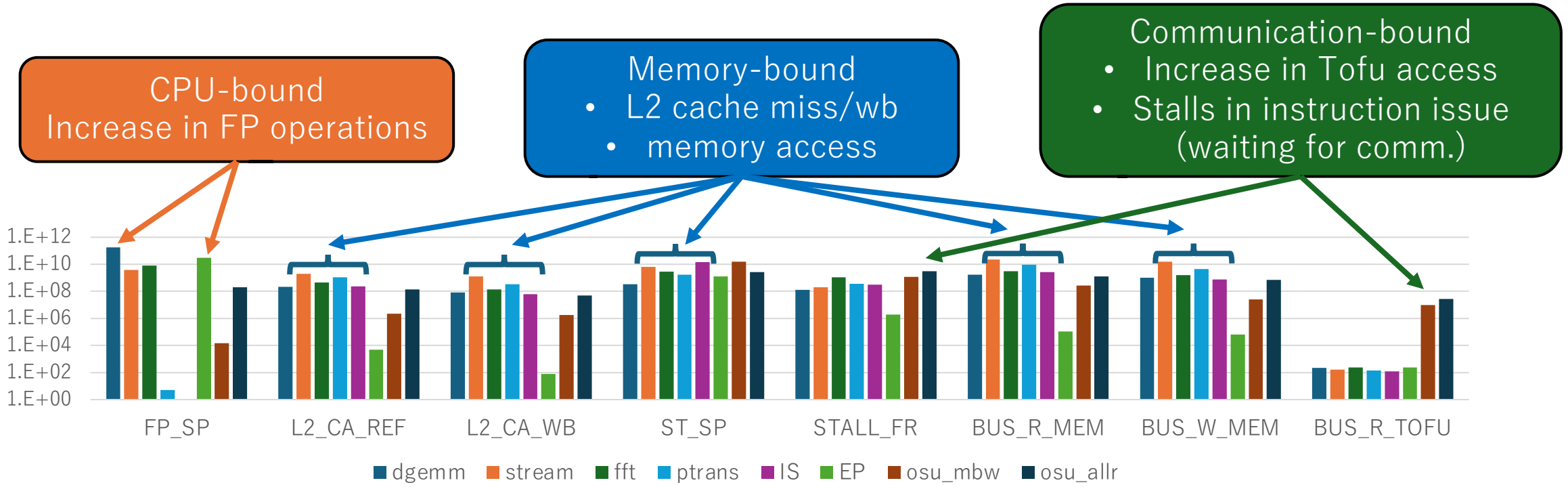
- All events: **eco values normalized by normal values**
- Run with 48 threads (1 node) or 96 processes (2 nodes)



Only dgemm and EP show uniform counter value reduction under eco

➢ Due to **FPU reduction by eco**

Counter values normalized by exec. time



Event Name	Description
FP_SPEC	# of executed FP instructions
L2D_CACHE_REFILL	L2 cache refill count
L2D_CACHE_WB	L2 cache write-back count
ST_SPEC	# of executed store instructions

Event Name	Description
STALL_FRONTEND	Cycles with no issuable instr. (frontend)
BUS_READ_TOTAL_MEM	CMG: Local memory read count
BUS_WRITE_TOTAL_MEM	CMG: Local memory write count
BUS_READ_TOTAL_TOFU	Reads from Tofu controller to CMG

Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Optimal power knob estimation

Based on the above analysis, the most important PMU events for understanding an application's power characteristics are:

① Floating-point operations
② L2 cache refill/write-back
③ Tofu accesses

We propose a method to compute key metrics from relevant PMU counter values and suggest the optimal power knob.

Event Name	Description
INST_SPEC	# of executed instructions
FP_SCALE_OPS_SPEC	# of executed FP instructions
FP_FIXED_OPS_SPEC	# of executed Advanced SIMD and scalar FP instructions
L2D_CACHE_REFILL	L2 cache refill count
L2D_SWAP_DM	Demand access hits in prefetch-prepared buffer
L2D_CACHE_MIBMCH_PRF	Prefetch hits in demand-allocated buffer
LD_SPEC	# of executed load instructions
ST_SPEC	# of executed store instructions
BUS_READ_TOTAL_TOFU	Reads from Tofu controller to CMG
BUS_WRITE_TOTAL_TOFU	Writes from Tofu controller to CMG

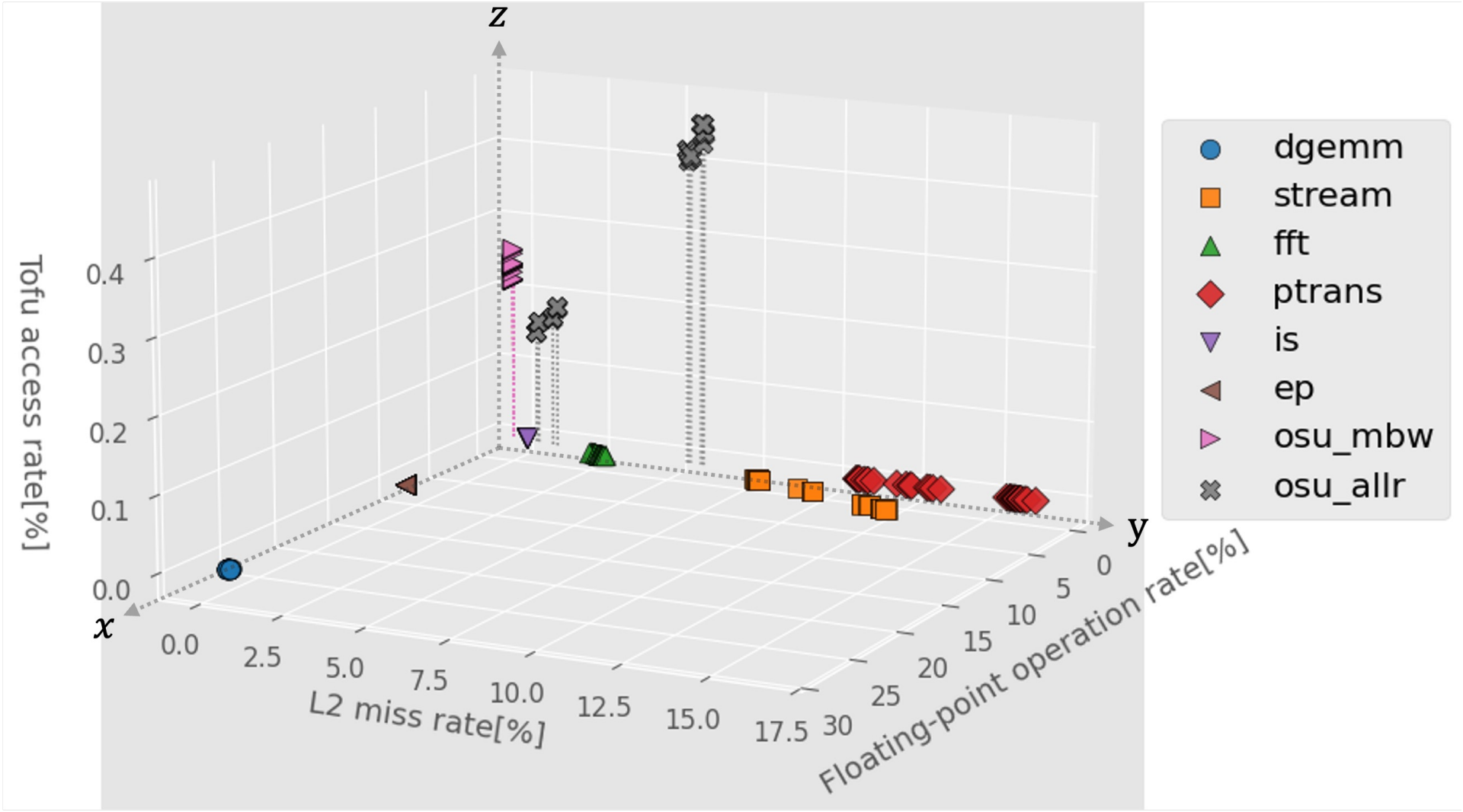
Metrics used for optimal knob estimation (based on Fujitsu Manual)

$$\textbf{Floating-point operation rate} = \frac{\mathrm{FP\_SCALE} \times 512/128 + \mathrm{FP\_FIXED}}{\mathrm{INST\_SP} \times 2 \times 2 \times 8} \quad (1)$$

$$\textbf{L2 miss rate} = \frac{\mathrm{L2\_CA\_REF} - \mathrm{L2\_SW} - \mathrm{L2\_PRF}}{\mathrm{LD\_SP} + \mathrm{ST\_SP}} \quad (2)$$

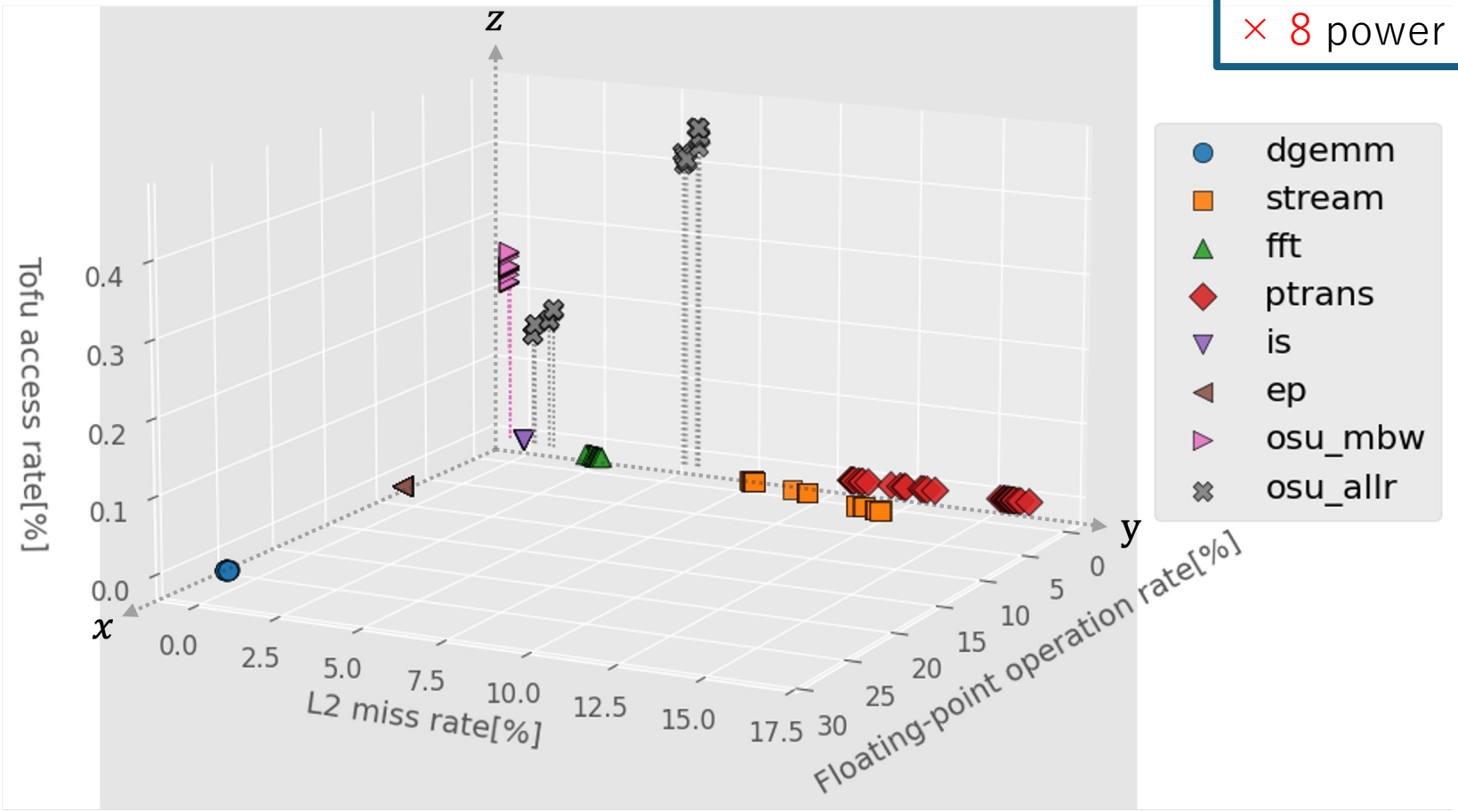
$$\textbf{Tofu access rate} = \frac{\mathrm{BUS\_R\_TOFU} + \mathrm{BUS\_W\_TOFU}}{\mathrm{INST\_SP}} \quad (3)$$

Metric calculation results



Metric calculation results

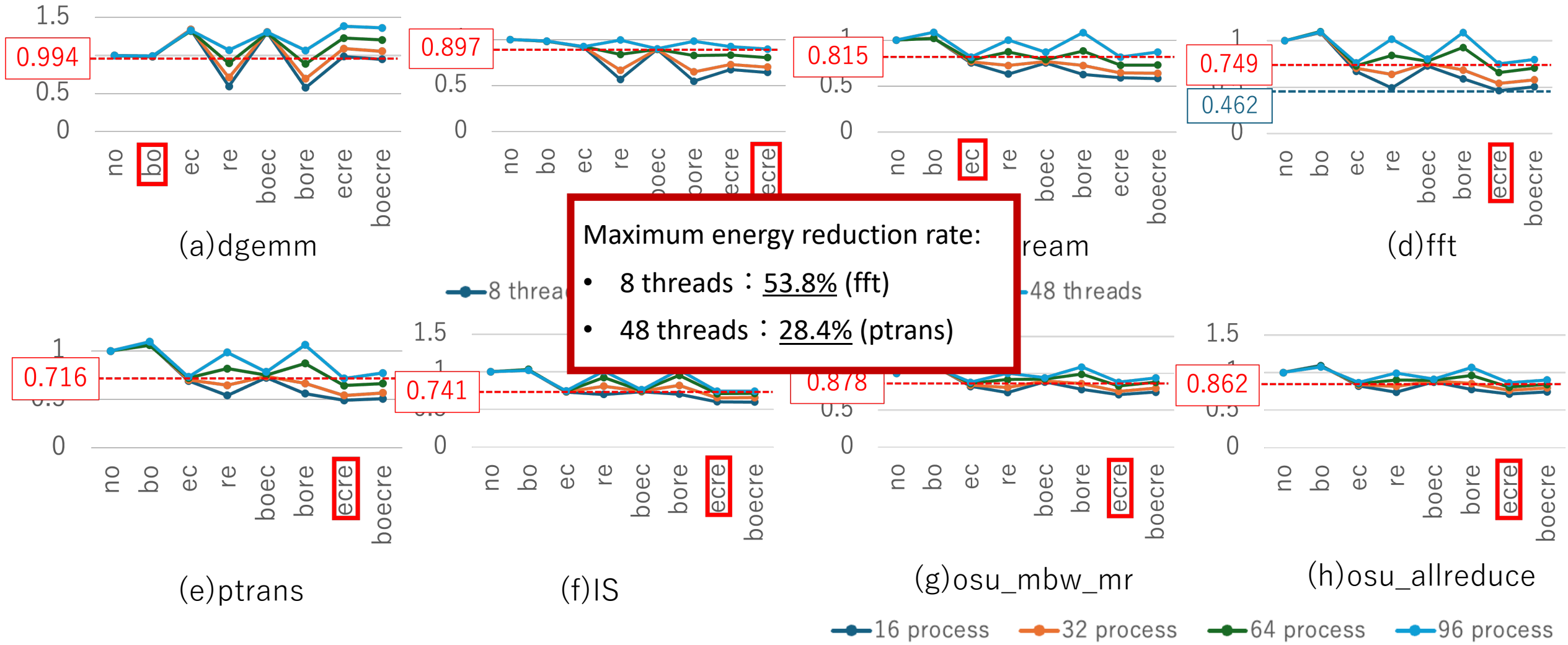
For each benchmark, measurements were taken for 32 combinations= 4 thread/process patterns × 8 power knob settings.



benchmark	optimal power knob
dgemm	boost
EP	boost+eco+retention
stream	eco
fft	eco+retention
ptrans	eco+retention
IS	eco+retention
osu_mbw_mr	eco+retention
osu_allreduce	eco+retention

Optimal Power Knob

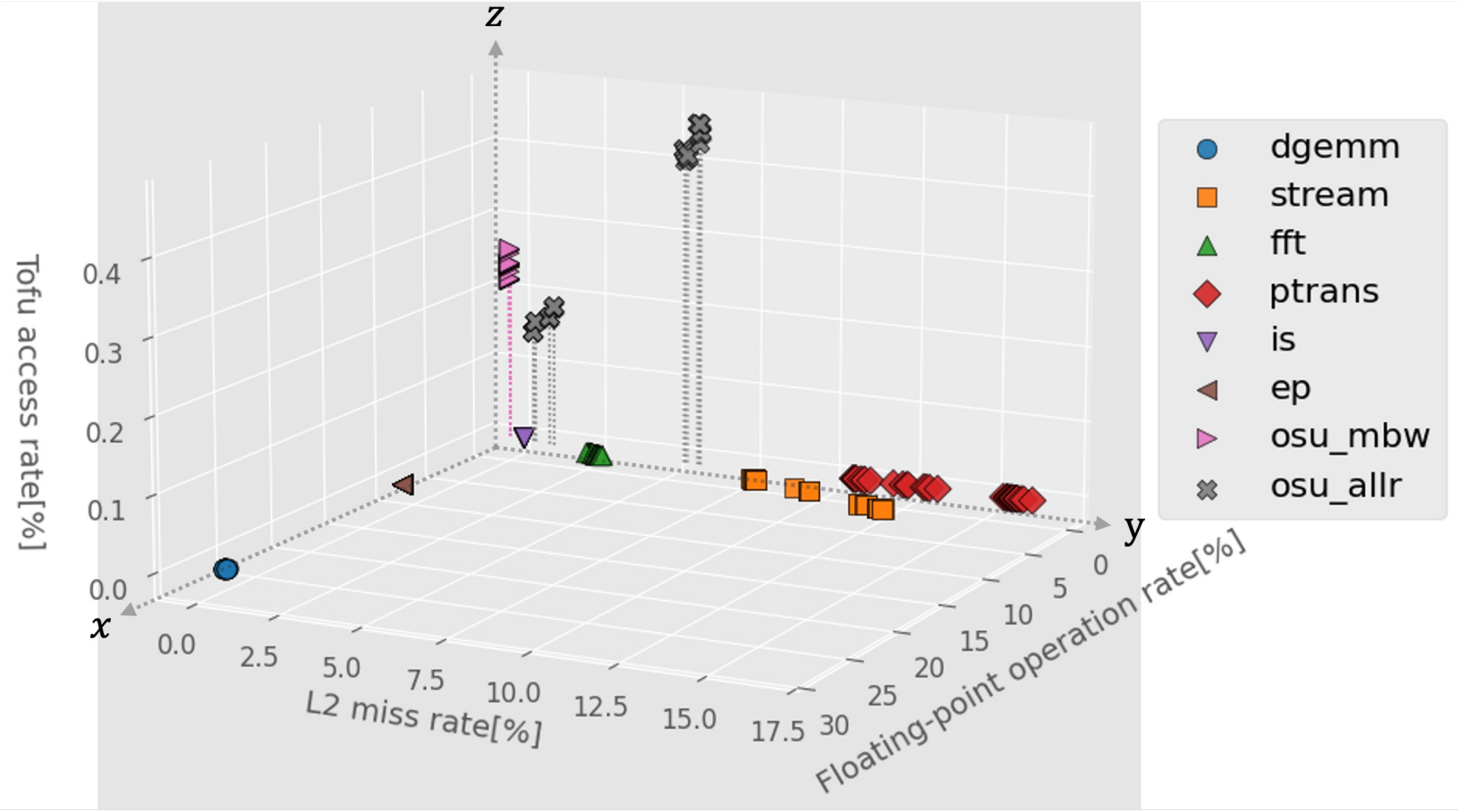
- **Energy consumption** normalized by normal value at each thread/process count
- **Lowest energy consumption** is defined as **optimal power knob** setting



Maximum energy reduction rate:

- 8 threads：53.8% (fft)
- 48 threads：28.4% (ptrans)

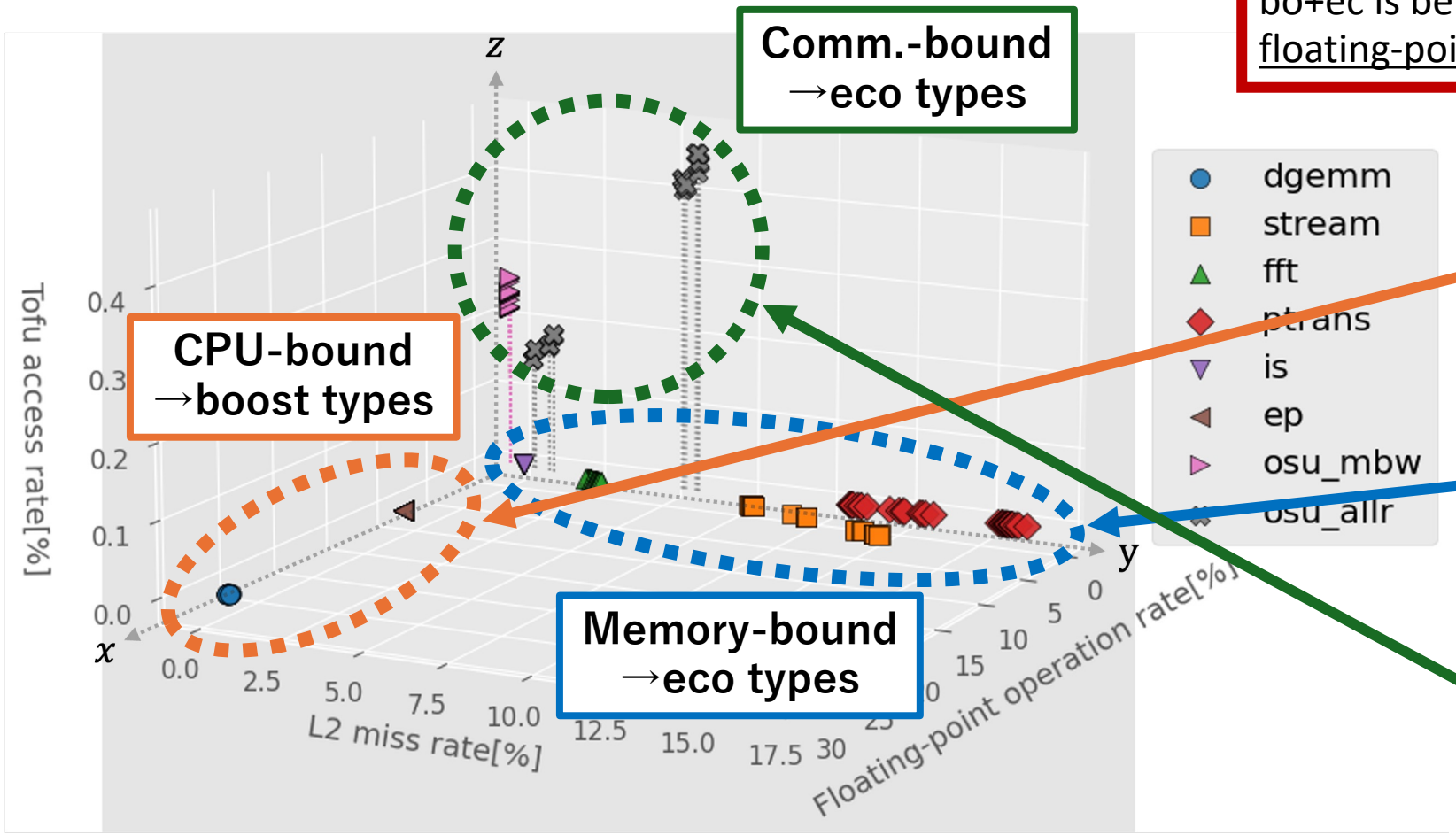
Metric calculation results



benchmark	optimal power knob
dgemm	boost
EP	boost+eco+retention
stream	eco
fft	eco+retention
ptrans	eco+retention
IS	eco+retention
osu_mbw_mr	eco+retention
osu_allreduce	eco+retention

Metric calculation results

EP is optimal with boost+eco, but whether bo or bo+ec is better likely depends on the density of floating-point operations (how large x values are).



benchmark	optimal power knob
dgemm	boost
EP	boost+eco+retention
stream	eco
fft	eco+retention
ptrans	eco+retention
IS	eco+retention
osu_mbw_mr	eco+retention
osu_allreduce	eco+retention

Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Related Work

	Papadimitriou et al. [7]	Kusaba et al. [8]	Fan et al. [9]	Our Study
Target	CPU: X-Gene 2/3, single node	CPU: A64FX system	GPU: AMD MI100, NVIDIA V100 systems	**A64FX, 1~2 node**
Approach	**Dynamic DVFS & core alloc.** during monitoring	**Node reduction** under power constraints	**Fine-grained freq. prediction by ML**	**App analysis for dynamic power knob adjustment**
Method	Optimize settings based on L3 cache access rate	Reduce nodes based on power variation	Static feature extraction at compile time.	**Determine the optimal power knob using only PMU counter values**
Objective	Energy, ED^2P ($ED^2P = Energy \times Delay^2$)	Peak system perf.	EDP, ED^2P, etc.	**Energy**

[7] Papadimitriou, G., Chatzidimitriou, A., & Gizopoulos, D. (2019, February). Adaptive voltage/frequency scaling and core allocation for balanced energy and performance on multicore cpus. In 2019 IEEE international symposium on high performance computer architecture (HPCA) (pp. 133-146). IEEE.
[8] Tomoya Kusaba, Yusuke Awaki, Kohei Yoshida, Shinobu Miwa, Hayato Yamaki, Toshihiro Hanawa, and Hiroki Honda. Power-efficiency variation on a64fx supercomputers and its application to system operation. In 2024 IEEE International Conference on Cluster Computing Workshops (CLUSTER Workshops), pp. 55–65. IEEE, 2024.
[9] Fan, K., D'Antonio, M., Carpentieri, L., Cosenza, B., Ficarelli, F., & Cesarini, D. (2023, November). SYnergy: Fine-grained Energy-Efficient Heterogeneous Computing for Scalable Energy Saving. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 1-13).

Outline

1. Introduction
2. Evaluation of Microbenchmarks
3. Analysis of PMU Counter Values
4. Estimation of Optimal Power Knob Setting
5. Related Work
6. Conclusion & Future Work

Conclusion

To improve energy efficiency on supercomputers, we conducted the following:

- **Analyzed correlations between PMU counter values, application power characteristics, and power knob settings for each benchmark.**
  - Demonstrated that using power knobs can reduce energy consumption by up to 28.4% (48 threads) and 53.8% (8 threads) compared to normal settings, highlighting their effectiveness per application.
  - Through PMU-based power analysis, we revealed correlations among PMU values, app characteristics, and power knob choices.

Towards optimal power knob estimation:
Based on the analysis, we selected key PMU events to estimate app characteristics and proposed a method to choose the optimal power knob using only PMU counter values.

Future Work

Practical Power Knob Estimation

- Increase the number of target applications
- Include more complex, real-world applications
- ➢ Apply machine learning approaches

Dynamic Power Knob Adjustment

- Use ML model to change power knobs dynamically during execution
- Build a system for real-time estimation and control

Portability

- Explore applicability to other systems, especially GPUs

Thank you for your listening!!