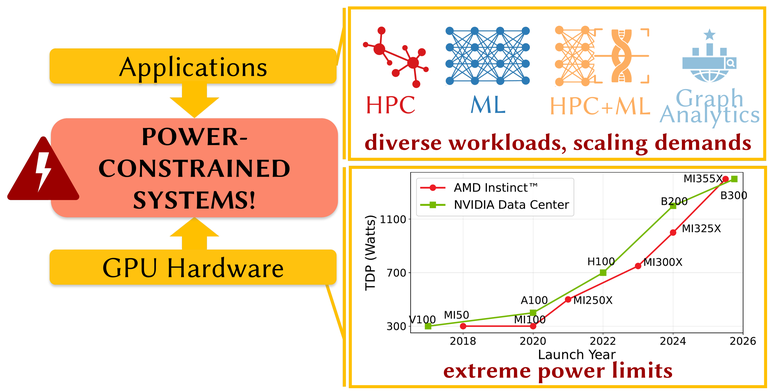
Embrace the Chaos: Harnessing Variability in Accelerator-Rich Systems

Rutwik Jain | Advisors: Matthew D. Sinclair, Shivaram Venkataraman

University of Wisconsin–Madison

SC26 Chicago, IL | hpc unites.

MOTIVATION



- **Power** is the dominant constraint in large-scale, accelerator-rich systems
- **My research: Co-optimizing power and performance at scale** through cross-stack approaches spanning hardware, Power Management (PM) firmware, systems, and applications

BACKGROUND



Quantifying variability

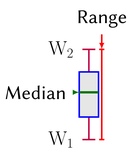
Interquartile range:
$\text{IQR} = Q_3 - Q_1$

Whiskers and range

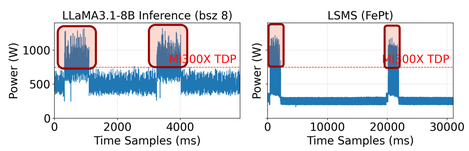
$W_2 = Q_2 + 1.5 \times \text{IQR}$
$W_1 = Q_1 - 1.5 \times \text{IQR}$
$\text{Range} = W_2 - W_1$

Variability

$$\text{Variability}_{\text{metric}} = \frac{\text{Range}_{\text{metric}}}{\text{Median}_{\text{metric}}}$$

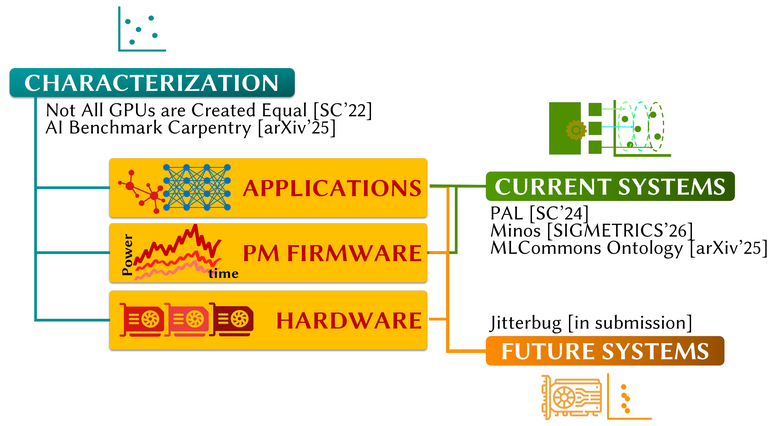


Defining power excursions



Power excursions: regions where power (temporarily) spikes over the GPU's TDP limit

OUR APPROACH



By characterizing and harnessing **performance variability** and **power excursions** in **current and future HPC systems**, we can improve efficiency at multiple scales within these systems.

RESULTS

Characterizing Variability [SC'22]

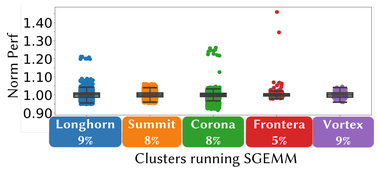
Context

- Acun, et al. [1], Zhang, et al. [5] : studied variability in CPU-based systems
- Scogland, et al. [4]: Motivated the need for more in-depth study

Goal

Understand and characterize GPU variability in large scale, accelerator-rich computing clusters

Results



- Consistent GPU perf variation (**avg 8%, up to 22% per app**, up to 1.5× slower outliers)
- Other metrics (freq, power, temp) vary too
- Consistent performance variation across **vendors, cluster size, GPU generations, cooling types**
- Variability is application-dependent – **compute-intensive applications see worst variability**, while memory-intensive applications show little to no variation.

Harnessing Variability [SC'24]

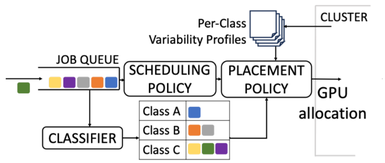
Context

- SC'22 study shows consistent, **application-dependent** variability
- For existing systems, underlying hardware or firmware can't be changed to mitigate this variability

Goal

Harness variability to our advantage by designing a **variability-aware scheduler**

Our Solution: PAL



Results

PAL improves geomean job completion time by **42%**, cluster utilization by **28%**, and makespan by **47%** over existing state-of-the-art schedulers.

Impact

- Heterogeneity-aware schedulers for Deep Learning [2]
- Collaborations with GPU vendors (AMD and NVIDIA) to tackle variability

Power Excursions [SIGMETRICS'26]

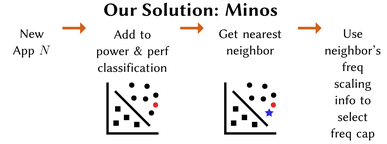
Context

- Applications evolve frequently ⟹ profiling overheads for new workloads
- Single metric like avg power insufficient to capture power dynamics (including excursions)

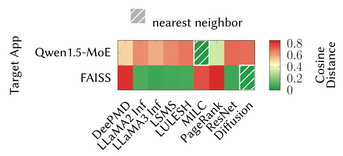
Goal

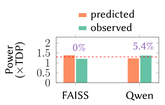
- Optimize **both performance and power**
- Reduce **profiling overheads** for **new apps**

Our Solution: Minos



Results





- High fidelity matches for unseen apps (**0-5%** error)
- Reduces profiling overheads by 89%

Excursion Management [arXiv, in submission]

Context

- **Power excursions** cause frequency throttling in critical application phases, degrading performance
- OCP spec [3] defines tolerable durations and magnitudes of excursions

Excursion Power (×TDP)	Duration
2×TDP	20 µs
1.75×TDP	2 ms
1.5×TDP	5 ms
1.1×TDP	20 ms
1×TDP	indefinitely

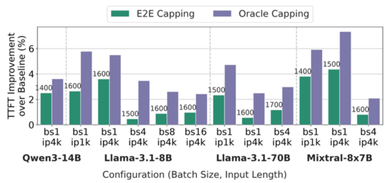
Our Solution: Jitterbug

Manage/mitigate excursions to get better + consistent performance

HW | Clamping

- Strictly limits excursions
- Loses sprinting benefits so performance degrades

FW | Frequency Capping



- 3-7% perf improv, reduced jitter
- Up to 14% perf improv for online serving

CONCLUSION

- Significant performance variability across accelerator-rich clusters – and worsening
- Power excursions exacerbate variability issues
- **Our Solution**: embrace/harness variability and excursions across the stack!

IMPACT

- **MLCommons Science/HPC WG**
  - Not All GPUs are Created Equal [SC'22] → variability in scientific benchmarking [arXiv'25, in submission]
  - Minos [SIGMETRICS'26] → Novel ontology to efficiently identify high quality benchmarks [arXiv'25, in submission]
- **NVIDIA**
  Mentee (Keting Chen) interned at NVIDIA – tackled variability in internal cluster
- **AMD Research and Advanced Development (RAD)**
  Ongoing collaboration on mitigating variability at-scale

REFERENCES

[1] Bilge Acun, Akhil Langer, Esteban Meneses, Harshitha Menon, Osman Sarood, Ehsan Totoni, and Laxmikant V. Kalé. Power, Reliability, and Performance: One System to Rule them All. *Computer*, 49(10):30–37, 2016.

[2] Zizhao Mo, Huanle Xu, and Wing Cheong Lau. Fast and Fair Training for Deep Learning in Heterogeneous GPU Clusters. In *Proceedings of the 39th ACM International Conference on Supercomputing*, ICS '25, page 324–338, New York, NY, USA, 2025. Association for Computing Machinery.

[3] OCP Open Accelerator Infrastructure (OAI) Workstreams. OCP Universal Baseboard (UBB) Design Specification v1.5, 2022. https://www.opencompute.org/documents/universal-baseboard-design-specification-v1p5-final-20220223-docx-pdf.

[4] Scogland, Thomas and Azose, Jonathan and Rohr, David and Rivoire, Suzanne and Bates, Natalie and Hackenberg, Daniel. Node variability in large-scale power measurements: perspectives from the Green500, Top500 and EEHPCWG. In *SC '15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11, 2015.

[5] Kaicheng Zhang, Seda Ogrenci-Memik, Gokhan Memik, Kazutomo Yoshii, Rajesh Sankaran, and Pete Beckman. Minimizing Thermal Variation Across System Components. In *2015 IEEE International Parallel and Distributed Processing Symposium*, pages 1139–1148, 2015.