Peak Efficiency Aware Scheduling for Highly Energy Proportional Servers

Daniel Wong
dwong@ece.ucr.edu
University of California, Riverside
Department of Electrical and Computer Engineering

Systems Optimization + Computer Architecture Lab

UC Riverside
University of California

Main Observations

- Servers are nearly energy proportional
- Peak energy efficiency does not occur at peak utilization
- Current data center scheduling techniques are unaware
- Peak Efficiency Aware Scheduling
  - Achieves better-than-ideal cluster-wide energy proportionality

Measuring Energy Proportionality

- Dynamic Range

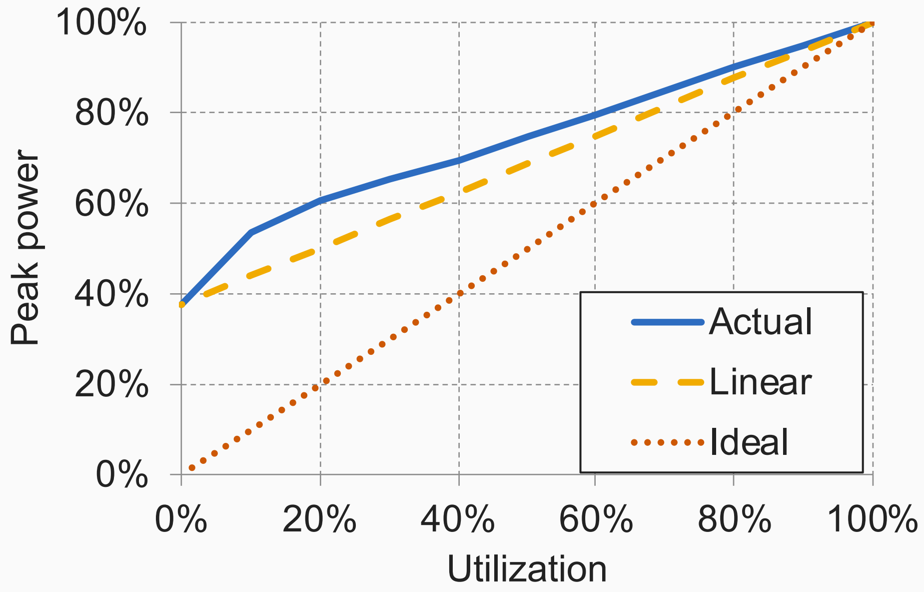
$$DR = \frac{Power_{peak} - Power_{idle}}{Power_{peak}}$$

- Energy Proportionality

$$EP = 1 - \frac{Area_{actual} - Area_{ideal}}{Area_{ideal}}$$

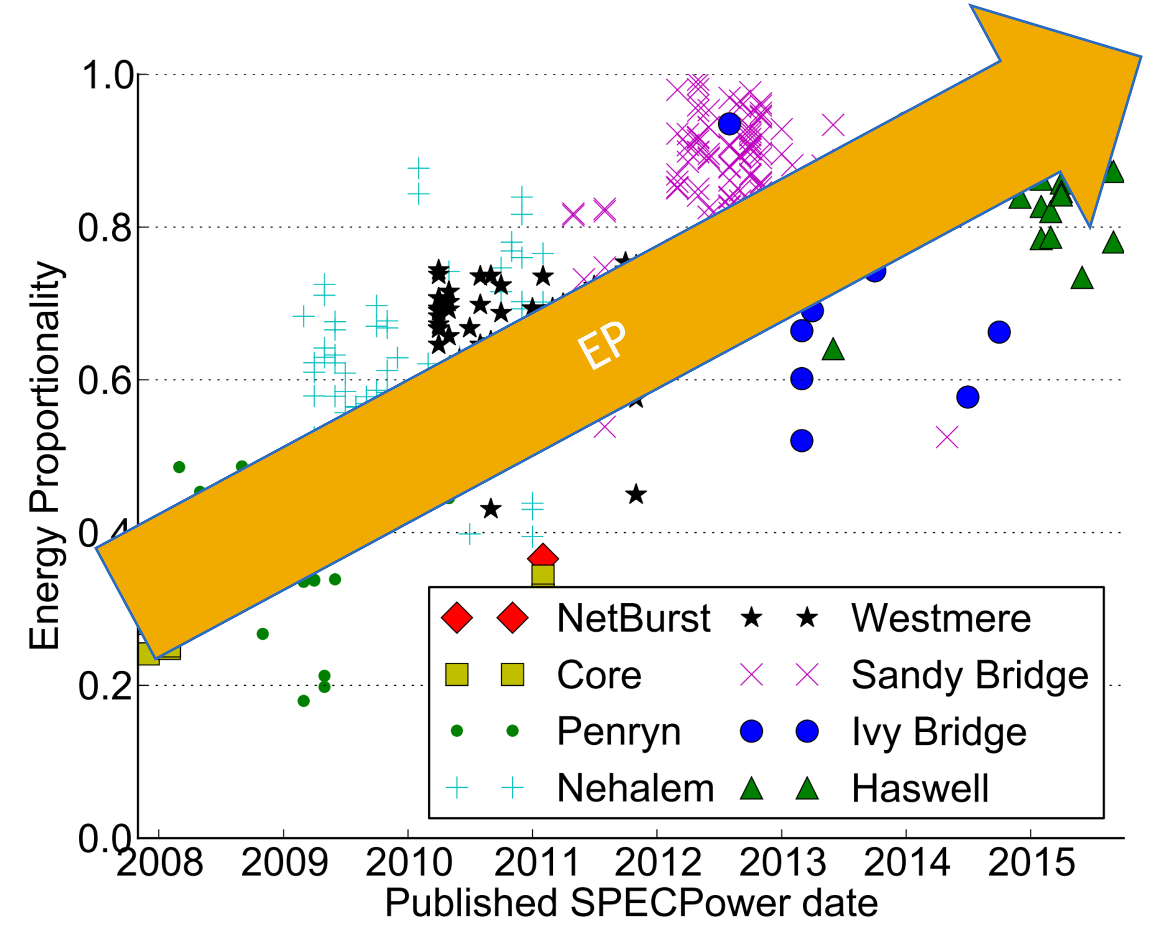
  - EP range: (0,2), 1 = Ideal EP, 0 = Energy disproportional

- More metrics in [1]



[1] D. Wong and M. Annavaram. "Knightshift: Scaling the energy proportionality wall through server-level heterogeneity." MICRO 2012.

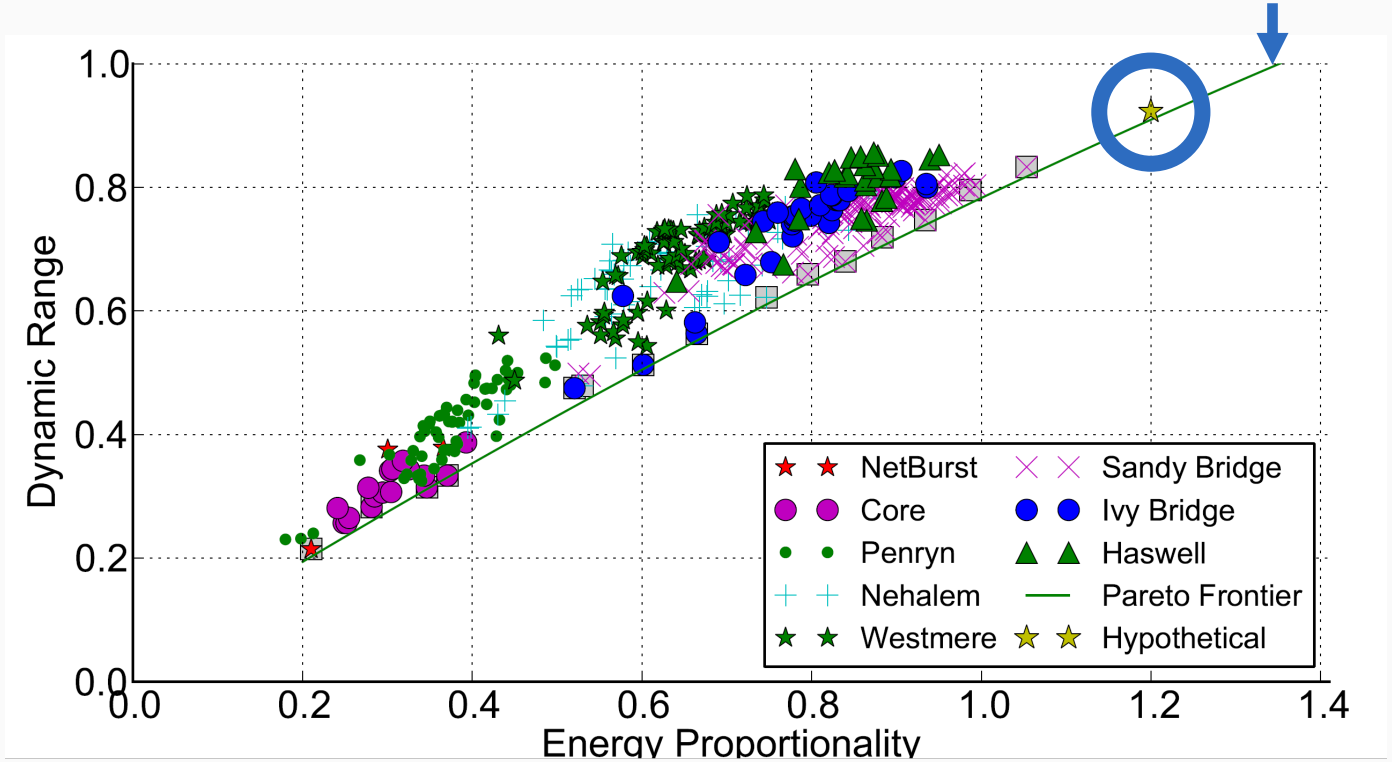
Servers are nearly energy proportional



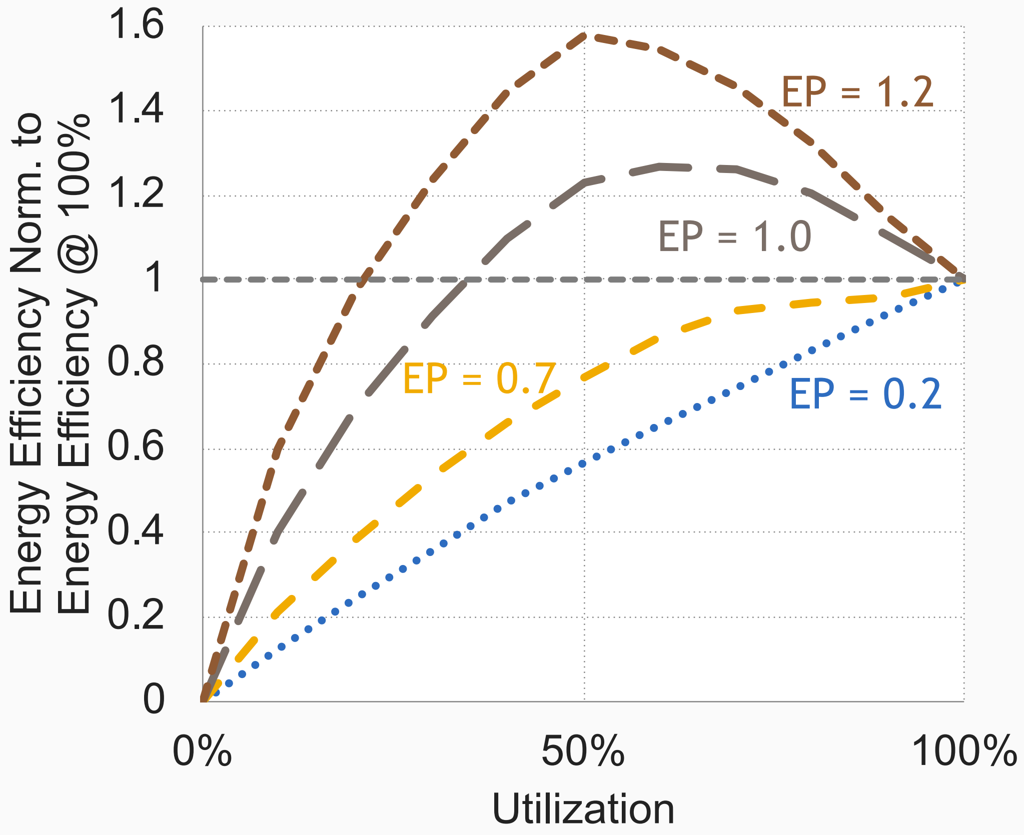
- Published SPECpower results
  - 426 servers
  - 12/2007 – 9/2015
- Most servers today are nearly energy proportional

What is the limit of EP?

- Identified Pareto frontier between DR and EP
- With ideal dynamic range, best possible EP = 1.35
- Hypothetical server where non-processor components are as proportional as processor
  - Pareto frontier still holds true for this extreme case
- Practical EP limit = 1.2



Peak Energy Efficiency ≠ Peak Utilization

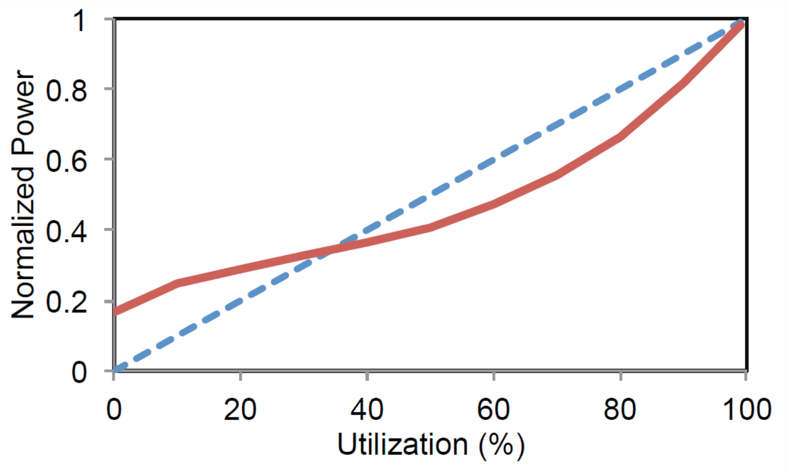


- EP = 1.0 servers achieve peak efficiency @ 60% utilization
- Future super EP servers (EP = 1.2) can achieve peak efficiency @ 50% utilization
- Peak Efficiency point shifts as EP improves

Schedulers are not peak efficiency aware[2]

Uniform scheduling

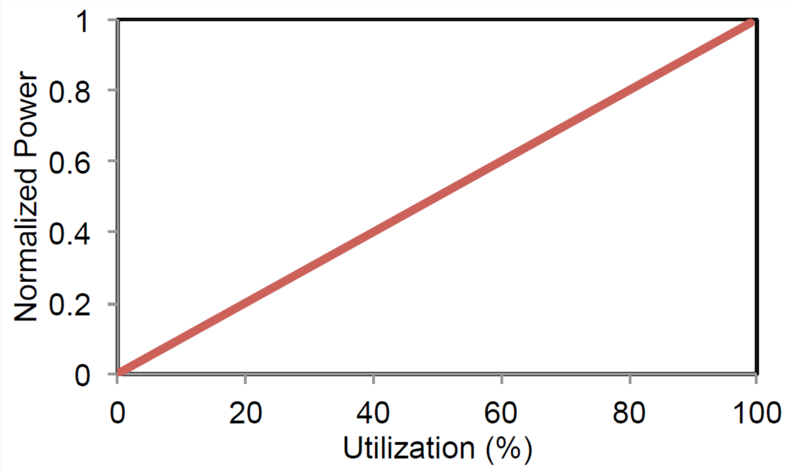
- Cluster-wide EP reflects underlying server's EP



- If server's EP is poor, then cluster's EP is poor

Packing Scheduling

- Have exact number of servers for load

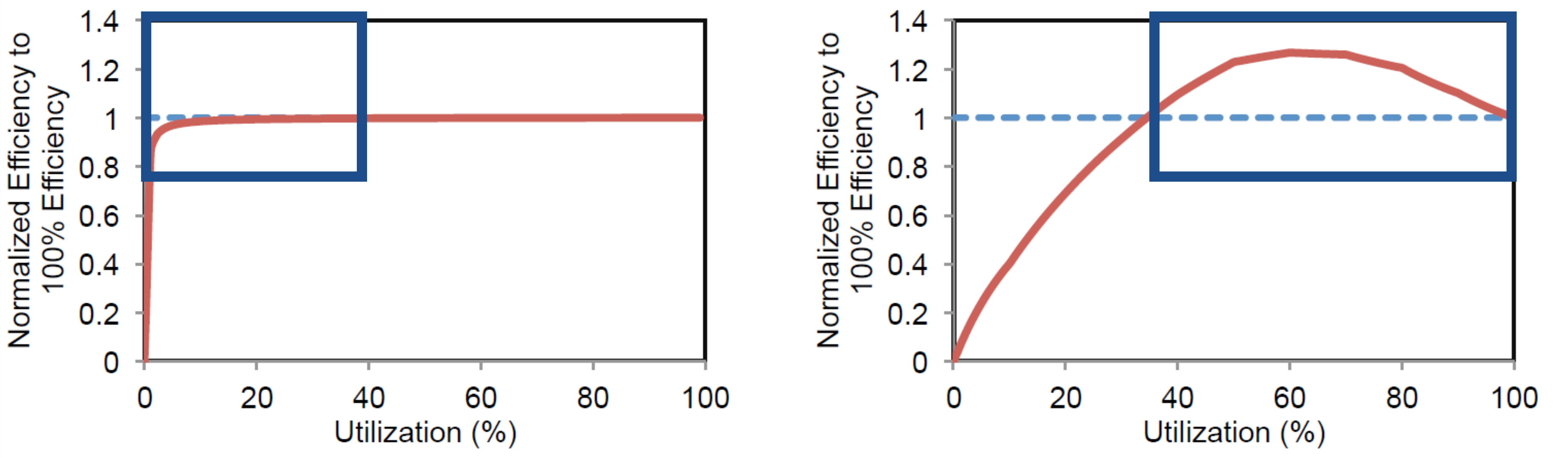


- Cluster's EP is ideal

[2] D. Wong and M. Annavaram. "Implications of high energy proportional servers on cluster-wide energy proportionality" HPCA 2014.

One-size does not fit all

- Prior work[2] identified that Packing is better for low EP servers, while Uniform is better for high EP servers
- We also identified that different utilization favors different scheduling policies



(a) Packing (b) Uniform

[2] D. Wong and M. Annavaram. "Implications of high energy proportional servers on cluster-wide energy proportionality“ HPCA 2014.

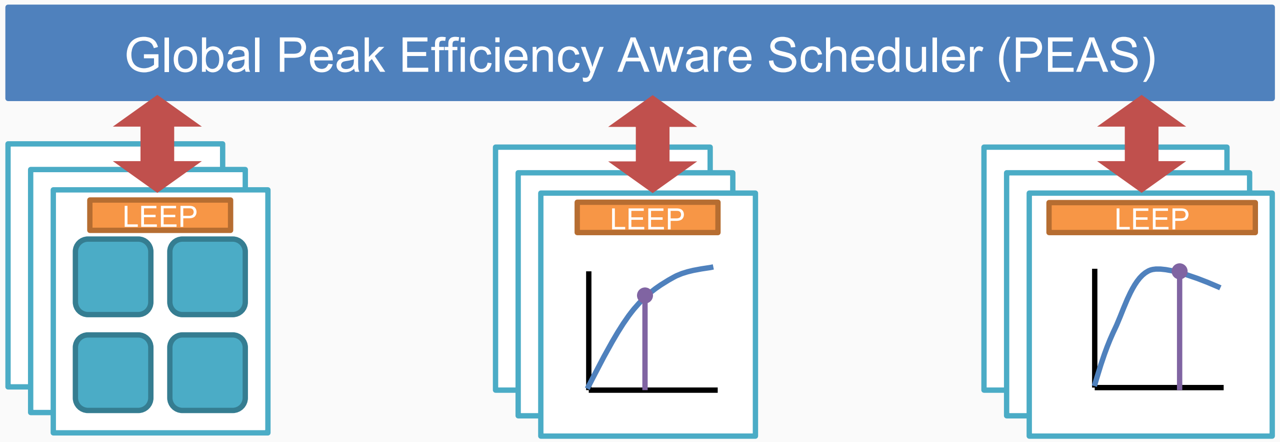
Peak Efficiency Scheduling (PEAS)

- Goal:
  - Capture behavior of both Packing and Uniform scheduling
  - 1. Pack servers up to peak efficiency point
  - 2. Then issue requests uniformly

- Intuition:
  - Quickly get servers to peak efficiency point
  - Move away from peak efficiency point as slowly as possible

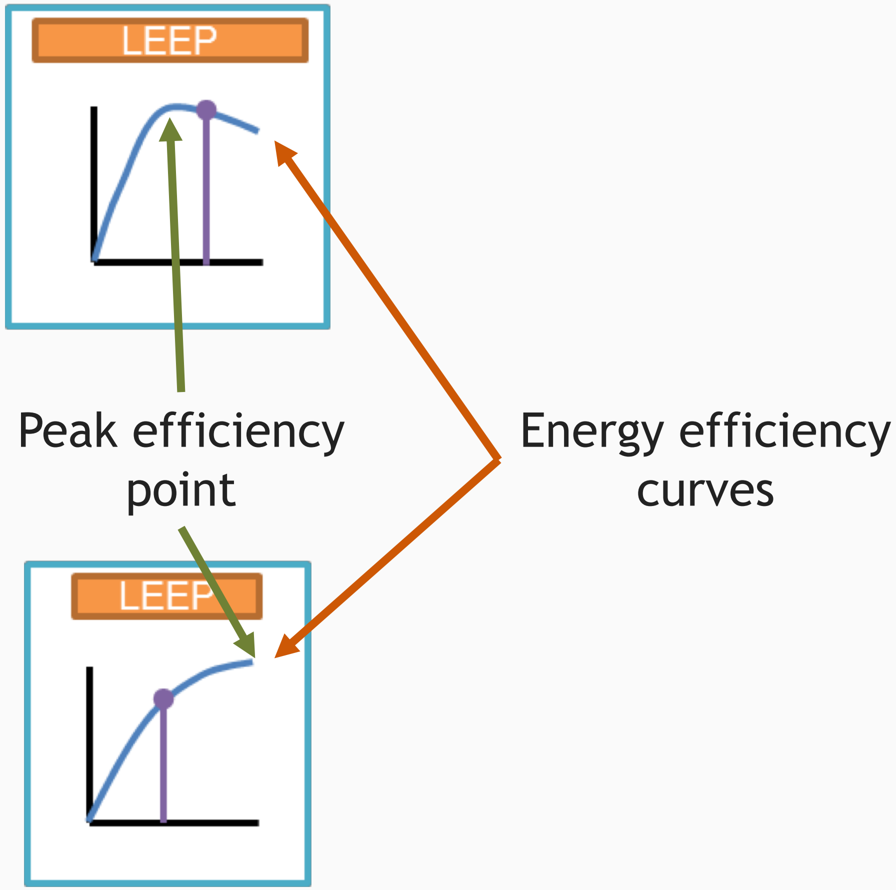
PEAS Design

- Per server local energy efficiency profiler (LEEP)
  - Identify peak energy efficiency point
- Global peak efficiency aware scheduler (PEAS)
  - Schedule workloads to server with highest energy efficiency



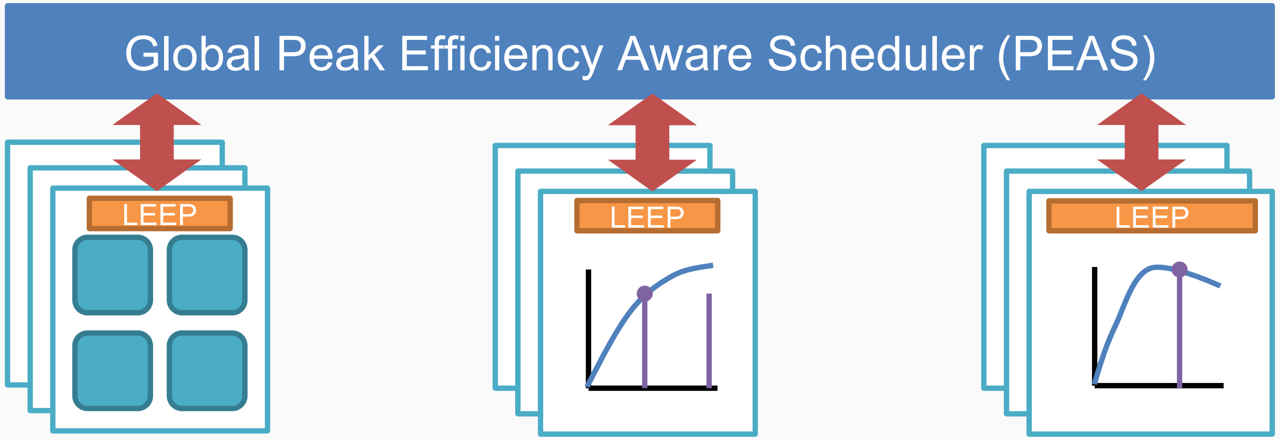
Local energy efficiency profiler (LEEP)

- Daemon periodically samples utilization and power consumption
- Dynamically captures energy efficiency curve of individual server configuration and workload
- Generates energy efficiency curve to identify peak efficiency point



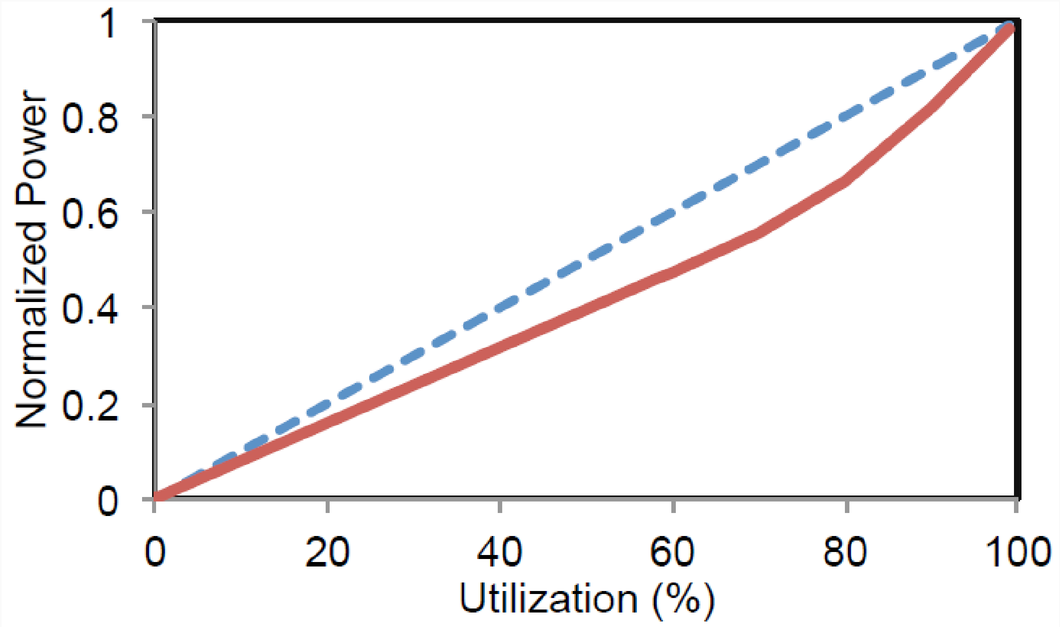
Global peak efficiency aware scheduler (PEAS)

- Scheduler maintain sorted list of servers based on peak energy efficiency
- Receives utilization update from servers
- Pack servers up to peak efficiency point, then issue requests uniformly



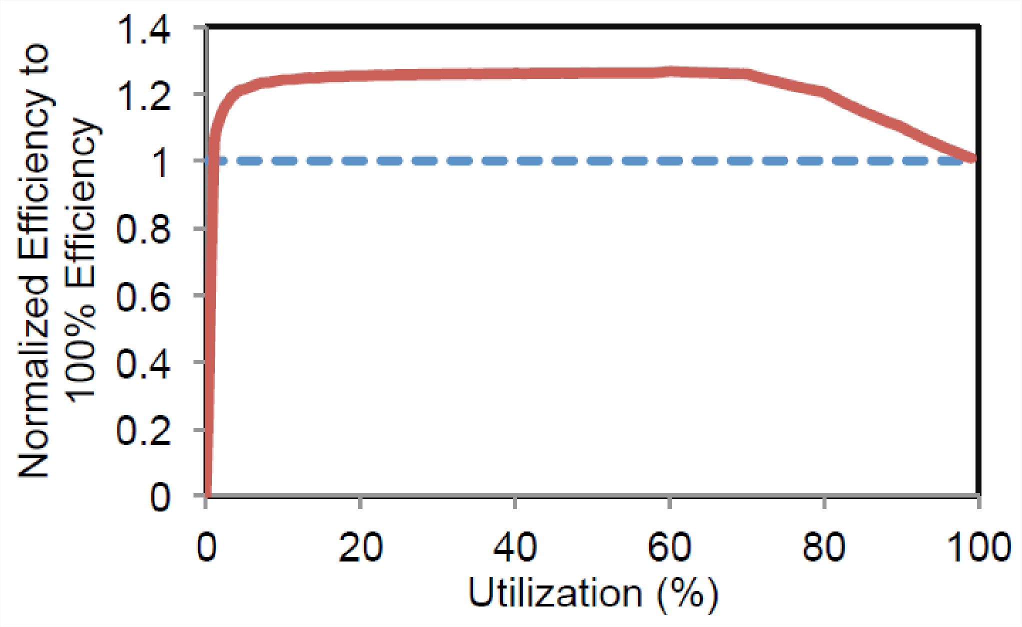
PEAS provide better-than-ideal EP and efficiency!

Energy proportionality



- Always outperform ideal EP

Energy efficiency



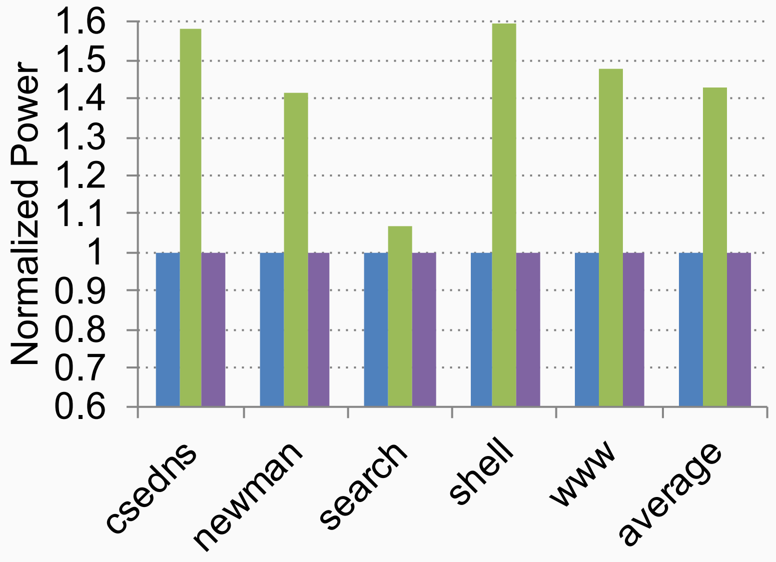
- Sustain peak energy efficiency

Evaluation Methodology

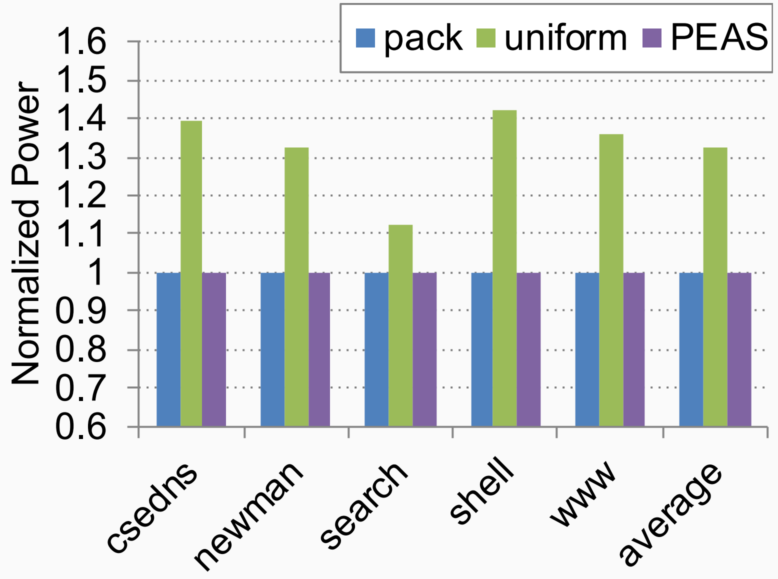
- BigHouse data center simulator
- 100 servers
  - Dual-socket 18-core processors
    (similar to recently reported SPECpower results)
- Four levels of EP: Low=0.24, Med=0.73, High=1.0, Super=1.2
- Evaluated 5 workloads
  - DNS (csedns), Mail (newman) , Apache (www), Search and Shell

Power Consumption

- Packing-based scheduling is most effective at low-med EP



Low EP

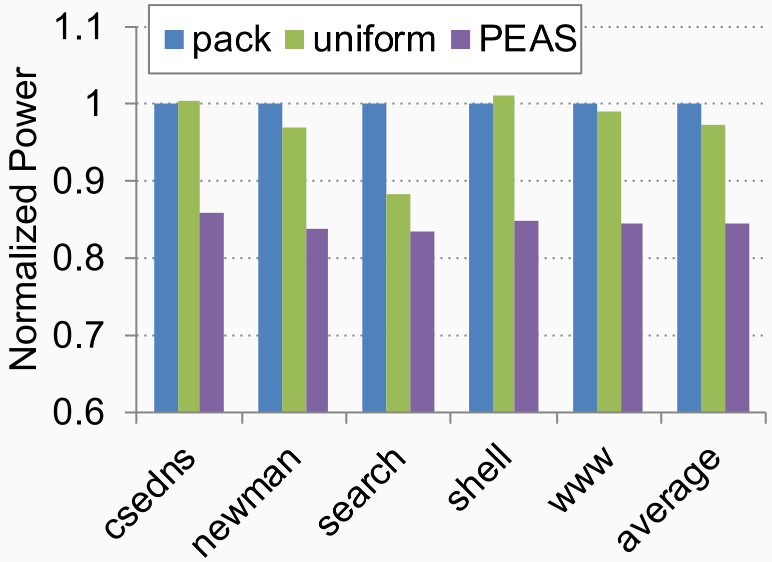


Med EP

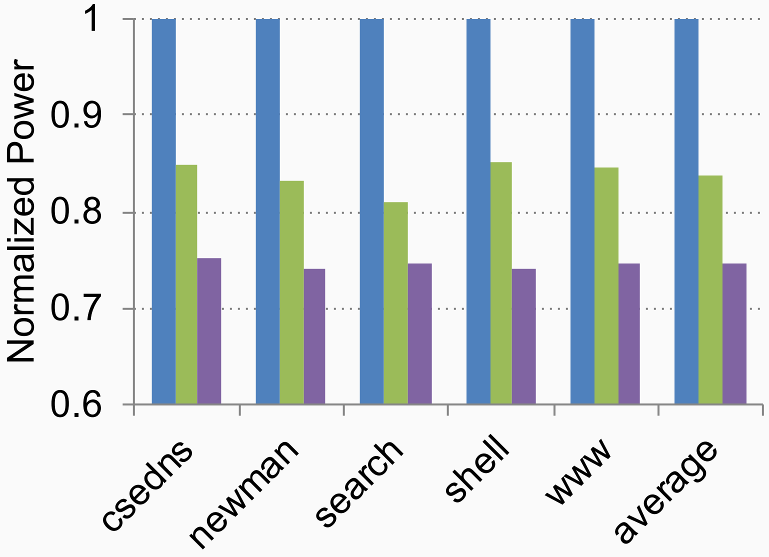
- PEAS matches performance of Packing at low-med EP

Power Consumption

- Uniform outperforms packing at high EP



High EP

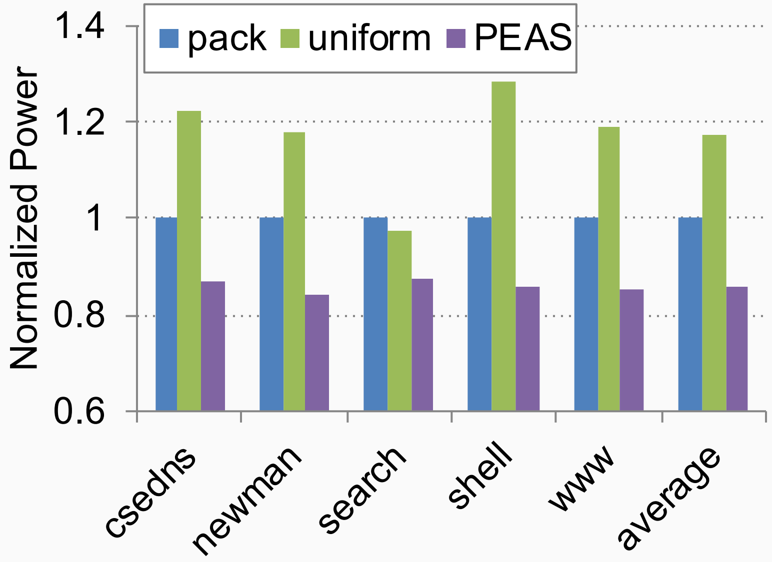


Super EP

- PEAS outperforms both uniform and packing!

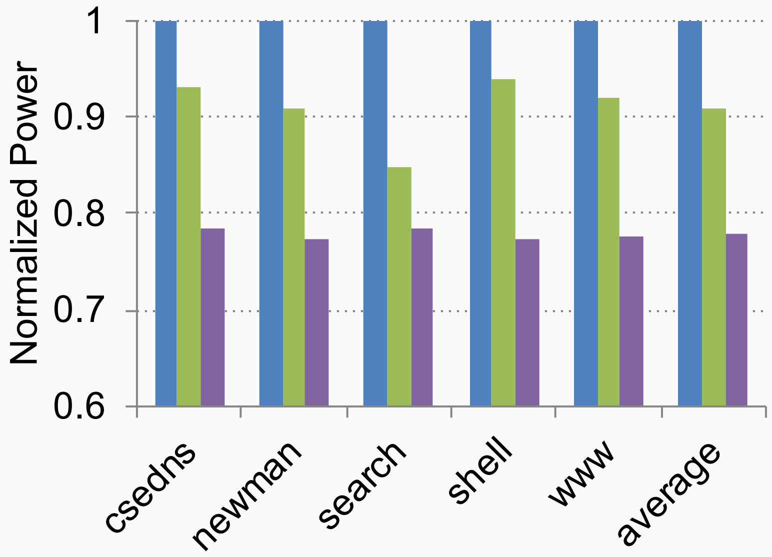
Heterogeneous Cluster

Mix of 25% Low, Med, High, and Super EP servers



- Uniform performs worst due to inability to mask low-med EP servers

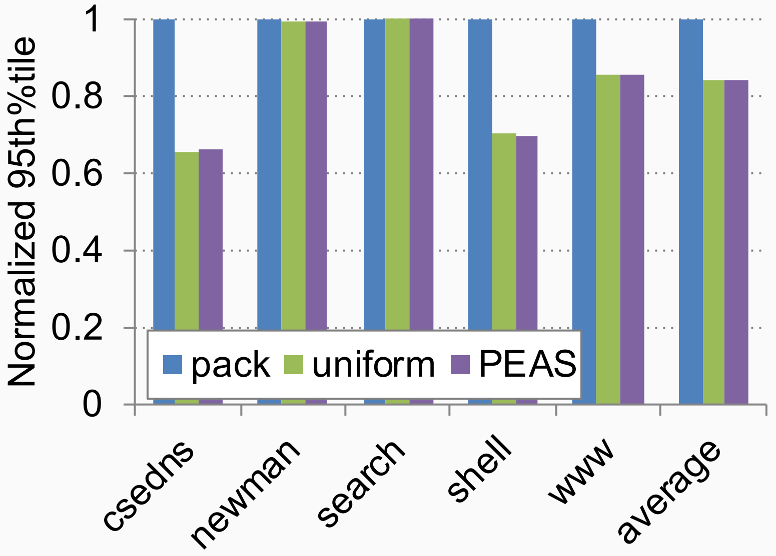
Mix of 50% High and Super EP servers



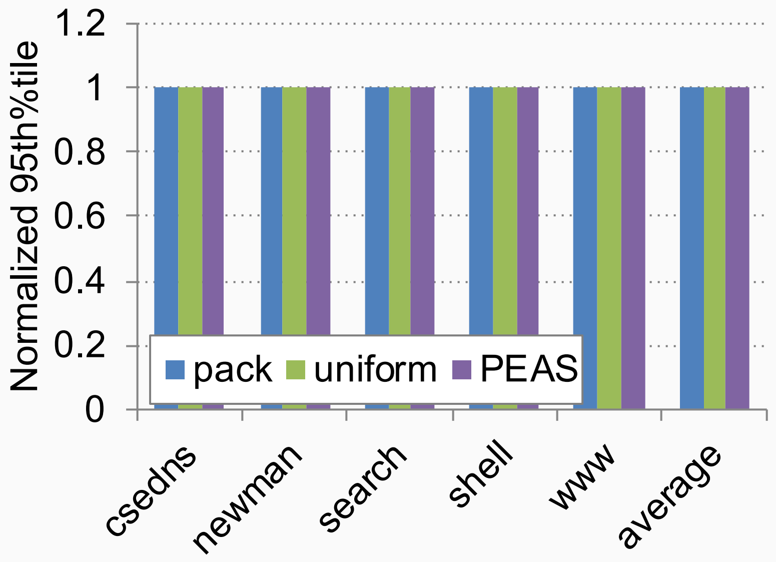
- PEAS consistently outperform other schedulers across various mixes of servers

Latency

- Observed tail latency similar to Uniform scheduling
  - Holds true across various sleep transition times



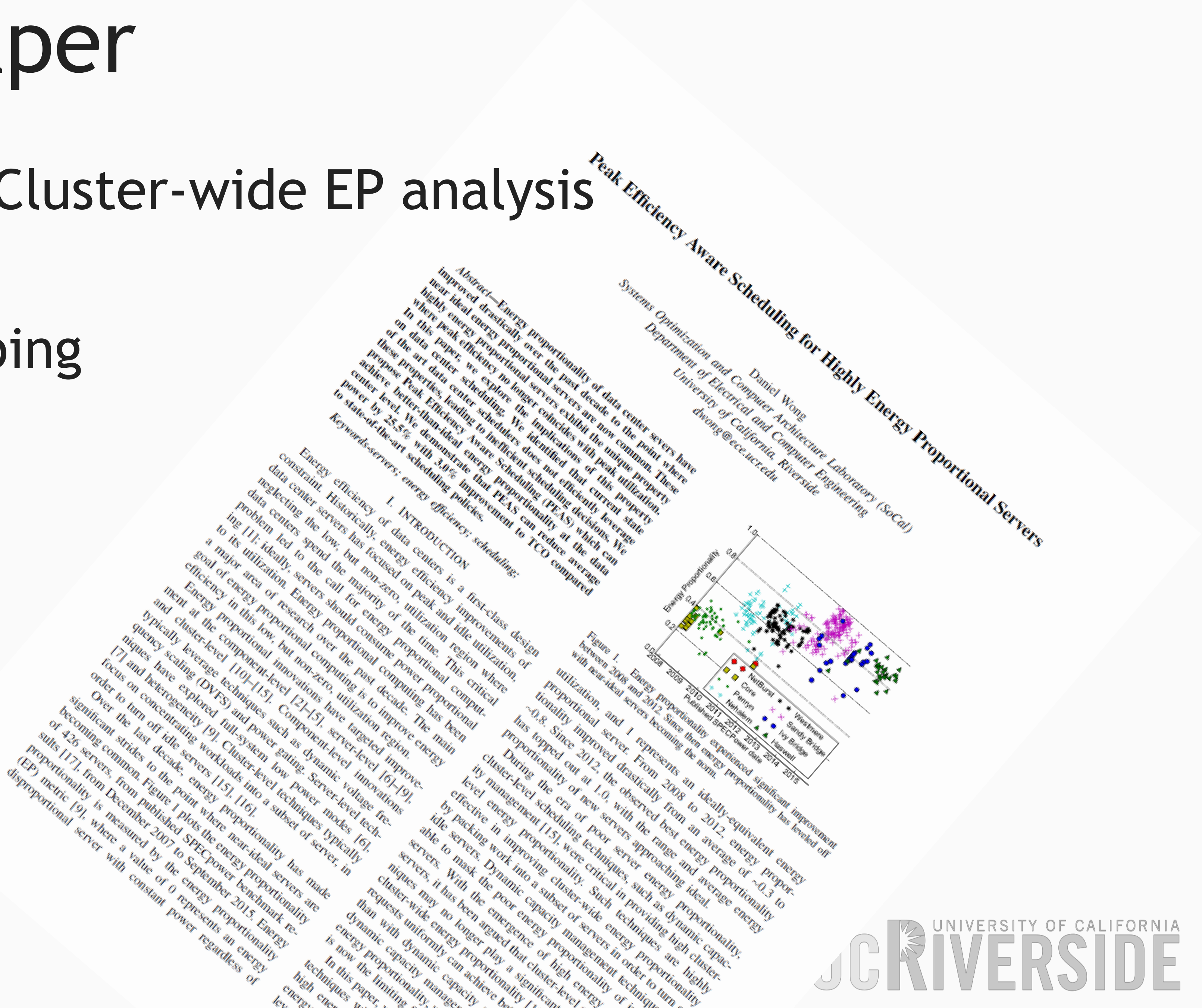
20s transition time



0s transition time

More in the paper

- Analytical Best-case Cluster-wide EP analysis
- TCO impact
- Effect on power capping



Peak Efficiency Aware Scheduling for Highly Energy Proportional Servers

Daniel Wong
Systems Optimization and Computer Architecture Laboratory (SoCal)
Department of Electrical and Computer Engineering
University of California, Riverside
dwong@ece.ucr.edu

Abstract—Energy proportionality of data center servers have improved drastically over the past decade to the point where near ideal energy proportional servers are now common. These highly energy proportional servers exhibit the unique property where peak efficiency no longer coincides with peak utilization. In this paper, we explore the implications of this property on data center scheduling. We identified that current state of the art data center schedulers does not efficiently leverage these properties, leading to inefficient scheduling decisions. We propose Peak Efficiency Aware Scheduling (PEAS) which can achieve better-than-ideal energy proportionality at the data center level. We demonstrate that PEAS can reduce average power by 25.5% with 3.0% improvement to TCO compared to state-of-the-art scheduling policies.

Keywords—servers; energy efficiency; scheduling;

Figure 1. Energy proportionality experienced significant improvement between 2008 and 2012. Since then energy proportionality has leveled off with near-ideal servers becoming the norm.

Conclusion

- Servers are nearly energy proportional
- Peak energy efficiency no longer occurs at peak utilization
- Peak Efficiency Scheduling (PEAS) can achieve better-than-ideal cluster-wide energy proportionality
  - Consistently outperforms Uniform and Packing scheduling

Thank you! Questions?

Peak Efficiency Aware Scheduling for Highly Energy Proportional Servers

Daniel Wong
dwong@ece.ucr.edu

University of California, Riverside
Department of Electrical and Computer Engineering

SOCAL Systems Optimization + Computer Architecture Lab

UC RIVERSIDE UNIVERSITY OF CALIFORNIA