Lessons Learned from Power-Saving Operations on Fugaku

Masaaki Terai[*1], Eiji Nagata[*2], Yoshitaka Furutani[*2], Fumichika Sueyasu[*2], Shin'ichi Miura[*1]
[*1)] RIKEN R-CCS, [*2)] Fujitsu Ltd.

Introduction

Backgrounds:
- Electricity accounts for the majority of OPEX in HPC data centers.
- The A64FX chip used in Fugaku provides power control mechanisms [1].
- Cooling equipment (such as chillers) cannot respond to rapid power fluctuations, making system-wide simultaneous implementation risky.

Objective:
Analyze the phased implementation of power-saving features from 2021 to 2025, and evaluate the implementation approach from the system operator's perspective, rather than the user-side incentive perspective (as in related work [2]).

Strategy:
- Power-saving approach in Fugaku:
  - Active nodes (nodes allocated to jobs during execution)
    - Method: Power knobs and core retention
    - Metric: Energy consumption (econ2) obtained via Power API from the job scheduler
  - Idle nodes (nodes not allocated to jobs)
    - Method: Node retention
    - Metric: Total system power consumption measured via facility-side power meters (Azbil)

[1] Kodama et al. 2020. Evaluation of Power Management Control on the Supercomputer Fugaku. In EE HPC SOP Workshop at IEEE International Conference on Cluster Computing (CLUSTER).
[2] Solórzano et al. 2024. Toward Sustainable HPC: In-Production Deployment of Incentive-Based Power Efficiency Mechanism on the Fugaku Supercomputer. In Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC24).

Power-knob Features in the A64FX Chip

The A64FX chip used in Fugaku is equipped with CPU power control mechanisms (power knobs). The available features are:

Table 1. A64FX power control features

Feature	Description
Normal/Boost modes	Adjustable CPU frequency: 2.0GHz / 2.2GHz
Eco mode	Disables one of the two floating-point pipelines
Retention	Places unused cores in standby state to reduce power consumption

These features are independent parameters that can be combined to customize power-saving modes. This study utilizes three power knob features. Four modes × retention (ON/OFF) = eight configurations.

Table 2. Power-knob mode configurations

Mode	Frequency	FP Pipelines	Retention
Normal	2.0GHz	Dual	ON/OFF
Normal-eco	2.0GHz	Single	ON/OFF
Boost	2.2GHz	Dual	ON/OFF
Boost-eco	2.2GHz	Single	ON/OFF

Power-saving Approach for Active Nodes

- **Method**: Power knobs and core retention during job execution
- **Results** (Fig. 1): Distribution of node-hours by power knob mode from Mar. 2021 to Dec. 2025
  - Started with opt-in retention for small jobs
  - Gradually expanded default coverage of core retention
  - Culminated in boost-eco mode as the default setting

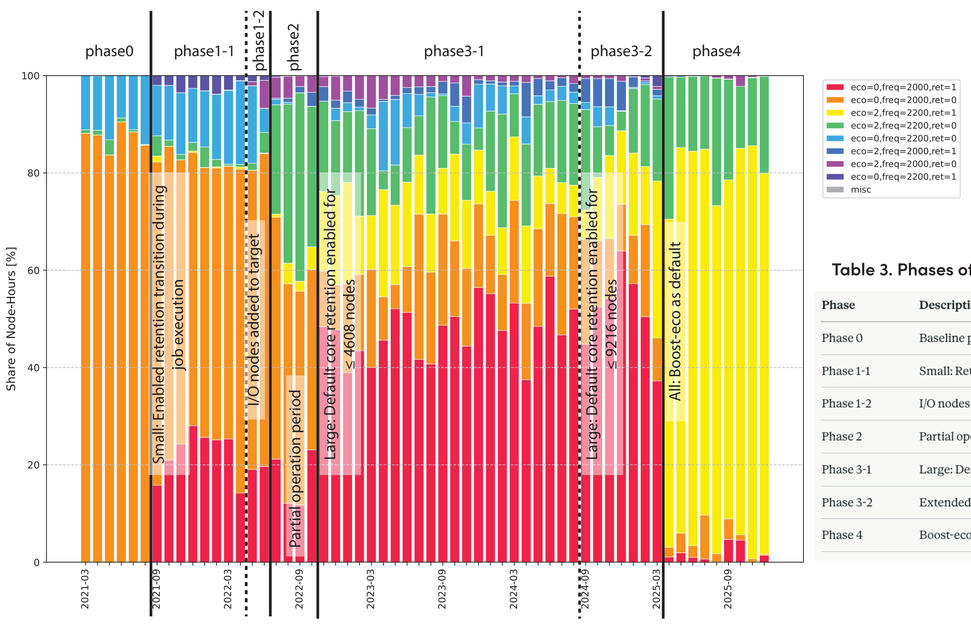

Table 3. Phases of power-saving implementation

Phase	Description
Phase 0	Baseline period
Phase 1-1	Small: Retention transition enabled during job execution
Phase 1-2	I/O nodes added to target
Phase 2	Partial operation: due to electricity price surge
Phase 3-1	Large: Default core retention for ⩽4608 nodes
Phase 3-2	Extended default core retention to ⩽9216 nodes
Phase 4	Boost-eco mode as default for all resources

Figure 1. Node-hour breakdown by power knob mode

- **Results** (Fig.2): Average power per node across implementation phases
  - Clear downward trend in average node power consumption from Phase0 to Phase4
  - Overall reduction of 29.4%
  - Variance (box size and whisker length) remains relatively consistent across phases, indicating diverse job workloads.
  - The median tends to be higher than the mean in later phases, suggesting a distribution skewed toward lower power values.

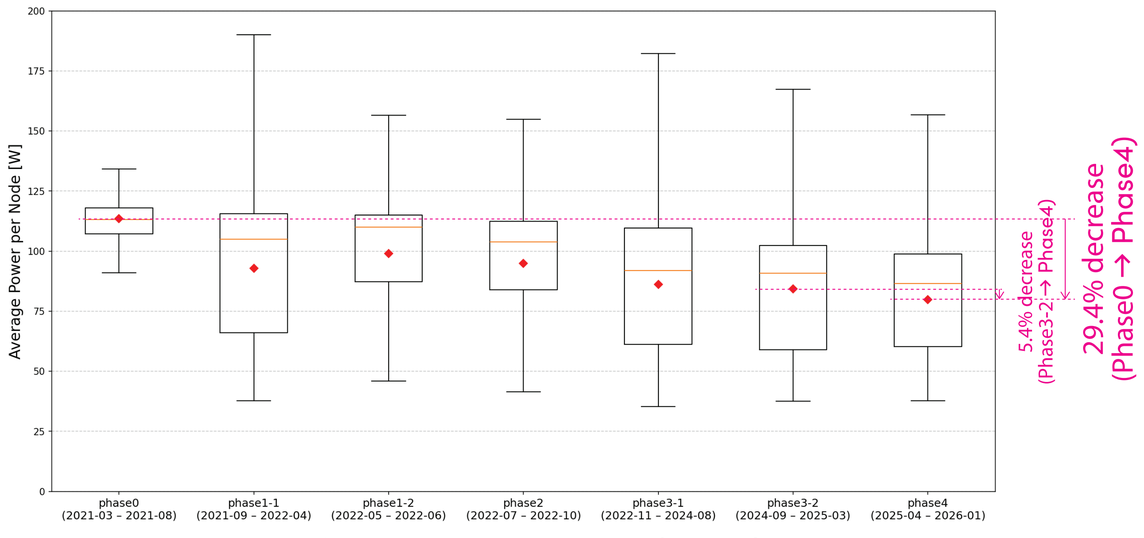

Figure 2. Average power per node over phases (orange line: median, red diamond: mean)

Power-saving Approach for Idle Nodes

- **Method**: Node retention - transition idle nodes to low-power state when not allocated to jobs
- **Measurement**:
  - Average node power including idle nodes is calculated from facility-side power meters (Azbil) on the computer building 3rd floor, dividing by the number of active (powered-on) nodes only.
  - Cross-check via two independent measurement systems (top-down vs bottom-up)
- **Results** (Fig. 3):
  - Facility-side power meters and Power API showed consistent trends
  - The 3rd floor includes loads other than compute racks. (Azbil > Power API/econ2)

Table 4. Node retention coverage expansion

Date	Racks	Coverage
2021-09	72	17%
2022-06	162	38%
2022-10	228	53%
2023-08	270	63%
2024-03	300	69%
2024-06	396	92%
2024-08	432	100%

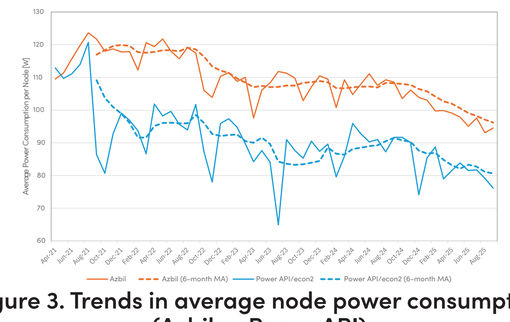

Figure 3. Trends in average node power consumption (Azbil vs Power API)

Evaluation of Boost-eco Mode (Phase3-2→Phase4)

Preliminary: All Jobs

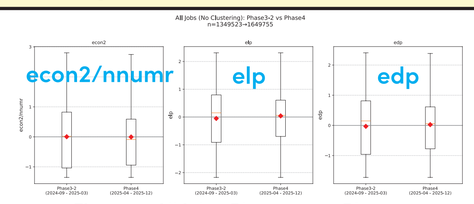

Figure 4. All jobs: Phase3-2 vs Phase4 (standardized values, orange line: median, red diamond: mean)

Table 7. Definition of evaluation metrics

Metric	Description	Significance
econ2/nnumr	Energy consumption per node	Directly measures the effect of power-saving measures. nnumr represents the number of request nodes.
elp	Elapsed time	Evaluates performance impact of power saving modes. Essential for understanding the trade-off between savings and performance.
edp	Energy-Delay Product per node (= econ2/nnumr × elp)	Captures the trade-off between power savings (operator's concern) and performance (user's concern).

- **Boost-eco mode:**
  - Reduce power consumption while minimizing performance impact
- **Results** (Fig. 4):
  - Node energy consumption per job
    - Slightly improved
  - Elapsed time
    - Slightly increased
  - Energy delay product
    - Slightly increased
- **Question**:
  - Did boost-eco mode improve energy efficiency across all job types?

Methods

Table 5. Clustering analysis workflow for boost-eco evaluation

Step	Process	Details
1	Preliminary + Sampling	Extracted target jobs (Period: 2024/09 - 2025/12); Filtered by normal termination jobs (pc=0); Normalization + log transformation + standardization; Sampling (#jobs: 13,510,587 → 2,999,715)
2	Feature extraction/reduction	28 Features (excluding econ2, elp, edp) → 11 Features
3	Clustering	kmeans++ (k=6, --n_init=100); Excluded elp, econ2, and edp from features (reserved for post clustering evaluation)
4	Comparison	Calculated average node power, elapsed time, and EDP for each cluster before and after the transition; Period 1 (Phase 2-2): 2024/09 - 2025/03; Period 2 (Phase 4): 2025/04 - 2025/12

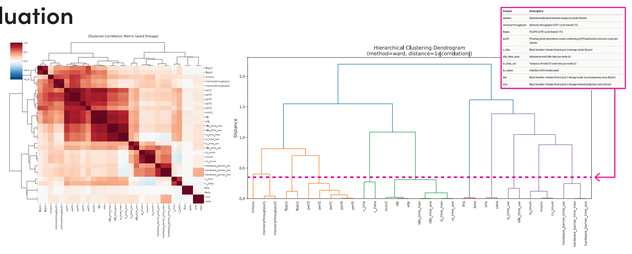

Figure 5. Feature selection via correlation analysis and hierarchical clustering

Results

Table 6. Job-type comparison (Phase3-2→Phase4)

Cluster	Estimated Job Type	n (P3-2→P4)	econ2	elp	edp	Trend
C0	I/O Cache-Locality	13,836→82,792	***↓	***↑	n.s.	Mixed
C1	Compute-Intensive	234,787→351,835	***↑	***↑	***↑	Degraded
C2	Lightweight / Low-Resource	491,565→565,446	***↓	***↓	***↓	Improved
C3	Large-Scale Parallel	8,211→16,793	***↓	**↓	***↓	Improved
C4	I/O Bottleneck	138,970→194,577	***↓	***↓	***↓	Improved
C5	Memory-Intensive	462,154→438,312	***↓	***↓	***↓	Improved

Welch's t-tests were conducted, and results showing no statistically significant difference are denoted as n.s. (not significant).

Key Findings:
- Boost-eco mode benefits most job types (77%)
- Compute-intensive jobs (C1, 19.6%) showed degradation
- C1 jobs increased by 1.5x in Phase4, impacting overall metrics

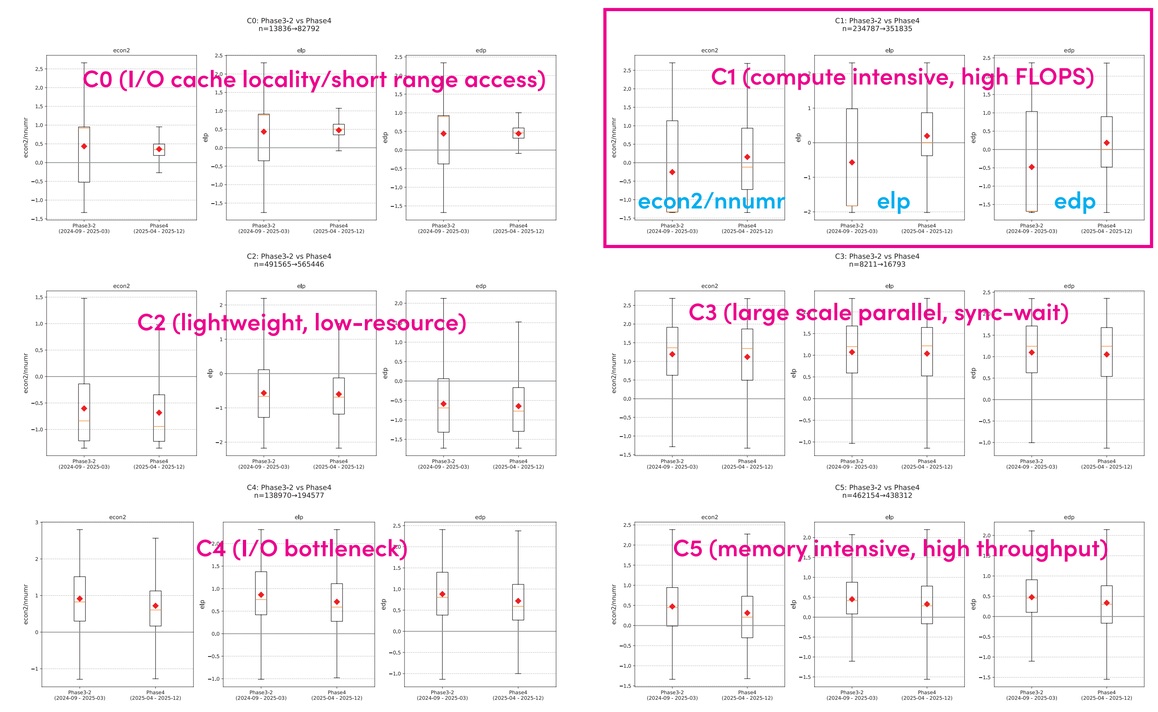

Figure 6. Phase3-2 to Phase4 comparison by job-type cluster (orange line: median, red diamond: mean)

Summary and Future Work

1. **Overall trends (entire period)**
   - Average node power decreased by 29.4% (113W → 79.8W)
   - Facility-side power meters and Power API showed consistent trends
2. **Boost-eco mode evaluation (Phase3-2 → Phase4)**
   - [All jobs (No clustering)]
     - Average node power: 84.3W → 79.8W (5.4% reduction)
     - Node energy consumption per job: slightly decreased
     - Elapsed time: slightly increased
     - EDP: slightly increased
   - [Job-type cluster analysis]
     - Compute-intensive jobs (C1, 19.6%): degraded
     - Other job types (C2–C5, 77%): improved
     - C1 jobs increased 1.5x from Phase3-2 to Phase4, impacting overall trends
3. **Future Work**
   - Optimize boost-eco application strategy for compute-intensive jobs