EESP 2026 Workshop

November 15, 2026

Variation-Aware Non-Uniform Node Power Capping for Exascale HPC Systems

Solomon Bekele[1], Brad Geltz[2], Siddhartha Jana[2], Swann Perarnau[1], Christopher M. Cantalupo[2], Servesh Muralidharan[1], Kevin Harms[1], Paul Rich[1], Brice Videau[1], and Kalyan Kumaran[1]

[1] Argonne National Laboratory, Lemont, IL, USA

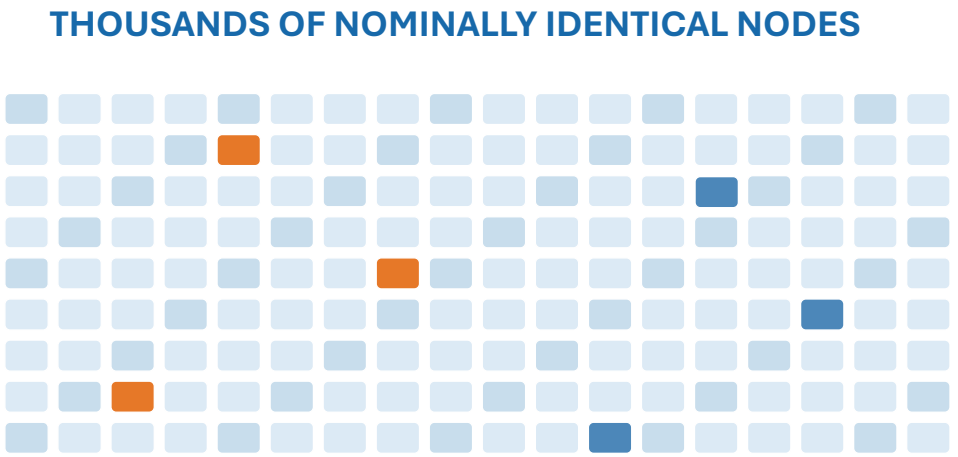
[2] Intel Corporation, Portland, OR, USA

U.S. DEPARTMENT of ENERGY

Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Thousands of nominally identical nodes still behave differently

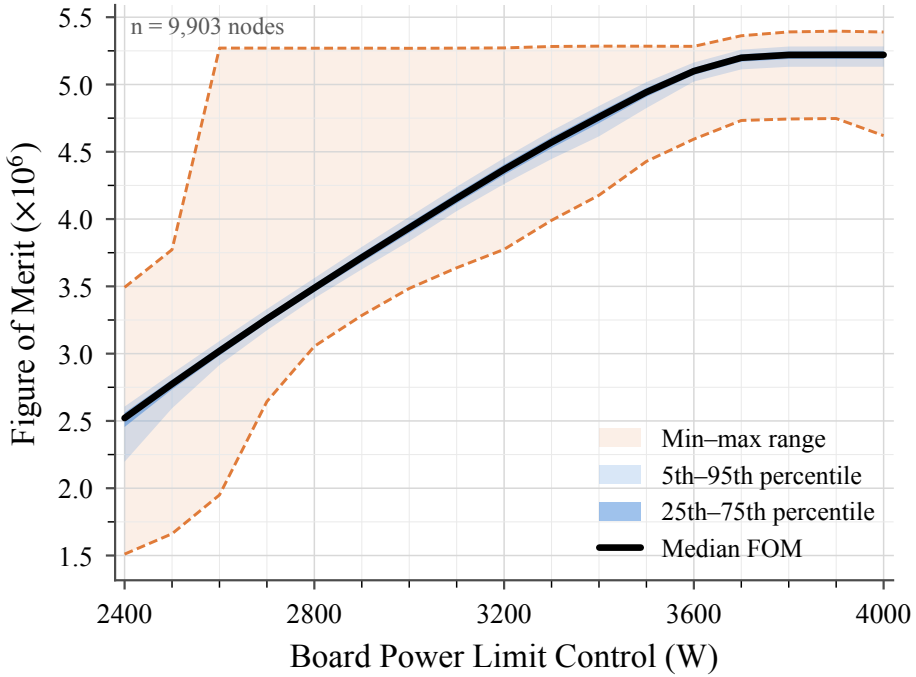
- **Process / manufacturing variation**
  - — Manufacturing differences accumulate across a very large machine
- **Aging**
  - — Node characteristics can drift over the lifetime of the system.
- **Thermal conditions**
  - — Local temperature and cooling conditions shift node behavior.
- **Other operating conditions**
  - — The same nominal configuration does not guarantee the same response



U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

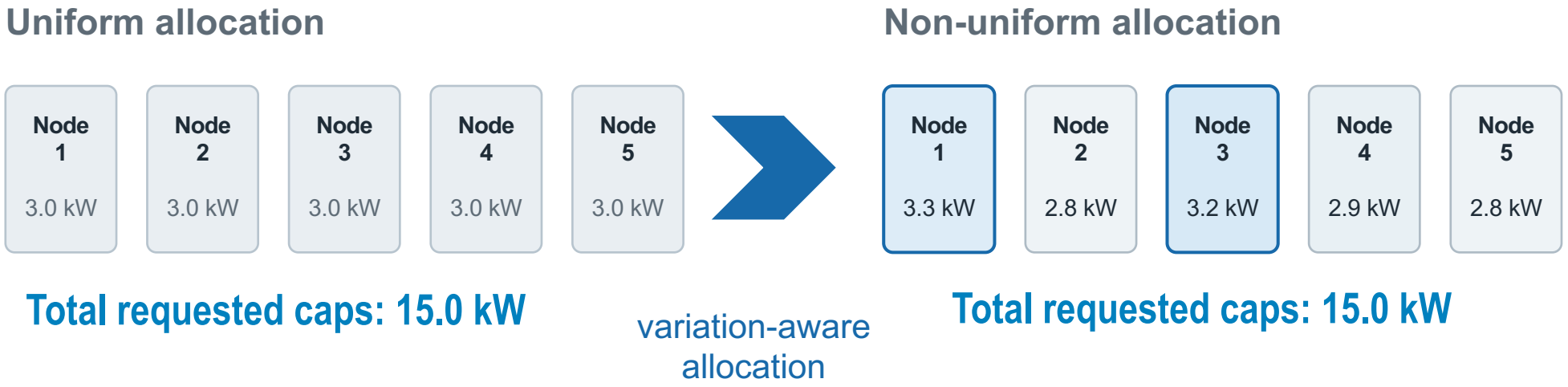
Uniform power caps assume identical node response

- Our measurements on Aurora show that assumption does not hold.
- Wide spread in Nekbone performance under the same node power limit.
- Bulk-synchronous jobs wait for the slowest node.
- A uniform cap can therefore waste part of a fixed job-level power budget.
- The issue is not total power alone
  - it is how that power is distributed across nodes.



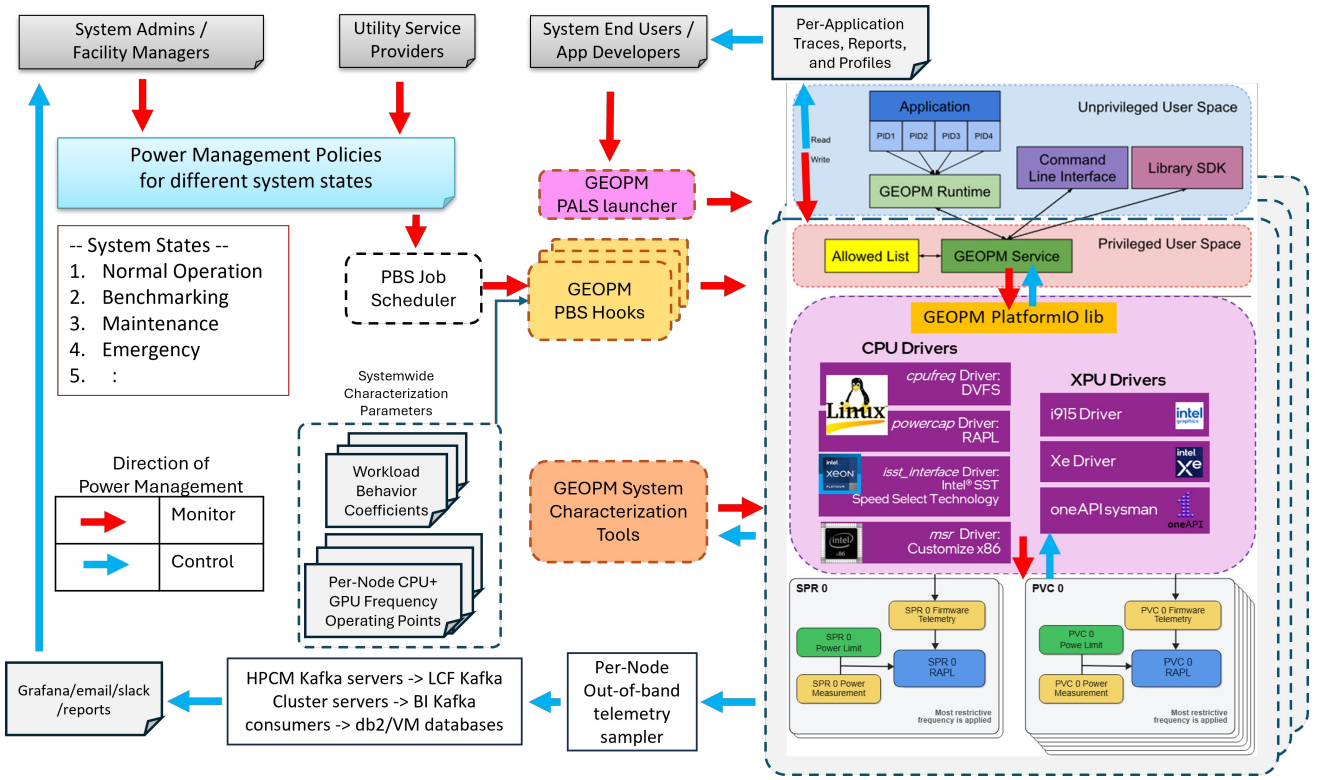
U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

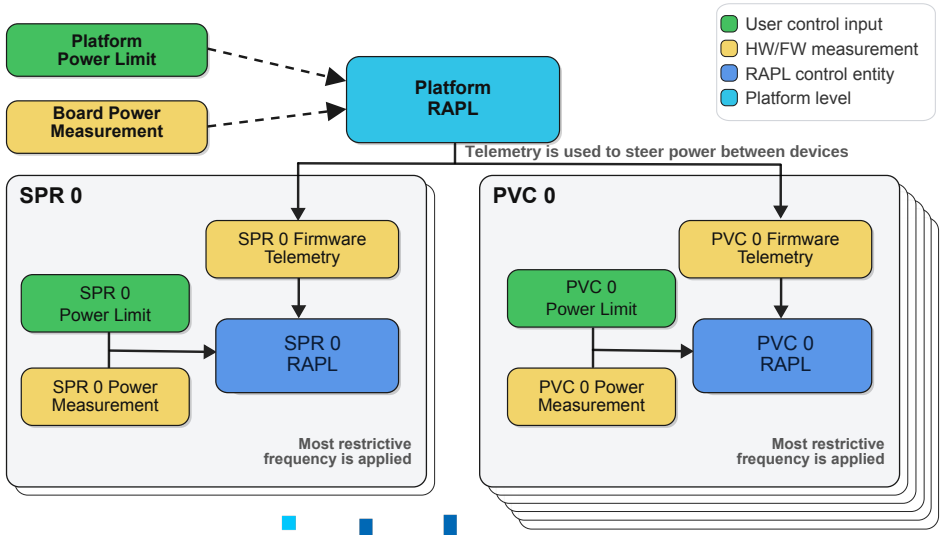
Our idea: keep the job budget fixed, change the per-node allocation



Can per-node characterization help us reduce the largest predicted slowdown without increasing the job's aggregate power budget?

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY





https://github.com/geopm/geopm

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
Argonne NATIONAL LABORATORY

End-to-End Model-Guided Power Allocation on Aurora via GEOPM

Offline modeling → PBS job integration → node-level control → safe restoration

Offline Characterization

- Measure power–performance tradeoffs
- Build allocation models
- Store model parameters

Job Launch (PBS Prologue)

- Read job power budget
- Select allocation model
- Compute per-node caps

Power Enforcement (GEOPM)

- Apply caps via PlatformIO
- Enforce node-specific limits
- Set limits before execution

Execution & Restore (PBS Epilogue)

- Run application under caps
- Invoke epilogue at completion
- Restore default settings

AURORA COMPUTE NODES

Platform power controls

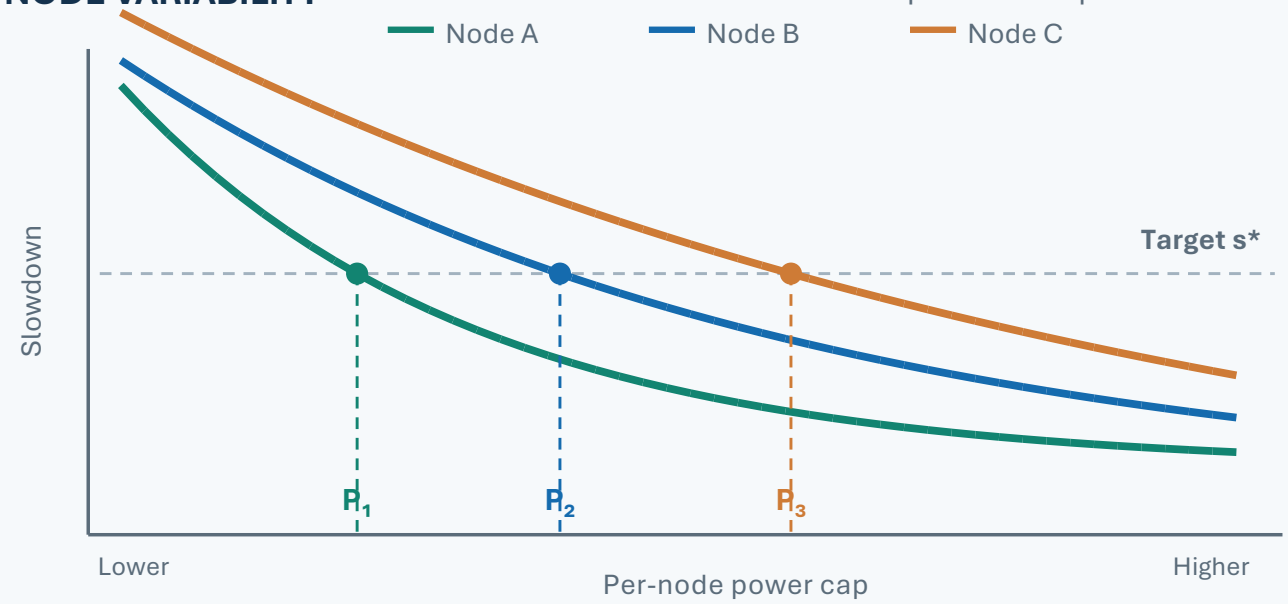
- **CPUs**
- **GPUs**
- **Memory / Uncore**

No application modifications required.

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

Variation-Aware Non-Uniform Power Allocation

A common slowdown target can require different power caps on different nodes.



ALLOCATE UNDER BUDGET B

FIT

$$S_i(P) \approx M_i(P)$$

Predict each node's slowdown.

INVERT

$$P_i(s) = M_i^{-1}(s)$$

Find the required power at slowdown s.

BISECT

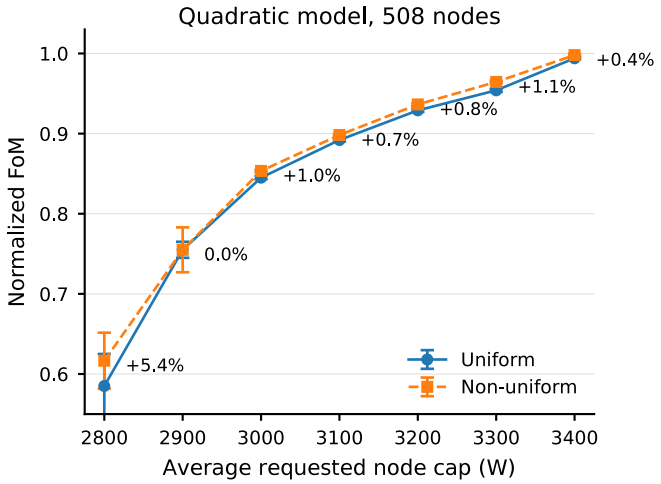
$$\Sigma\, P_i(s^*) \leq B$$

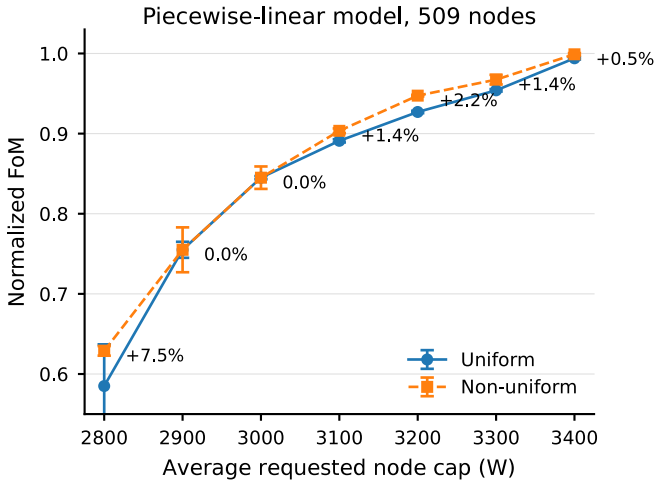
Find the smallest feasible s*.

Result: non-uniform per-node caps minimize the predicted worst-node slowdown within the budget.

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

At ~512 nodes, non-uniform allocation improves Nekbone FOM





7.5%

best gain
piecewise-linear @ 2,800 W

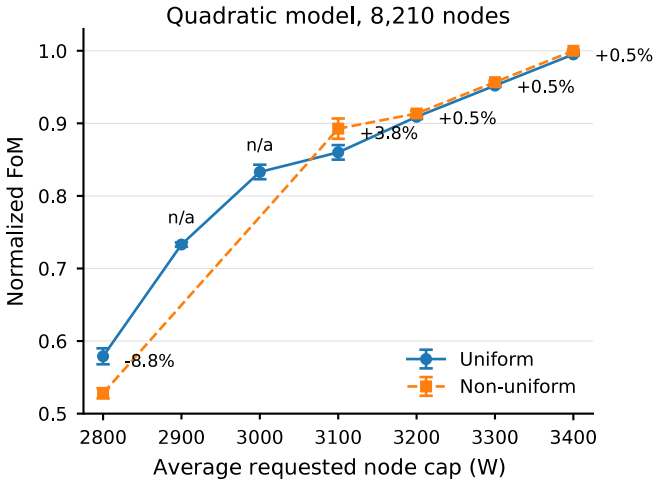
5.4%

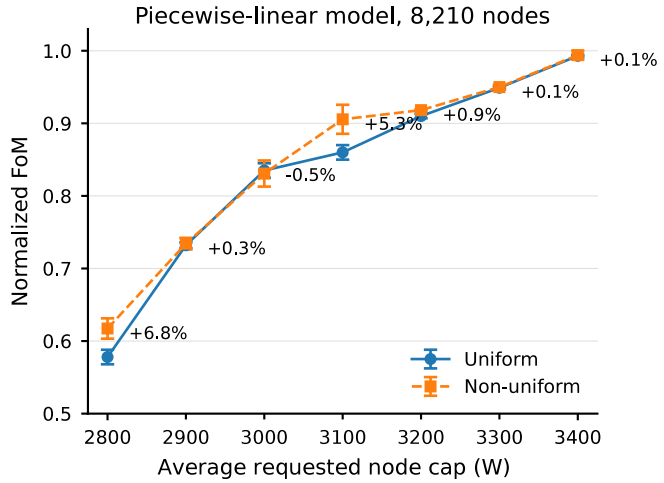
best gain
quadratic @ 2,800 W

At moderate caps, both model choices generally track the uniform baseline closely while producing small positive gains.

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

At 8,210 nodes – mixed observation





+6.8%

piecewise-linear @ 2,800 W

+5.3%

piecewise-linear @ 3,100 W

−8.8%

quadratic @ 2,800 W

+3.8%

quadratic @ 3,100 W

Same allocator. Different representation of the characterized response.

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

Takeaways

- **Node variability matters:** Identical power caps do not guarantee identical performance.
- **Variation-aware allocation** can improve application FOM without increasing the aggregate requested power budget.
- **At ~512 nodes:** Both models improve Nekbone FOM, achieving up to **+7.5%** (piecewise-linear) and **+5.4%** (quadratic).
- **At 8,210 nodes:** Piecewise-linear achieves **+6.8%,** while quadratic regresses by 8.8% at 2,800 W/node, highlighting the importance of model accuracy.

Deployment Workflow: Offline characterization → Model selection and allocation (PBS prologue) → Power cap enforcement (GEOPM) → Restore defaults (PBS epilogue).

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

THANK YOU!

Argonne NATIONAL LABORATORY

U.S. DEPARTMENT of ENERGY

This research used resources of the Argonne Leadership Computing Facility, a U.S. Department of Energy (DOE) Office of Science user facility at Argonne National Laboratory and is based on research supported by the U.S. DOE Office of Science-Advanced Scientific Computing Research Program, under Contract No. DE-AC02-06CH11357.

intel

What the results show — and limits

Established by the paper

- Development of the node power limit control and PBS-hook based job-level power control
- Node-to-node variation is large enough to matter for a tightly coupled job.
- Redistributing a fixed power budget can improve Nekbone FOM.
- The same allocator can use different node models.
- Piecewise-linear modeling was more robust at the tightest large-scale cap.

Not established

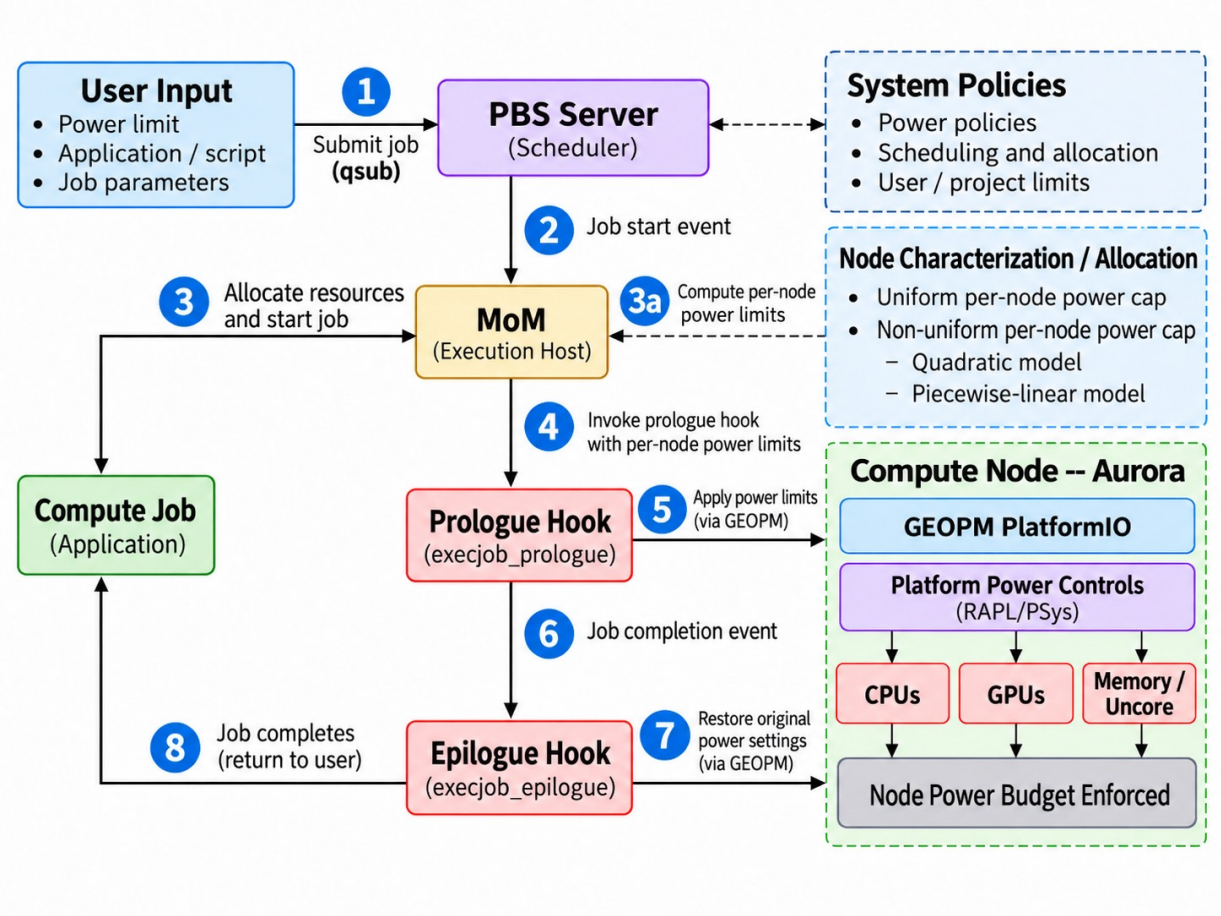
- Generalization to substantially different workload classes.
- Dynamic or phase-aware redistribution during execution.
- A universal model that transfers across workloads without re-characterization.
- Benefits on applications that are not bulk-synchronous not covered.

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

End-to-End Model-Guided Power Allocation on Aurora via GEOPM

- **Offline Characterization**
  - Measure power–performance tradeoffs across Aurora nodes
  - Develop non-uniform power allocation models.
- **At job launch**
  - PBS prologue reads the job-level power budget and selected model.
  - The allocator computes a per-node cap.
  - GEOPM **PlatformIO** applies the cap before the application starts.
- **At job completion**
  - PBS epilogue restores the original platform settings.

No application changes are required.



U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
intel
Argonne NATIONAL LABORATORY

Example Quadratic Model


```
{
  "max_power": 3700.0,

  "node_profile_name": "jobtype_a",
  "profiles": {
    "jobtype_a": {
      "hosts": {
        "host1": {
          "model":
          {
            "x0": 1.0,
            "A": 8e-08,
            "B": 7e-05,
            "C": 0.0
          }
        },
        "host2": {
          "model":
          {
            "x0": 1.0,
            "A": 8e-08,
            "B": 6e-05,
            "C": 0.0
          }
        }
      }
    }
  }
}
```


U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
Argonne NATIONAL LABORATORY