EESP WORKSHOP

Adaptive Control with Offline Reinforcement Learning for HPC Power Management

Elias Klett and Christian Simmendinger

Hewlett Packard Enterprise



SC26 hpc unites | Chicago

Introduction

Objective

- Save energy by adjusting socket power caps with minimal performance loss
- Saving potential depends on the application's bottleneck
- Memory-bound workloads often tolerate lower caps than compute-bound workloads

Introduction

Objective

- Save energy by adjusting socket power caps with minimal performance loss
- Saving potential depends on the application's bottleneck
- Memory-bound workloads often tolerate lower caps than compute-bound workloads

Current Solution

- *PowerSched* uses Golden-Section Search on a performance-oriented EDP objective
- Search starts over for each job and may converge slowly
- Repeated probing of suboptimal settings

Introduction

Objective

- Save energy by adjusting socket power caps with minimal performance loss
- Saving potential depends on the application's bottleneck
- Memory-bound workloads often tolerate lower caps than compute-bound workloads

Current Solution

- *PowerSched* uses Golden-Section Search on a performance-oriented EDP objective
- Search starts over for each job and may converge slowly
- Repeated probing of suboptimal settings

Can we learn suitable caps instead of repeating the search?

Offline RL from Historical Telemetry

~13M transitions | ~15,000 production jobs | Historical GSS decisions

No application instrumentation or online exploration. Existing PowerSched safeguards.

Historical Data

Only tried settings have outcomes.
Search actions are not optimal labels.
Unfamiliar states remain a risk.

Offline RL from Historical Telemetry

~13M transitions | ~15,000 production jobs | Historical GSS decisions

No application instrumentation or online exploration. Existing PowerSched safeguards.

Historical Data

Only tried settings have outcomes.
Search actions are not optimal labels.
Unfamiliar states remain a risk.

Telemetry and Control

State: counters, clock, power, cap, objective
Cap: 140-400 W/socket, 5 W steps
Cycle: 5 s wait + 10 s measurement

Reward: r = -log(M), next-interval objective | 90/10 training/validation split by job

Offline RL from Historical Telemetry

~13M transitions | ~15,000 production jobs | Historical GSS decisions

No application instrumentation or online exploration. Existing PowerSched safeguards.

Historical Data

Only tried settings have outcomes.
Search actions are not optimal labels.
Unfamiliar states remain a risk.

Telemetry and Control

State: counters, clock, power, cap, objective
Cap: 140-400 W/socket, 5 W steps
Cycle: 5 s wait + 10 s measurement

Implicit Q-Learning

1. Upper-expectile V from logged Q

$$\tau = 0.7$$

2. Bellman update

$$Q\ \text{target} = r + 0.9\ V(s')$$

3. Advantage-weighted policy learning

$$A(s,a) = Q(s,a) - V(s)$$

Reward: r = -log(M), next-interval objective | 90/10 training/validation split by job

Energy Savings and HPCG Control

One Hunter node, 2 x AMD EPYC 9374F | 10 benchmarks | 3 runs/configuration

Mean changes vs. 400 W/socket static cap: whole-node energy and benchmark throughput

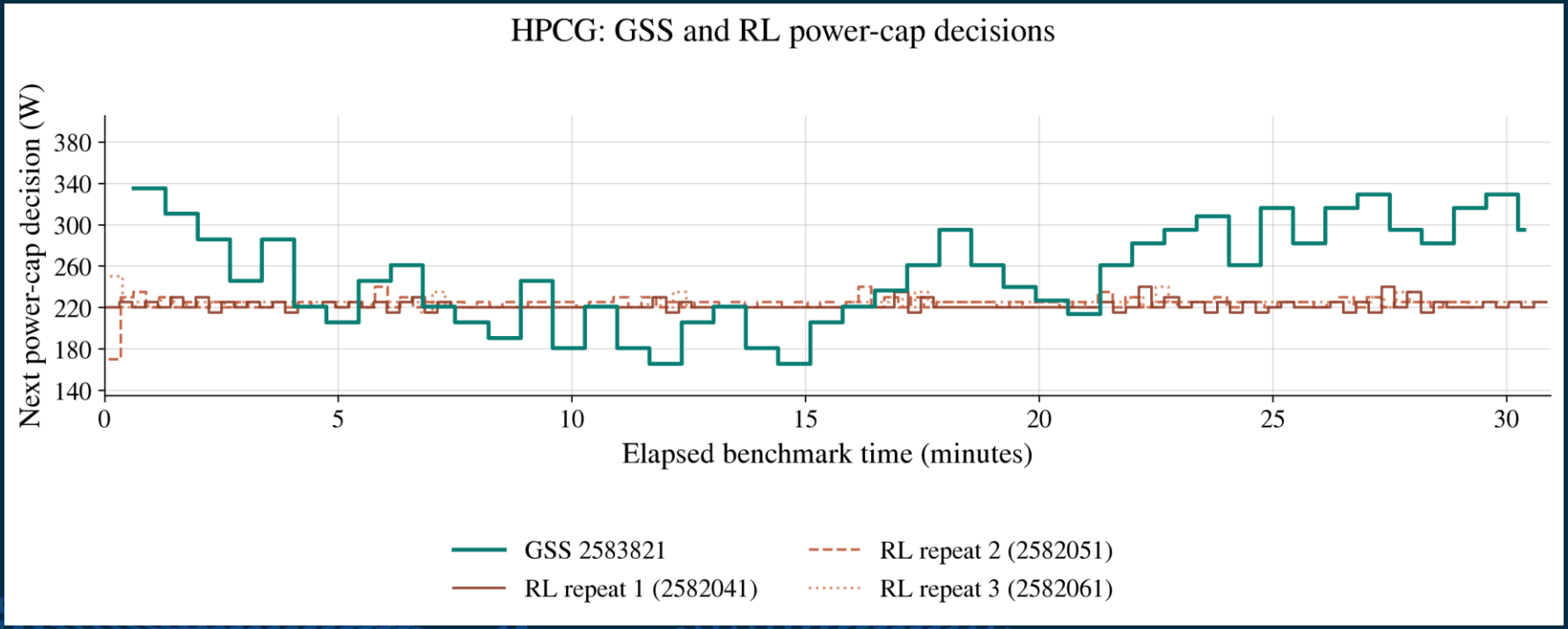
Energy Savings and HPCG Control

One Hunter node, 2 x AMD EPYC 9374F | 10 benchmarks | 3 runs/configuration

Mean changes vs. 400 W/socket static cap: whole-node energy and benchmark throughput

All ten benchmarks

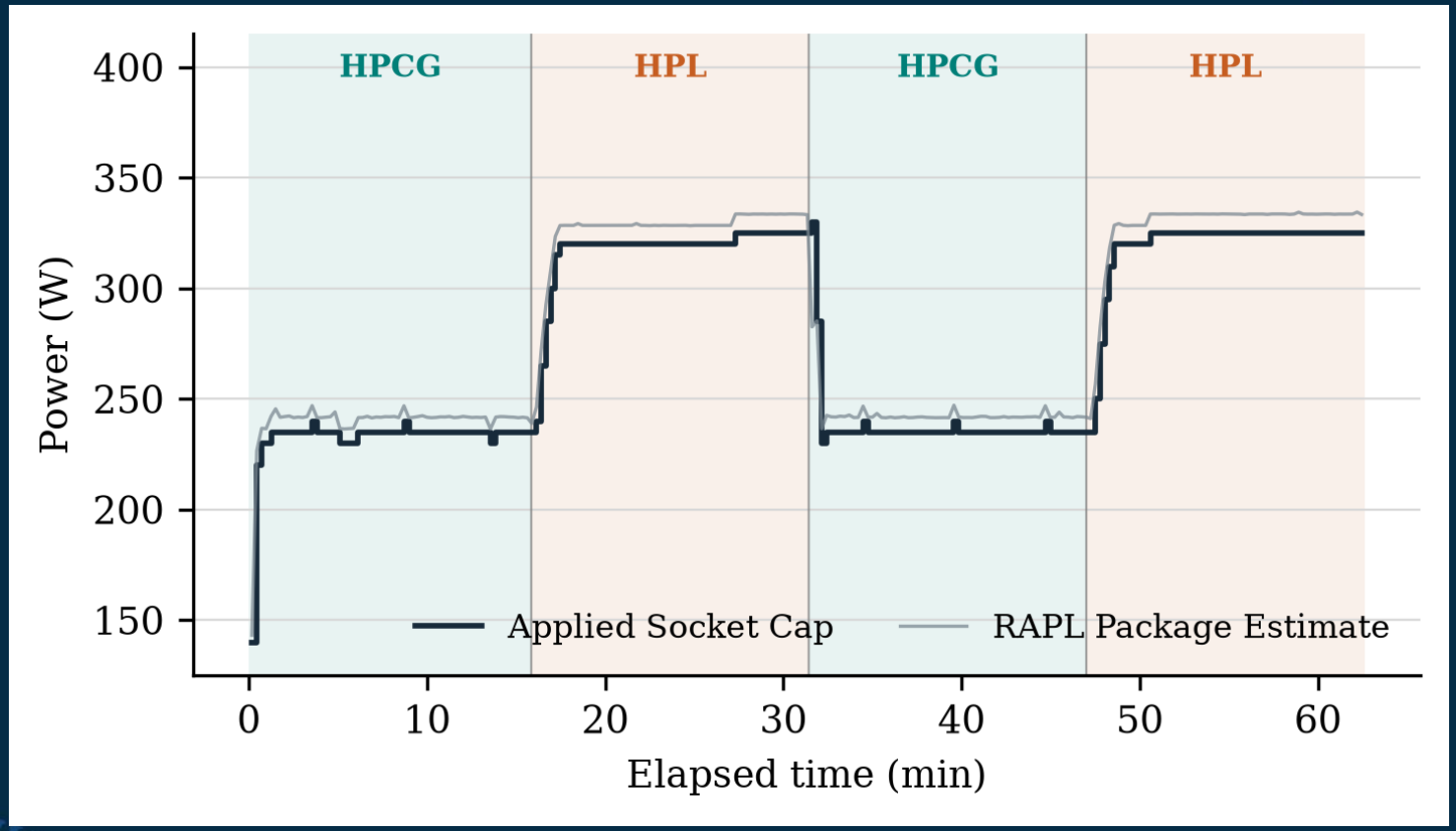
7.10% less node energy | 1.03% lower throughput



HPCG: one GSS run, three RL repeats, not a matched timing study. EP has high repeat variability.

Workload Adaptation without Repeated Probing

4 HPCG-HPL cycles | ~15-minute phases | First 2 cycles shown



HPCG | MEMORY-BOUND

230-240 W

43.5 +/- 6.1 s

~3 decisions to settle

HPL | COMPUTE-BOUND

320-325 W

77.8 +/- 20.3 s

~5 decisions to settle

Settling: +/-5 W for 3 consecutive samples. Mean +/- SD across phases. No matched GSS speed comparison.

Key Takeaways

Production telemetry supports offline power control

IQL in PowerSched, without application instrumentation or online exploration.

7.10% less node energy, 1.03% lower throughput

All ten benchmarks, relative to the 400 W/socket static cap.

Workload-adaptive caps in a few decisions

HPCG: 43.5 s (~3 decisions) | HPL: 77.8 s (~5 decisions)

Future Work

- Unfamiliar-workload detection
- Multi-node and more extensive evaluation
- Multi-dimensional CPU/GPU steering