Hewlett Packard
Enterprise

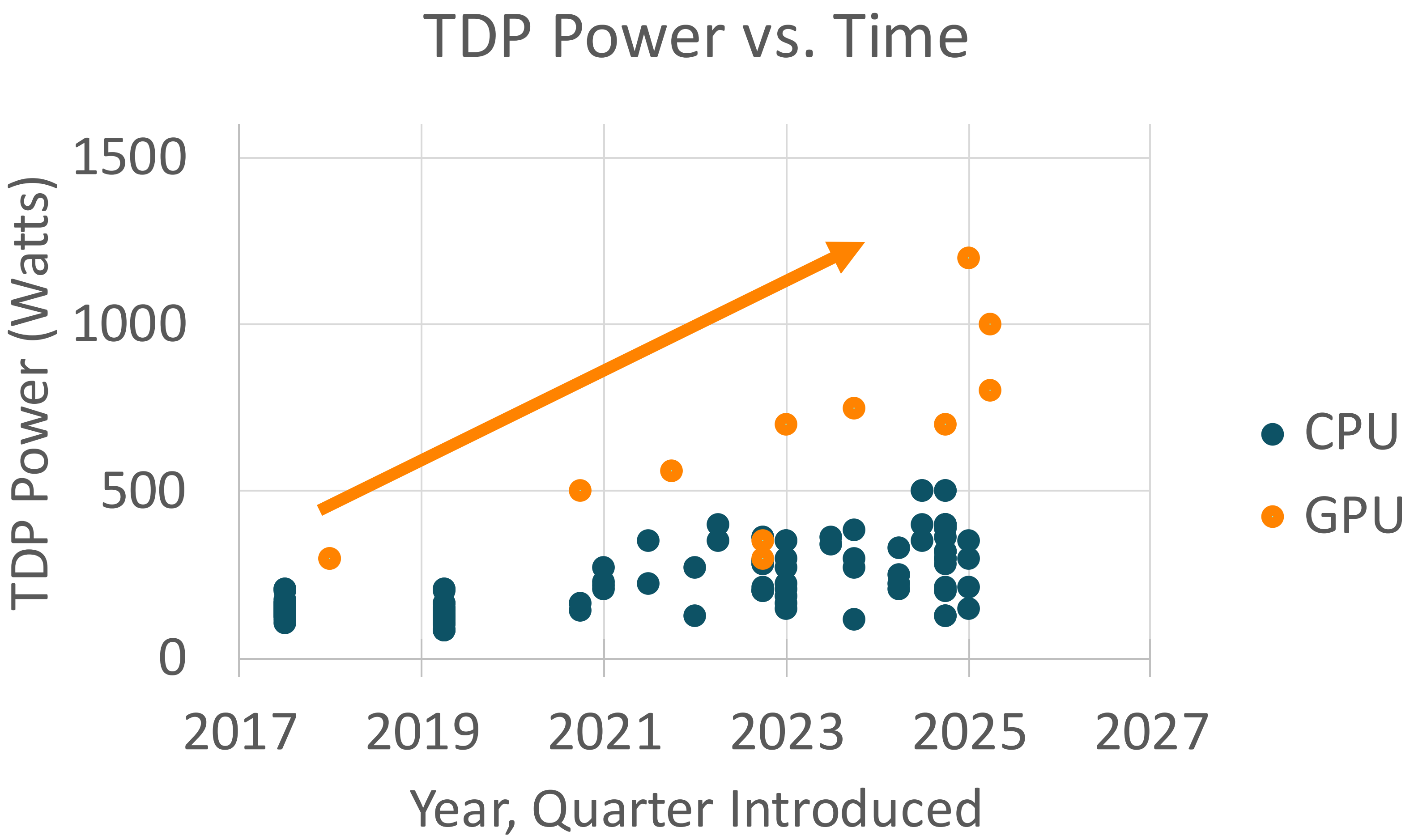
EVeREST: An Effective and Versatile Runtime Energy Saving Tool for GPUs

Anna Yue, Sanyam Mehta (Ph.D.), Torsten Wilde (Dr. rer. nat.)

Presenter: Barbara Chapman (Ph.D.)

May 7, 2025

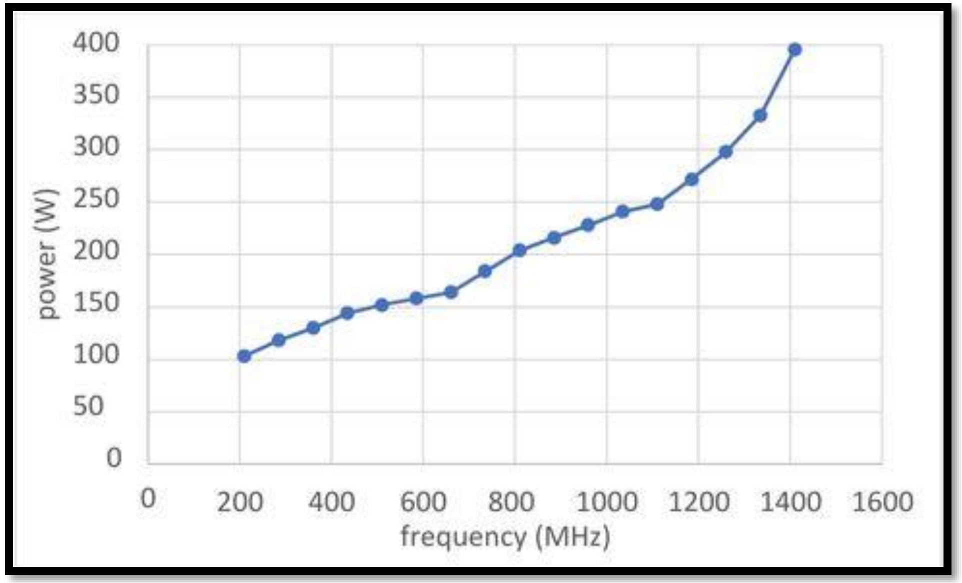
Power Trends



Opportunities to Save GPU Power/Energy

- Default settings: Run at max frequency

- (Dynamic) Power ~ **CV^2f**
- Dynamic Voltage Frequency Scaling (DVFS)
  - Also affects performance

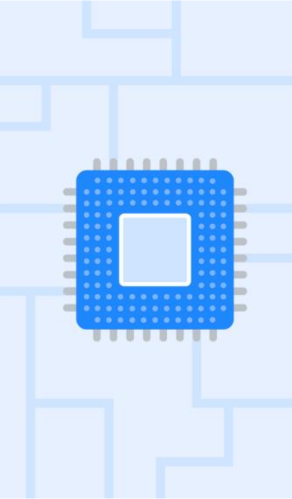


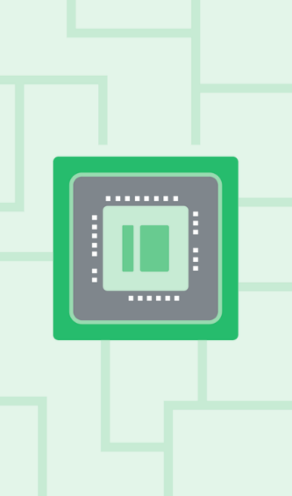
- Application performance is only as fast as the slowest component
  - Memory
  - Communication

Previous Work

- Everest for CPUs
  - Exploit DVFS during memory-bound and communication-bound phases
  - Determine phase compute-boundedness by measuring instructions per second (IPS) at different frequencies
  - Meet user-defined performance limit

- How to apply to GPUs?





https://blog.google/technology/ai/difference-cpu-gpu-tpu-trillium/

Challenges

- **#1: Predicting GPU performance at runtime**

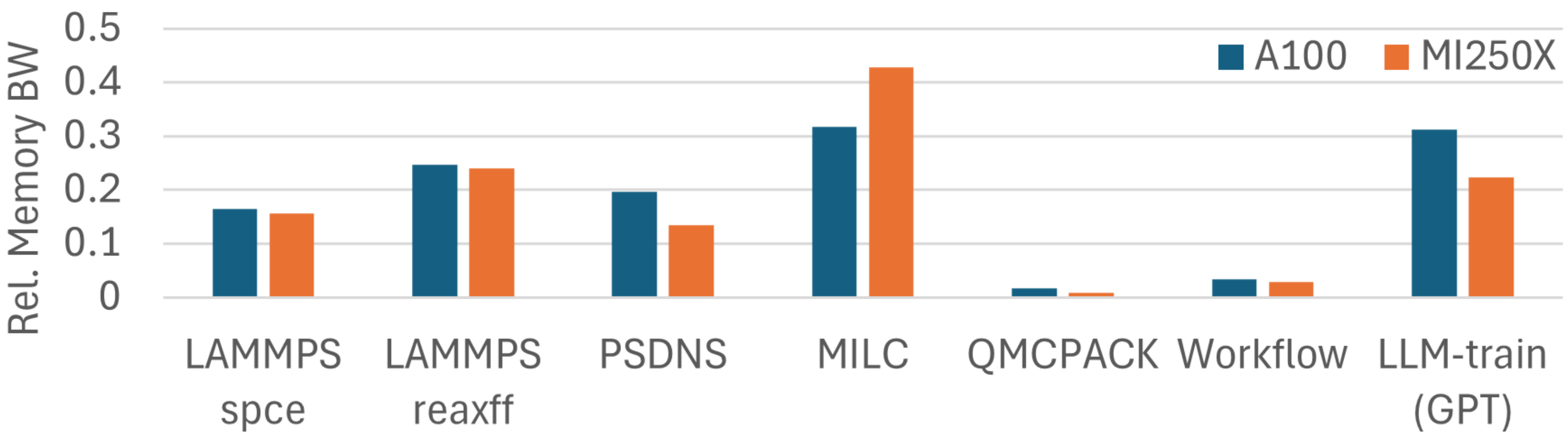
- Limited performance counters at runtime
  - Differences among GPU vendors
- Profiling associated with significant overhead (1.5x to >3x)
  - Stack walking, kernel serialization
- Prior work
  - Profile beforehand
  - Code instrumentation
  - Hardware changes

Challenges

- **#2: Need to identify more opportunities for saving energy on GPUs**

- HBM has vastly improved memory performance

Average Memory Bandwidth Utilization, Relative to Theoretical Peak



EVeREST for GPUs

Proof-of-concept (PoC) tool

Solution

- We propose EVeREST, an **E**ffective and **Ve**rsatile **R**untime **E**nergy **S**aving **T**ool.

EVeREST dynamically characterizes workloads with a lightweight and portable algorithm and uses DVFS to achieve power/energy savings while meeting a specified performance guarantee.

- EVeREST addresses the associated challenges on the GPU by
  - #1: Predicting performance using a single, accurate, and portable metric
  - #2: Identifying effective opportunities for energy savings

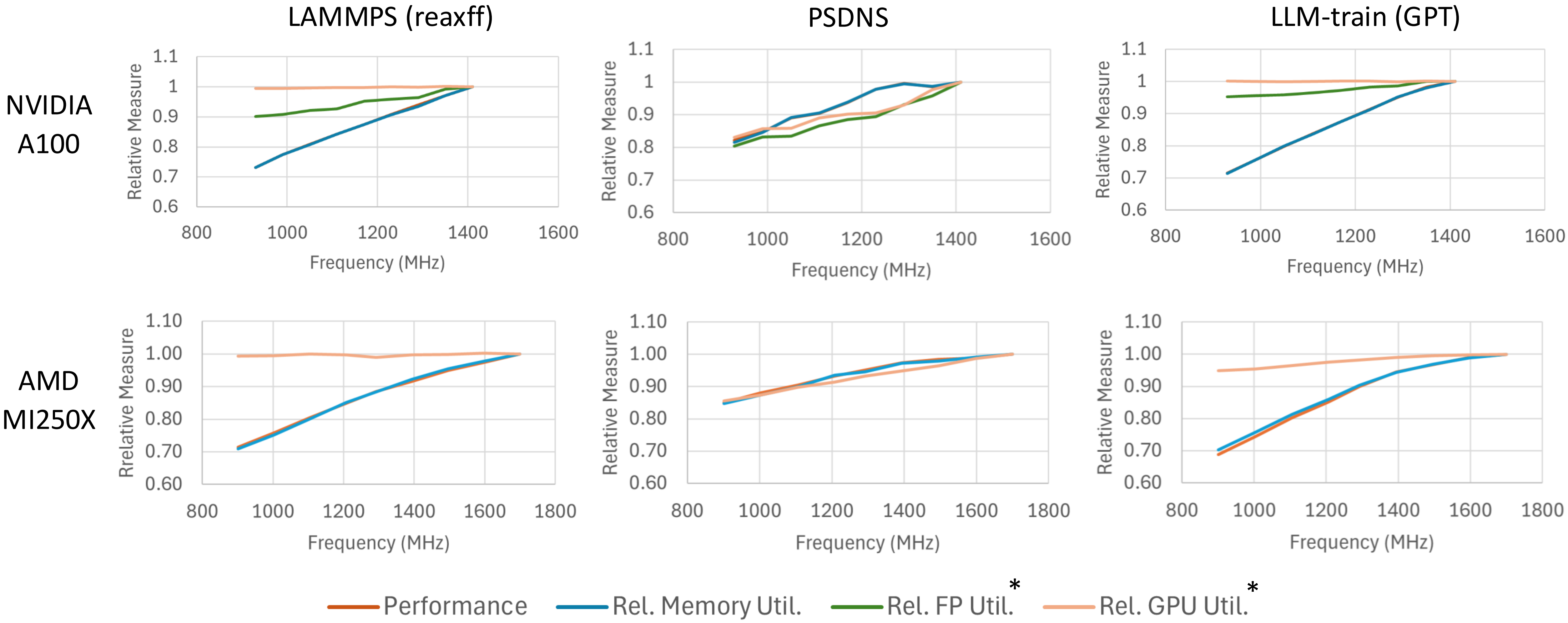
#1: Predicting Performance

- Observation from CPUs: Need metric to inform performance at different frequencies
  - Allows us to calculate frequency-sensitivity (FS)
    - –100% FS means performance changes in **same ratio** as frequency
    - –0% FS means performance does not change at all with frequency

$$\%FS = 100\% * \frac{\frac{?_{high}}{?_{low}} - 1}{\frac{Freq_{high}}{Freq_{low}} - 1}$$

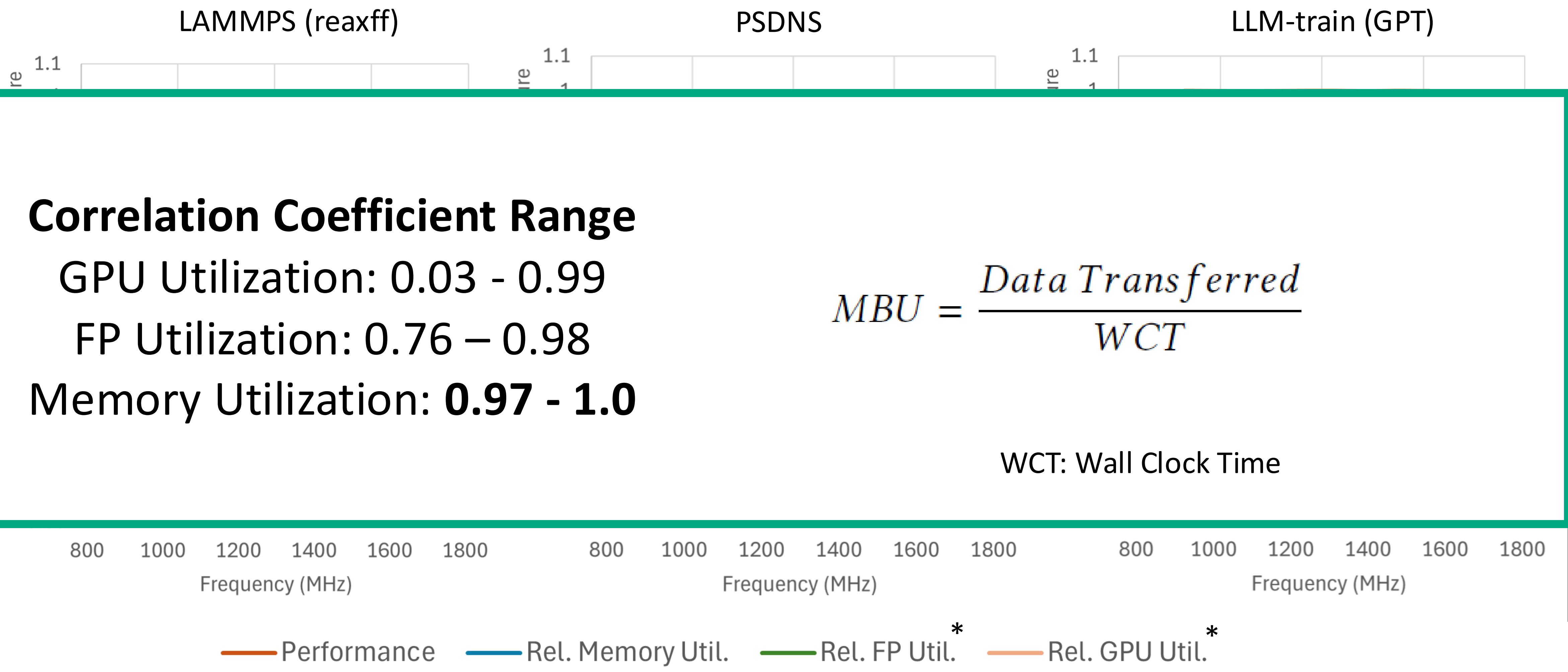
- Characterization study of available metrics
  - GPU Utilization
  - Memory (Bandwidth) Utilization
  - FP Utilization (NVIDIA only)

#1: Predicting Performance



*inverse

#1: Predicting Performance on GPU



Correlation Coefficient Range

GPU Utilization: 0.03 - 0.99

FP Utilization: 0.76 – 0.98

Memory Utilization: **0.97 - 1.0**

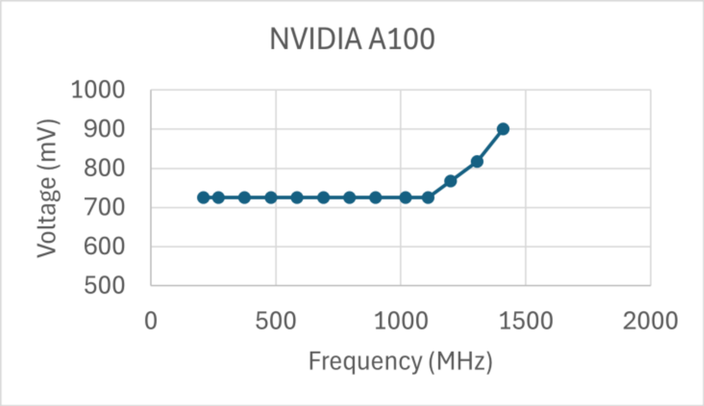
$$MBU = \frac{Data\ Transferred}{WCT}$$

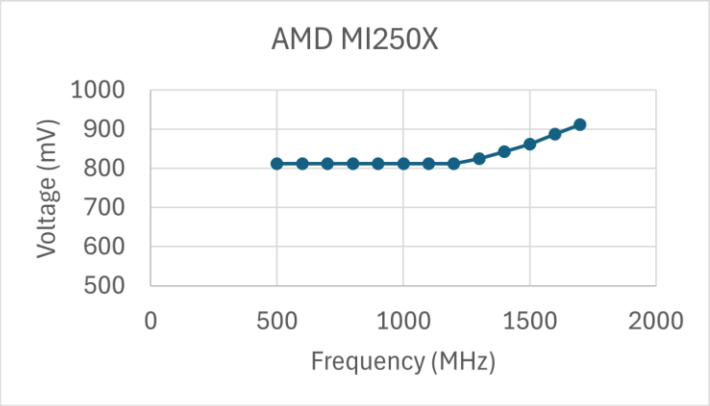
WCT: Wall Clock Time

*inverse

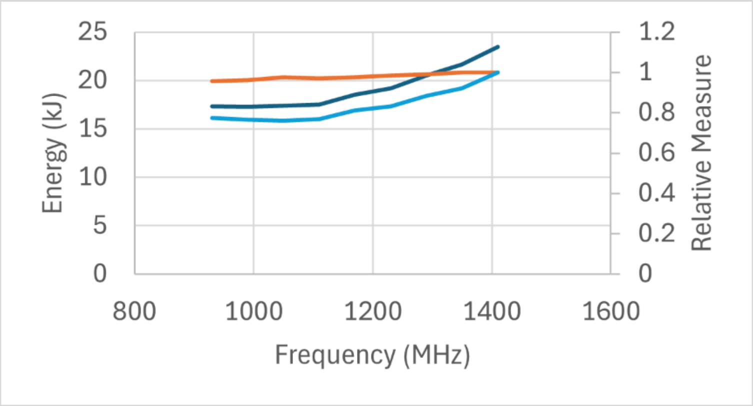
#2: New Opportunities for Energy Savings

- Voltage frequency curves allow for superlinear savings

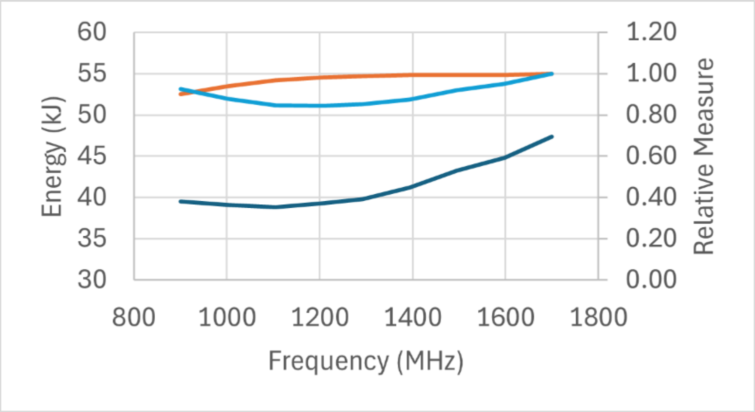




- Application performance may be bottlenecked by the CPU



GROMACS, NVIDIA A100



WarpX, AMD MI250X

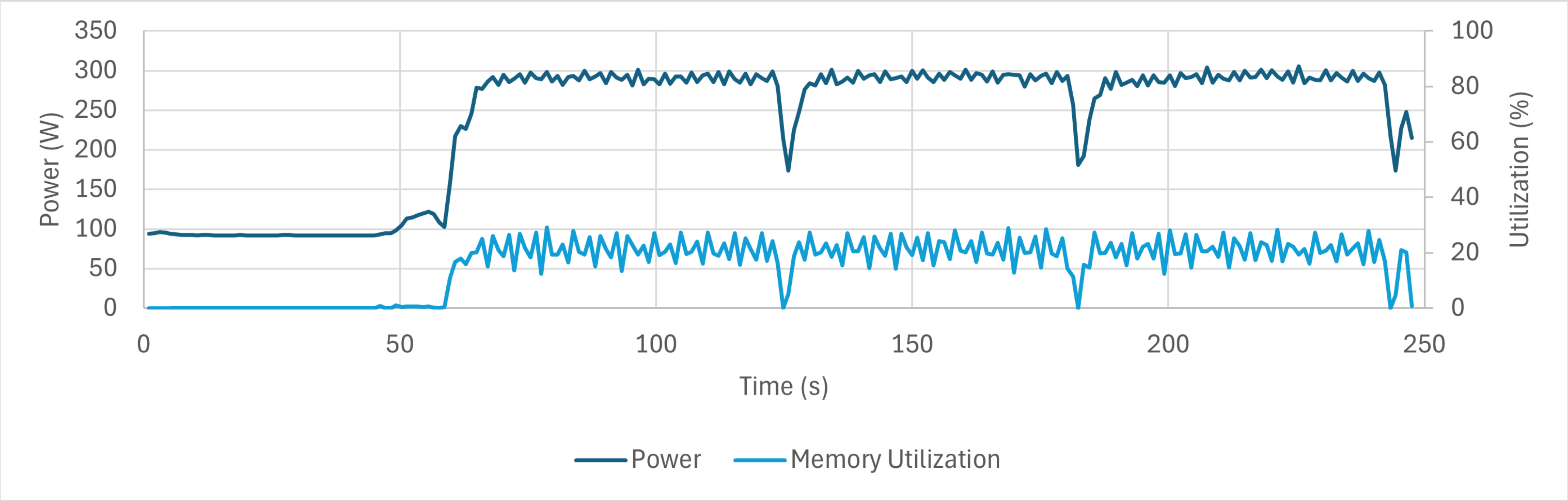
— Energy — Performance — Rel. EDP

How Everest Works

Using GPU & memory utilization

Phase Identification

Usi



ow

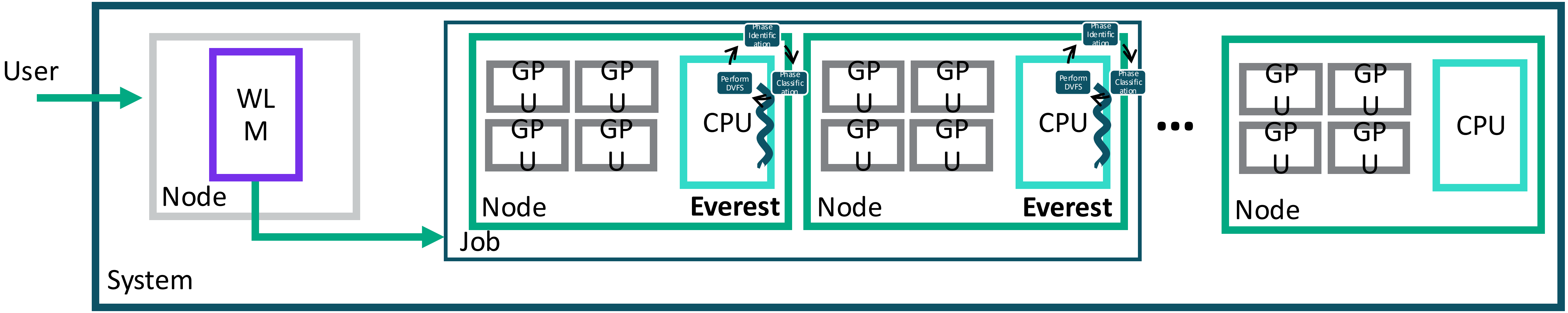
e

nge

/

How Everest Works

- Integration with workload manager
- One Everest process per node

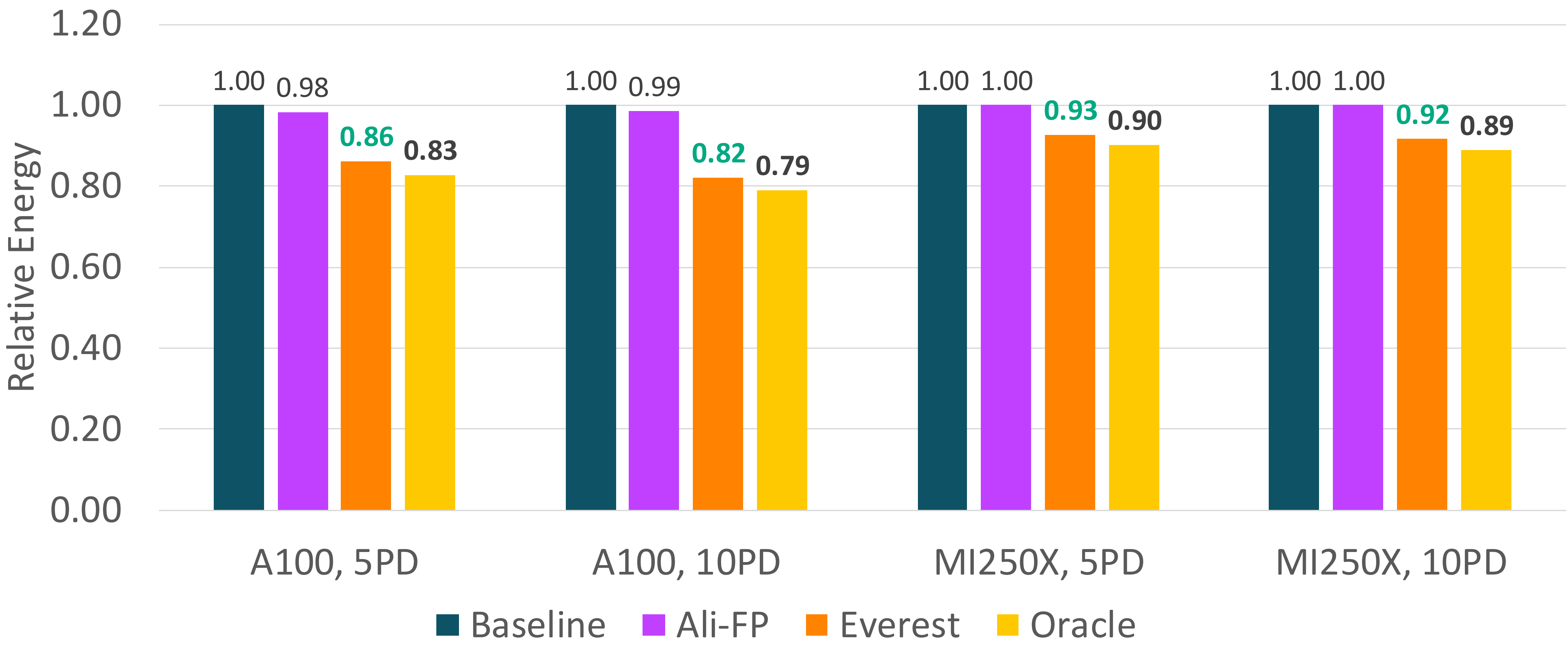


Results

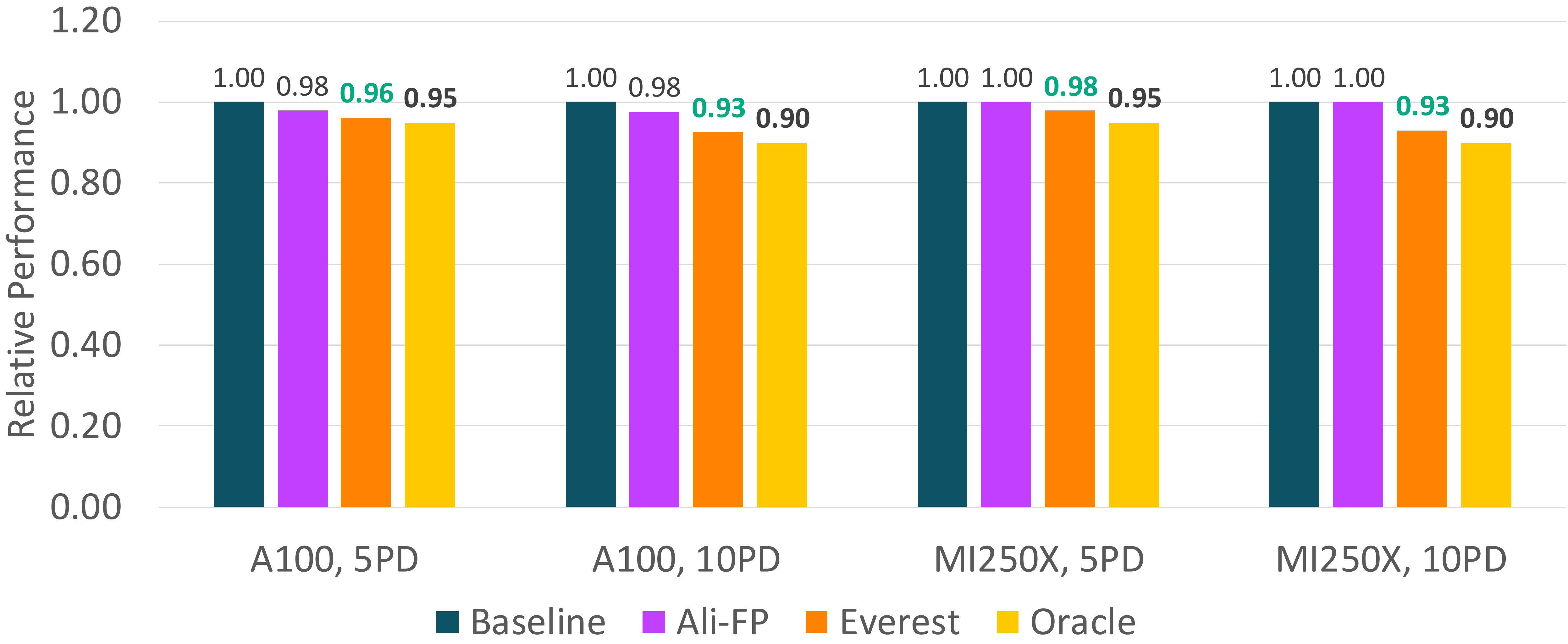
Setup

- 4x NVIDIA A100 nodes
- 4x AMD MI250X (8x GCDs) nodes
- 9 HPC and AI applications (OLCF)
- Different levels of acceptable performance loss (5%, 10%)
- Comparison to other works + static oracle

Energy (Geomean)



Performance (Geomean)



Conclusion

- Lightweight solution for dynamic optimization of application according to power/performance/energy tradeoffs
- Compute vendor agnostic
- Portable
  - Runtime-only, integration with user code not required
- Phase awareness
  - Can extract maximum power/energy savings without requiring user input
- Opportunity for collaboration and influencing product roadmap

Contact us!

anna.yue@hpe.com
wilde@hpe.com
barbara.chapman@hpe.com