Z6110X0035: Introduction to Cloud Computing - Power Management

Lecturer: Prof. Zichen Xu

DC Is The "SUV"

DCs in the world use up to 250 billion kWh
2% of US electricity consumption goes to DC

(the 51st state?)

The impact:

- Social movement, e.g. greenpeace.com
- Use the energy wisely, e.g. power-aware workload management
- Try alternatives, e.g. renewable energy

[1] Growth in data center electricity use 2005 to 2010 by Jonathan Koomey
[2] EPA report to Congress (EPA 2007)

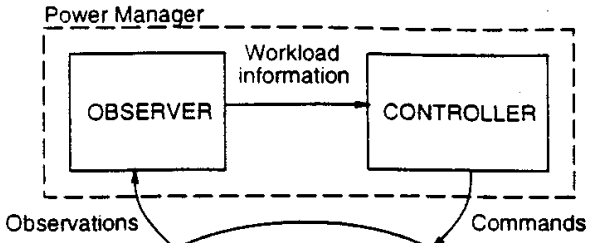
Power-Aware Workload Management

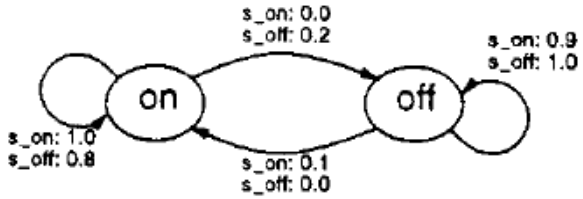
- Predictive schemes

  e.g. static model/adaptive model (the first paper)

- Stochastic optimum control
  - Deal with the existence of uncertainty in the data

  e.g. Markov process





- Others

  management on specific system (the second paper)

What Else We Need to Learn Today

- What's the novelty in their ideas?
- Comparing to the classic ones, what technique used?
- What's new in their environment setup?
- How they measure their performance?

Renewable and Cooling Aware Workload Management for Sustainable Data Centers

Zhenhua Liu[1], Yuan Chen[2], Cullen Bash[2], Adam Wierman[1], Daniel Gmach[2], Zhikui Wang[2], Manish Marwah[2], and Chris Hyser[2]

[1] :California Institute of Technology [2] :HP Lab

Courtesy to Mr. Liu

SIGMETRICS 2012

Making Data Centers Sustainable

Significant amount of work on improving energy efficiency

Power Usage Effectiveness

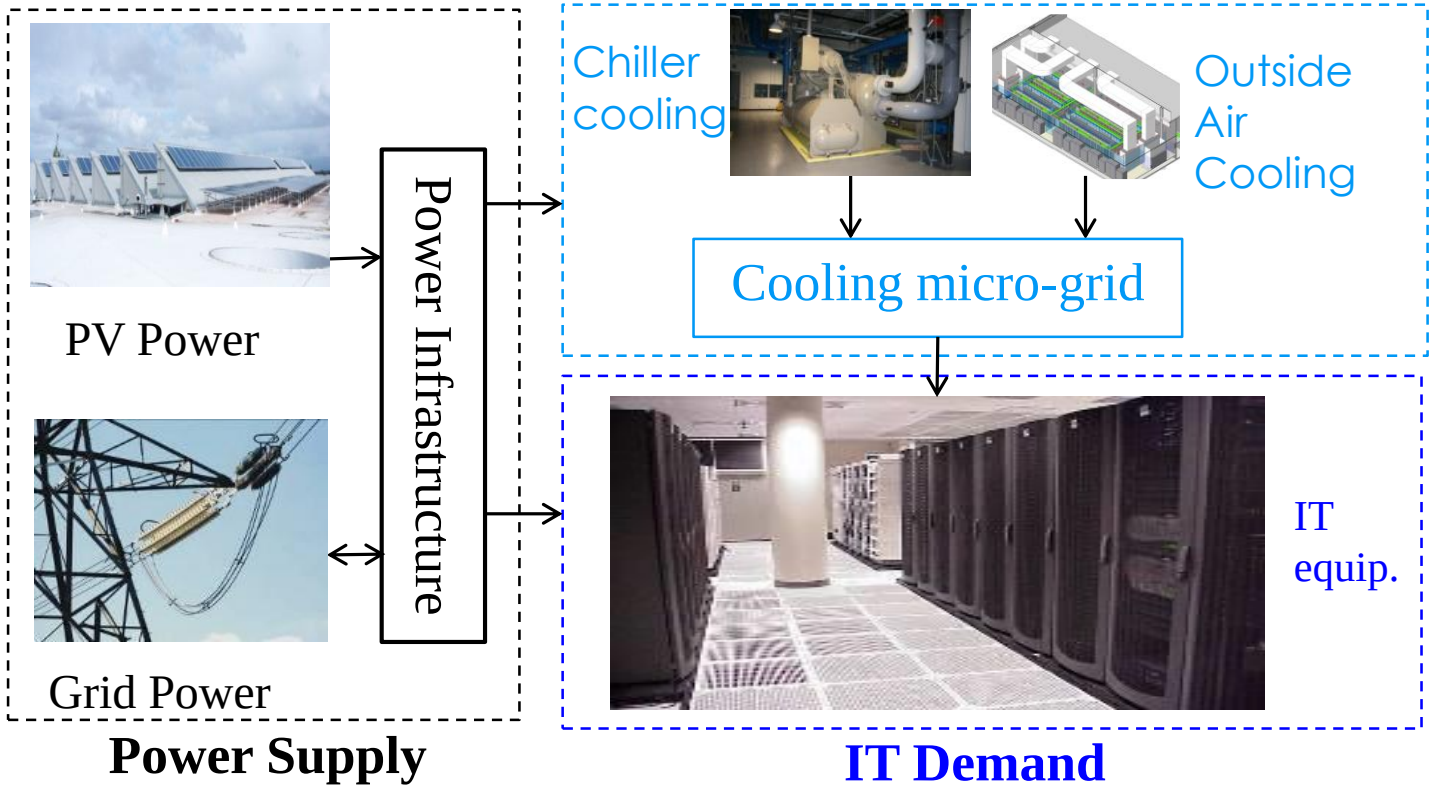
$$\text{PUE} = \frac{\text{Total facility power}}{\text{IT equipment power}}$$

Sustainability requires renewable supplies

E.g. Apple's new North Carolina data center

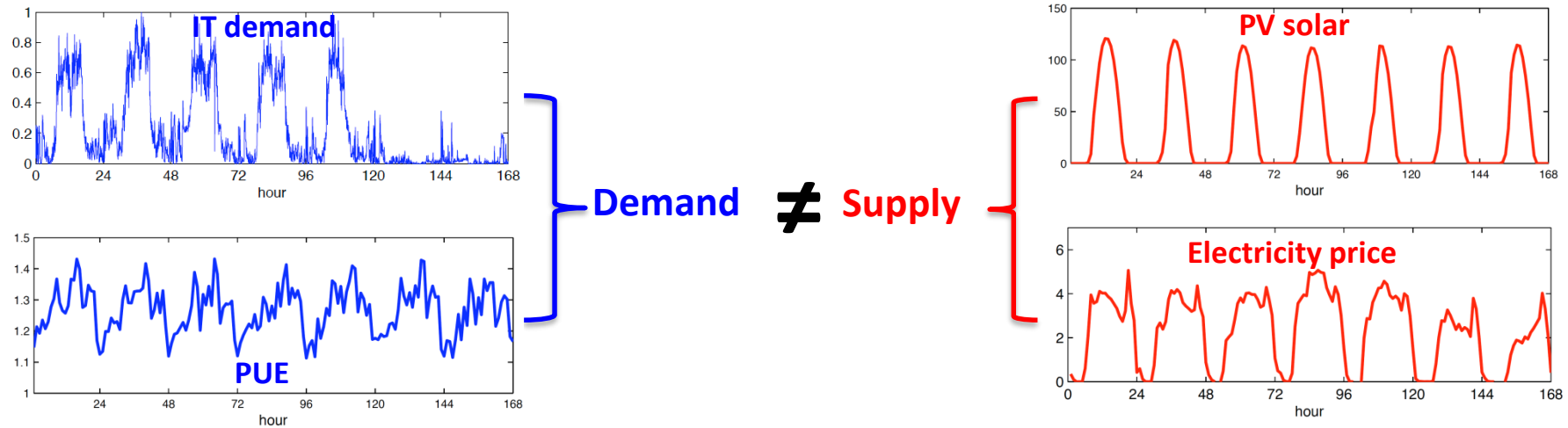
But **HOW** did the solution efficiently use the renewable generation?

Sustainable Data Center



Work Proposed In This Paper

- Manage data centers to better use renewable supplies



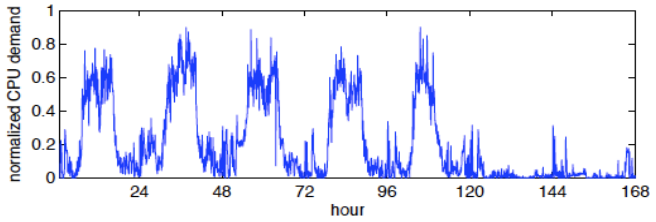
Novelty: Optimizing both cooling and computing power under the constraints of renewable energy supplies and price.

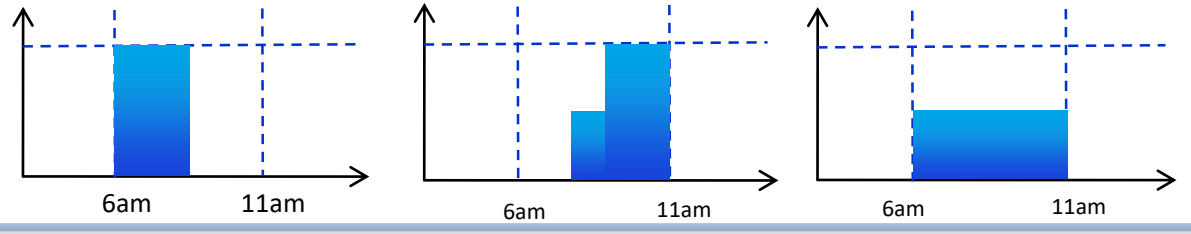
System Study: Observation 1

Most data centers support two types of IT workloads:

critical interactive applications that run 24x7

delay tolerant, ba



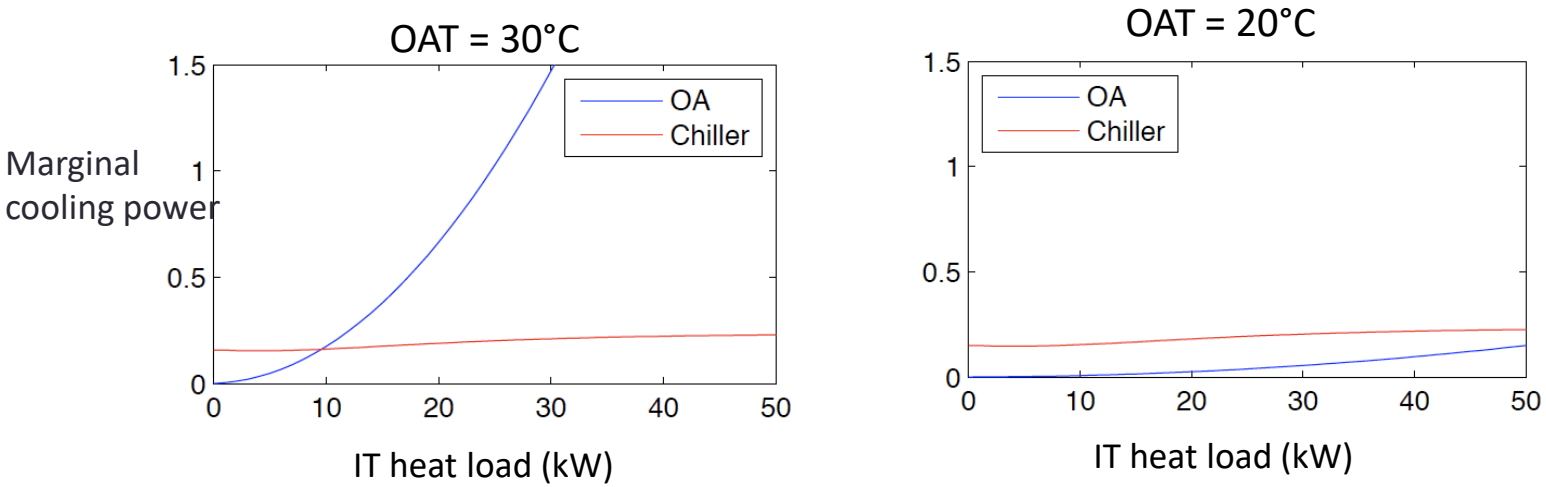


Opportunity 1: load shifting

Schedule delay tolerant workloads to better use renewable generations and efficient cooling supplies.

System Study: Observation 2

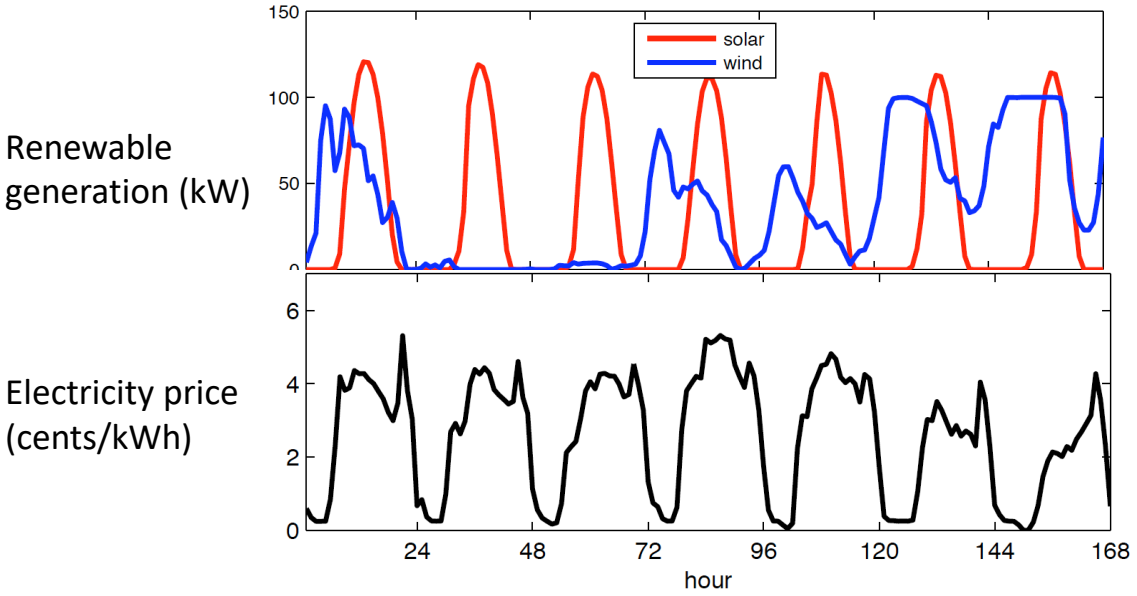
- Different cooling supplies have different cooling efficiencies
- Dynamic over time depend on external conditions and load



Notes:
OA: outside air
OAT: outside air temperature
Critique: no quantification of IT head load. The concepts of IT demand and head load are messed, though they are correlated.

System Study: Observation 3

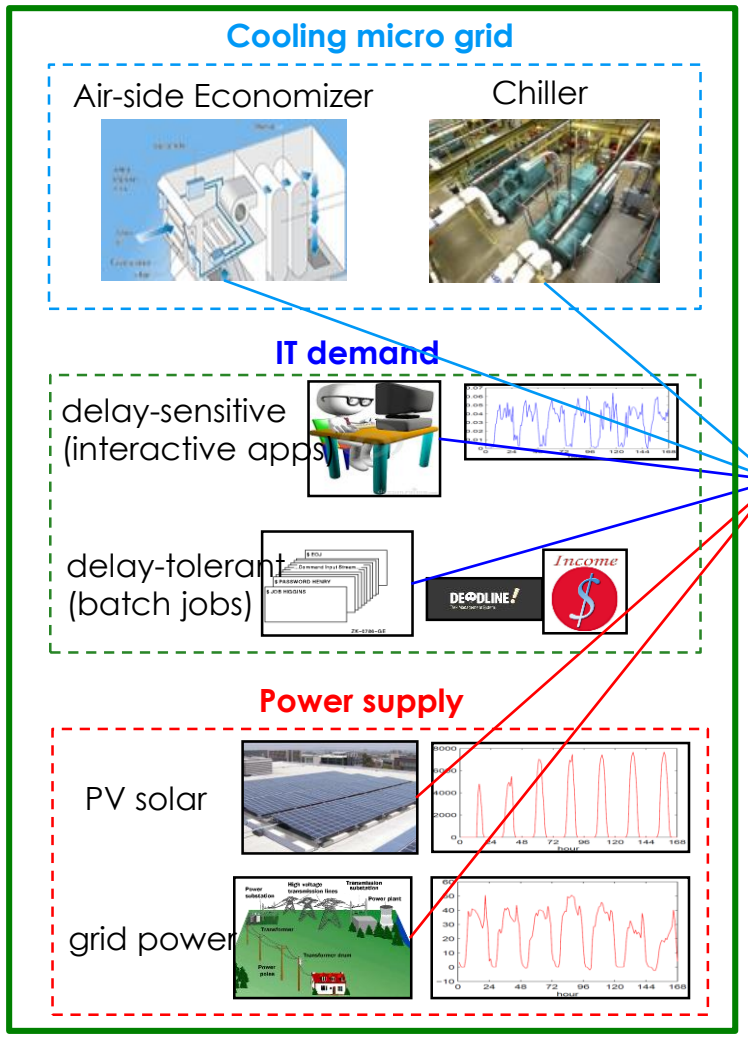
- The availability and cost are often dynamic over time



Opportunity 3: supply-aware demand shaping

Follow the availability of renewable generation.
Utilize timeslots with low electricity prices.

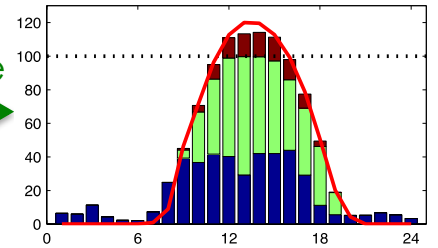
System Overview



Critique: missing energy storage part (e.g. UPS) in the implementation, which is inconsistent with their modeling work.

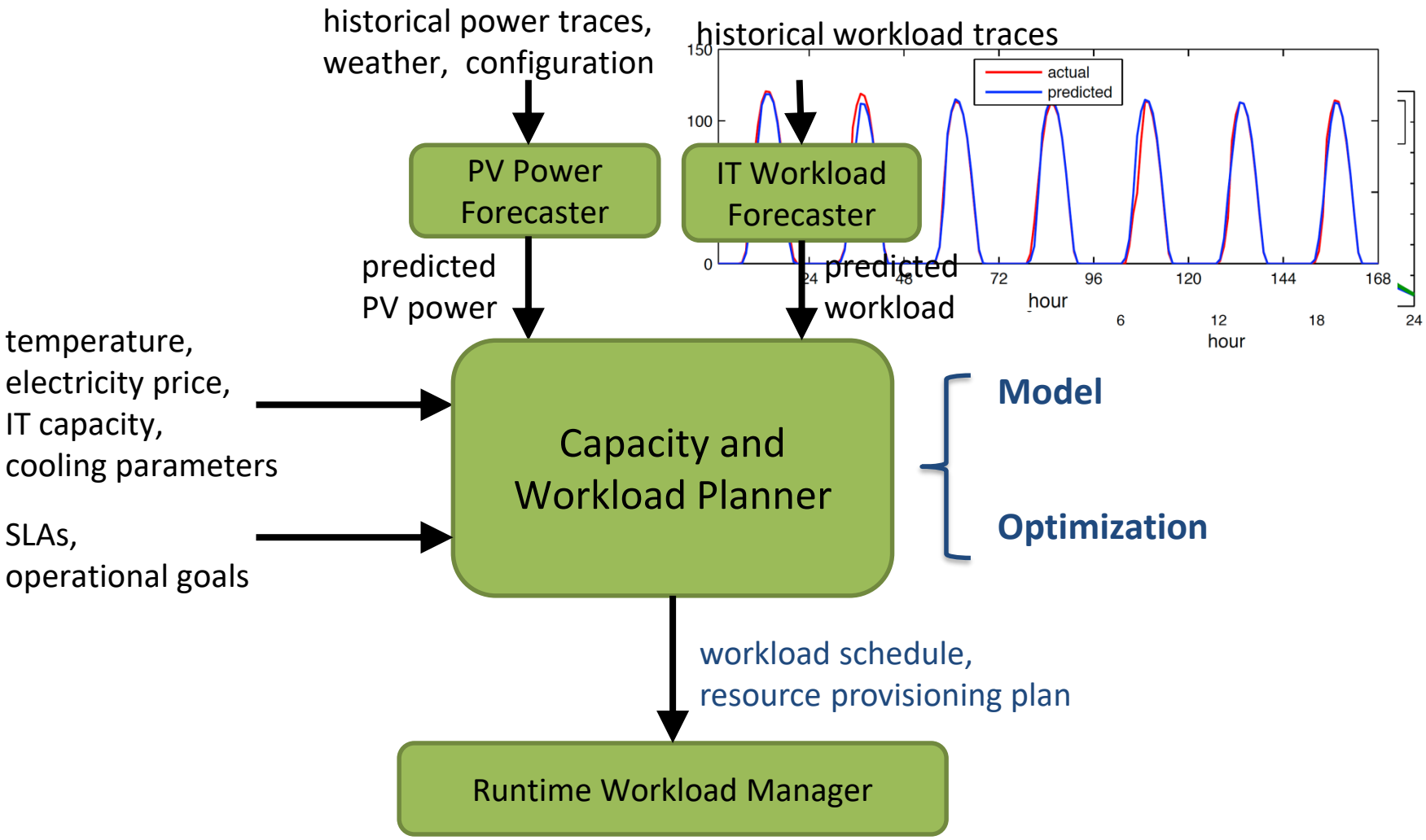
Capacity and Workload Planner

schedule



Min environment impact
Min energy cost
Meet SLAs

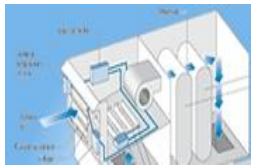
System Architecture



Model

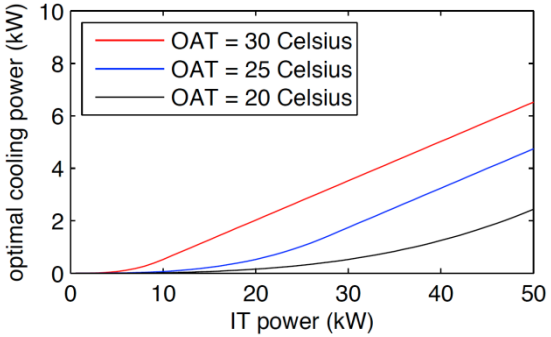
Cooling micro grid

Air-side Economizer



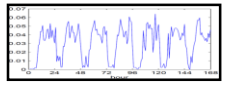
Chiller

$$f_a(d) = kd^3, 0 \le d \le \overline{d}$$
$$f_c(d) = \gamma d$$
$$c(d) = \min_{d_2 \in [0,\overline{d}]} \gamma(d - d_2)^+$$

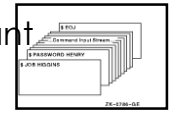


IT demand

delay-sensitive (interactive apps)



delay-tolerant (batch jobs)



$a_i(t)$

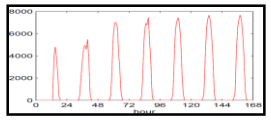
$b_j(t)$

$$d(t) = \Sigma_i a_i(t) + \Sigma_j b_j(t)$$

Starting time S_j, deadline E_j, revenue R_j

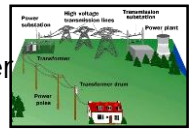
Power supply

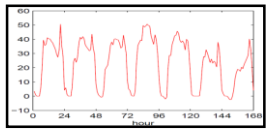
PV solar



$r(t)$

grid power





$p(t)$

Implementation: a detailed modeling on many aspects that enriches technique significance of this paper.

Workload Planner Optimization

$$\min_{\mathbf{b},\mathbf{e}} \Sigma_t p(t)\underbrace{(d(t) + c(d(t)) - r(t) - e(t))^+}_{\text{price, IT, cooling, renewable, energy storage}} - \underbrace{\Sigma_j R_j \left(\Sigma_{t\in[S_j,E_j]} b_j(t)\right)}_{\text{revenue, batch job}}$$

Subject to

$$\Sigma_t b_j(t) \leq B_j,$$

Perform

$$es(t+1) = \rho(es(t) - e($$

$$0 \leq b_j(t) \leq MP_j,$$

$$0 \leq d(t) \leq D,$$

$$0 \leq es(t) \leq ES.$$

Critique:
comparing with the carbon planning paper (SIGMETRICS 2012), this model is over-simplified.
Why only batch jobs generate revenue?

Control variables

b: batch job scheduling

e: energy storage usage

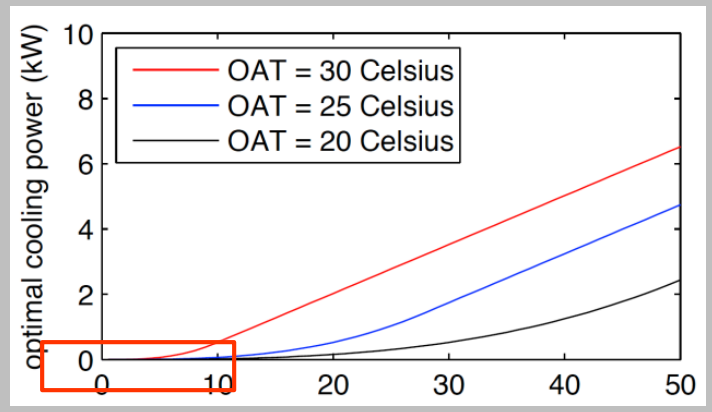
c(d): cooling micro-grid optimi

Convex optimization

Note: convex optimization subfield of optimization
convexity eases optimization

Properties

Convex Analysis 101:



IT Demand d

Cooling Cost:

$$c(d) = \min_{d_2 \in [0,\bar{d}]} \gamma(d - d_2)^+ + k d_2^3$$

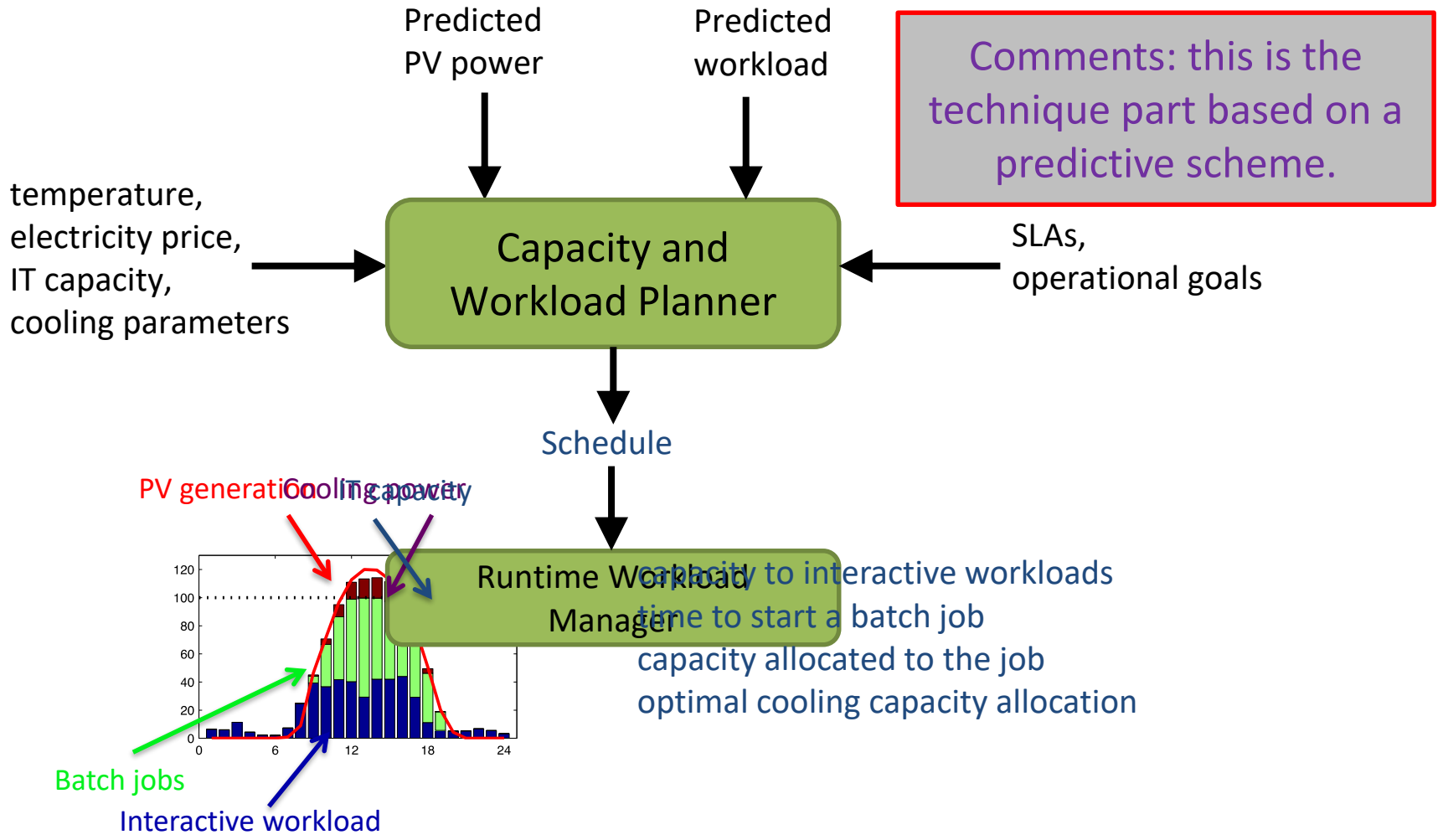
Energy Usage:

$$(d(t) + c(d(t)) - r(t) - e(t))^+$$

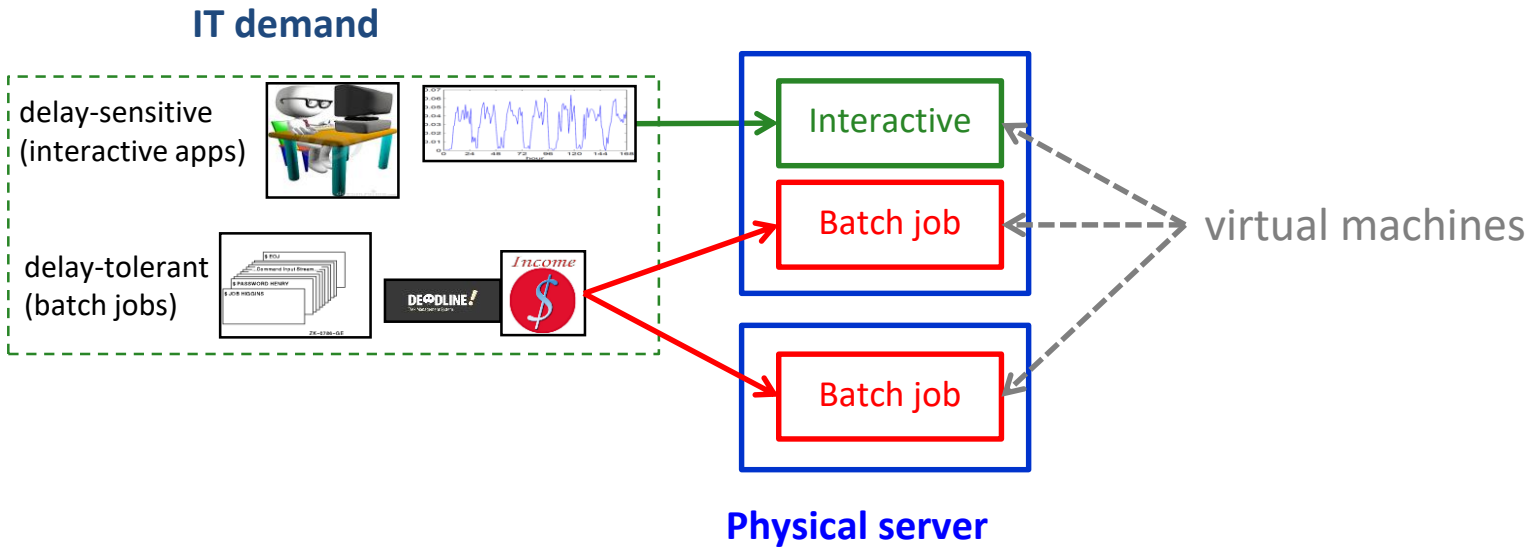
Marginal Power Cost:

$$\partial \left(p(t)(d(t) + c(d(t)) - r(t) - e(t))^+ \right) / \partial (d(t))$$

The Protocol



Runtime Workload Manager



Cgroup for prioritize interactive workload

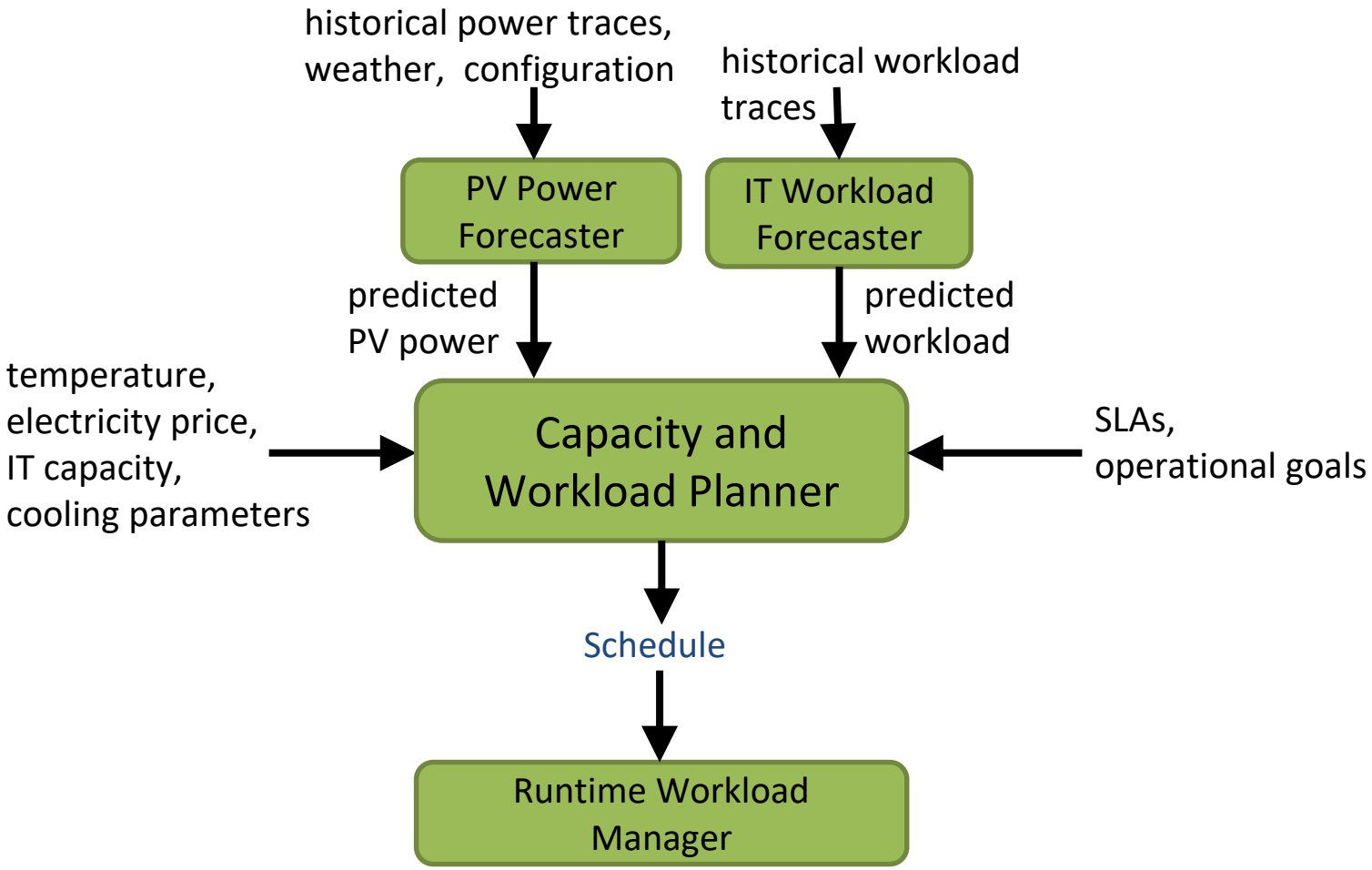
VCPUPIN for CPU capacity allocatic

Live migration for consolidation

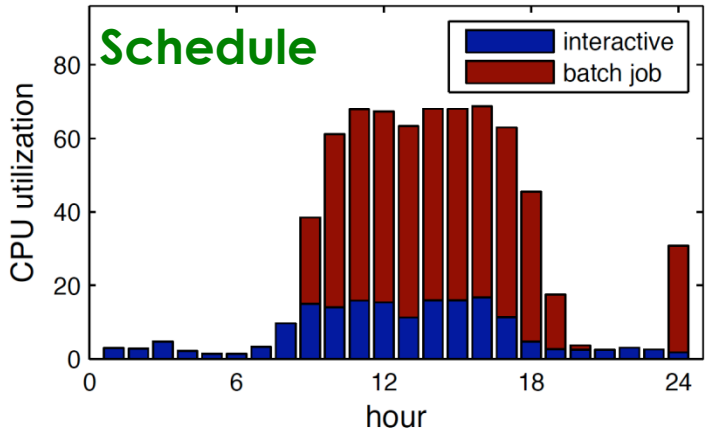
Workload scheduler

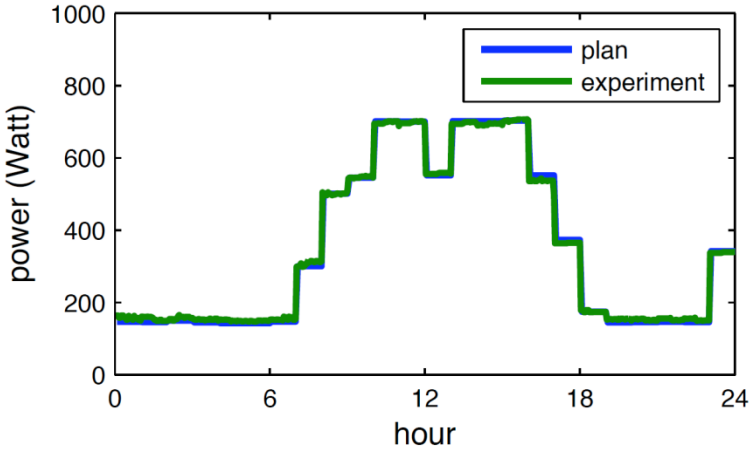
Critique:
Grouping could create lots of noises in the scheme.
Live migration is costly.

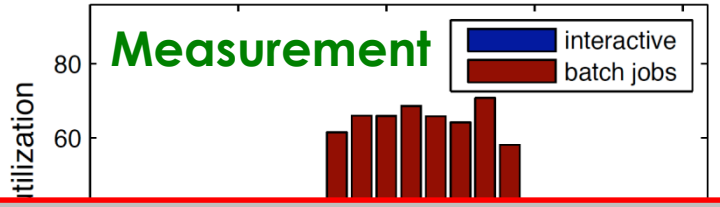
System Architecture



Small-scale Implementation





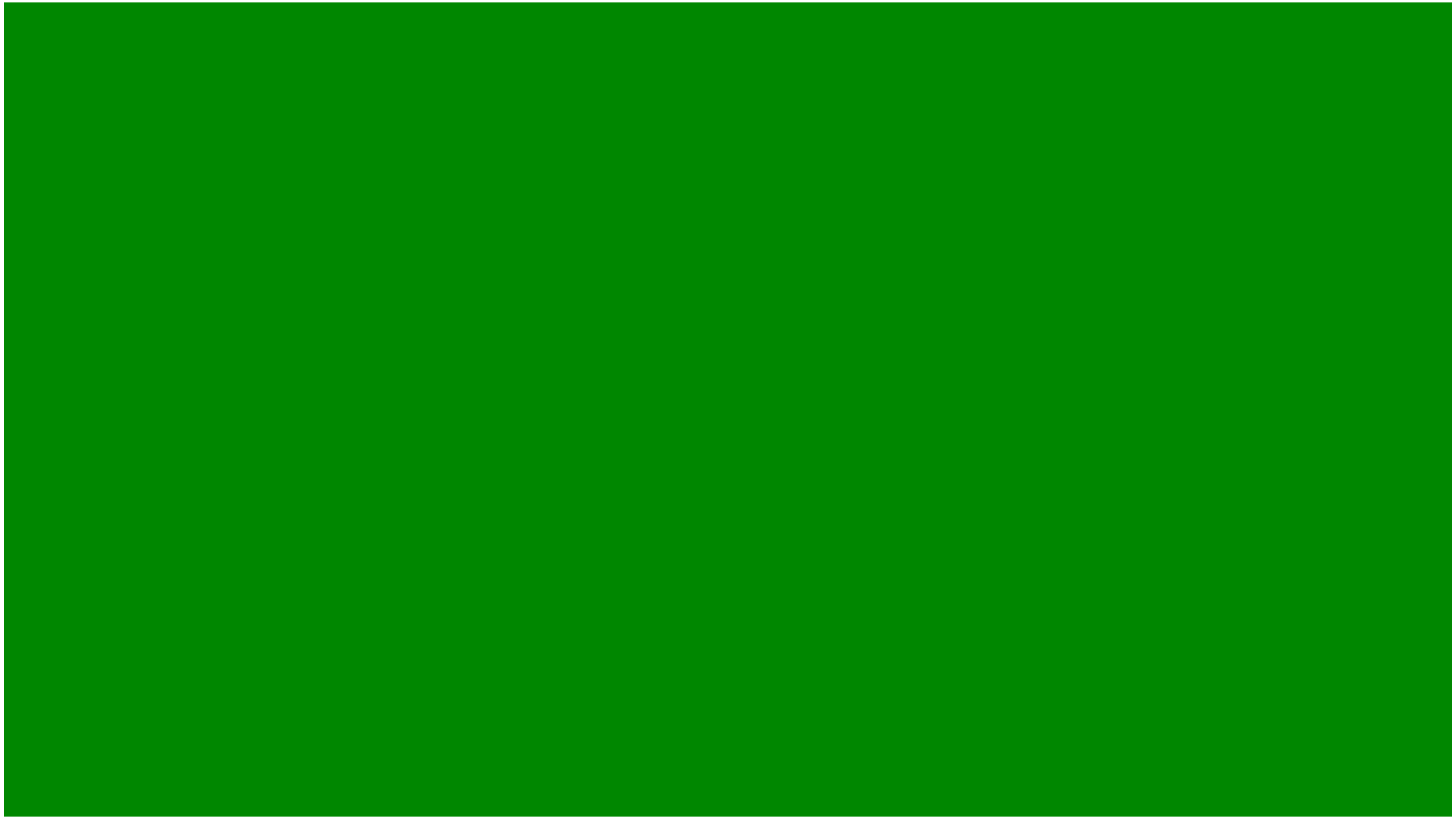


Comments: unlike their GLB paper (SIGMETRICS '11), this one works on temporal scale.

Comments: like many others, they prove their model based on physical and simulation results.

Results show high prediction accuracy regarding power consumption and the optimal beats the "Night" baseline

A Runtime Simulation Of The WM

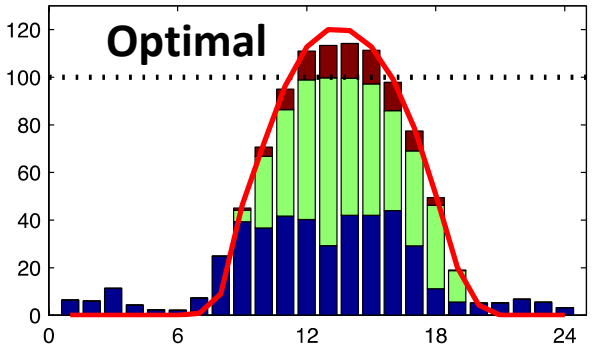


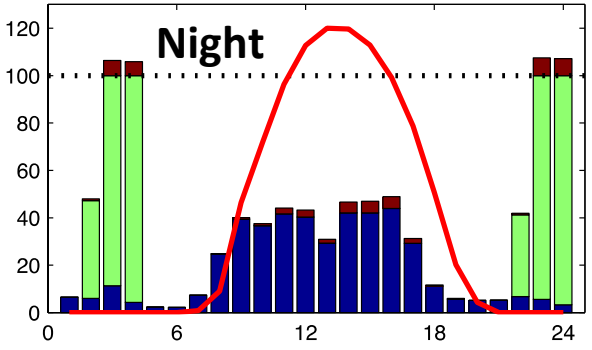
Performance Evaluation

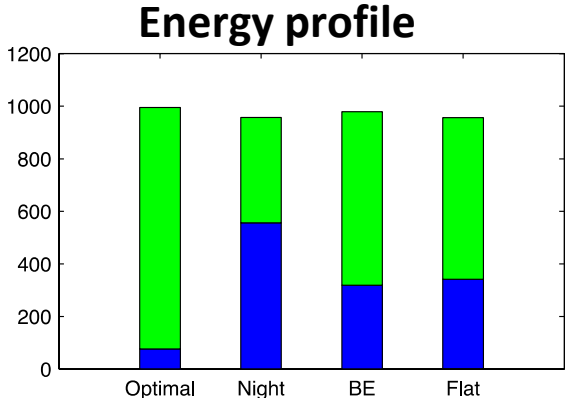
- How valuable is their workload management?
- Is net-zero energy consumption possible?

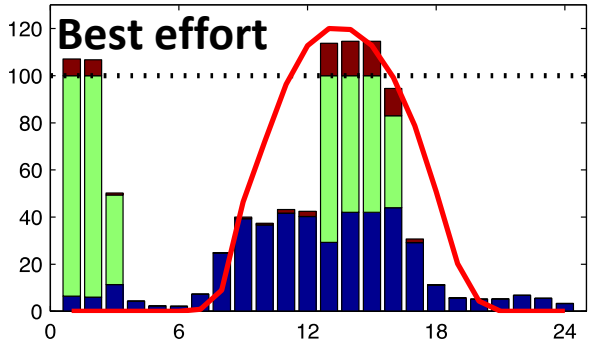
Result presentation: not only like others show how their solution works, but also hand out tech insight and possible impact.

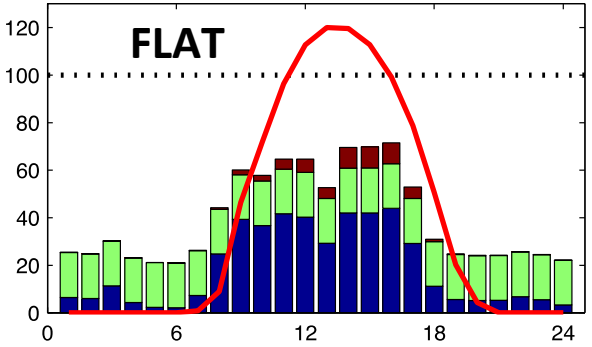
How Valuable Is This Solution?

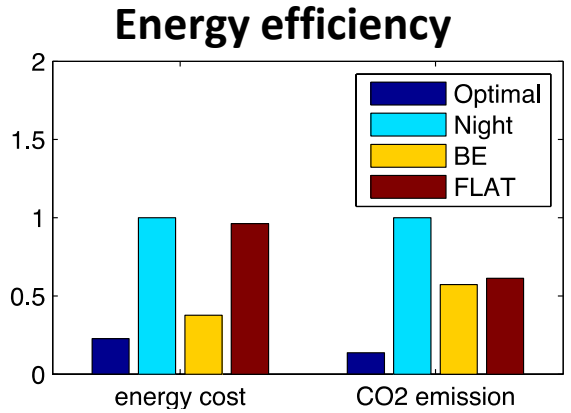






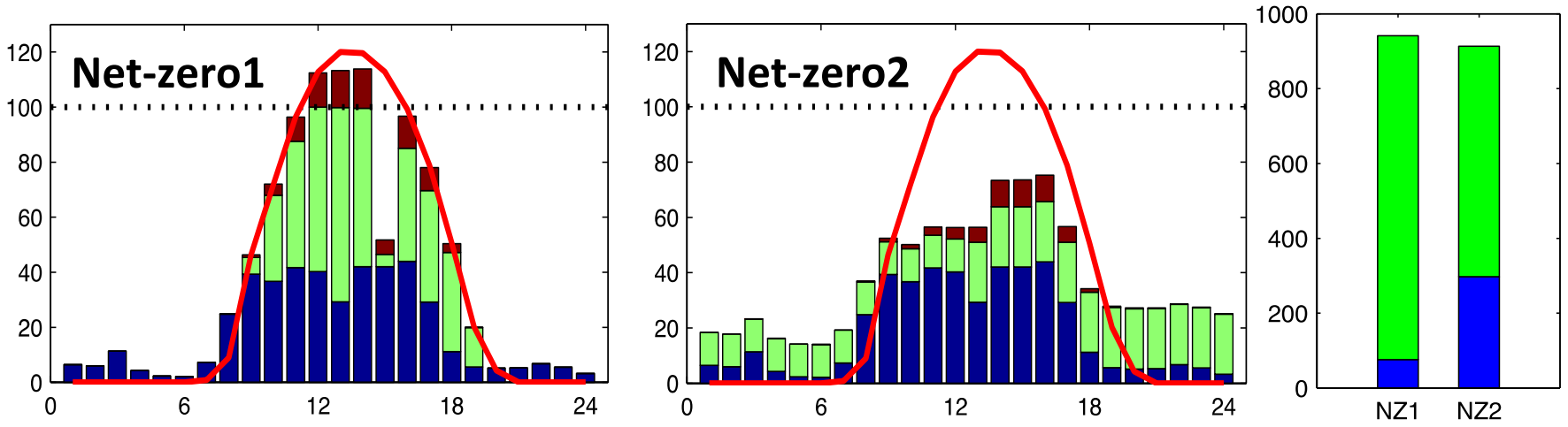






Power cost minimization while finishing all jobs. Optimal beats other at 63% energy savings.

Is Net-zero Energy Consumption Possible?



Energy storage needed to be off-grid: 82kWh vs 330kWh ($99,200)

HP Unveils Architecture for First Net Zero Energy Data Center

HP Unveils Architecture for First Net Zero Energy Data Center

PALO ALTO, CA -- (MARKETWIRE) -- 05/30/12 --

HP (NYSE: HPQ) today unveiled research from HP Labs, the company's central research arm, that illustrates the architecture for a data center that requires no net energy from traditional power grids.

The research shows how the architecture, combined with holistic energy-management techniques, enables organizations to cut total power usage by 30 percent, as well as dependence on grid power and costs by more than 80 percent.(1)

With the HP Net-Zero Energy Data Center research, HP aims to provide businesses and societies around the world the potential to operate data centers using local renewable resources, removing dependencies such as location, energy supply and costs. This opens up the possibility of introducing IT services to organizations of all sizes.

"Information technology has the power to be an equalizer across societies globally, but the cost of IT services, and by extension the cost of energy, is prohibitive and inhibits widespread adoption," said Cullen Bash, distinguished technologist, HP, and interim director, Sustainable Ecosystems Research Group, HP Labs. "The HP Net-Zero Energy Data Center not only aims to minimize the environmental impact of computing, but also has a goal of reducing energy costs associated with data-center operations to extend the reach of IT accessibility globally."

New In This Paper

Idea: temporal workload management with cooling and renewable energy constraints.

Implementation: a comprehensive model of cooling, IT and power supply.

Environment: physical test bed, real traces from industry

Experiment plan: multiple baselines comparison, cooling analysis, social impact analysis.

Power Management of Online Data-Intensive Services

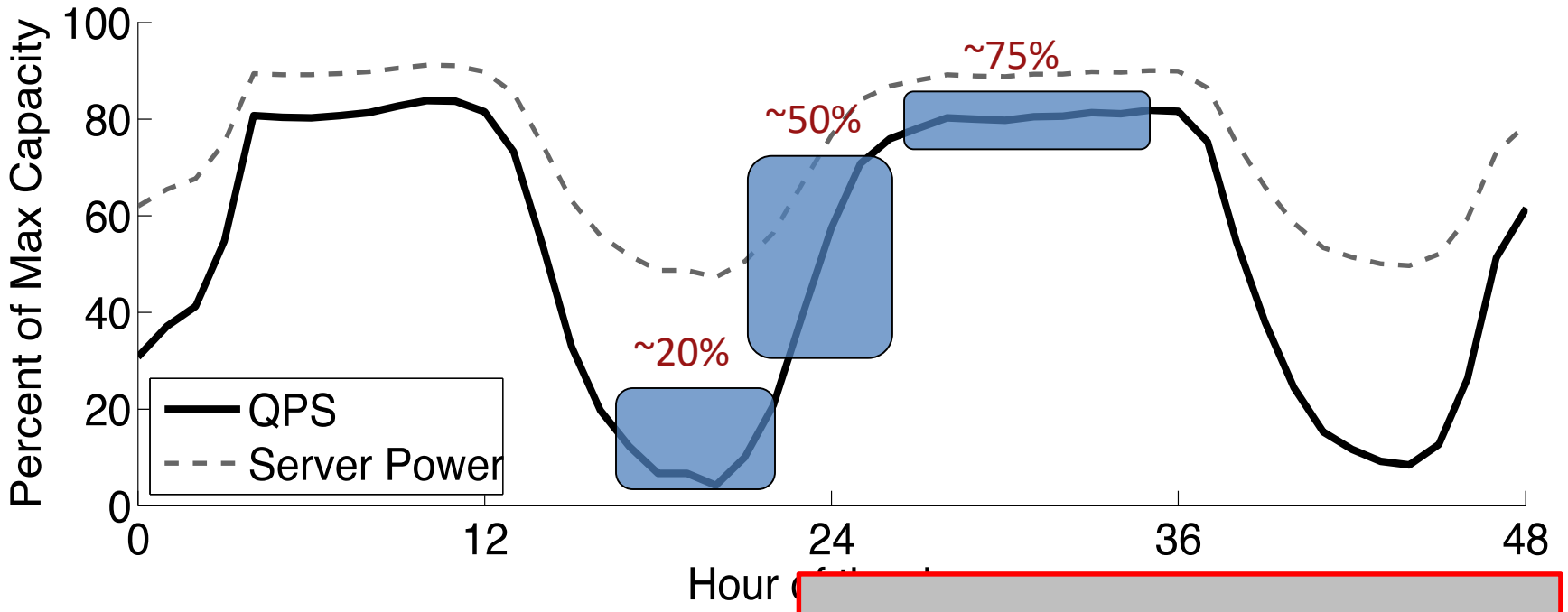
David Meisner[†], Christopher M. Sadler[‡], Luiz A. Barroso[‡], Wolf-Dietrich Weber[‡], Thomas F. Wenisch[†]

[†]The University of Michigan [‡]Google, Inc.

Courtesy to Dr. Meisner

ISCA 2011

The Need For Energy-proportionality



How to achieve energy-prop
query per second(QPS) lev

Comments: unlike the first paper, this one is more fine-grained at server level.

Online Data-Intensive Services

Challenging workload class for power management

- Process TBs of data with O(ms) request latency
- Tail latencies critical (e.g., 95th, 99th-percentile latency)
- Provisioned by **data set size and latency** *not* **throughput**
- Examples: web search, machine translation, online-ads

Case Study: Google Web Search

- First study on power management for OLDI services
- Goal: Identify which power modes are useful

Two-part Study of Web Search

Part 1: Cluster-scale study

- Web Search on O(1,000) node cluster
- Measured per-component activity at leaf level
- Use to derive upper-bounds on power savings
- Determine power modes of interest/non-interest

Part 2: Single-node study

- Allows intrusive, latency-altering experiments
- Evaluate power-latency tradeoffs of power modes
- Can we achieve energy-proportionality with SLA “slack”

Need coordinated, full-system active low-power modes

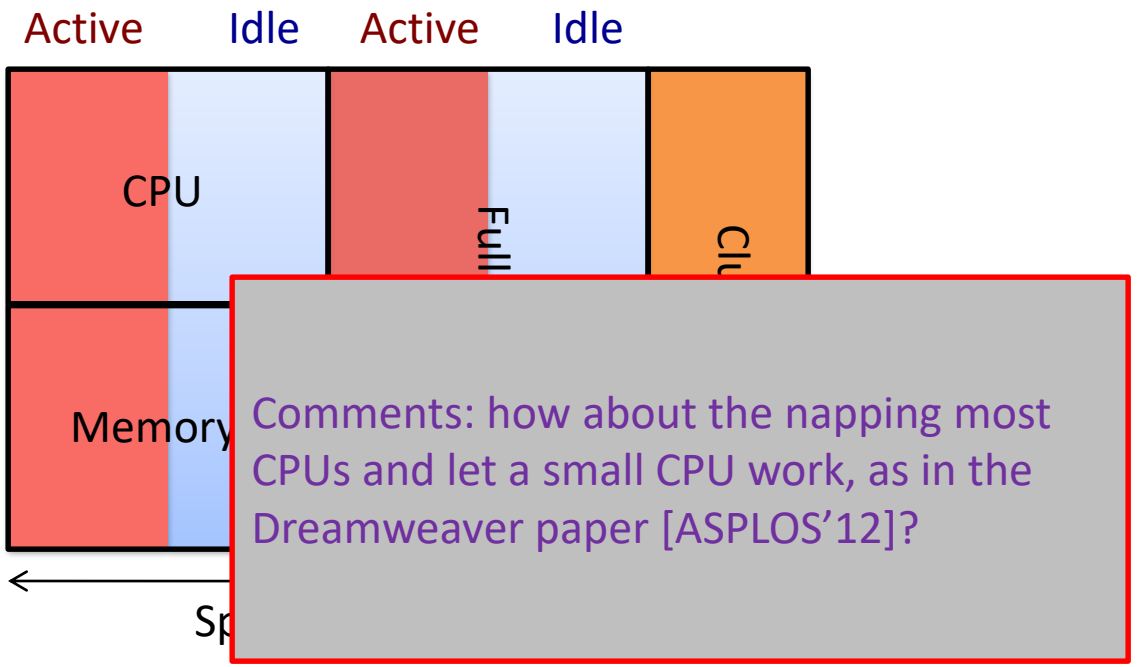
Low-power Mode Taxonomy

Active Modes – Reduce speed of component

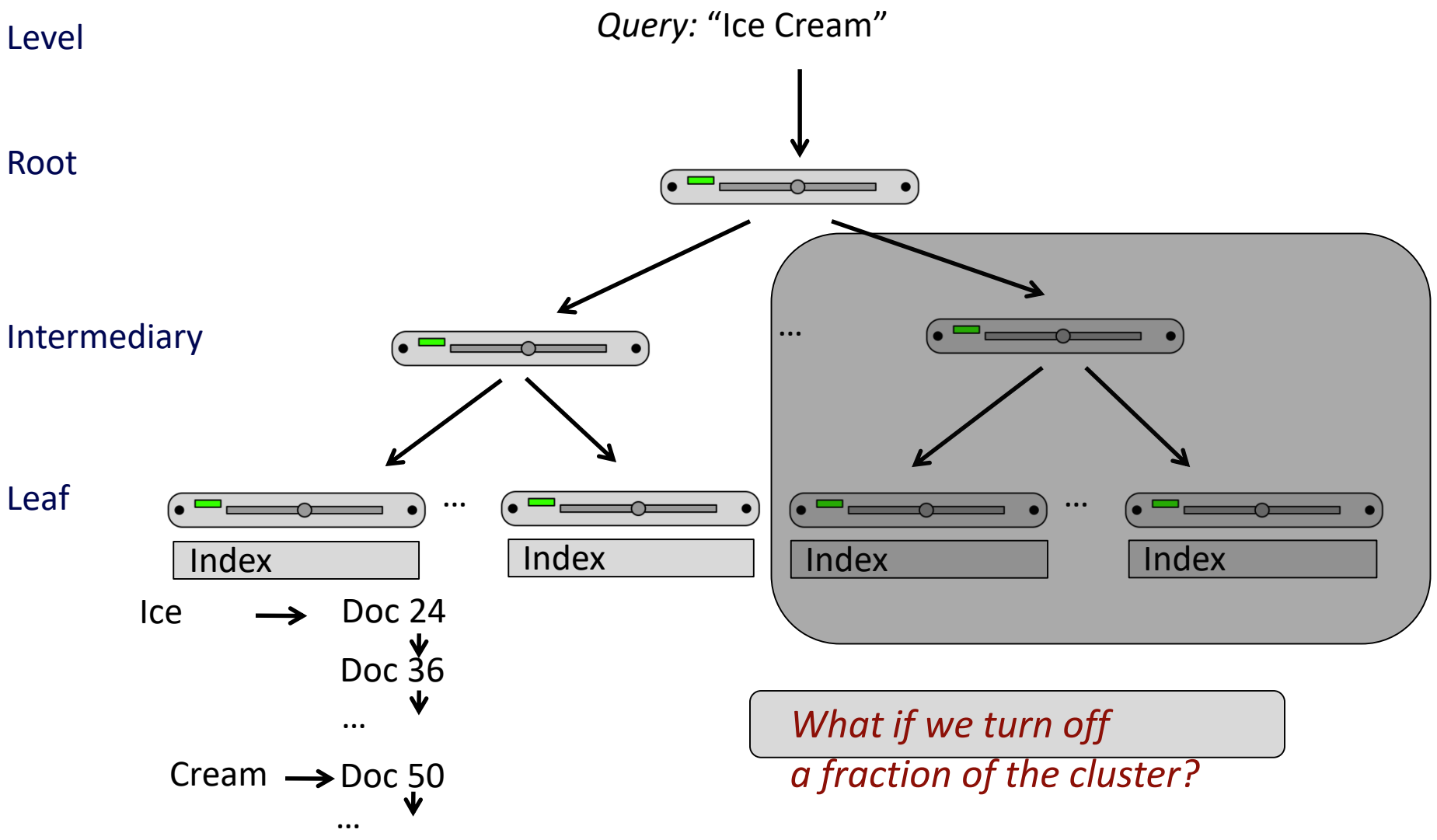
- Continues to do work (e.g., DVFS)

Idle Modes – Put component into sleep mode

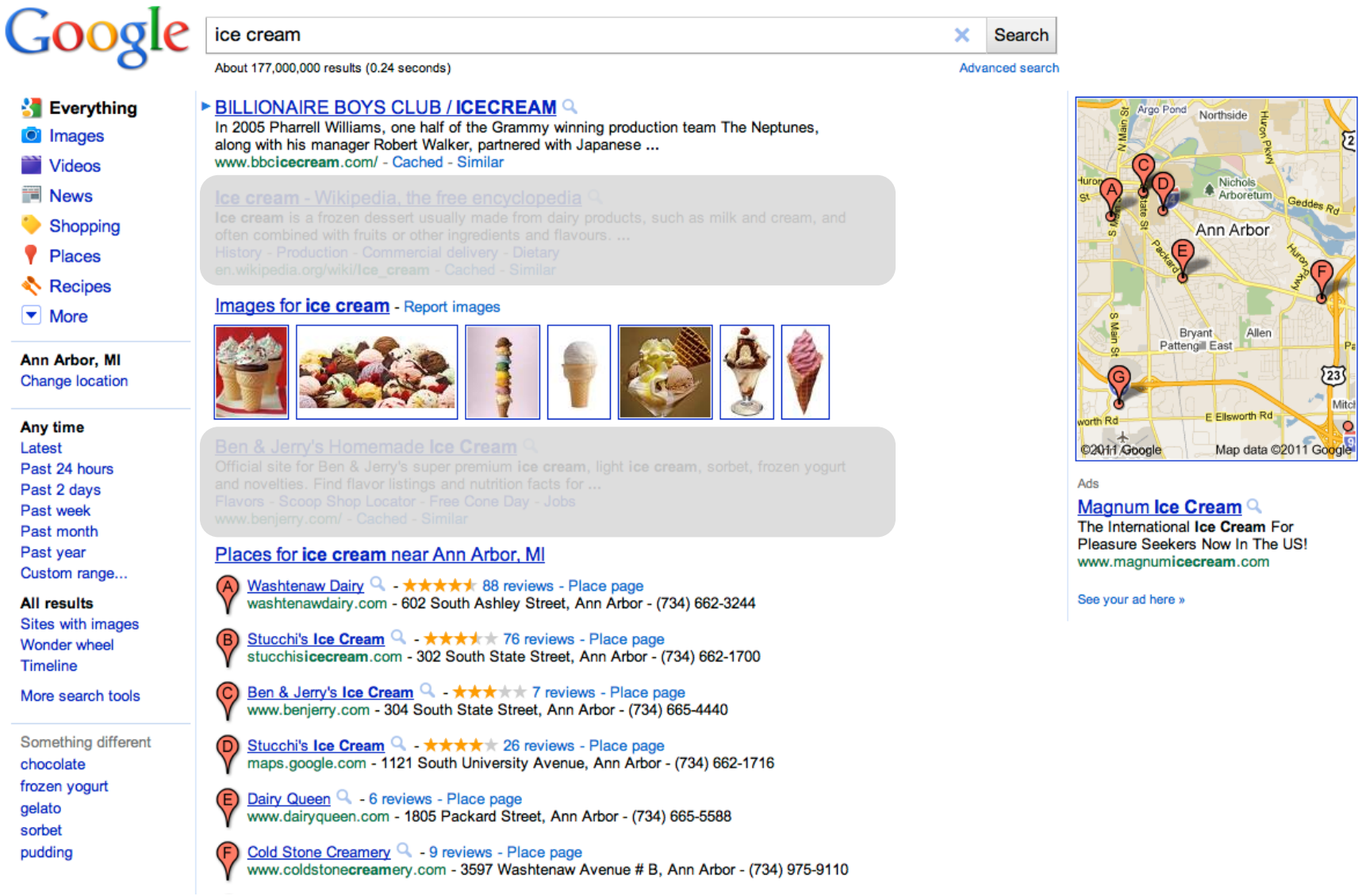
- No useful processing (e.g., ACPI C-states)



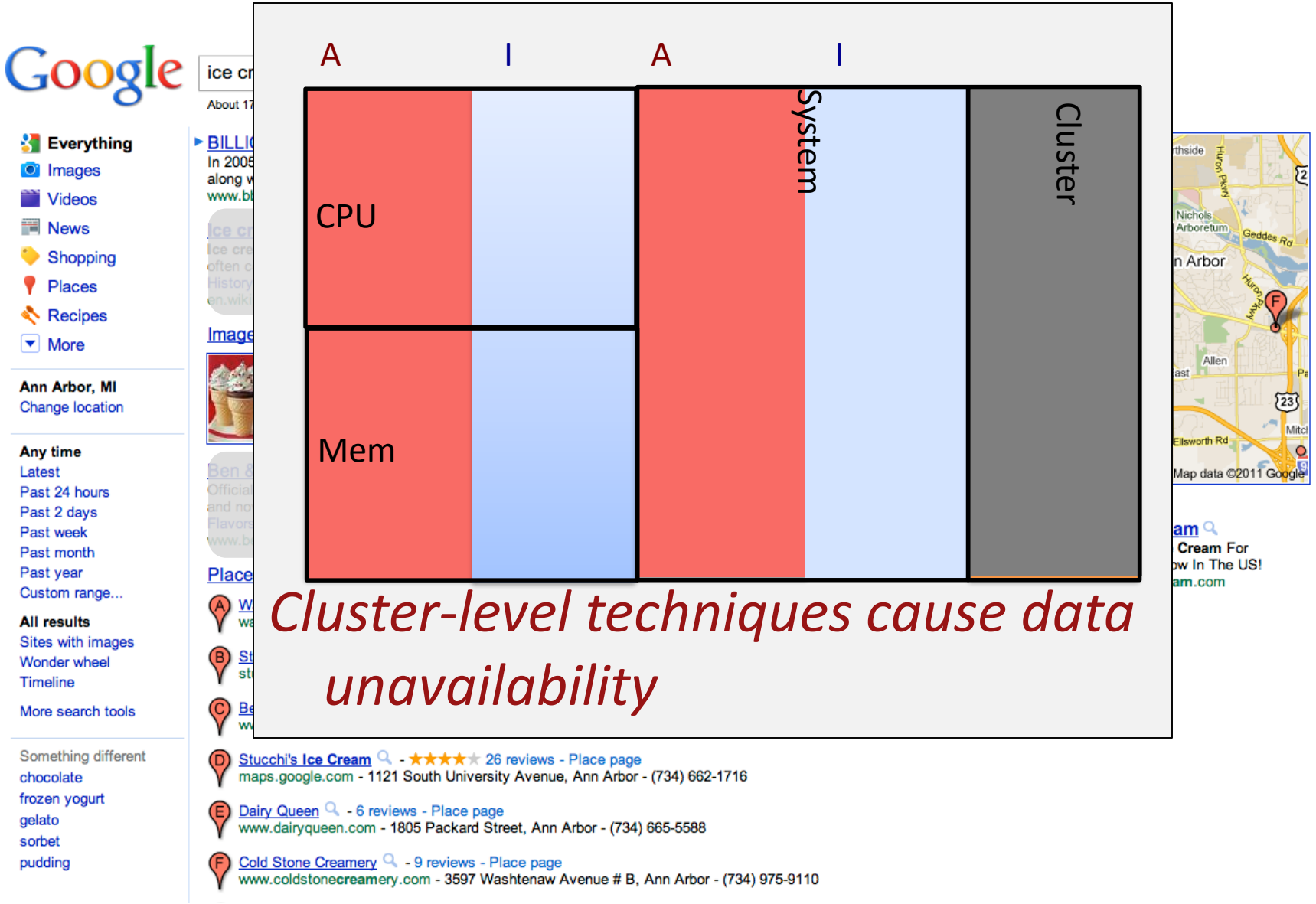
Web Search Operation



Web Search Operation



Web Search Operation



Cluster-scale Throughput Study

Web Search experimental setup

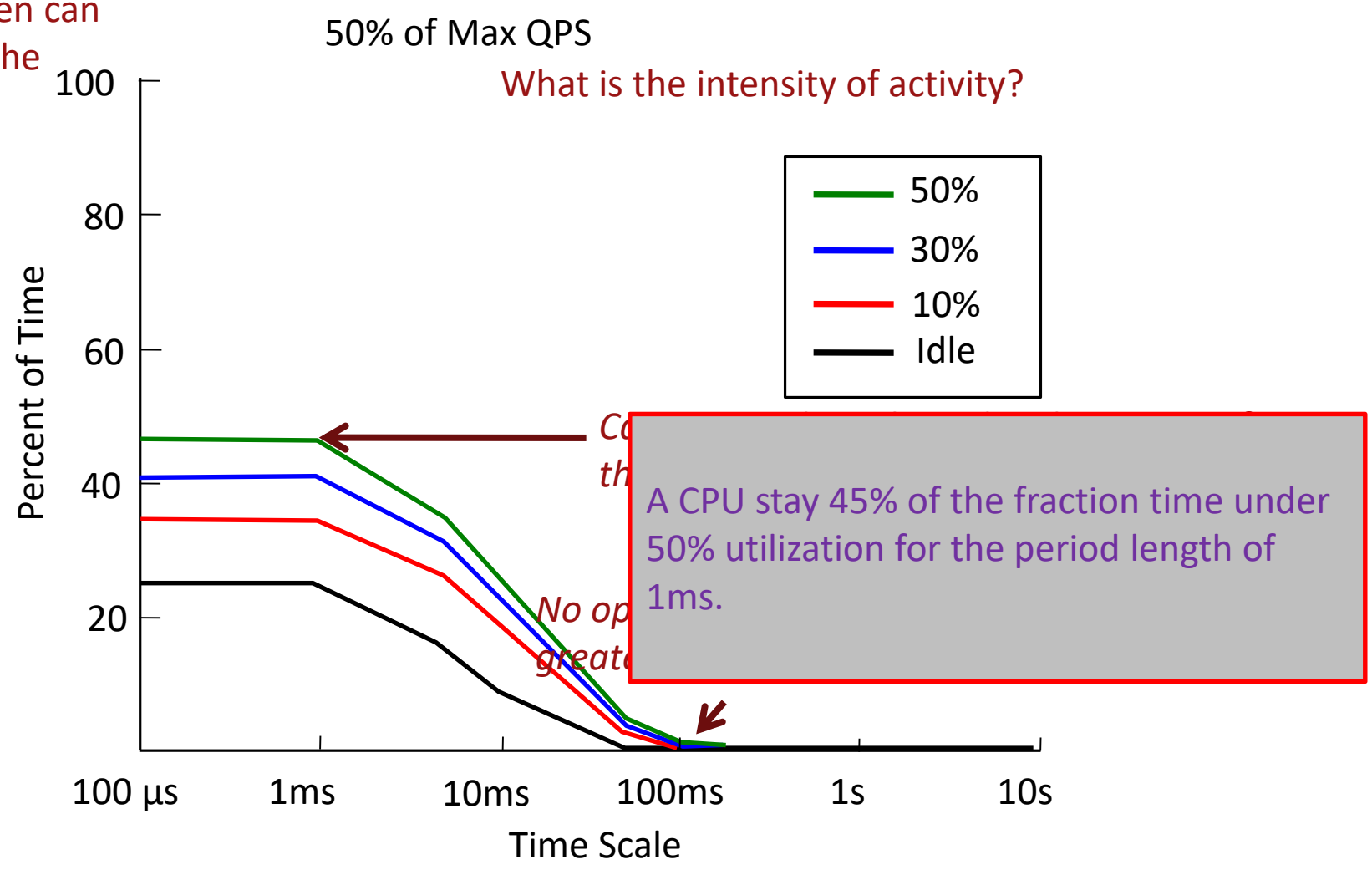
- O(1,000) server system
- Operated at 20%, 50%, 75% of peak QPS
- Traces of CPU util., memory bandwidth, disk util.
- Current disk power modes too slow (see paper)

Characterization

- Goal: find periods to use low-power modes
- Understand *intensity* and *time-scale* of utilization
- Analyze using *activity graphs*

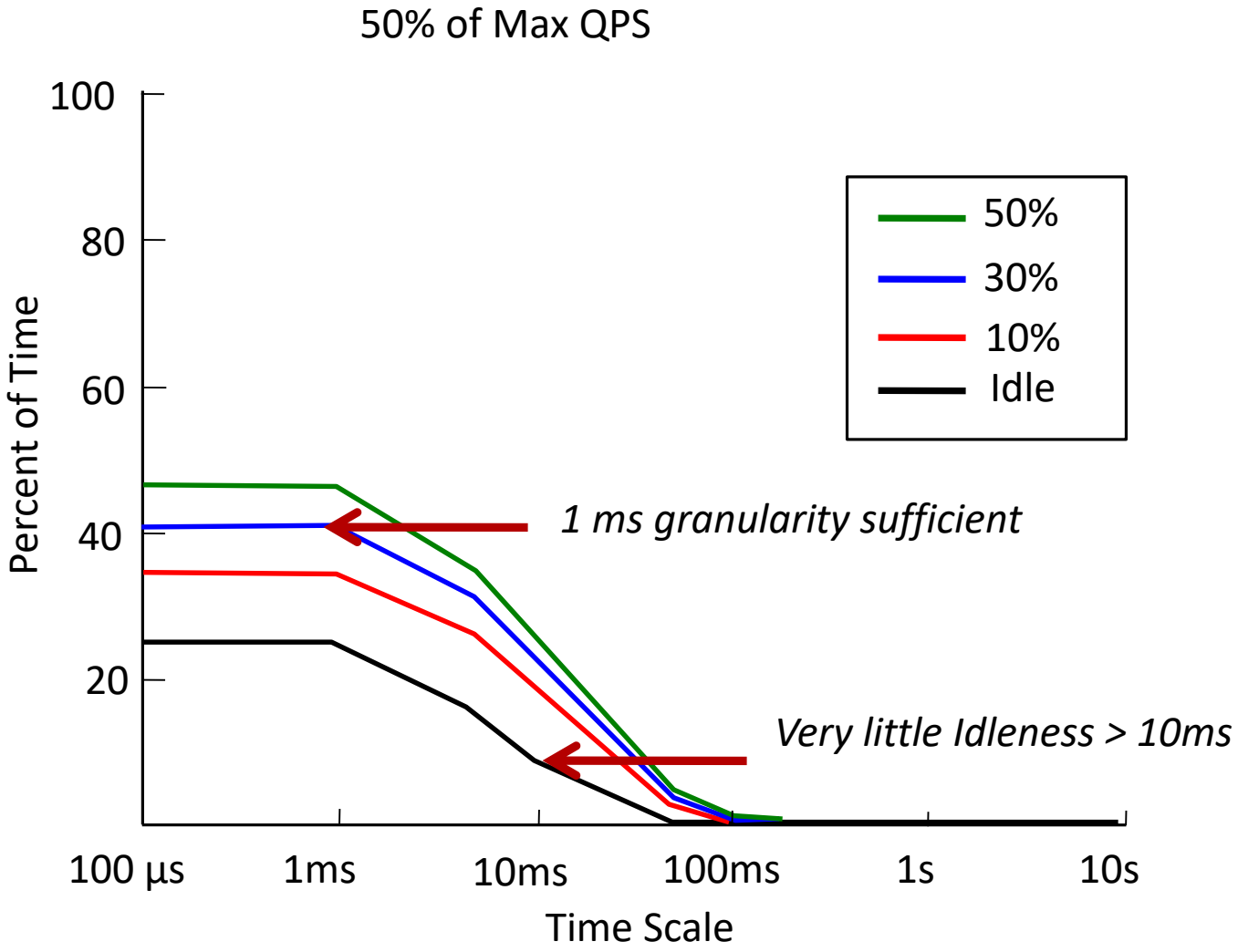
CPU Utilization

How often can we use the mode?

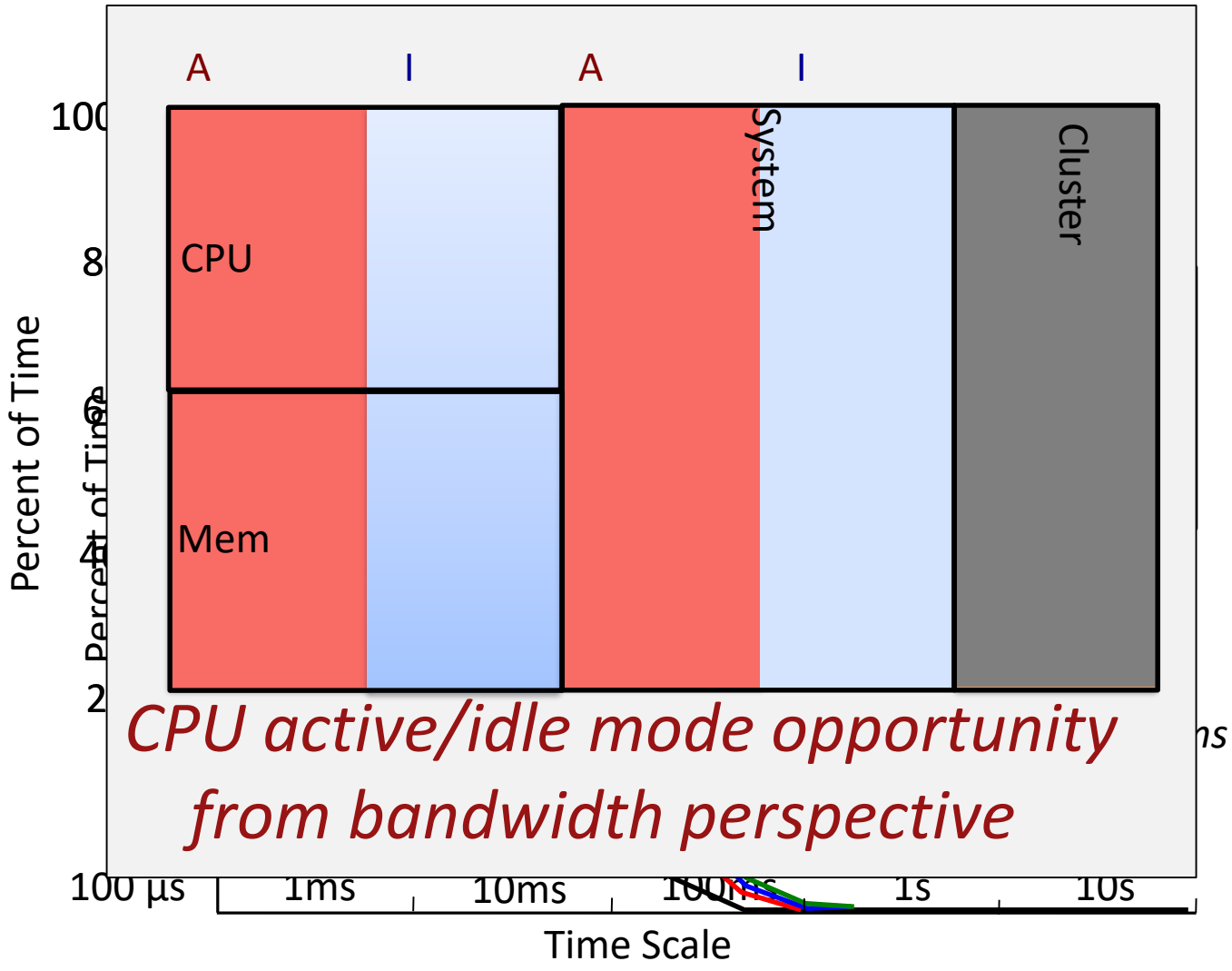


A CPU stay 45% of the fraction time under 50% utilization for the period length of 1ms.

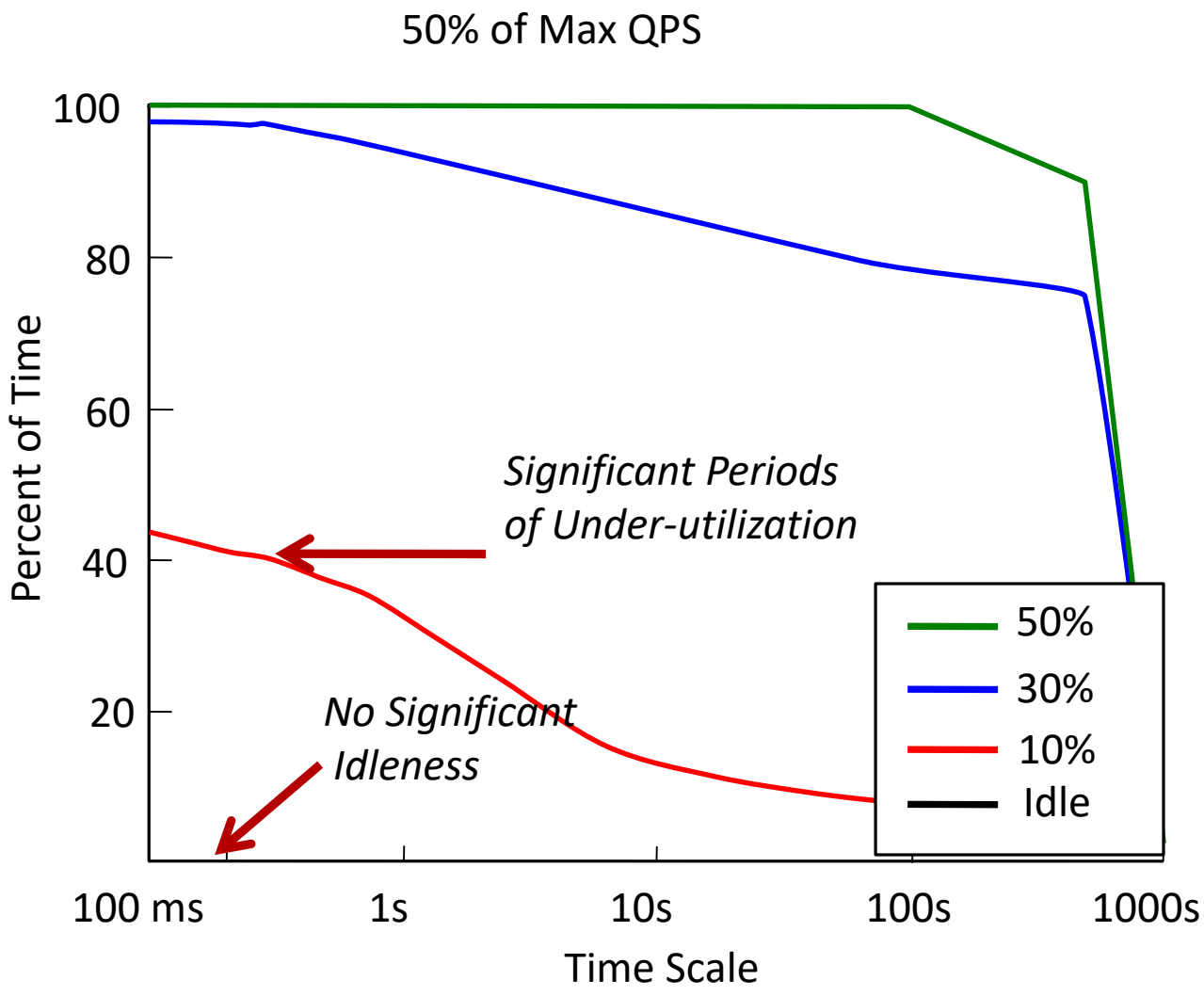
CPU Utilization



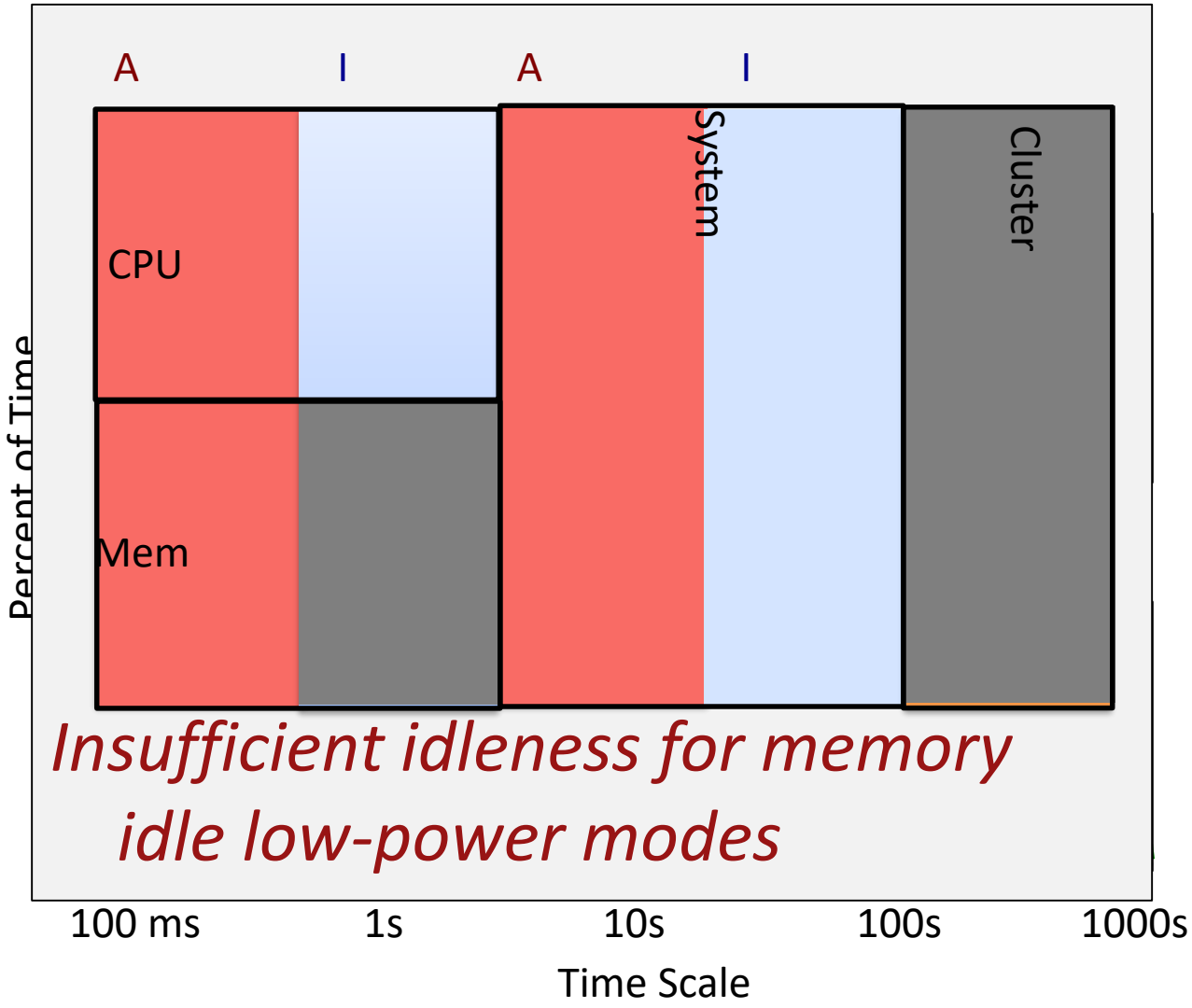
CPU Utilization



Memory Bandwidth Utilization



Memory Bandwidth Utilization



Leaf Node Load Test Methodology

Goal: Understand latency effect of power modes

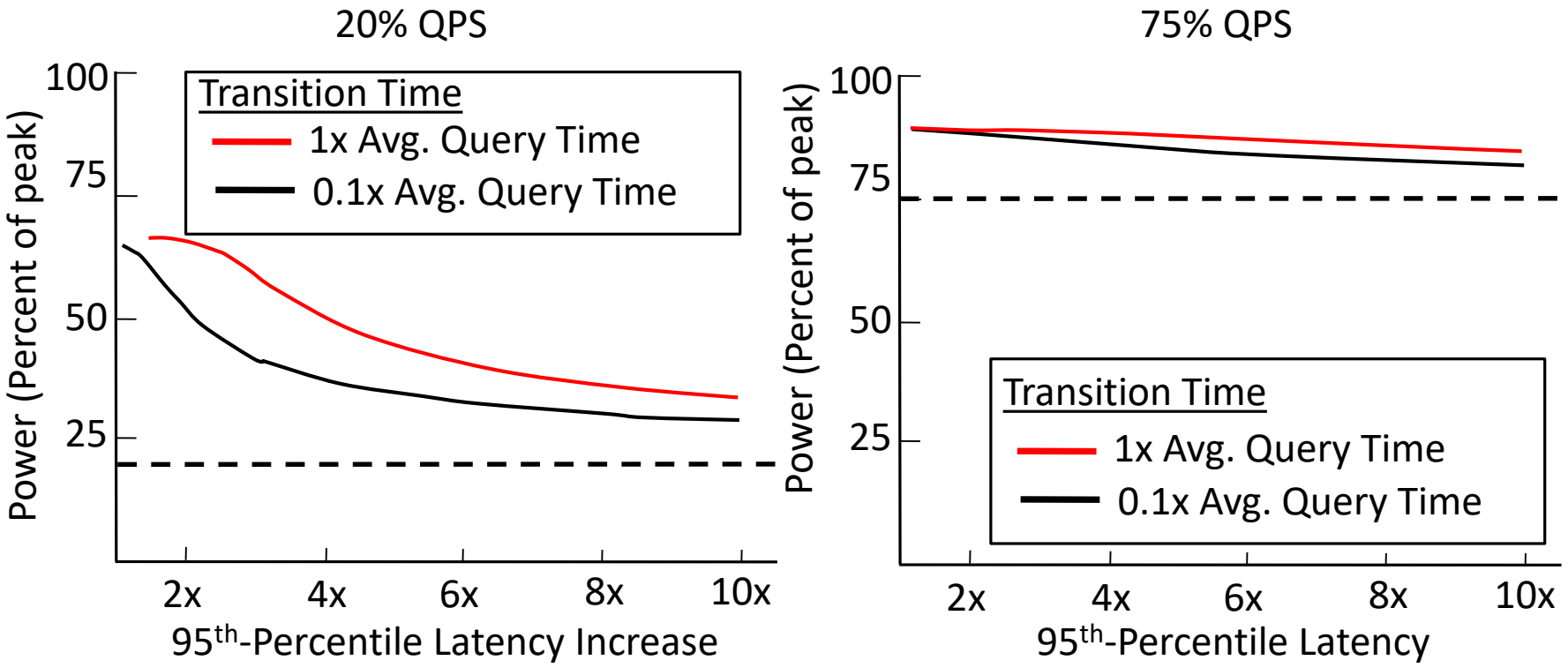
Leaf node testbed

- Faithfully replicate production queries at leaf node
- Arrival time distribution critical for accurate modeling

Validated power-performance model

- Characterize power-latency tradeoff on real HW
- Evaluate power modes using Stochastic Queuing Simulation (SQS) [EXERT '10]

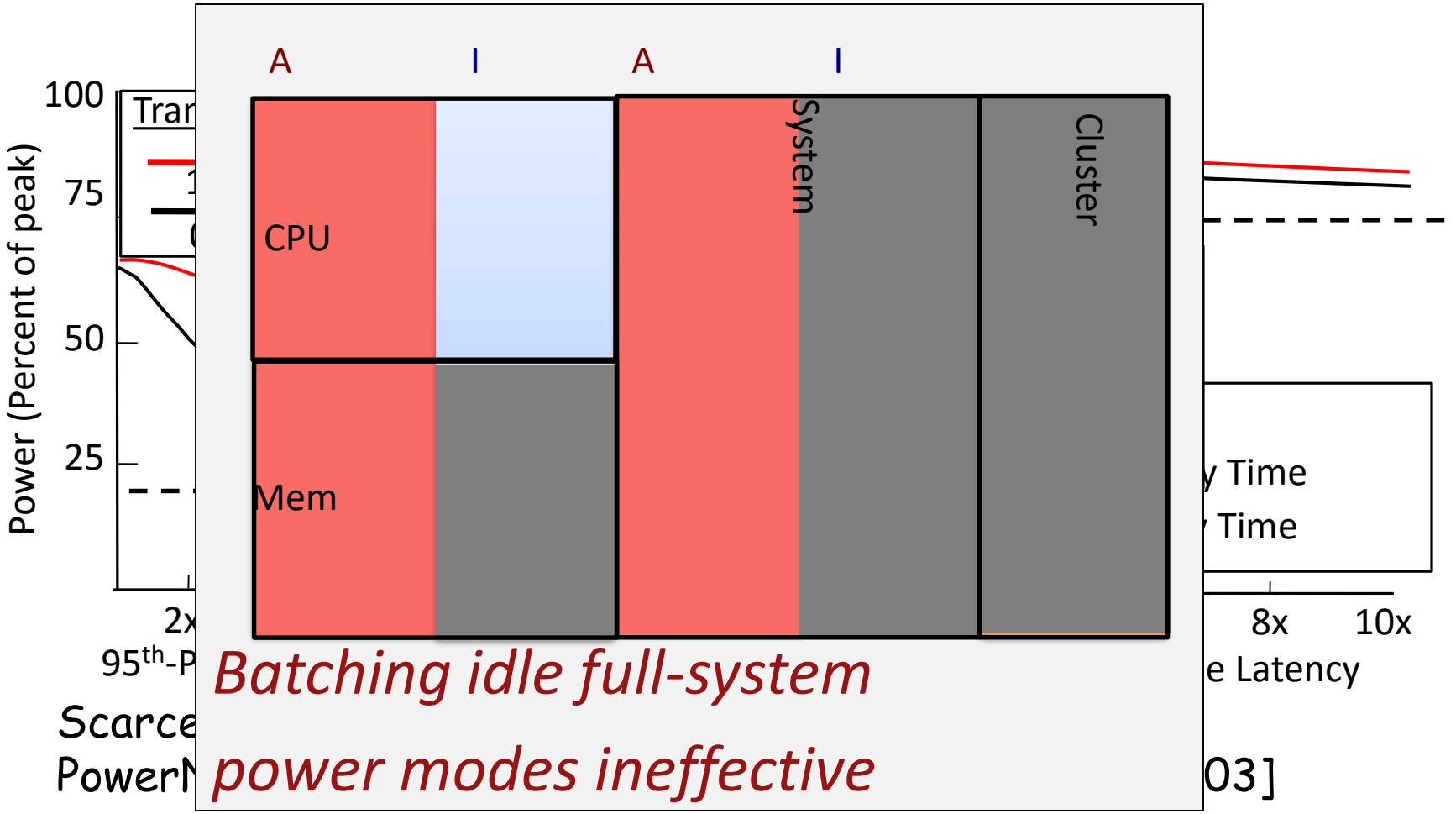
Full-system Coordinated Idle Scaling



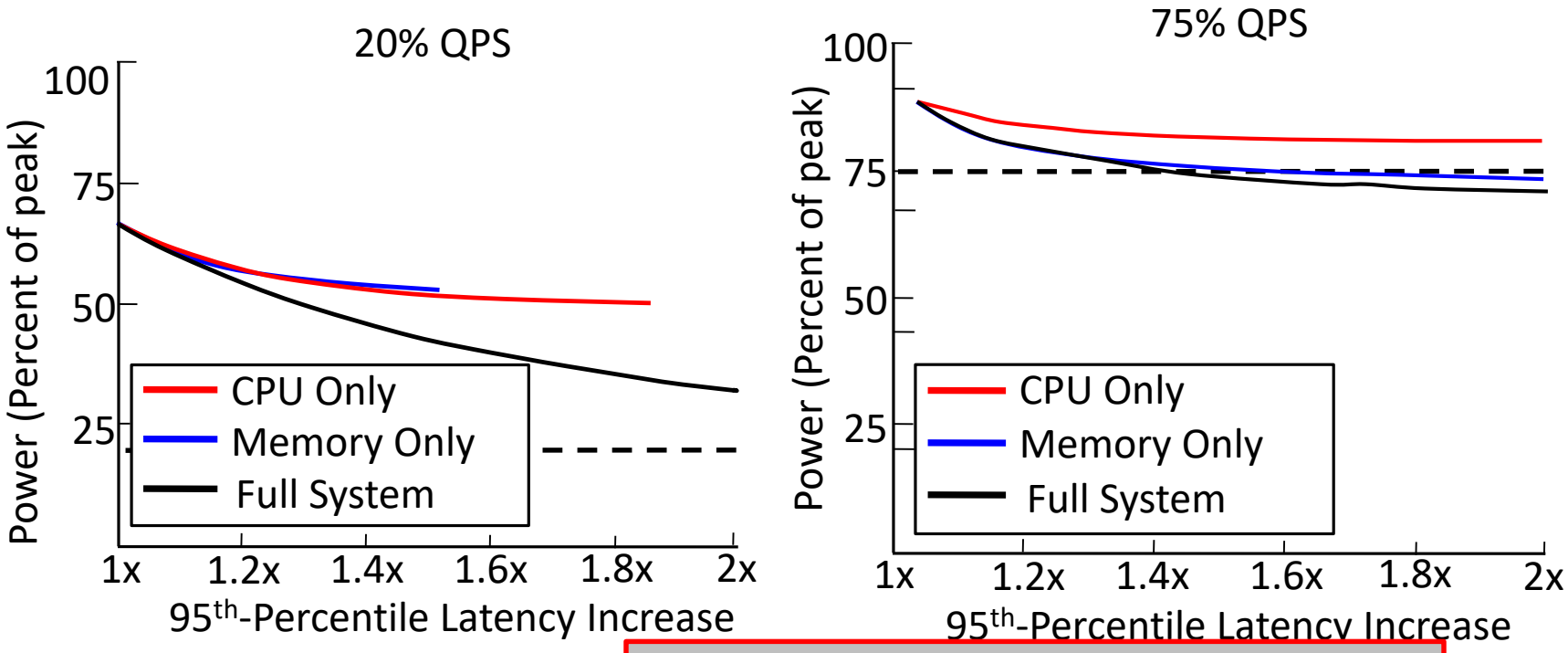
Scarce full-system idleness for multicore

PowerNap [ASPLOS '09] with batching [Elnozahy et al '03]

Full-system Coordinated Idle Scaling



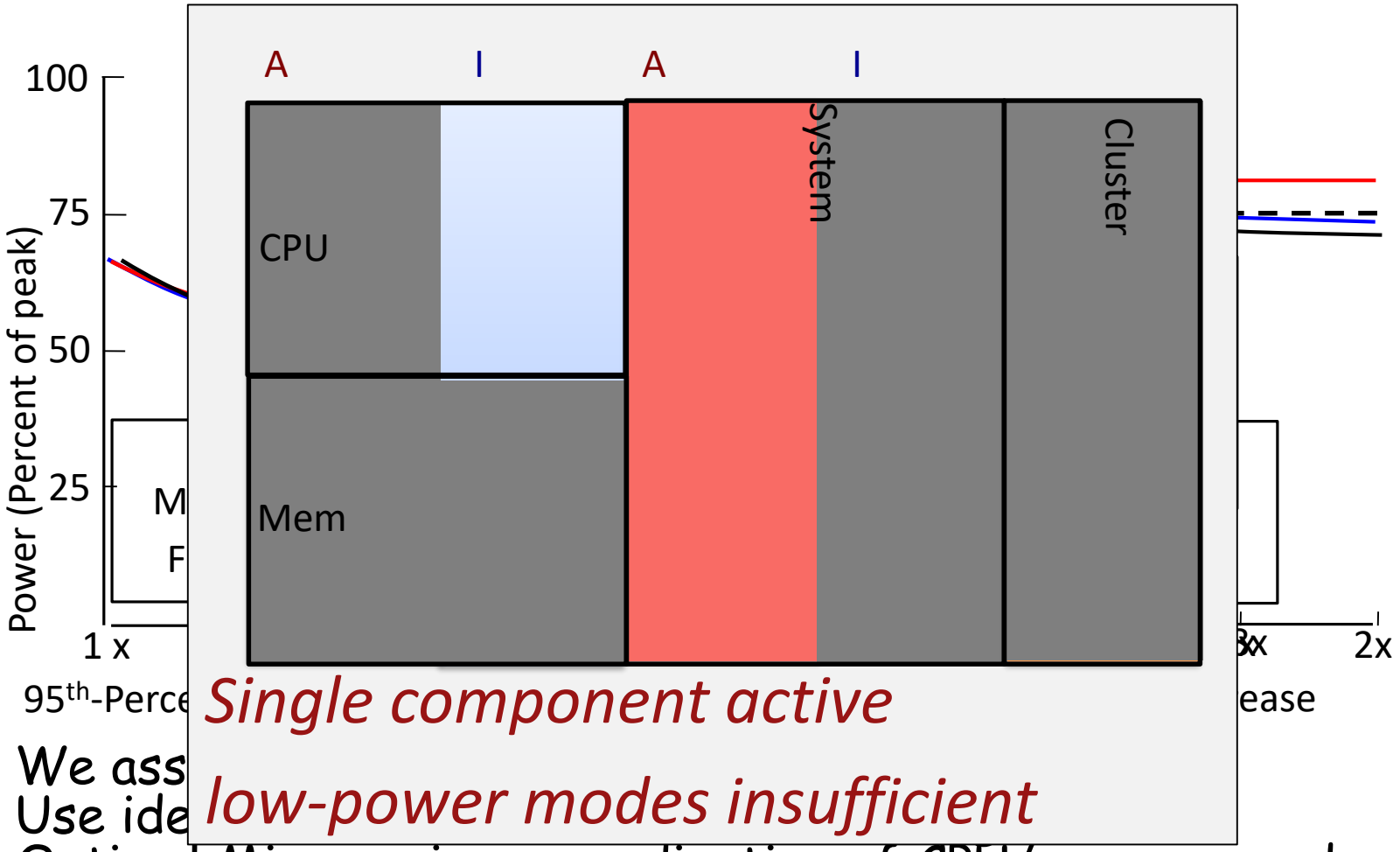
Full-system Coordinated Active Scaling



We assume CPU scaling
Use idealized version of
Optimal Mix requires co

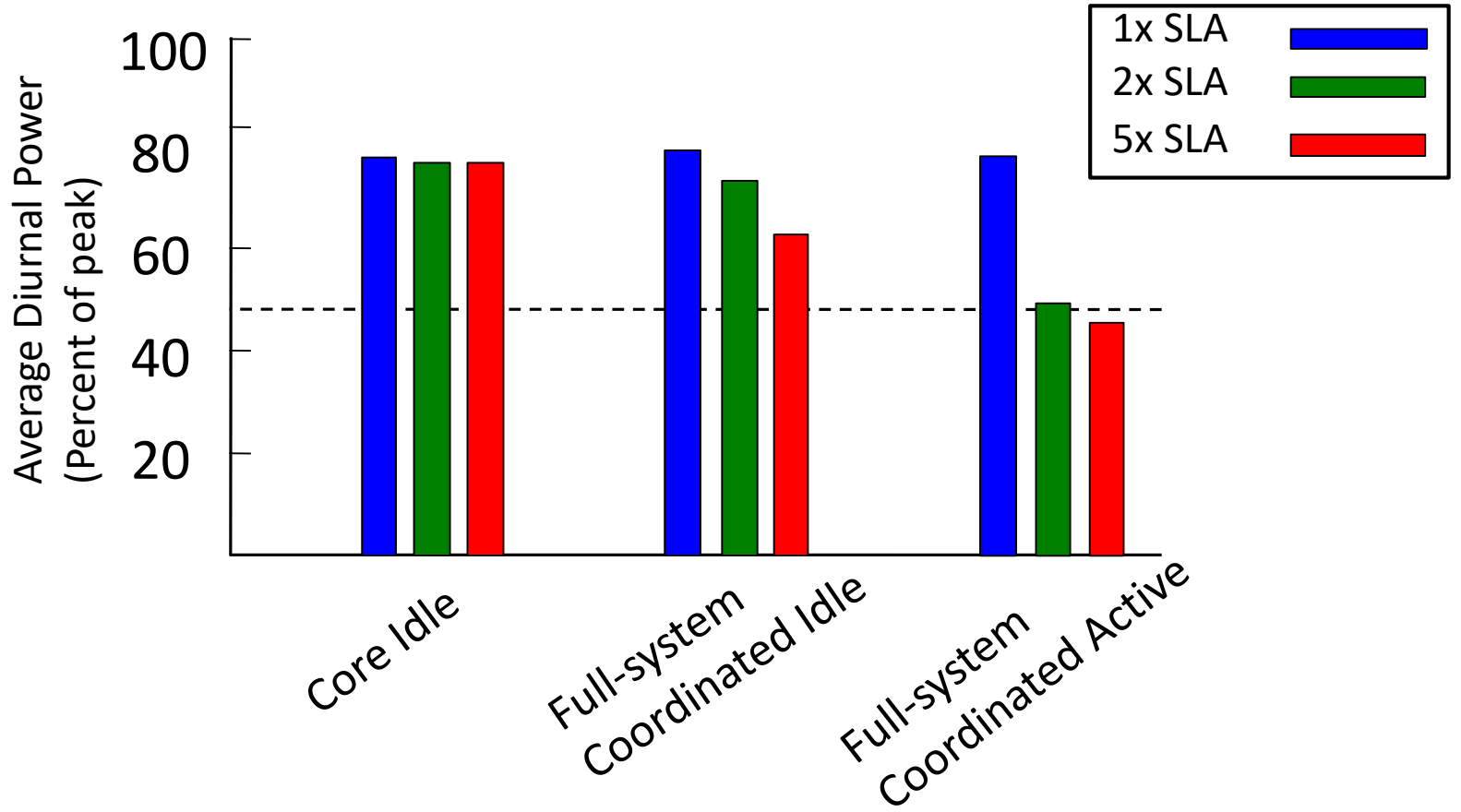
Critique: unrealistic assumption of the memory power mode and awake time.

Full-system Coordinated Active Scaling



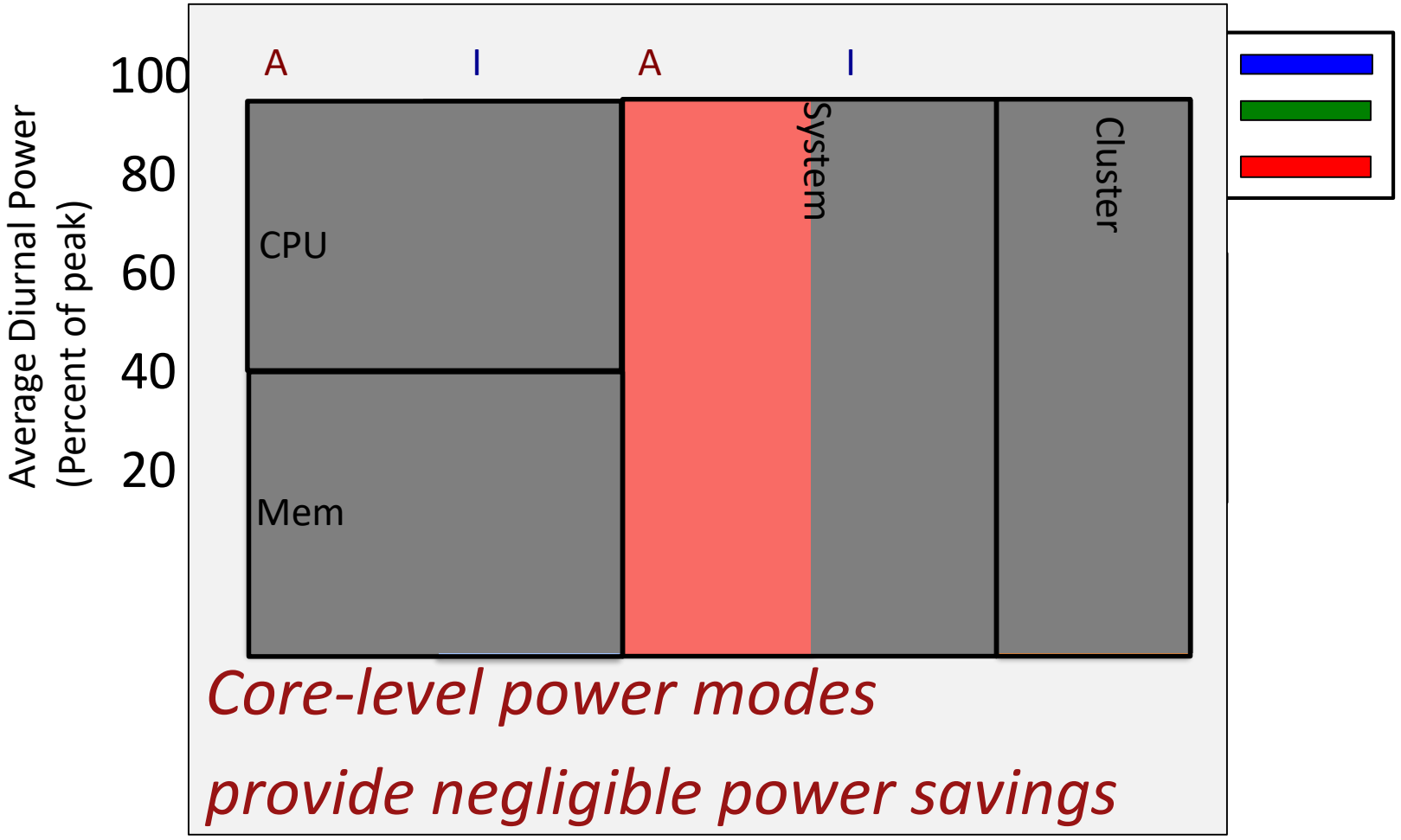
We ass
Use ide
Optimal Mix requires coordination of CPU/memory modes

Comparing Power Modes



Allow SLA "slack" – deviation from 95th-percentile latency

Comparing Power Modes



- Allow SLA "slack" – deviation from 95th-percentile latency

Summary

OLDI workloads challenging for power management

Cluster-scale study

- Current CPU power modes sufficient
- Significant opportunity for active modes for memory
- Need faster idle and active modes for disk

…Useful, but insufficient for proportionality

Single-node study

- Individual idle/active power modes do not achieve proportionality
- PowerNap + batching provides poor latency-power tradeoffs

Need coordinated, full-system active low- power modes

to achieve energy proportionality for OLDI workloads

Comparison

		SIGMETRICS2012	ISCA2011
Similarity	Topics	Renewable/Cooling aware workload management	Exploring characteristics for OLDI workload
	Environment	Data Center	Data Center
	Motivation	IT, cooling and PV study	OLDI in single node and cluster study
	Implementation	Physical test + simulation	Physical test
Differences	Goal	Efficient renewable energy usage	Energy proportionality on OLDI
	Scale	Data center scale analysis	Empirical study from single node to cluster
	Methodology	Optimization modeling on predictive scheme	DVFS, low power active mode
	Workload	Web service + batch workload	OLDI

Critiques: SIGMETRICS2012

The paper uses IT demand as the IT heat load, which are two different concepts.

No revenue analysis on the critical interactive workload.

The prediction of green energy availability could be inaccurate when involving the weather model.

The predictive scheme assumes a full knowledge of the workload, which is not realistic.

Critiques: SIGMETRICS2012 (Cont.)

The estimation errors from these two predicted values could result in failure in workload management.

The overhead of migration/scheduling is ignored.

The optimization metric would be more accurate with the one in the provision paper in SIGMETRICS 2012

Critiques: ISCA2011

The cooling energy consumption for processing OLDI is not considered in this paper.

Shut down partial cores in multi-core CPU would be a better solution than put the CPU in active low state.

The hardware state-transition power cost/performance loss is ignored in the analysis.

The coordination in the full-system low power mode requires correlation knowledge from application.


Thank
You

Take Home Message

- Novelty of building new systems
- The procedure of identifying and understanding a metric-based system
- Propose tempo/spatial-difference in realistic design