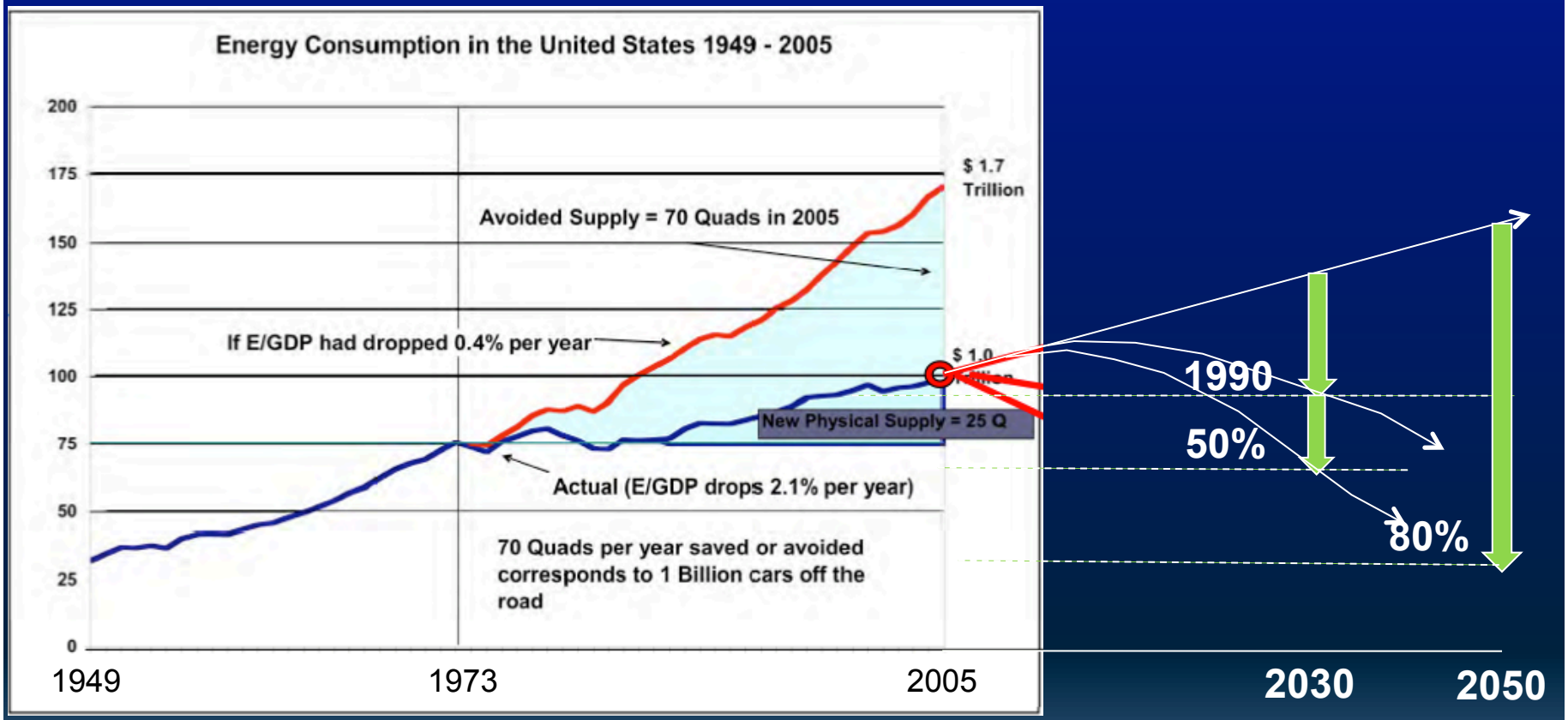
Doing Nothing Well – the essence of Green computing

2nd International Green Computing Conference

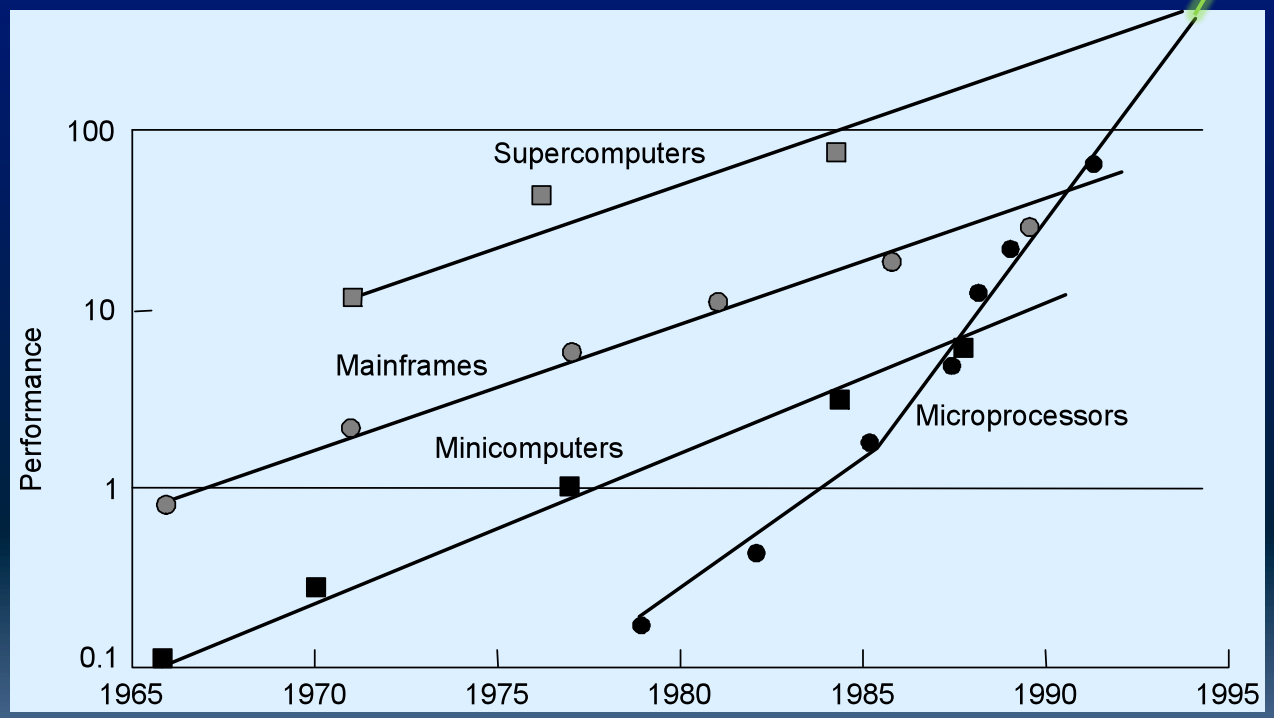
David E. Culler
University of California, Berkeley
July 26, 2011

Energy …

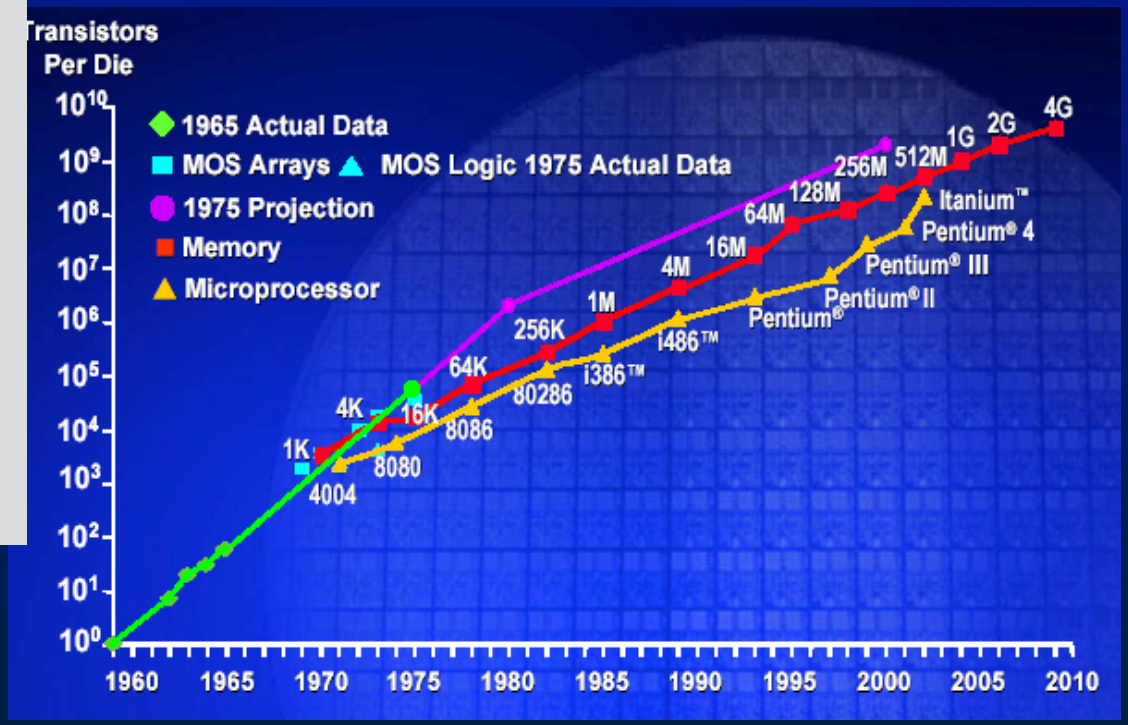


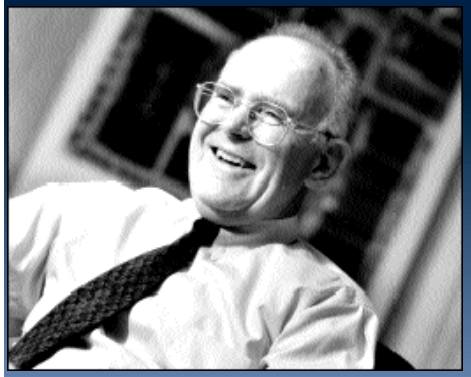
Source:LBNL

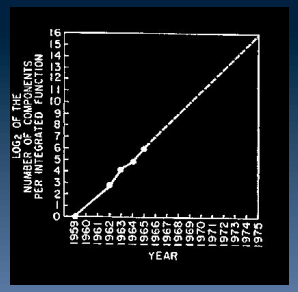
We've been 'doing something' very well



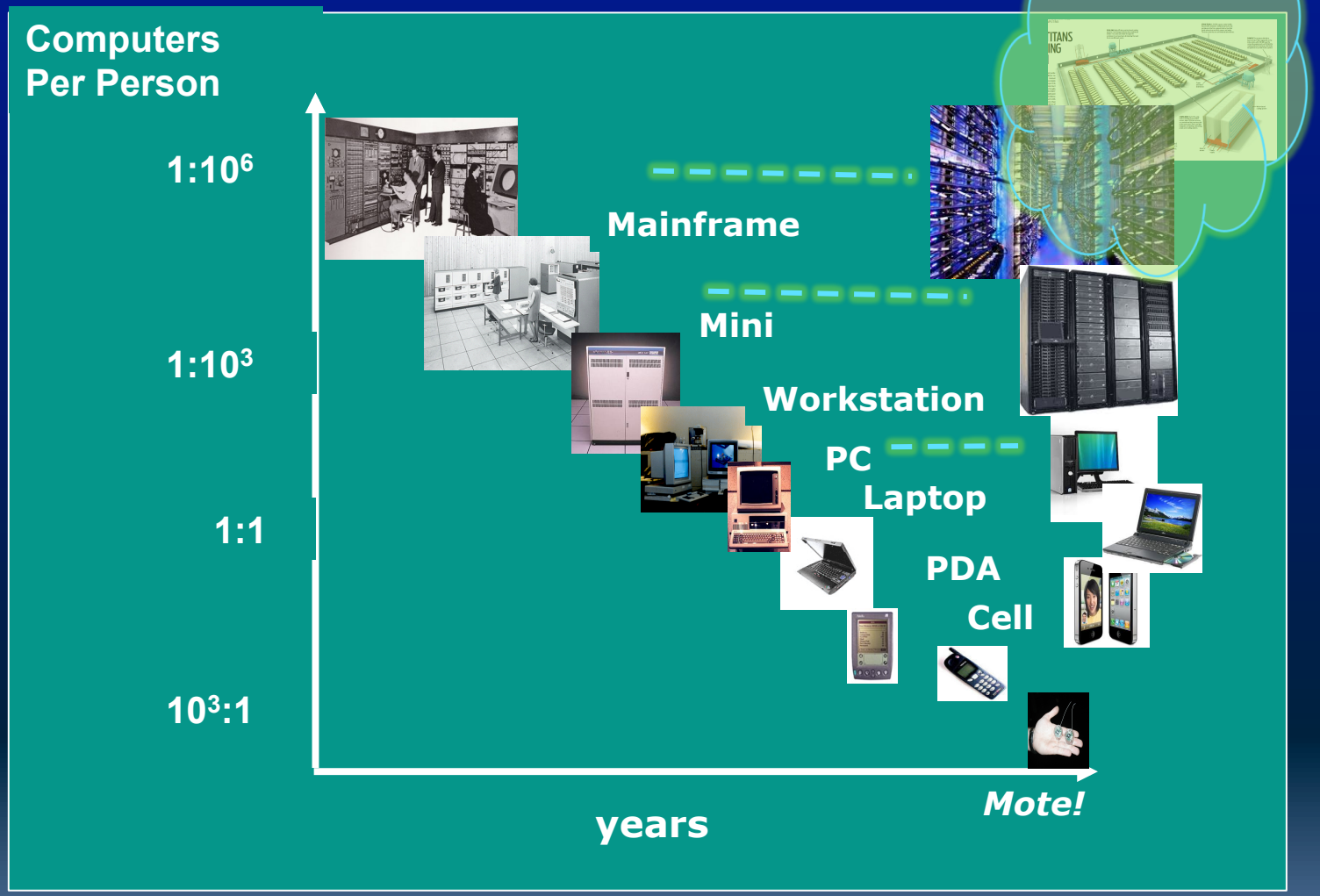
with more and more to do it with …







in more and more ways …



Bell's Law: new computer class per 10 years

for more and more people …

ARPANet

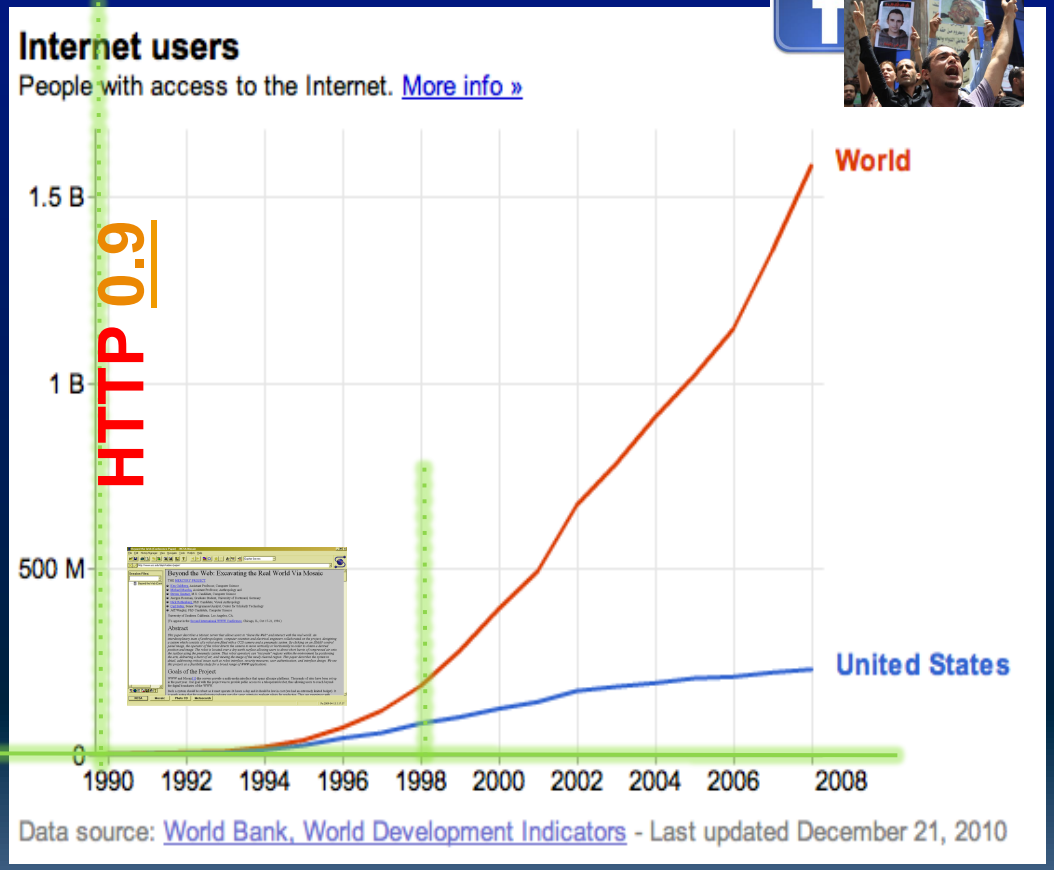
RFC 675 TCP/IP

Internet

WWW

HTTP 0.9

2.0 B 1/26/11



1969 1974

1990

2010

and times they are a changing

Global temperature change (relative to pre-industrial era)

0°C 1°C 2°C 3°C 4°C 5°C

Food — Crop yields fall

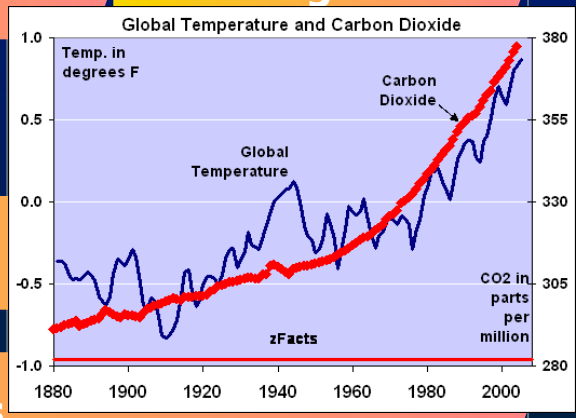
Water — Glaciers melt; Water shortages; Rising seas

Ecosystems — Reefs damaged; Species extinction

Weather — Storms, droughts, fires, heat waves

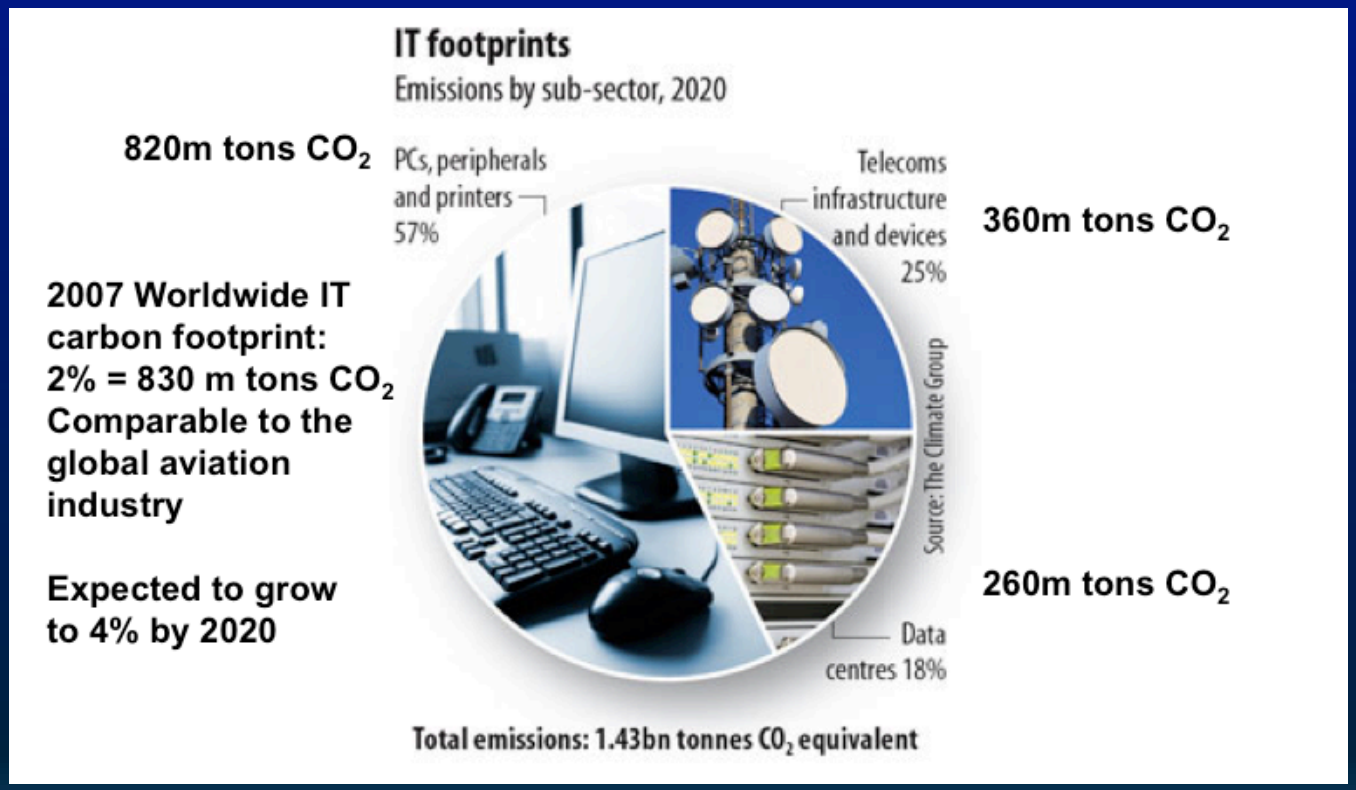
Feedback — Abrupt climate change

Today



Information Technology

- Moderate Part of the problem (2-4%)



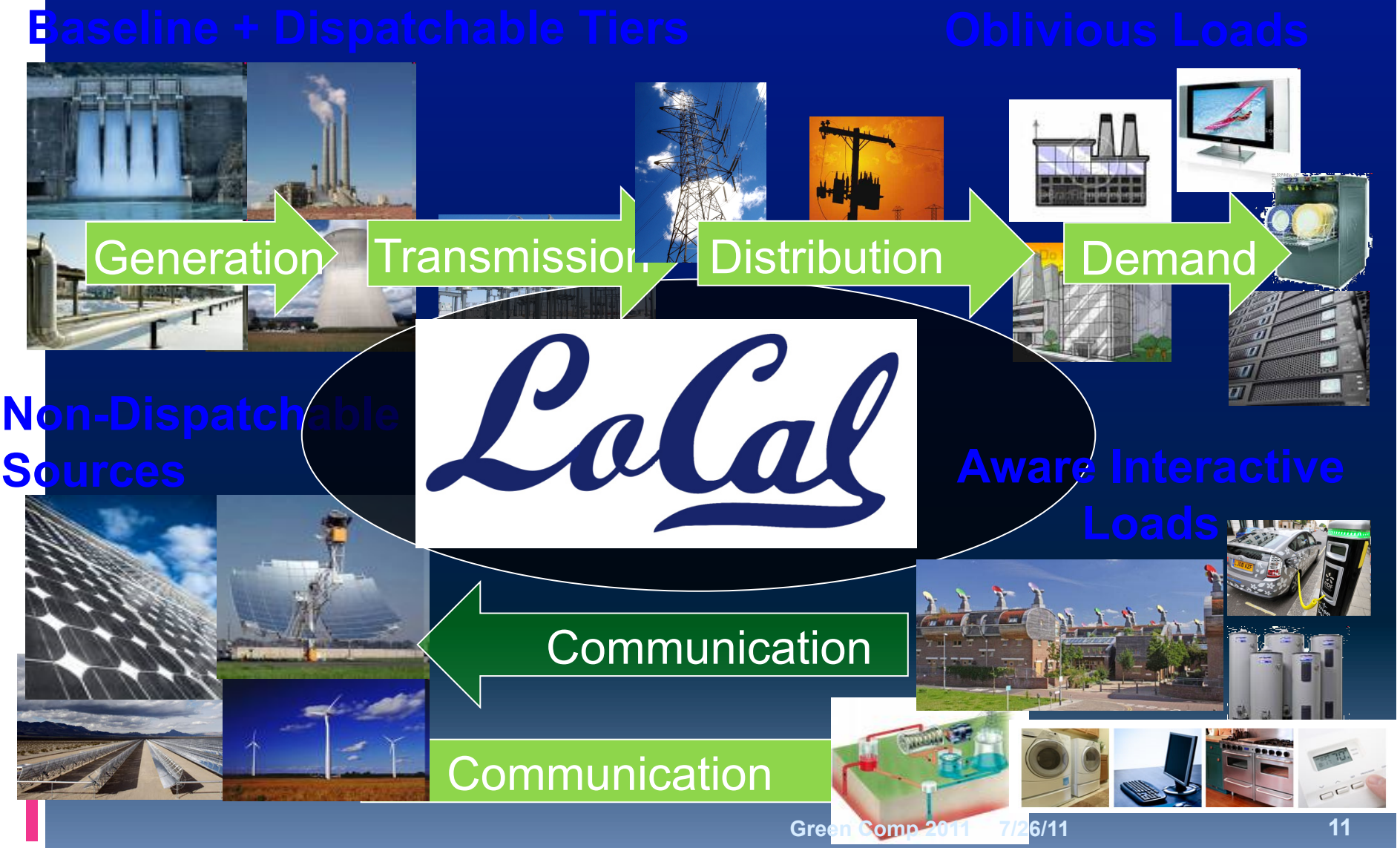
- Big part of the solution

‘Doing Nothing’ well

The Industrial Age Grid

Baseline + Dispatchable Tiers **Oblivious Loads**

Towards an "Aware" Energy Infrastructure



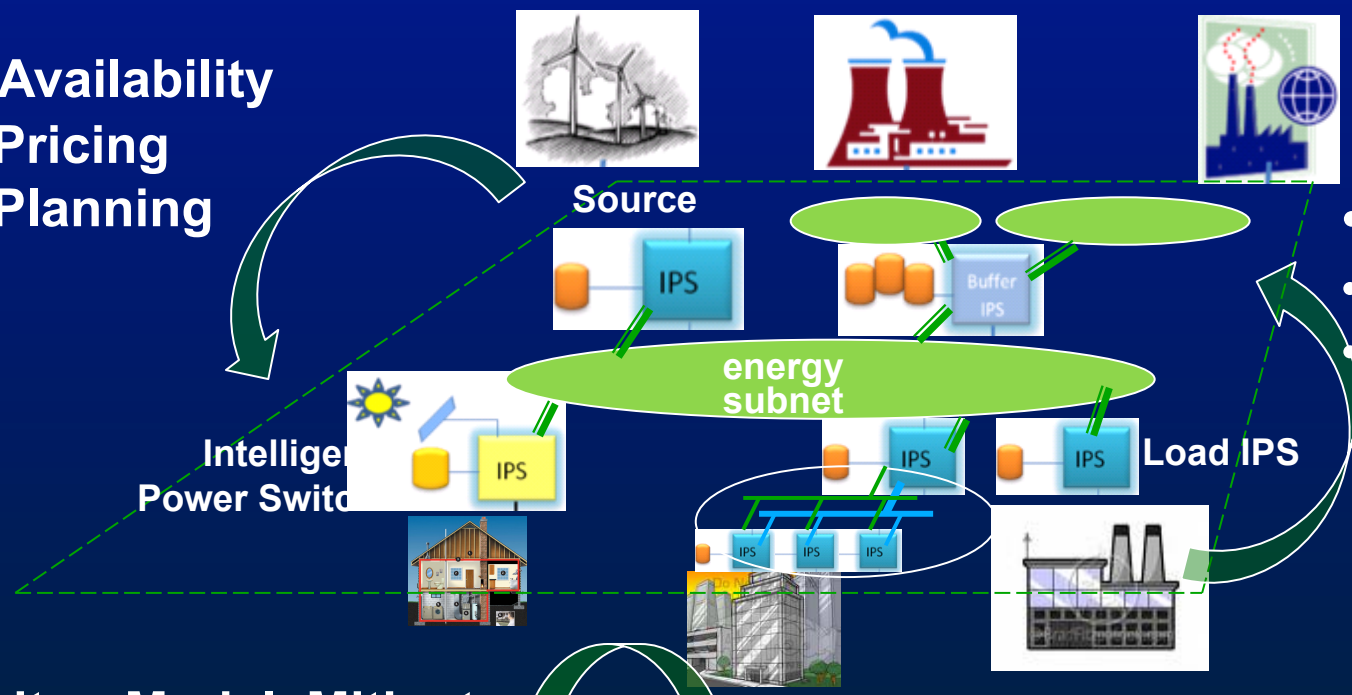
Limits to Renewable Penetration

- Variability, Intermittency of Supply
- Visibility into Availability of Supply
- Ability of Loads to Adapt
- Algorithms and Techniques for Reactive Load Adaptation

- Capability of the Infrastructure to maintain the match

A Cooperative Grid

- Availability
- Pricing
- Planning



- Forecasting
- Tracking
- Market

- Monitor, Model, Mitigate
  - Deep instrumentation
  - Waste elimination
  - Efficient Operation
- Shifting, Scheduling, Adaptation

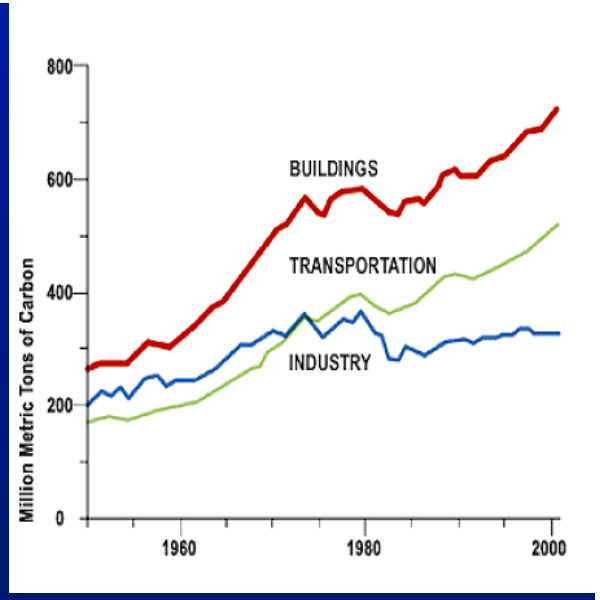
Where to Start?

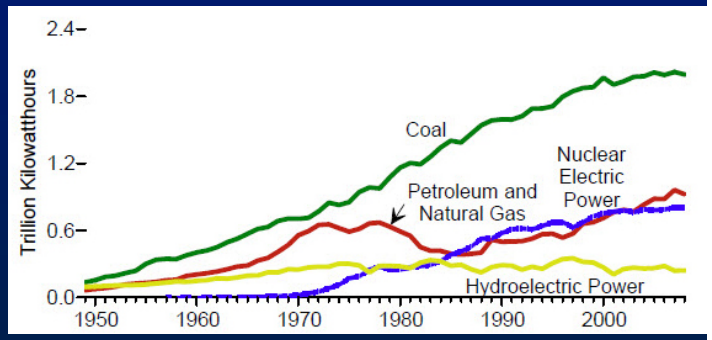
Buildings

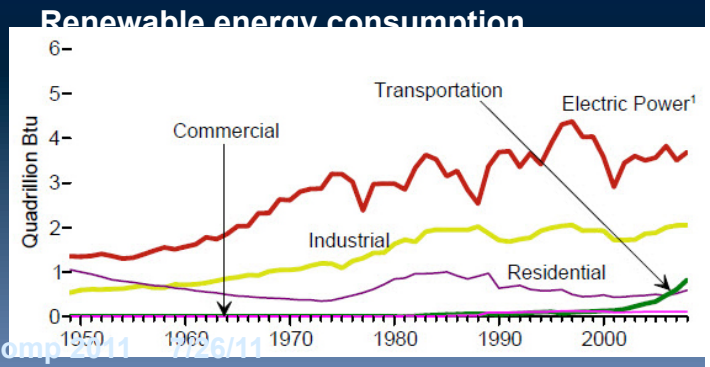
- 72% of electrical consumption (US),
- 40-50% of total consumption,
- 42% of GHG footprint
- US commercial building consumption doubled 1980-2000, 1.5x more by 2025 [NREL]

Where Coal is used

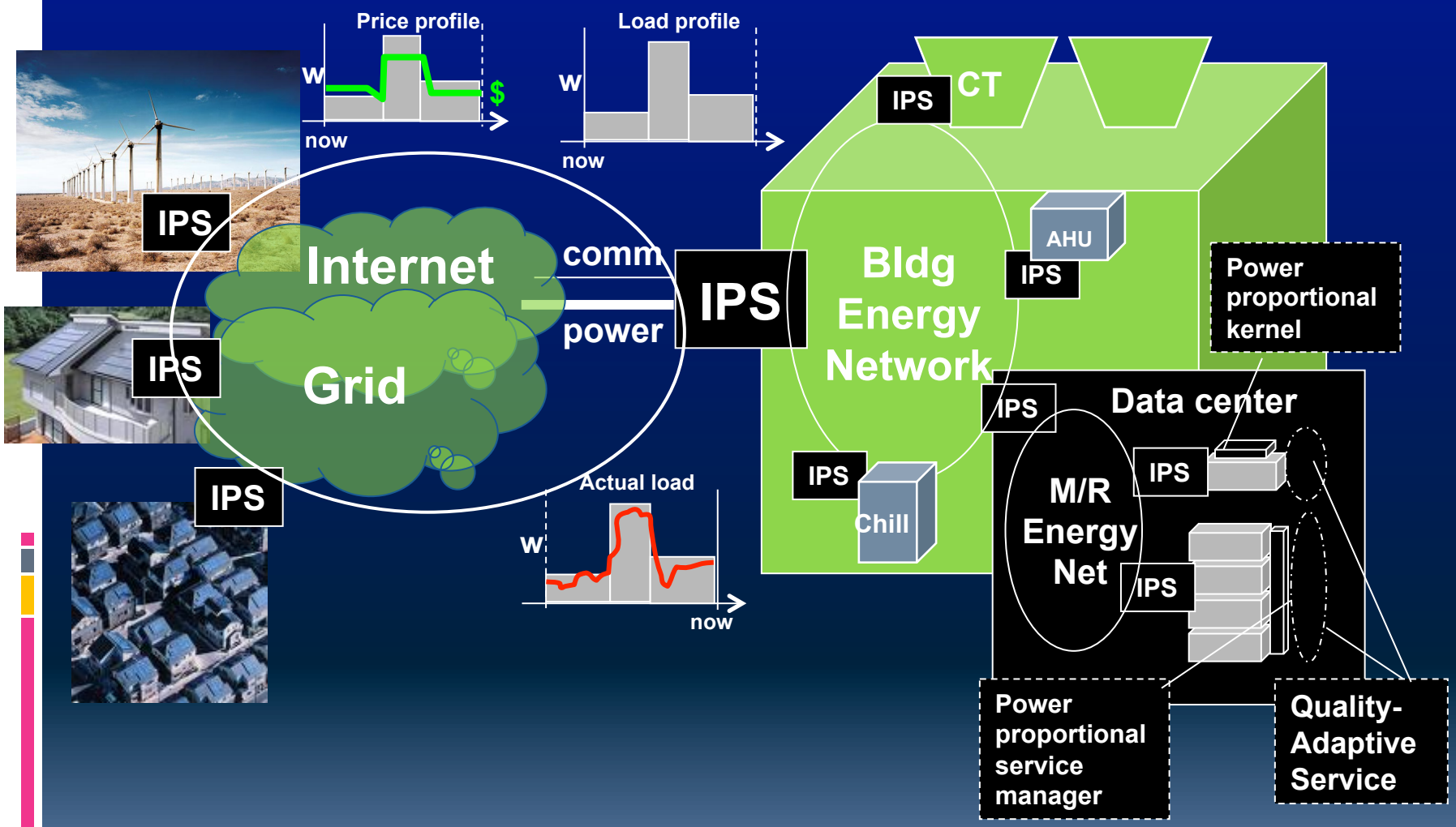
Prime target of opportunity for renewable supplies



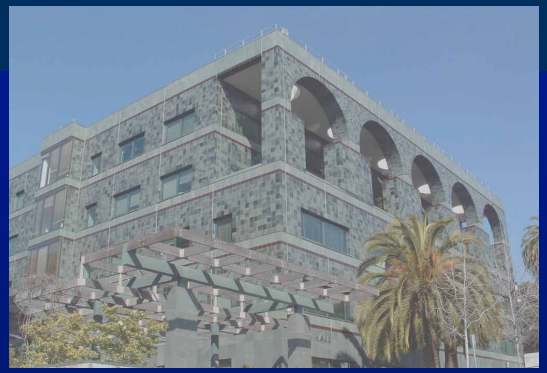


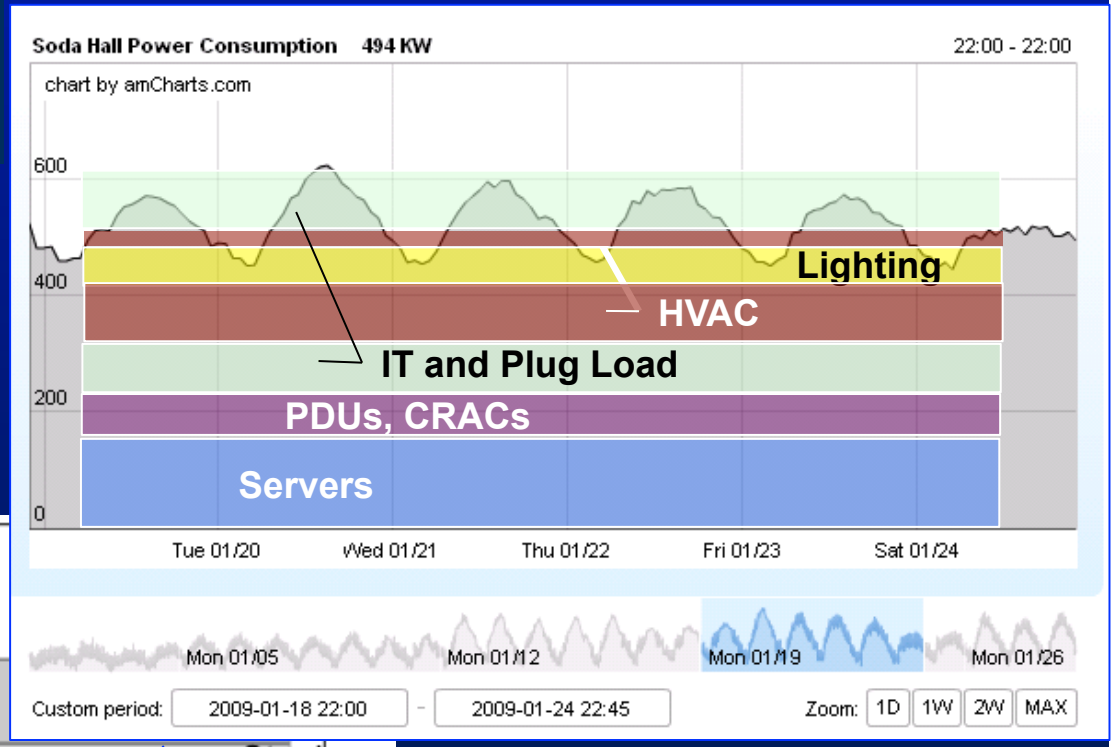


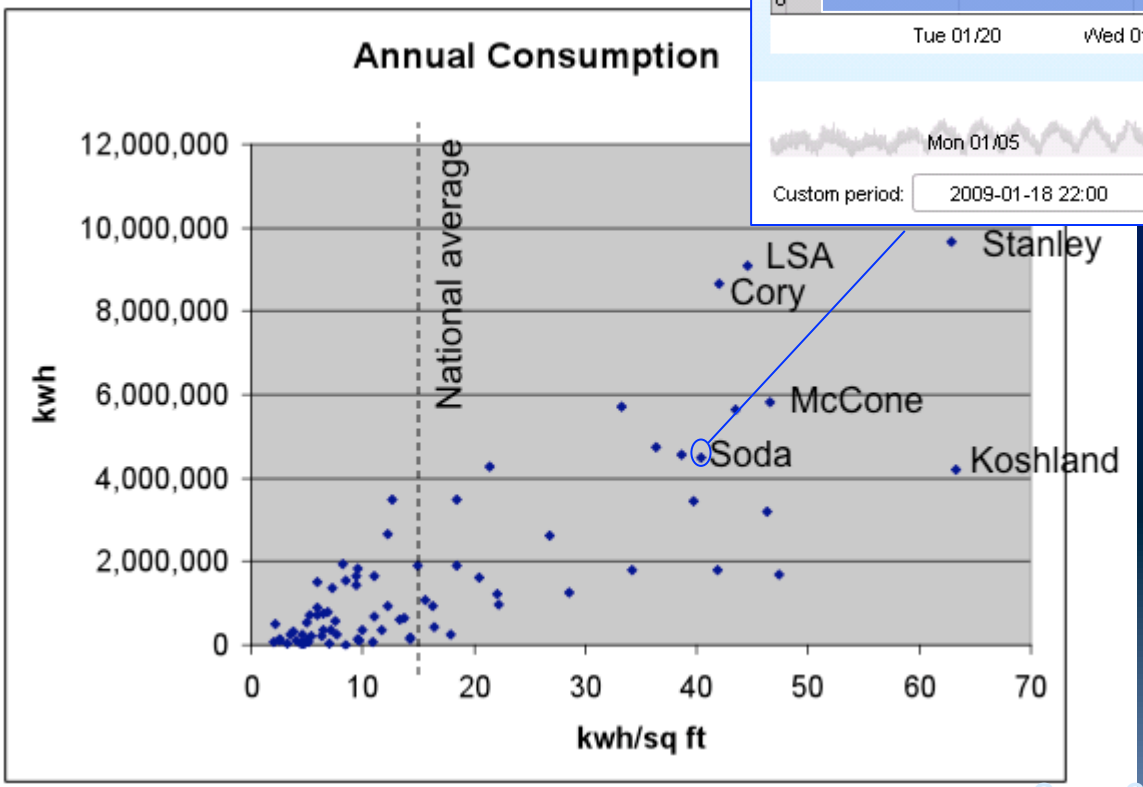
Interesting Building within Building



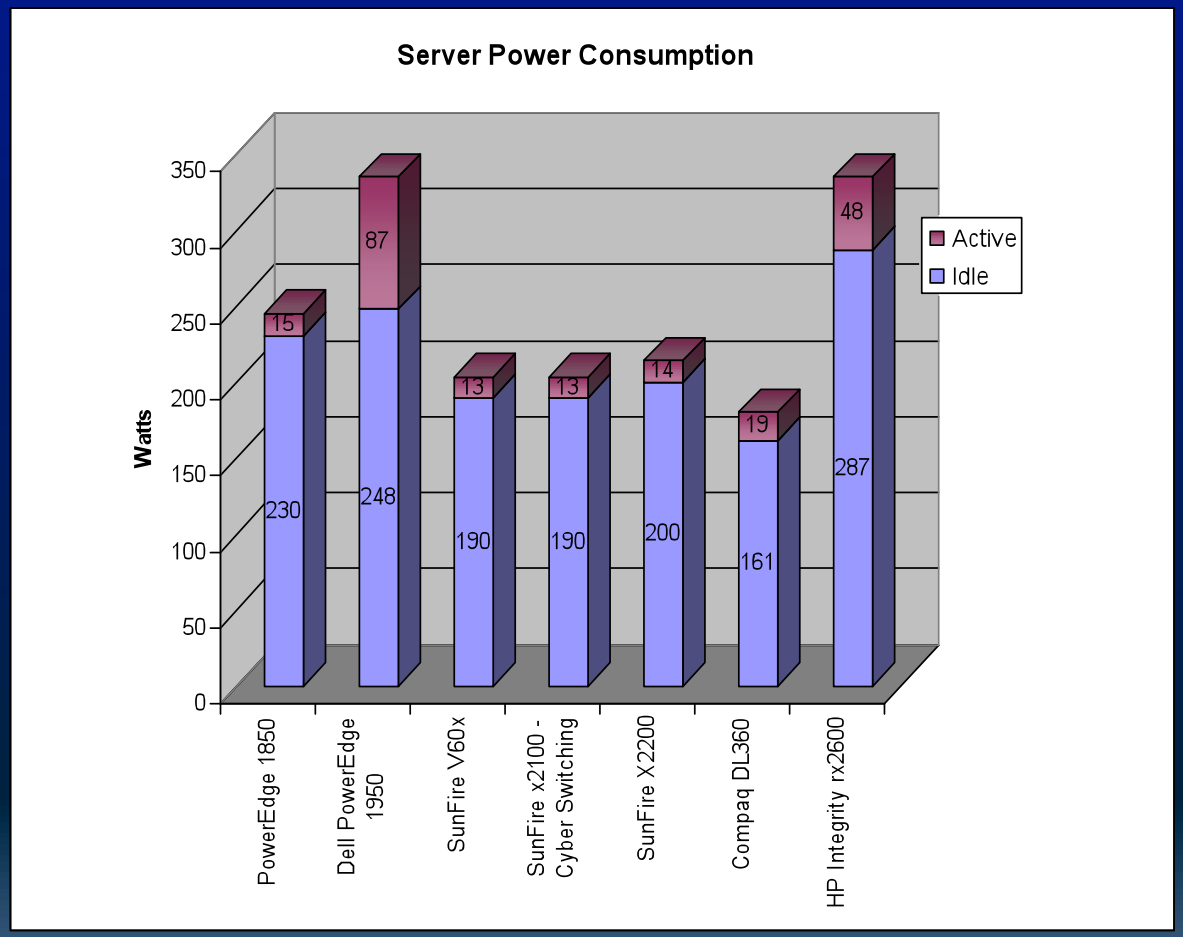
Our Building







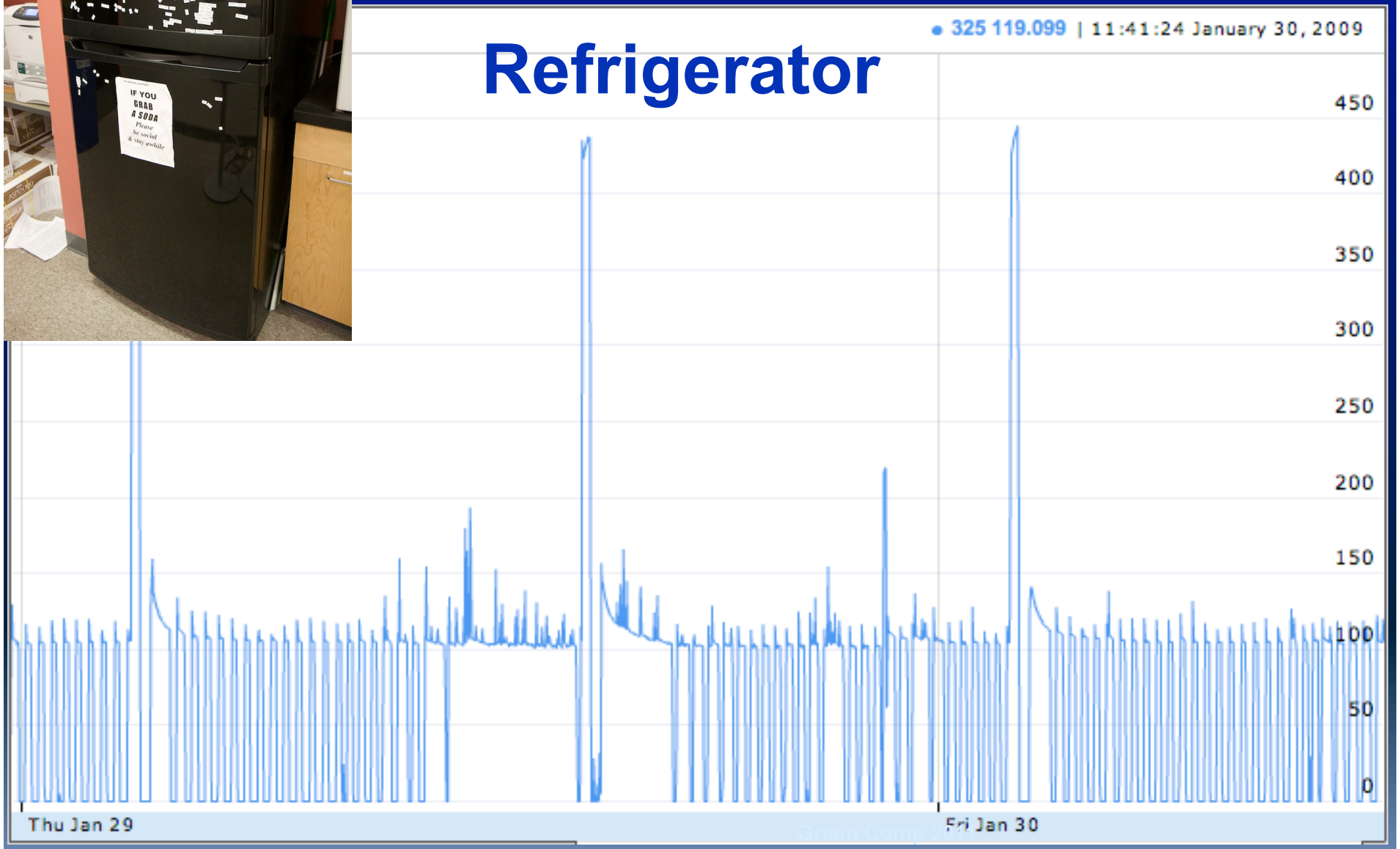
Doing nothing poorly



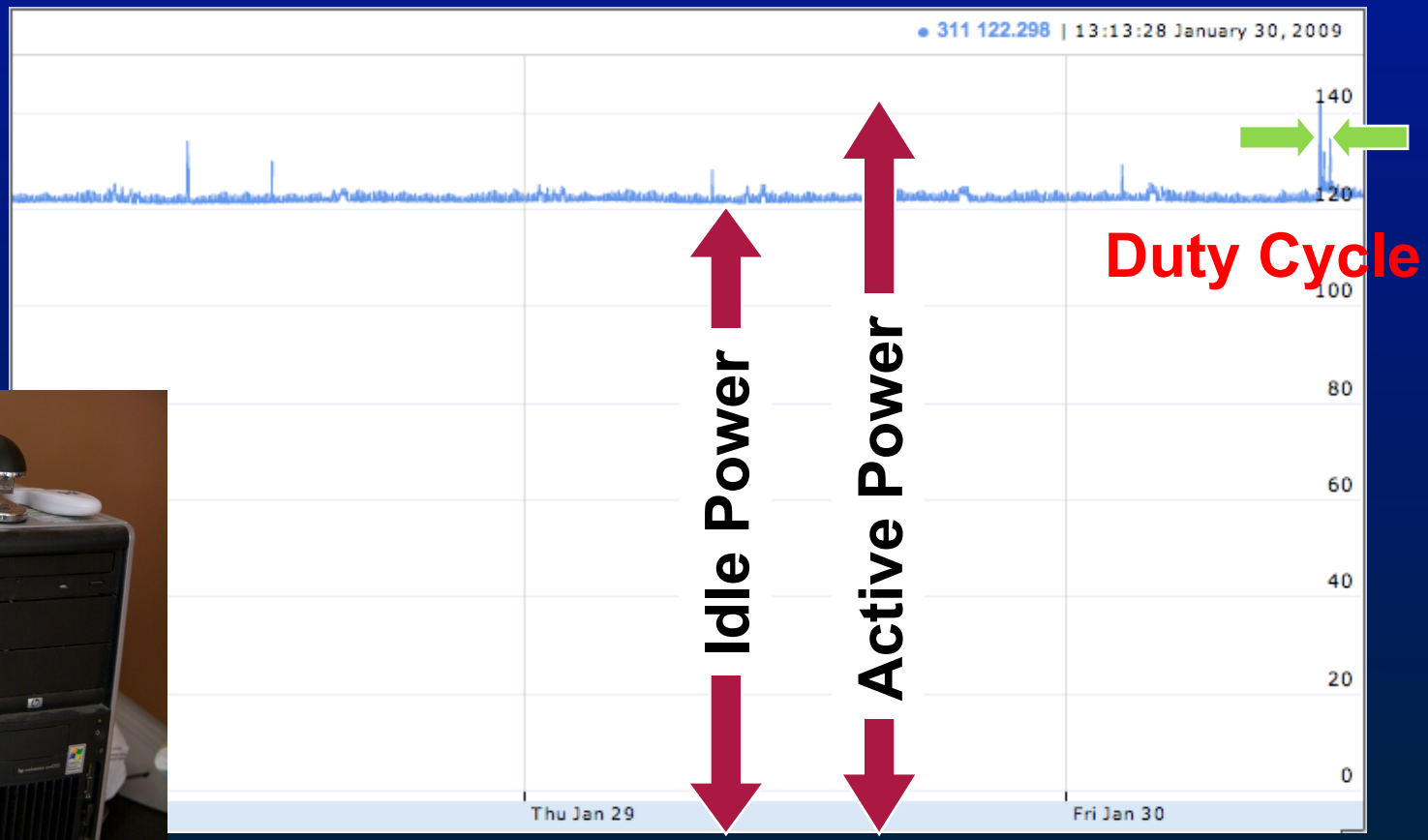
Our Other Building

s this Load …

Refrigerator



Guess this load…

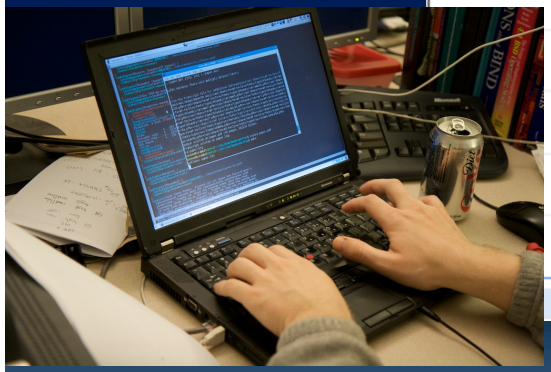


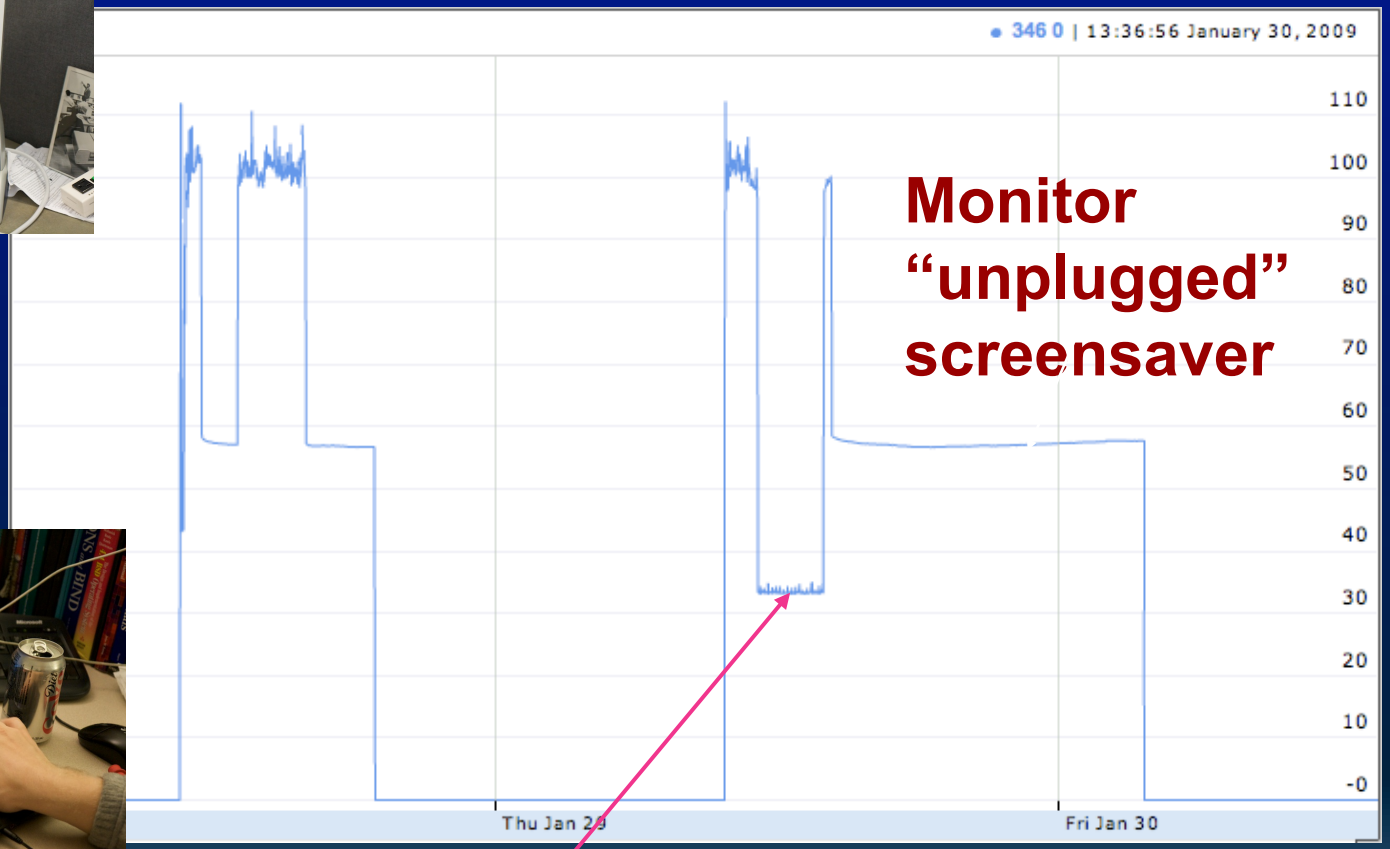
HP desktop Workstation

Guess this Load …

Dell Desktop + monitor

Guess this Load … Laptop + monitor





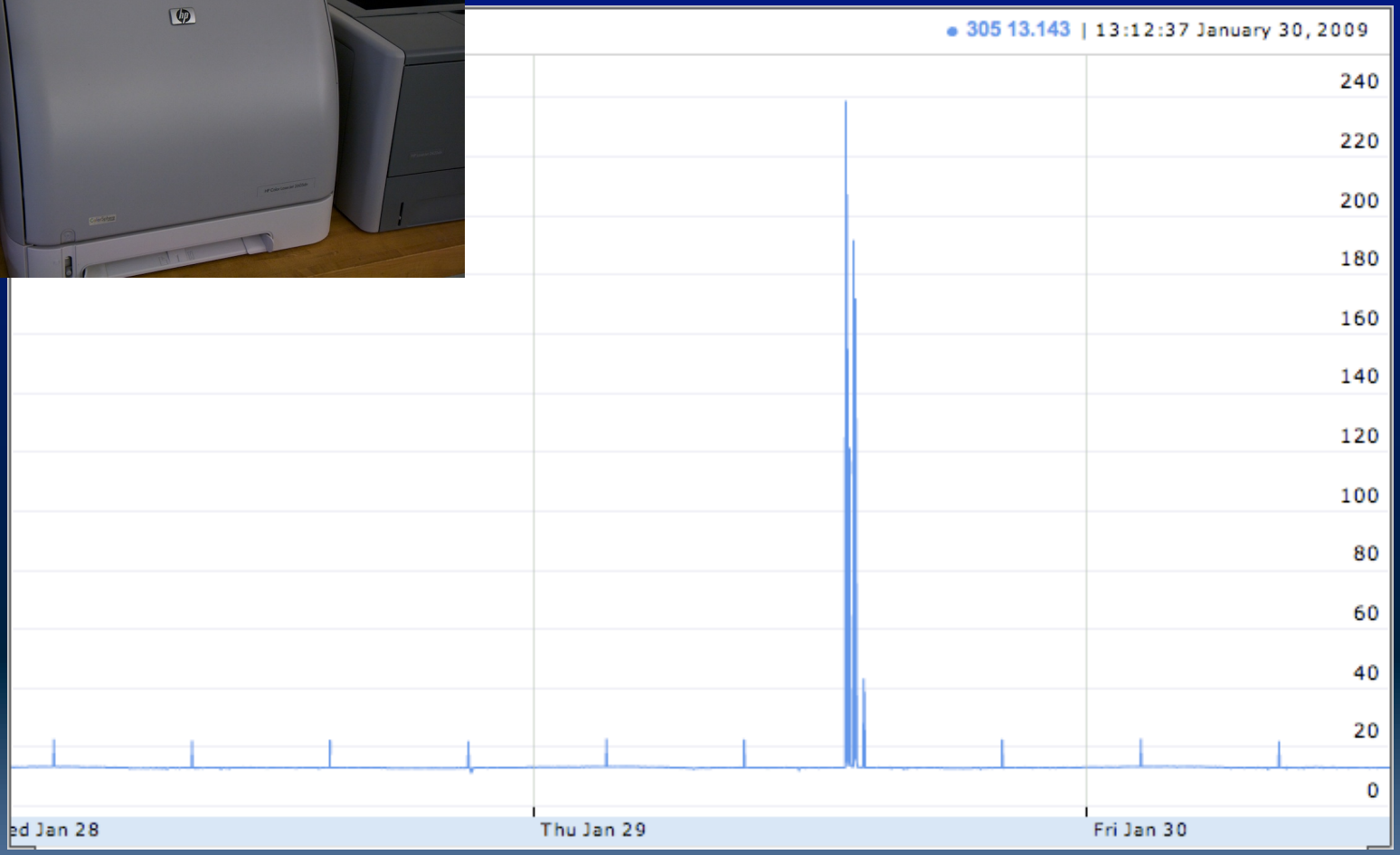
Windows "turns off" monitor

Guess this Load …

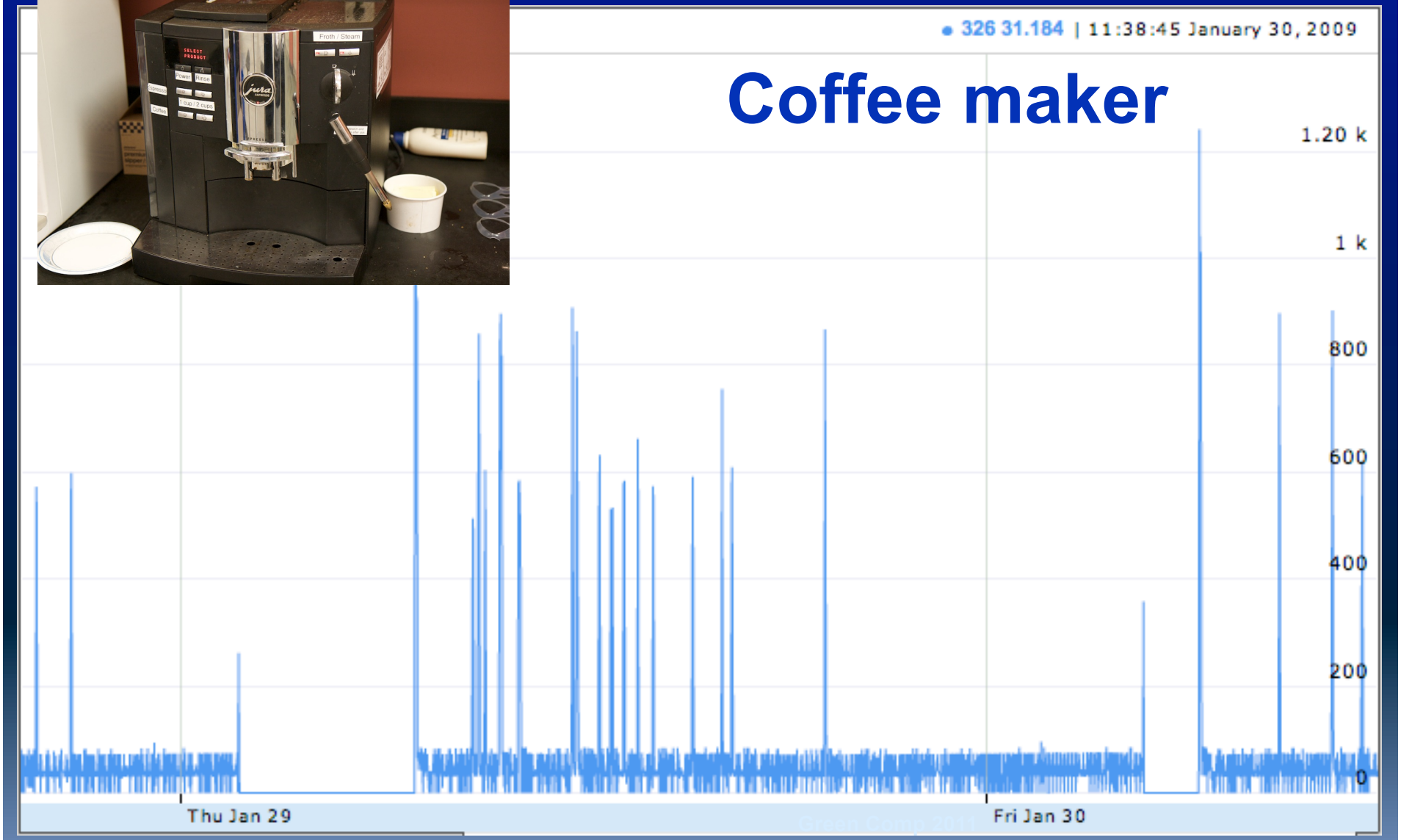
HP B&W laser printer

Guess this Load …

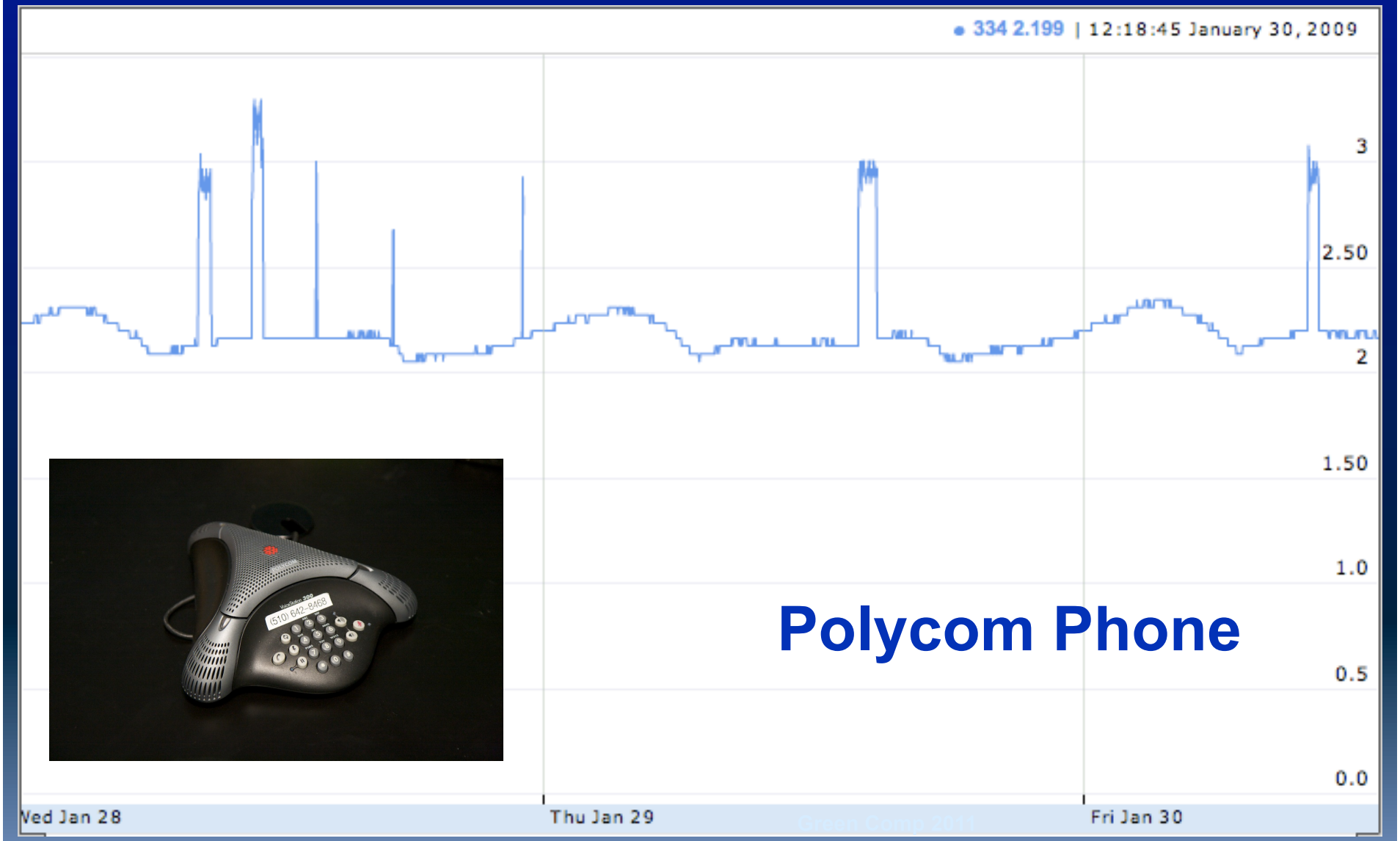
HP color laser printer



Guess this Load …



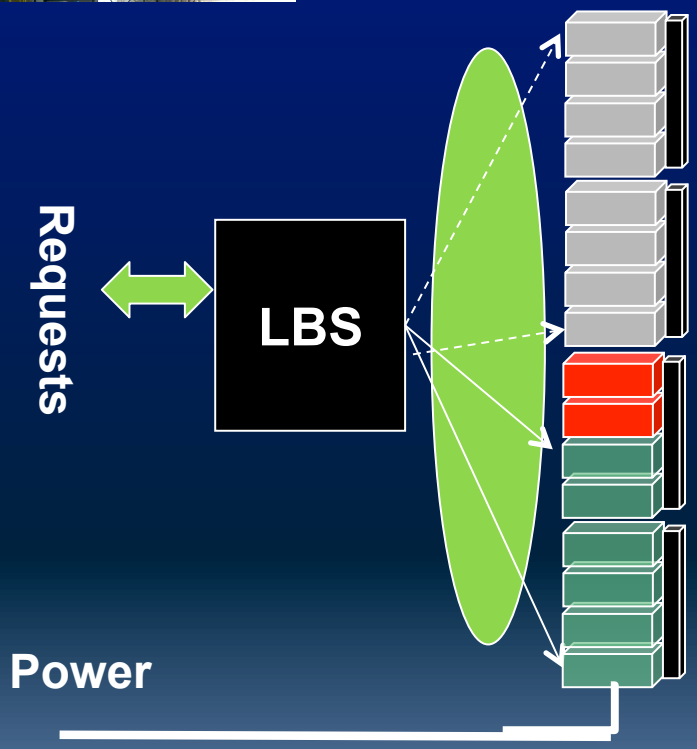
Guess this Load …

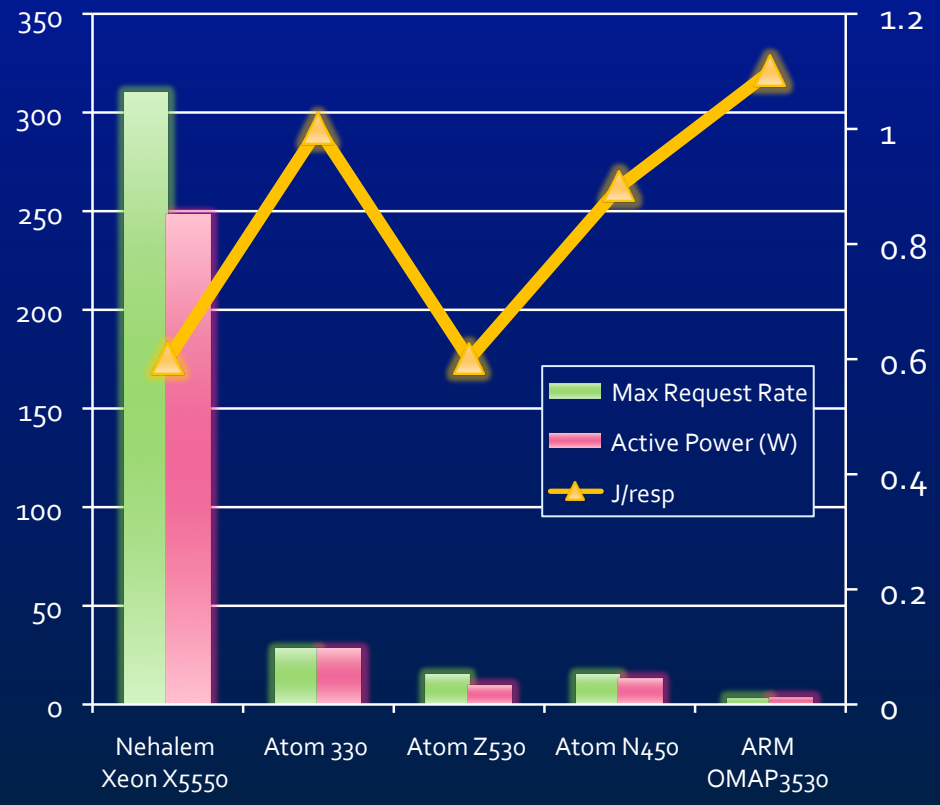


Polycom Phone

Doing Nothing

Brawny or Scrawny



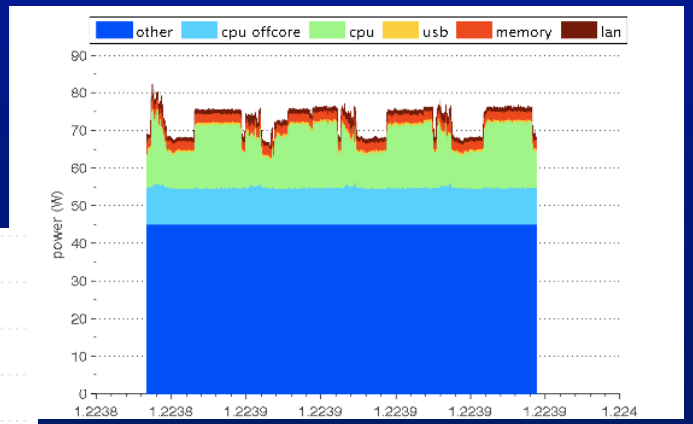


Something or Nothing

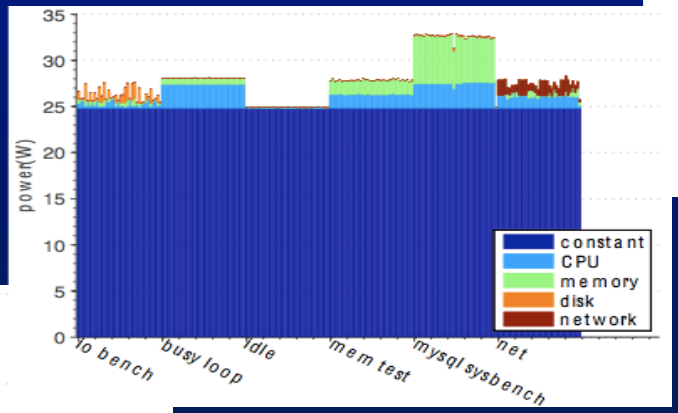
Platform	Max Request Rate	J/resp	Active Power (W)	Idle Power (W)	Idle / Active	Power in S3	Sleep / Active	Wake Time from S3	Wake Time from Off
Nehalem Xeon X5550	310	0.6	248	149	60%	-	NA	-	60
Atom 330	28	1	28	22	79%	1	4%	2.4	40
Atom Z530	15	0.6	9.5	7.7	81%	1.7	18%	0.8	35
Atom N450	15	0.9	12.8	9.7	76%	1.9	15%	5.2	45
ARM OMAP3530	3	1.1	3.4	3.4	100%	-	NA	-	33

Where does the Power go?

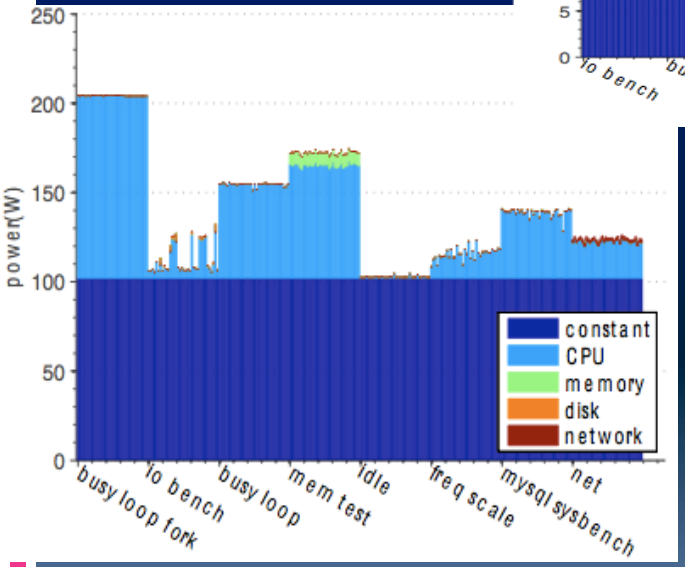
- The Glue, not the processor



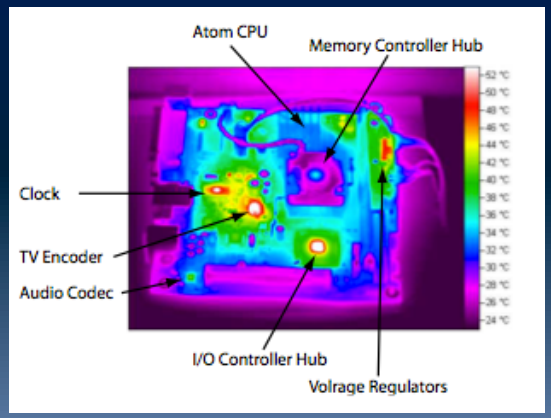
Westmere



Atom 333



Core i7



Doing something a bit slower ?

- $P \approx C \cdot V^2 \cdot f$?

$$P_{total} = p_t (C_L * V * V_{dd} * f_{clk}) + I_{sc} * V_{dd} + I_{leakage} * V_{dd} \quad \text{(EQ 1)}$$

Low Power CMOS Digital Design

Anantha P. Chandrakasan
Samuel Sheng
Robert W. Brodersen†

EECS Department,
University of California at Berkeley

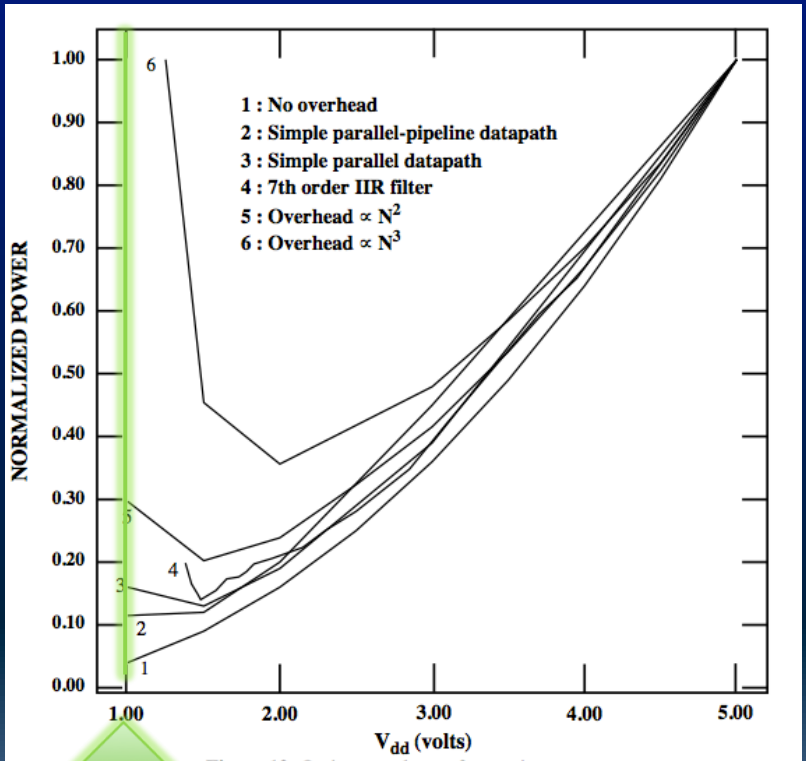
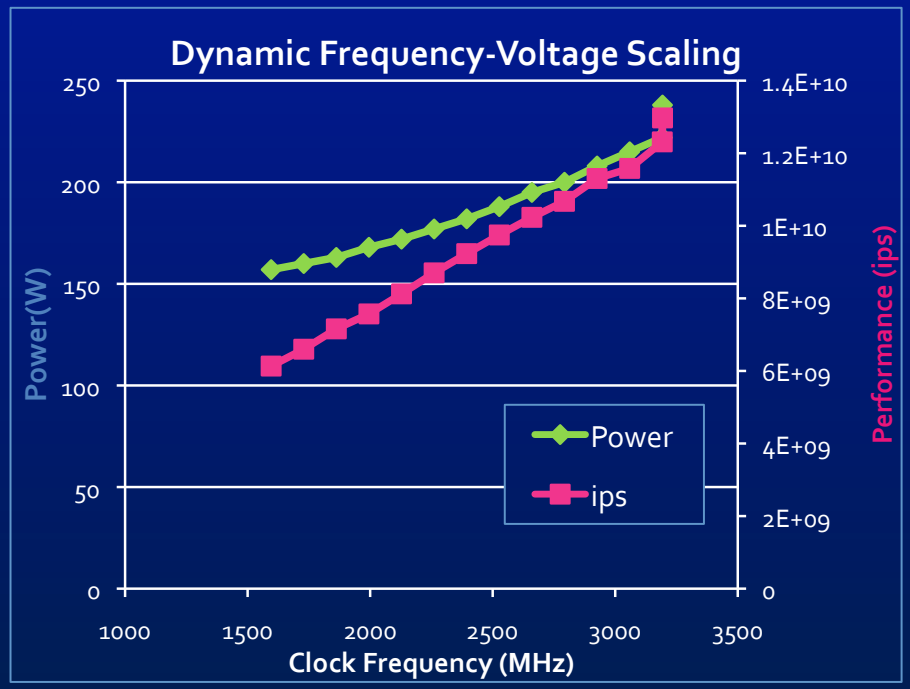


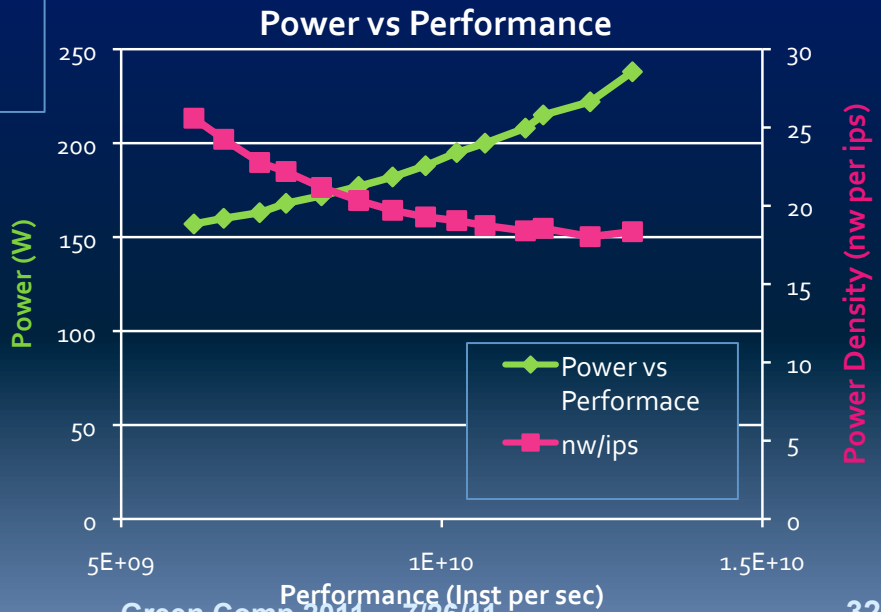
Figure 10: Optimum voltage of operation.

Reduced threshold devices tends to lower the optimal voltage; however, as seen in Section 3.3, at thresholds below 0.2V, power dissipation due to the subthreshold current will soon start to dominate and limit further power improvement.

DVFS really

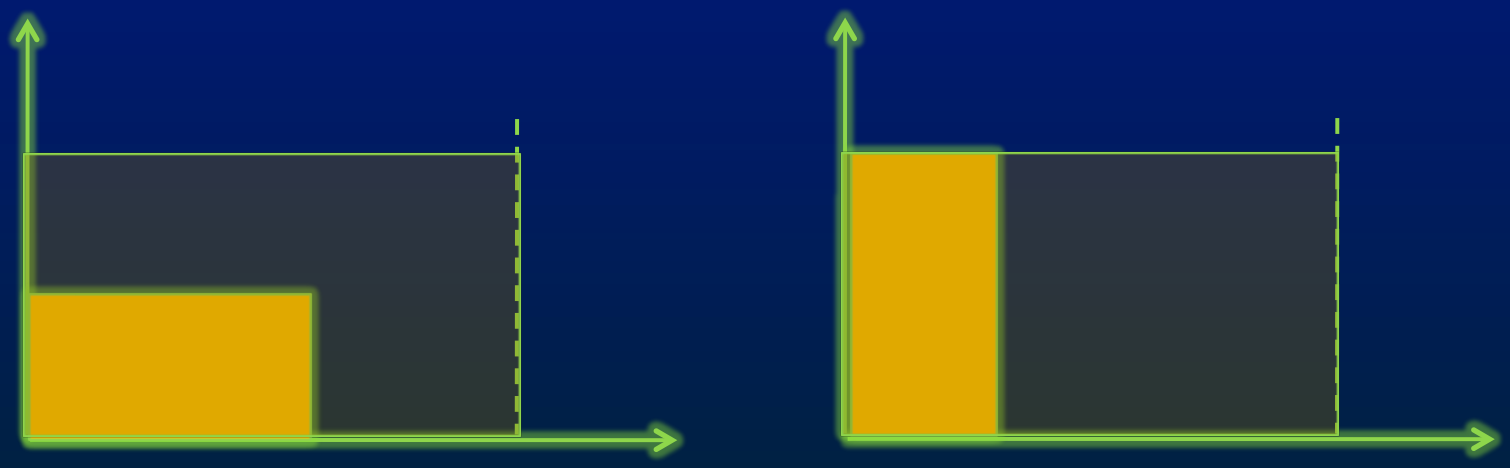


Xeon Core i7 extreme



Remember …

- Energy = Power x Time



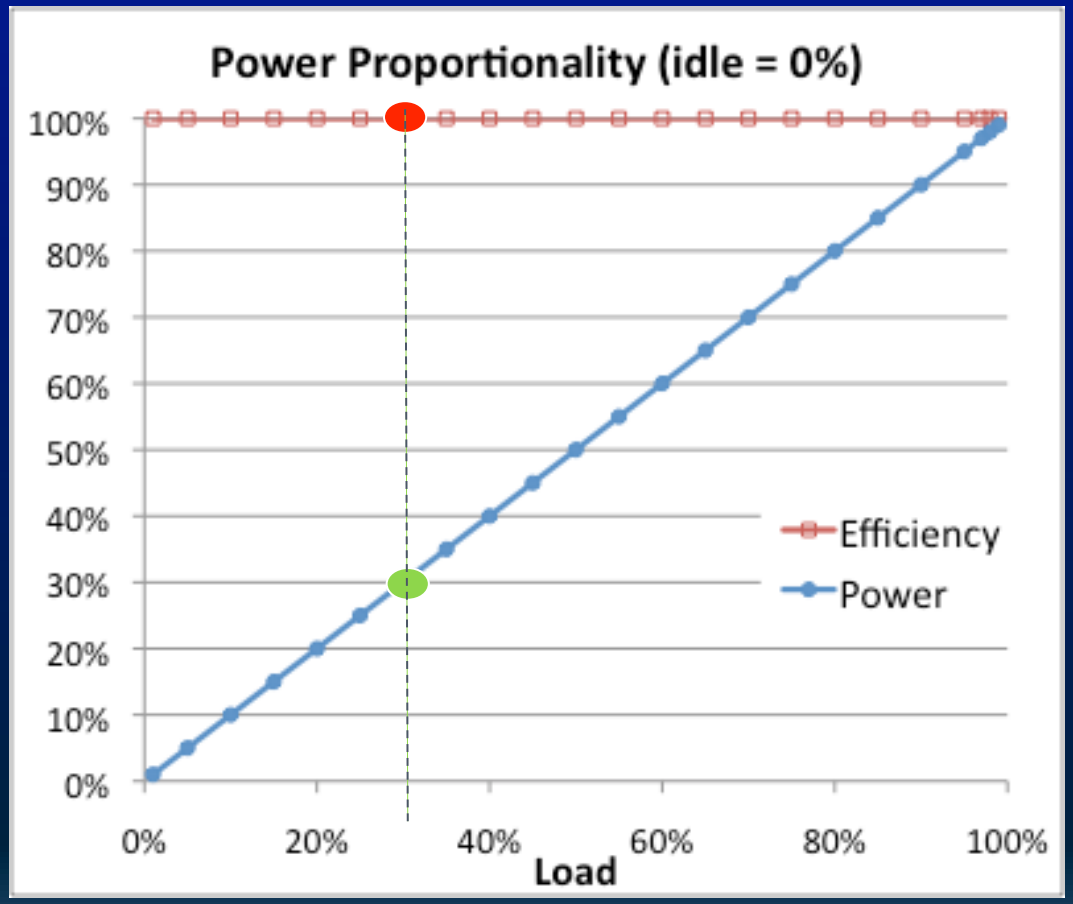
Something or Nothing

Platform	Max Request Rate	J/resp	Active Power (W)	Idle Power (W)	Idle / Active	Power in S3	Sleep / Active	Wake Time from S3	Wake Time from Off
Nehalem Xeon X5550	310	0.6	248	149	60%	-	NA	-	60
Atom 330	28	1	28	22	79%	1	4%	2.4	40
Atom Z530	15	0.6	9.5	7.7	81%	1.7	18%	0.8	35
Atom N450	15	0.9	12.8	9.7	76%	1.9	15%	5.2	45
ARM OMAP3530	3	1.1	3.4	3.4	100%	-	NA	-	33

Engineering 101

- Design, Plan, Size, and Test at **Full Load**
  - Performance measured at full Load
- Add headroom and safety margin
- Operate at Partial Load

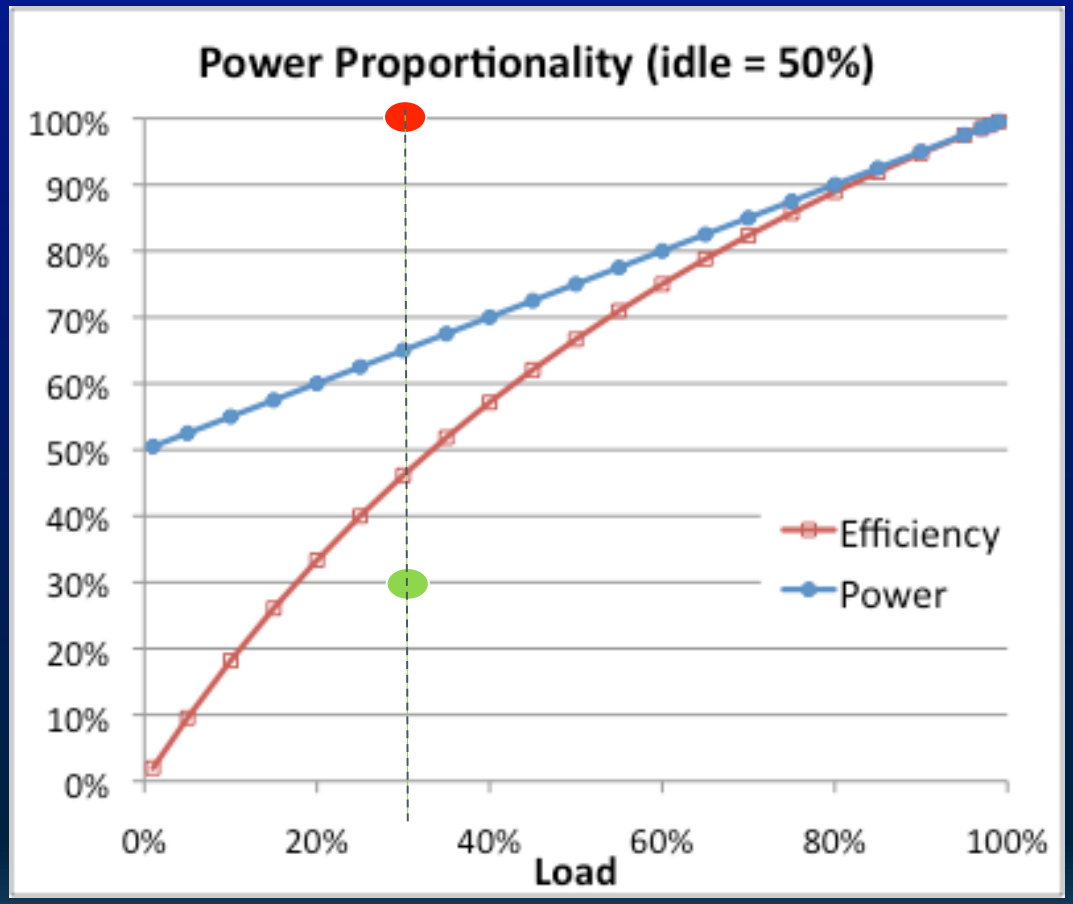
Power Proportionality



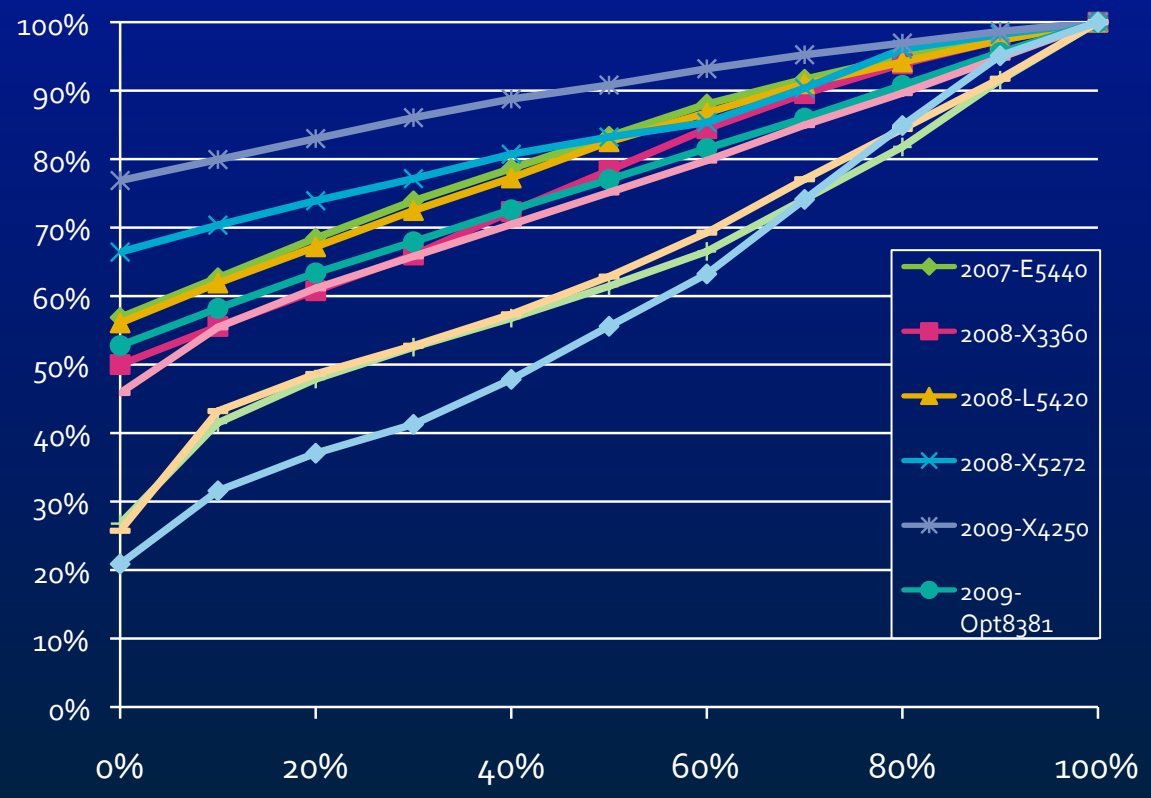
Consumption

Productivity

Power Proportionality



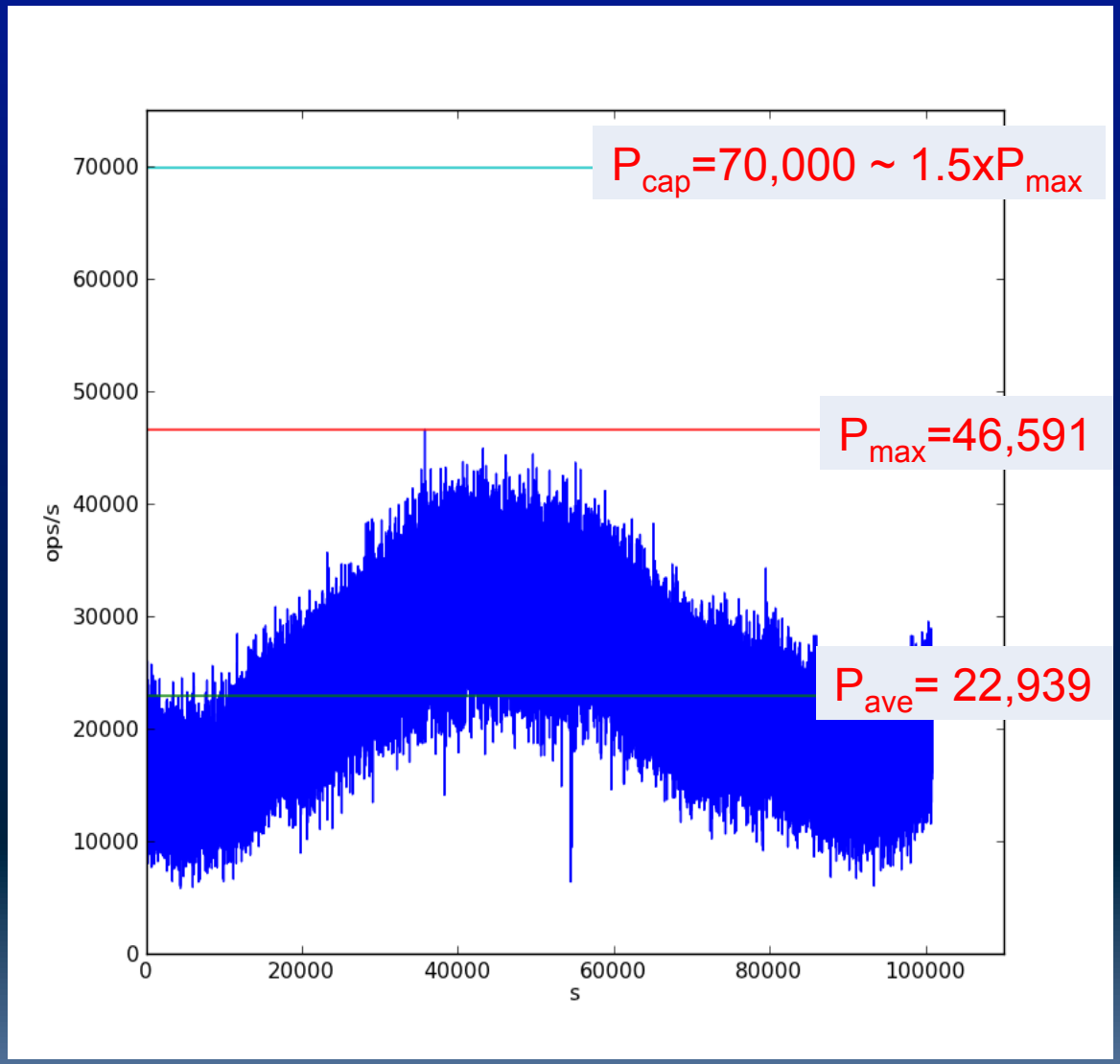
Doing less better and better

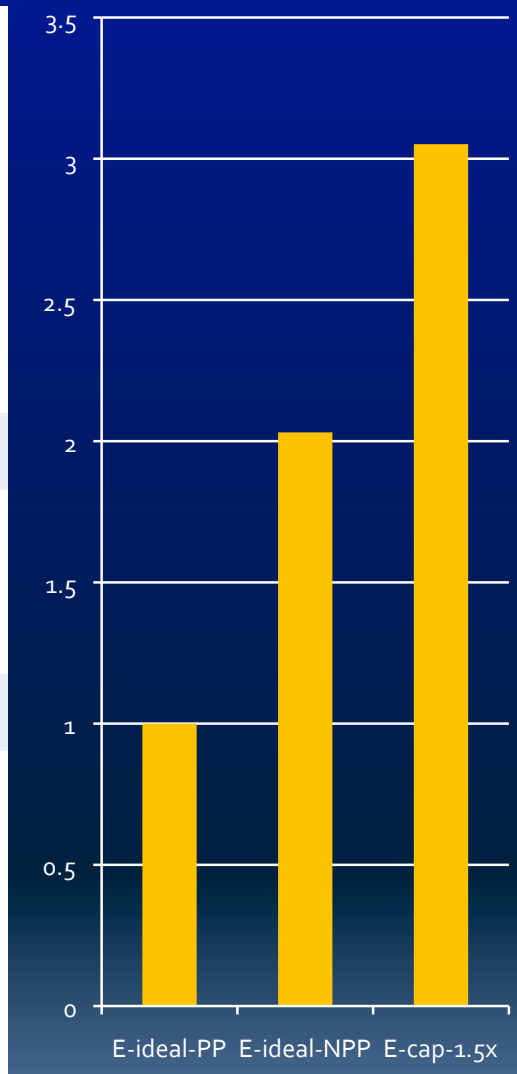


- http://www.spec.org/power_ssj2008

Opportunity in Doing Nothing Well?

A Day in the life of Wikipedia (12/7/2007)

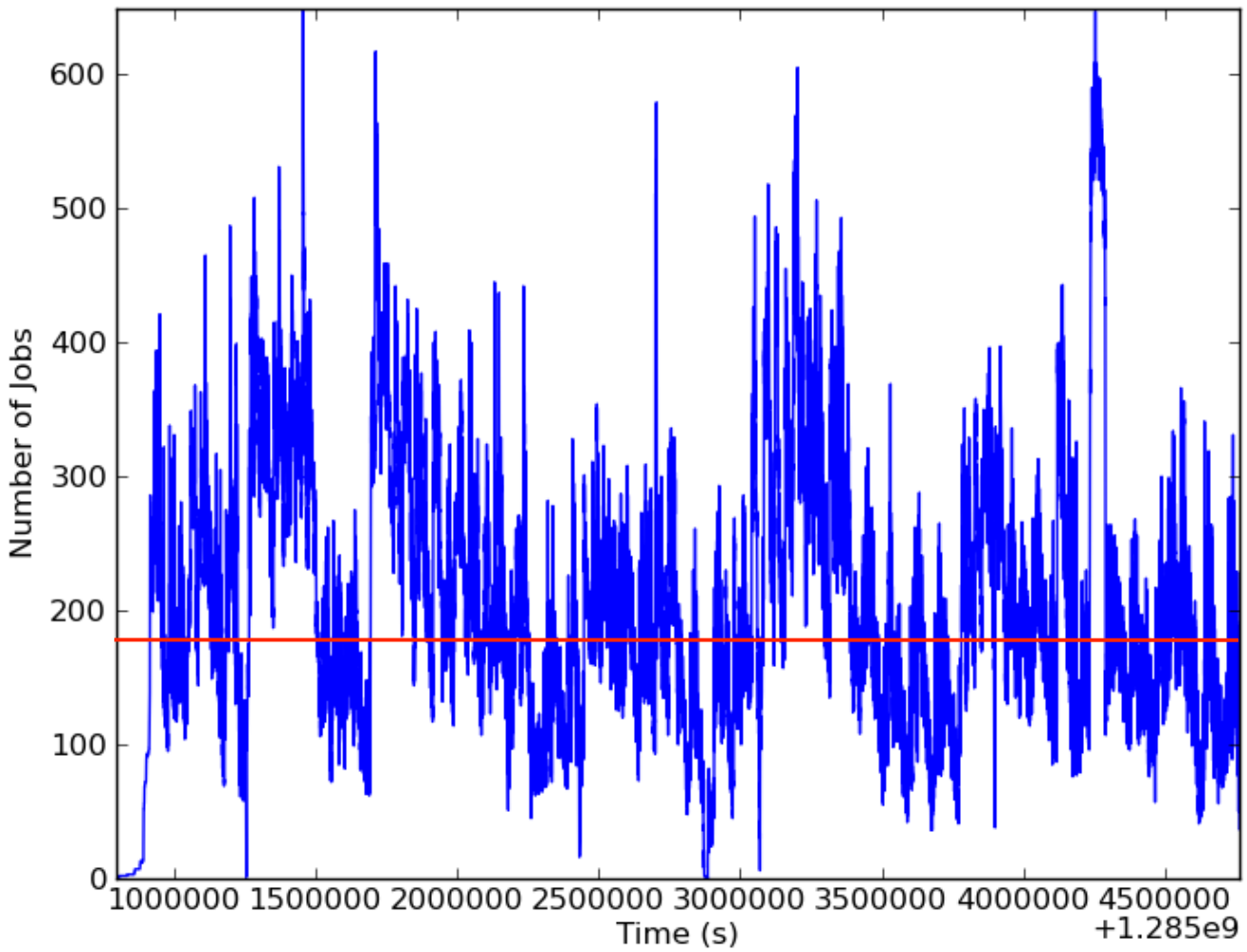




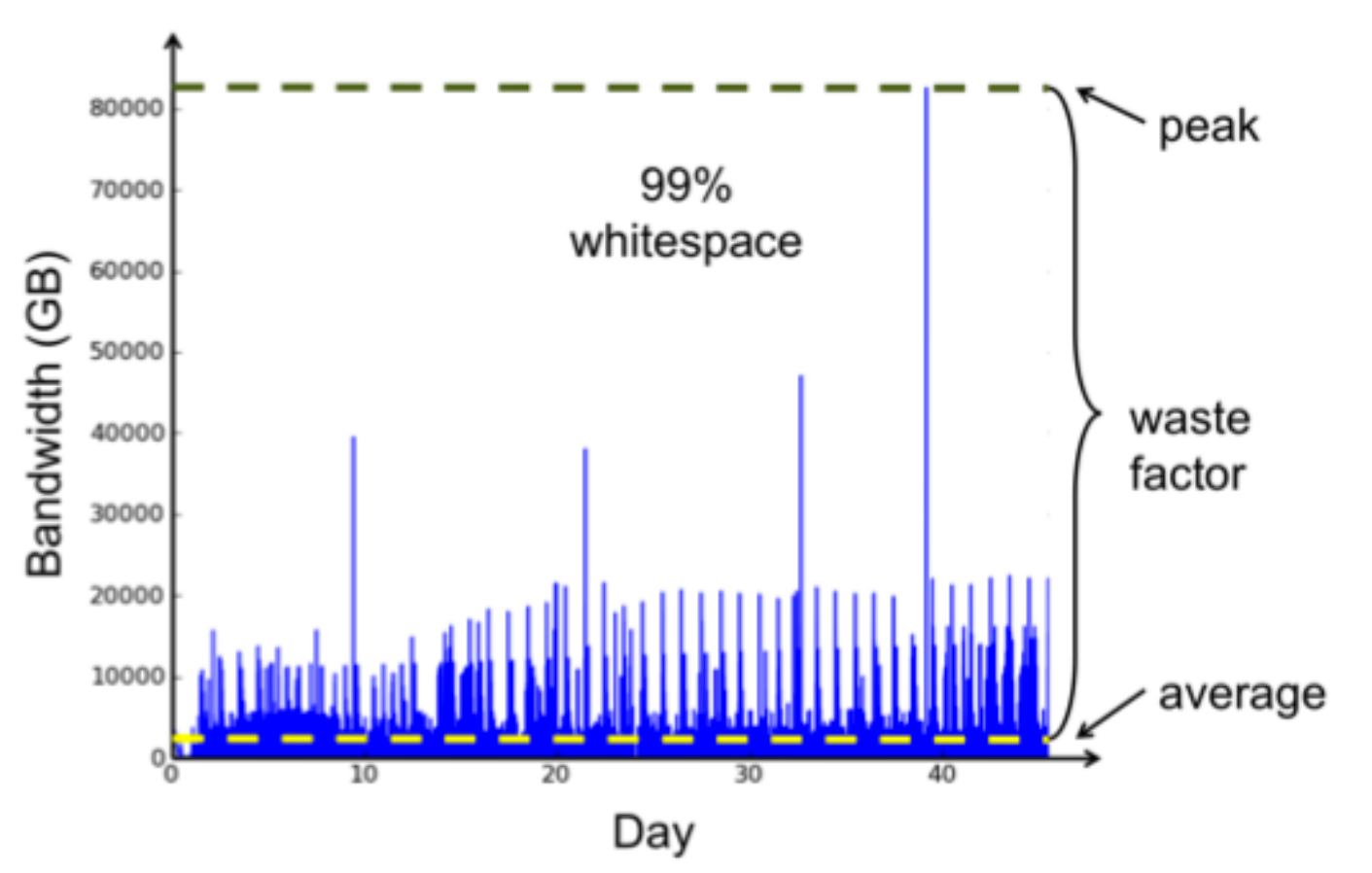
The Opportunity

$$\frac{\text{Energy non-PP}}{\text{Energy ideal PP}} = C * \frac{\text{Peak}}{\text{Ave}}$$

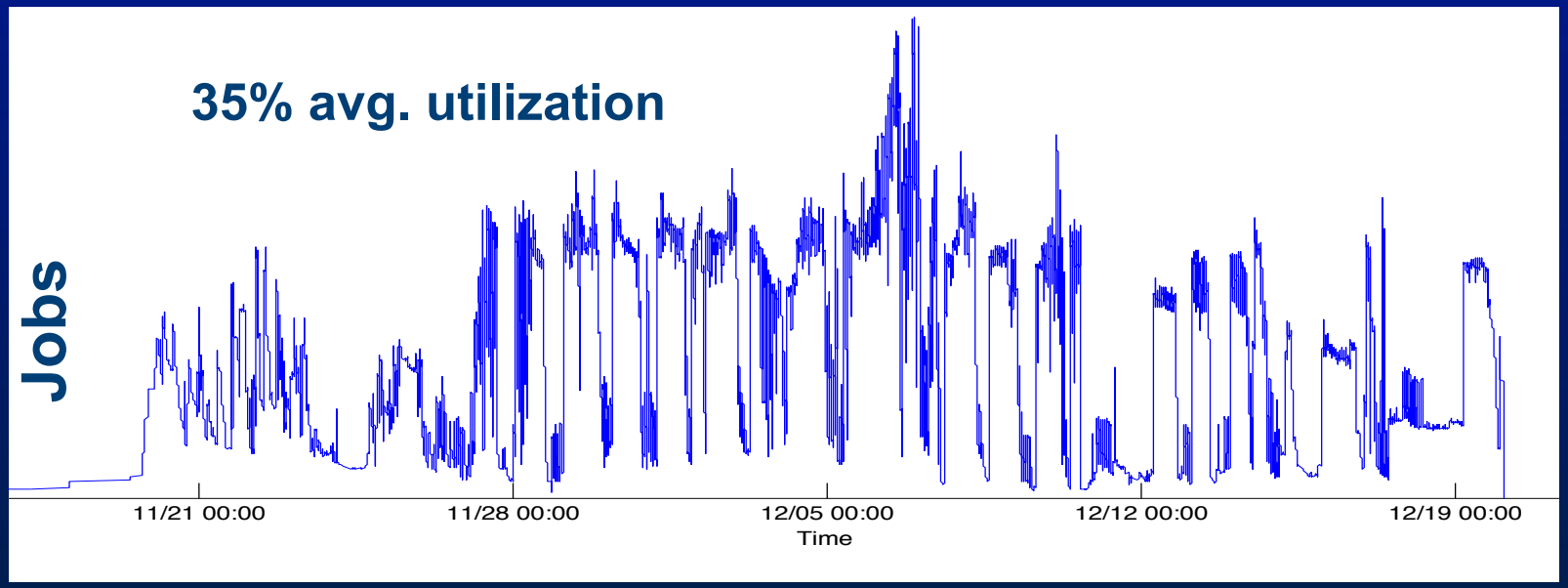
MapReduce Processing @ …



Facebook data analytics

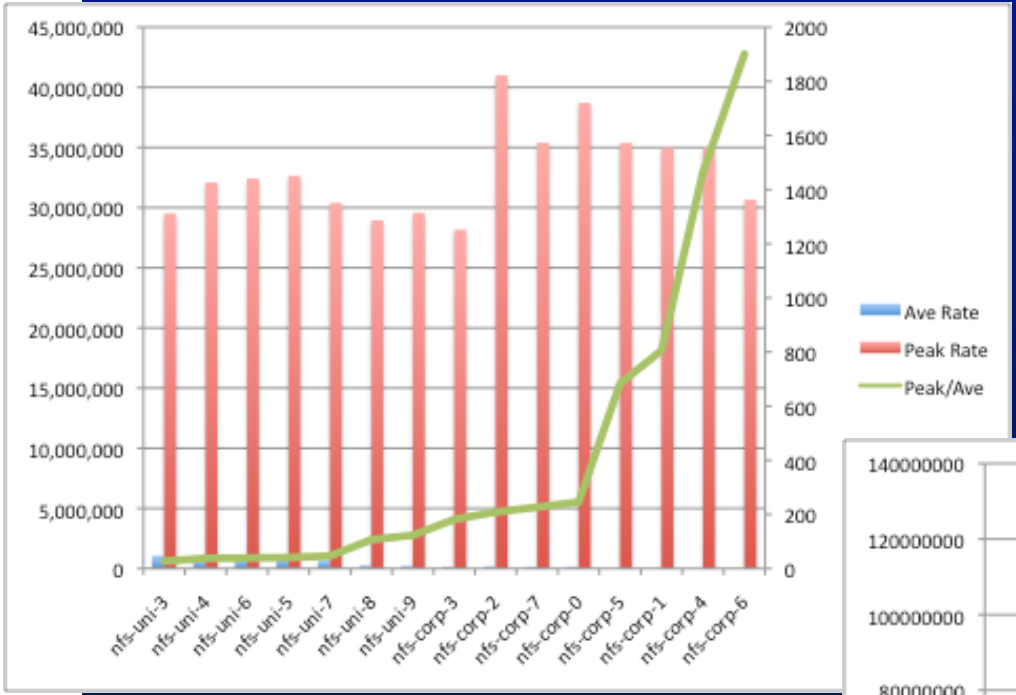


Scientific Computing



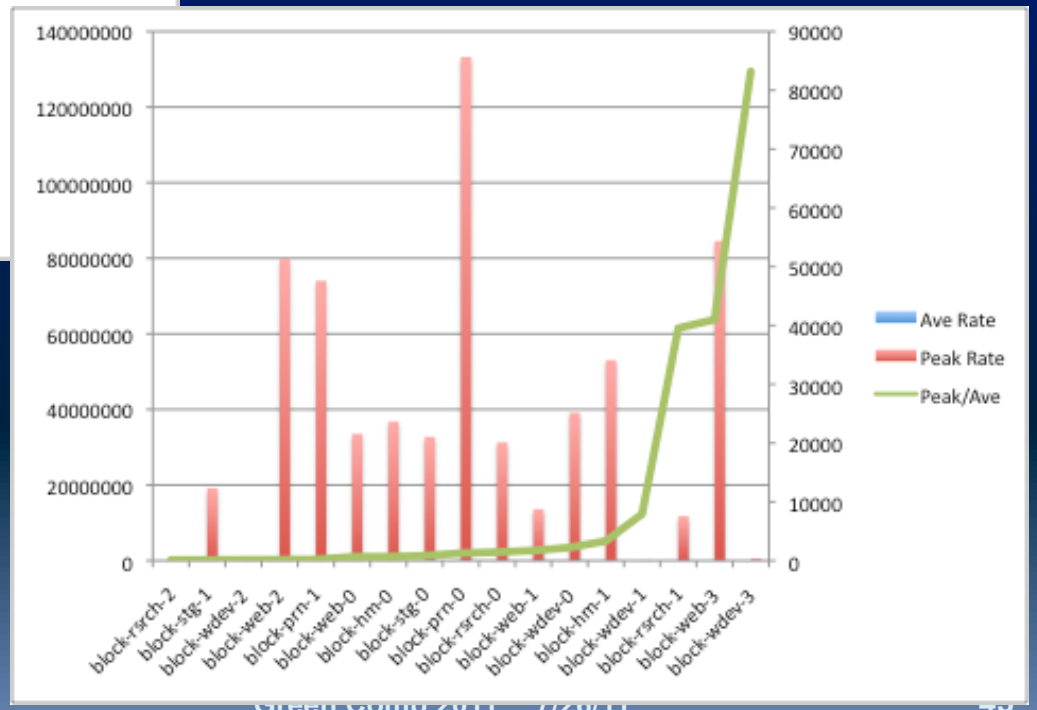
NERSC
112 node (576 core) cluster
~1 month trace, 128,915 jobs

Storage



Enterprise NFS traces

MSR file server traces



Engineering ~~101~~ 102

- Design, Plan, Size, and Test at **Full Load**
  - Performance measured at full Load
- Add headroom and safety margin
- Operate at Partial Load

- Repeat, designing for efficiency at Typical Load

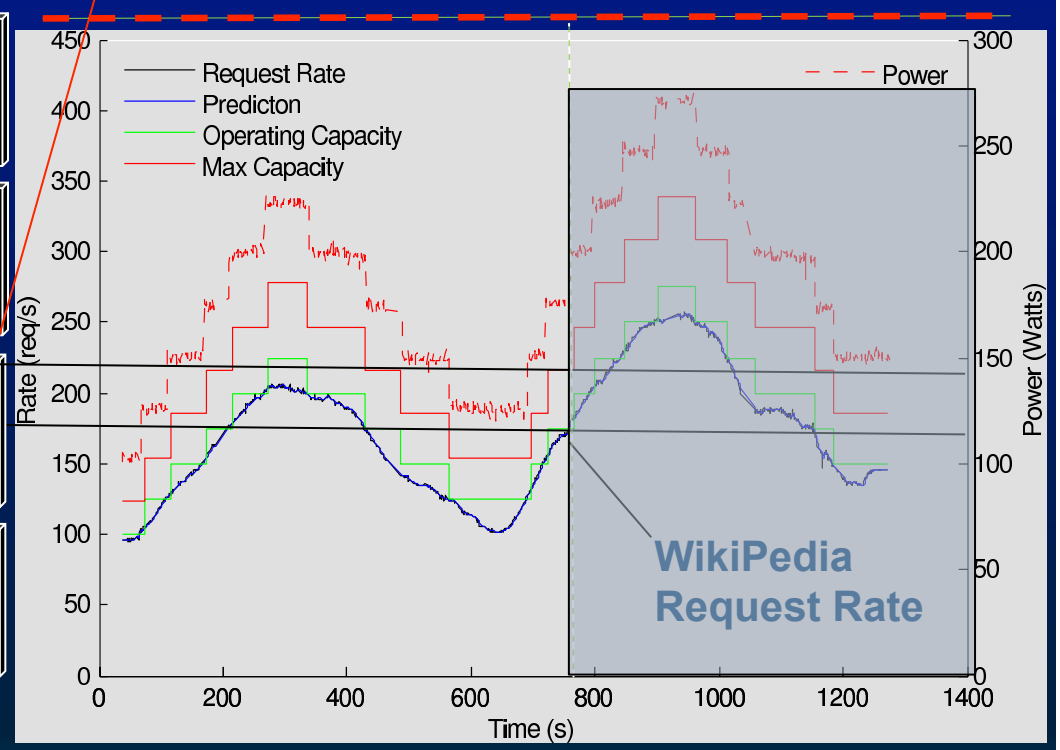
PP Systems of non-PP pieces

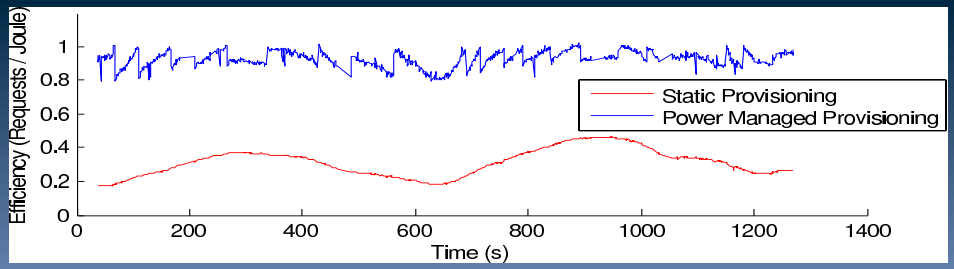
Requests

IPS

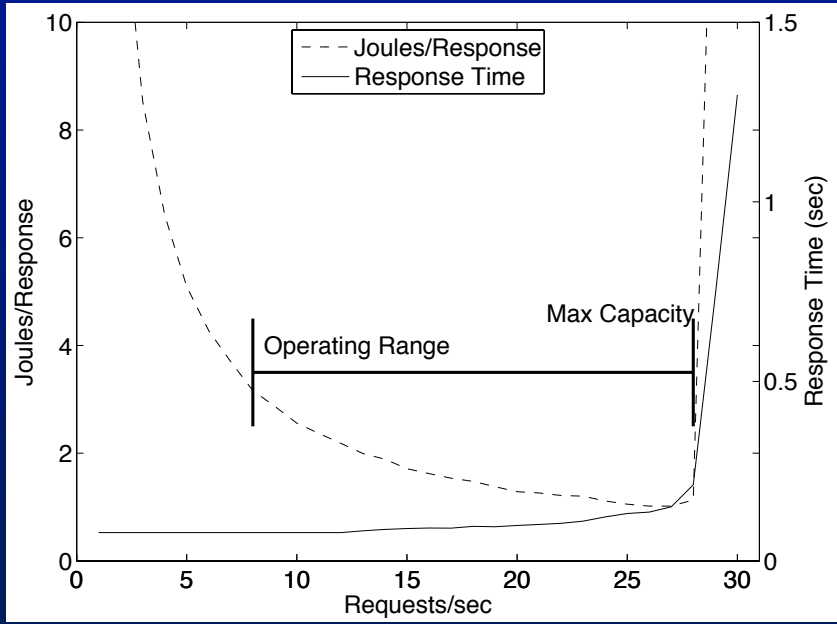
Load Distribution
Scheduling
Power Management

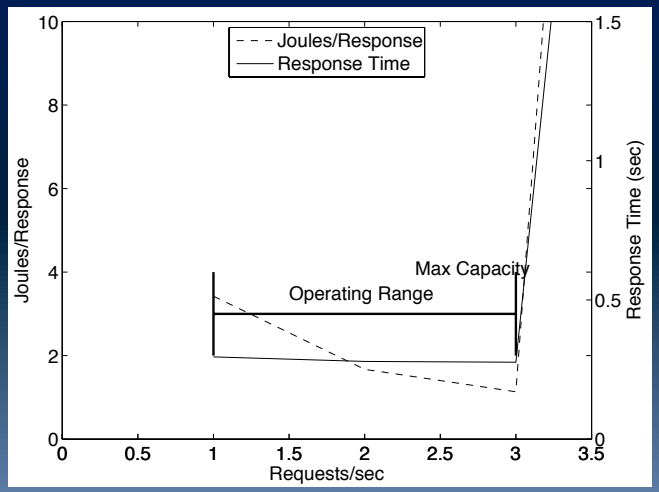
Power

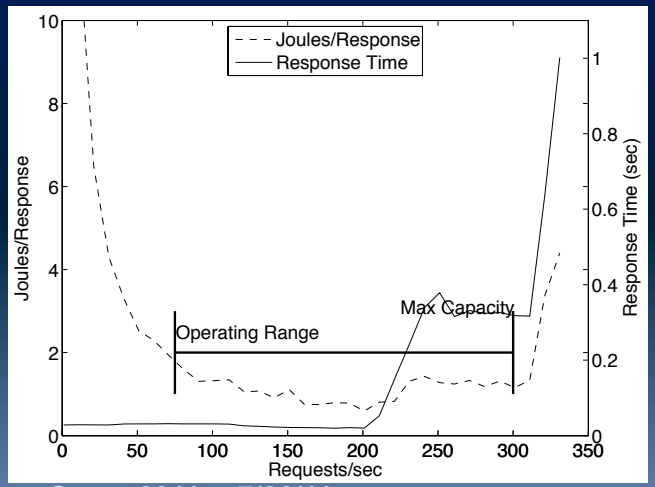




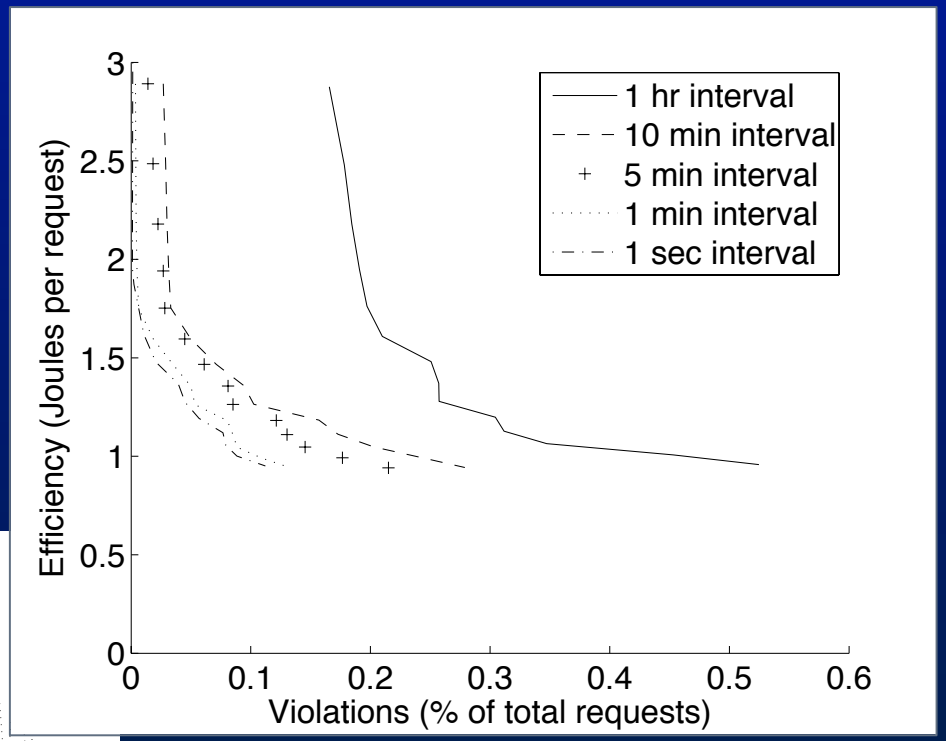
Power/Performance Sweet Spot

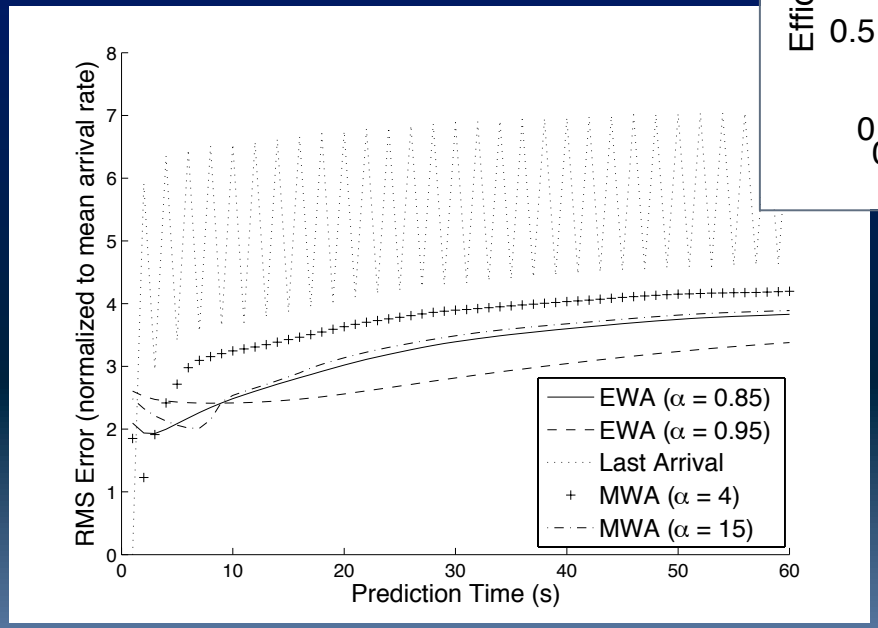






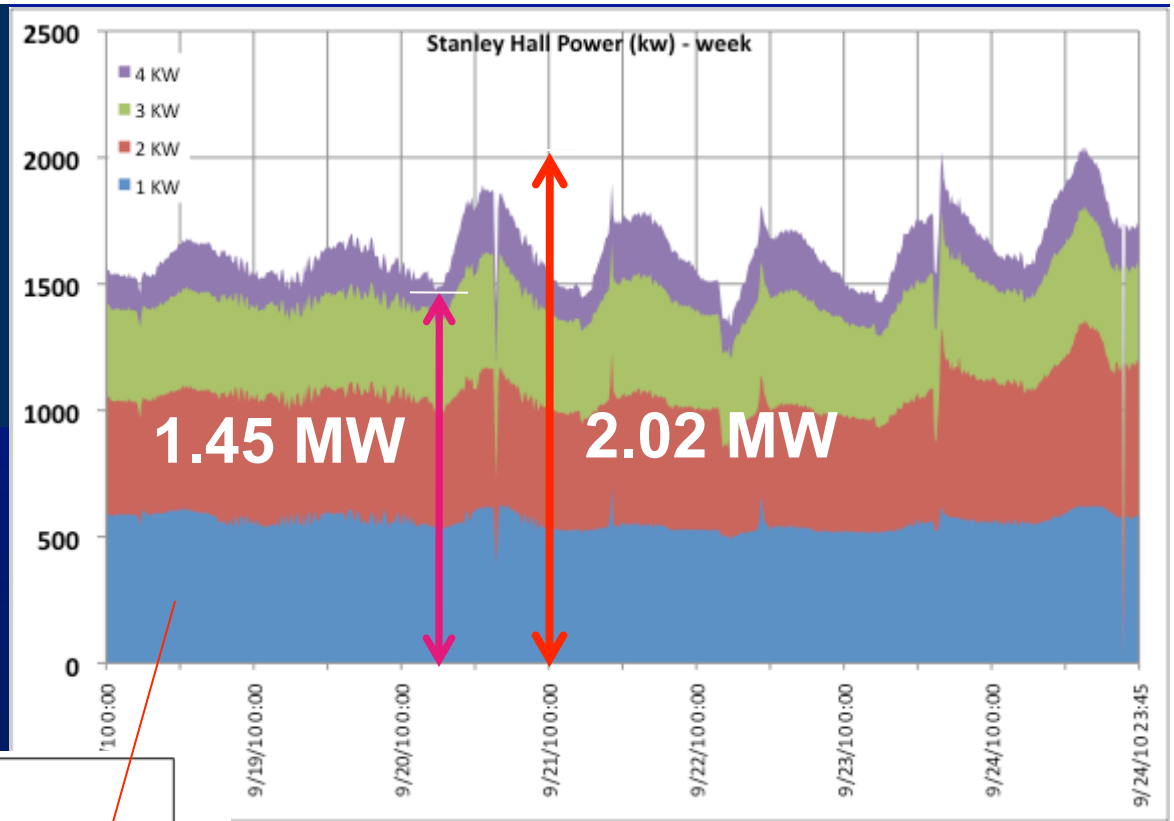
Prediction and Penalties

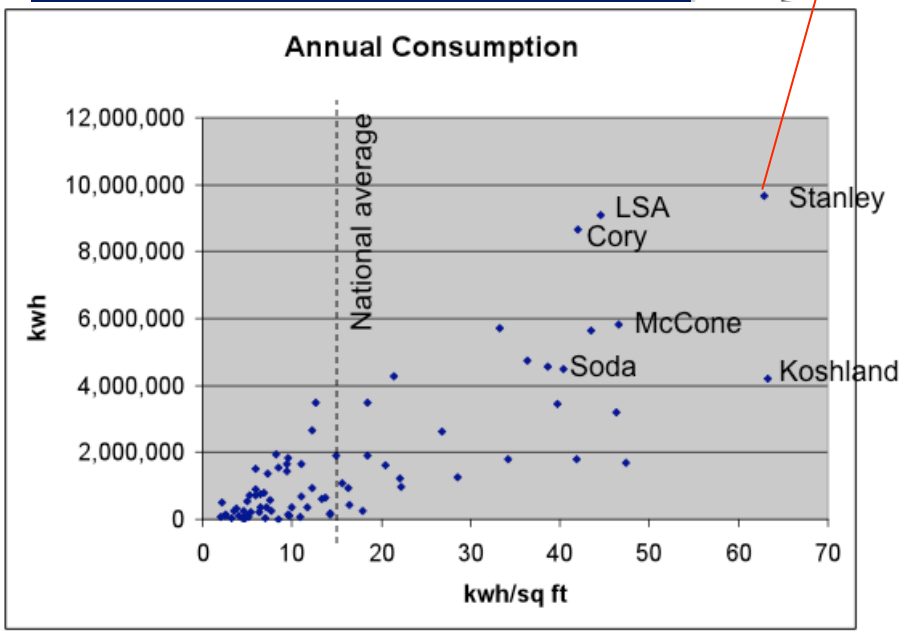




Power-Proportional Buildings ?

Stanley Hall:
Office + BioScience
- 13 NMRs

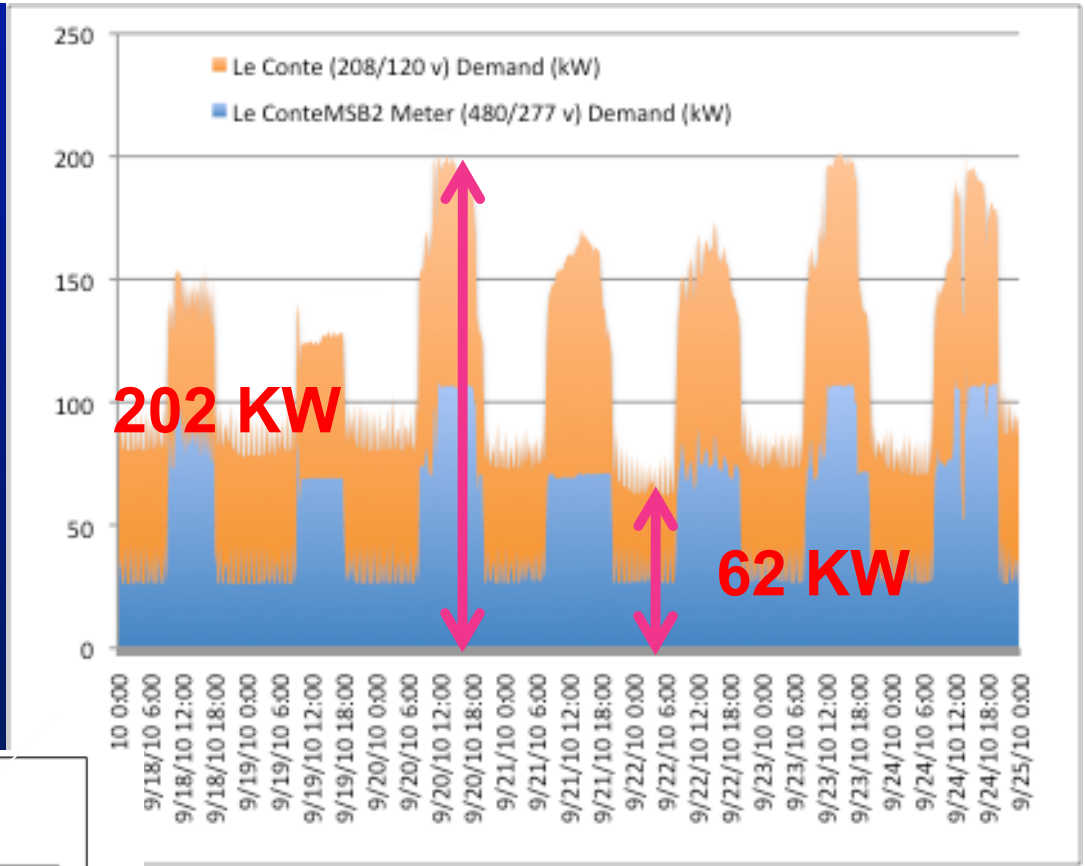


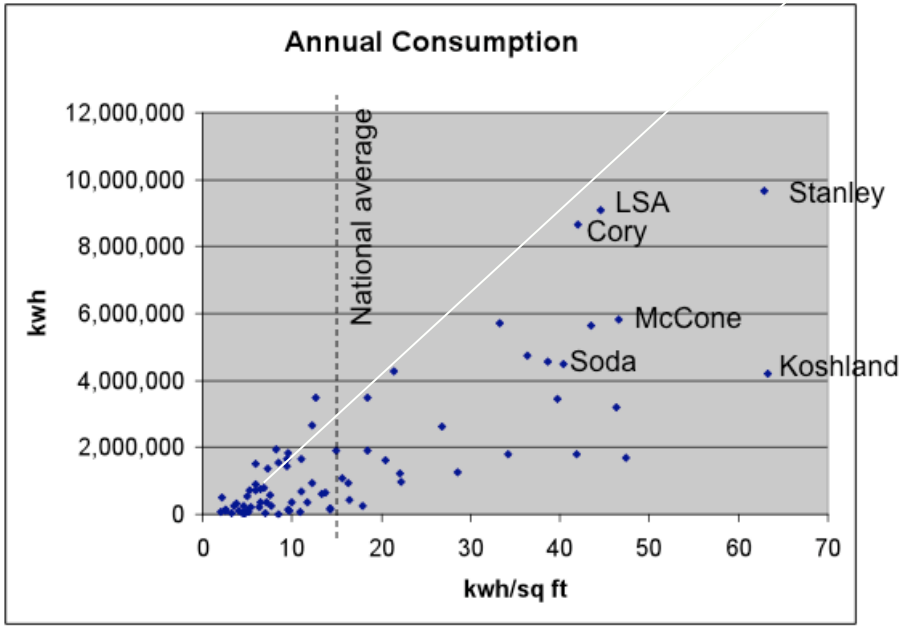


Min = 72% of Max

Power-Proportional Buildings ?

LeConte Hall: Office





Min = 31% of Max

Power Proportional Buildings?

50 Ton Chiller

200 Ton Chiller

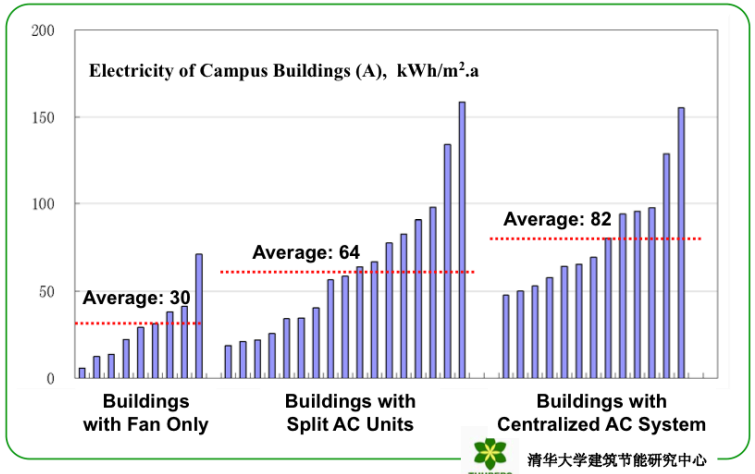
10 months

2 months

Scott McNally Bldg Manager

Power Proportional Buildings?

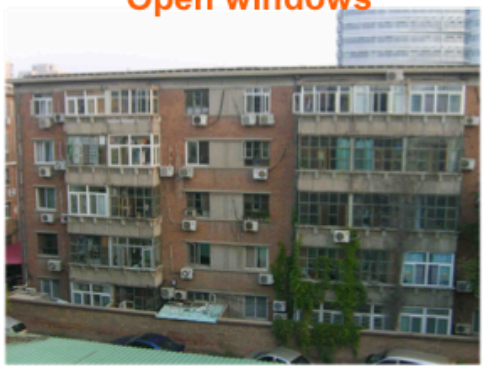
Observations on a Campus in Beijing

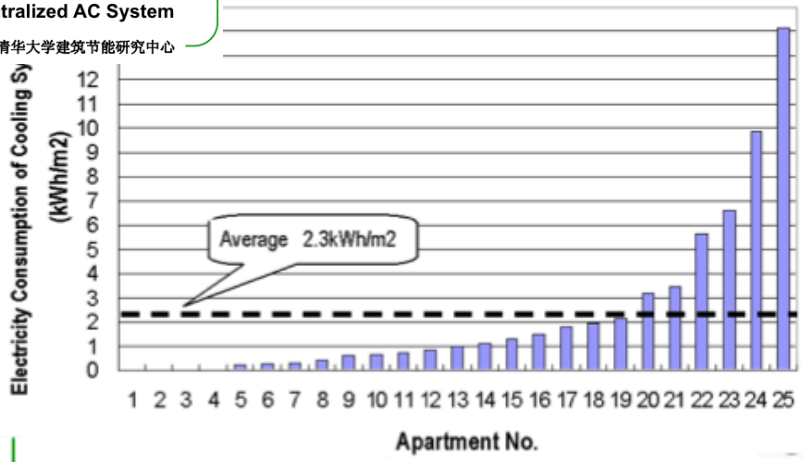


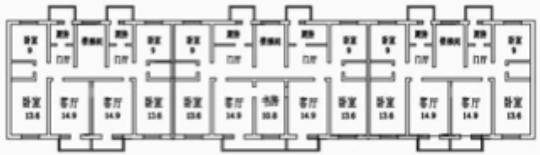
se study: AC in Residential buildings

neasured energy consumption
: in every units of a residential
ing in Beijing,2006, split unit

Key reason:
- Part time
- Part space
- Open windows







Centre AC for residential buildings in Beijing: 19.8 kWh/m²,a

Yi Jiang

Building Energy Research Centre
Tsinghua University P R China

Towards Supply-Following Loads

SUPPLIES:

provide power

communicate renewable availability, price

LOADS:

adapt demand

communicate forecast

Internet

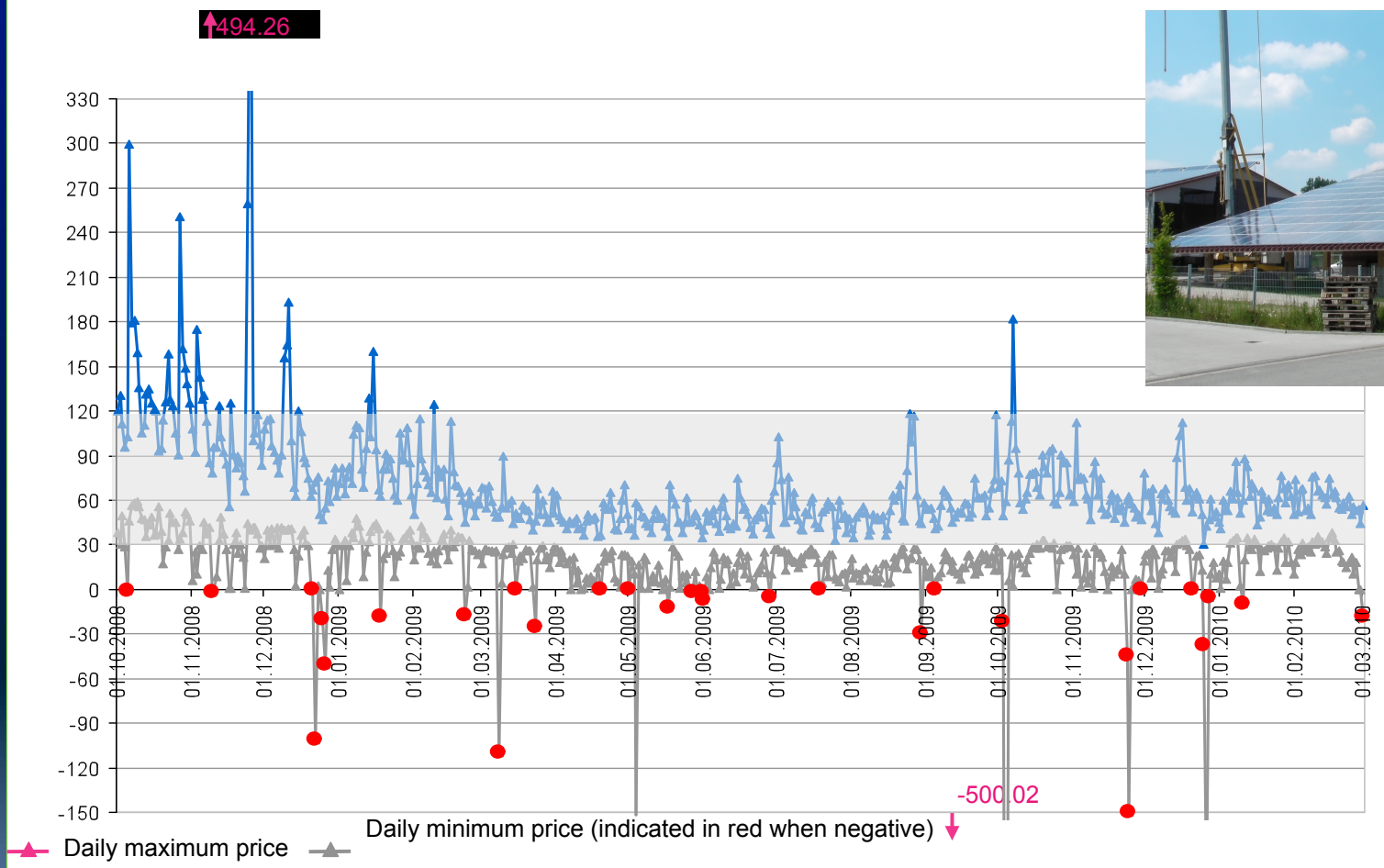
Grid

electricity

information

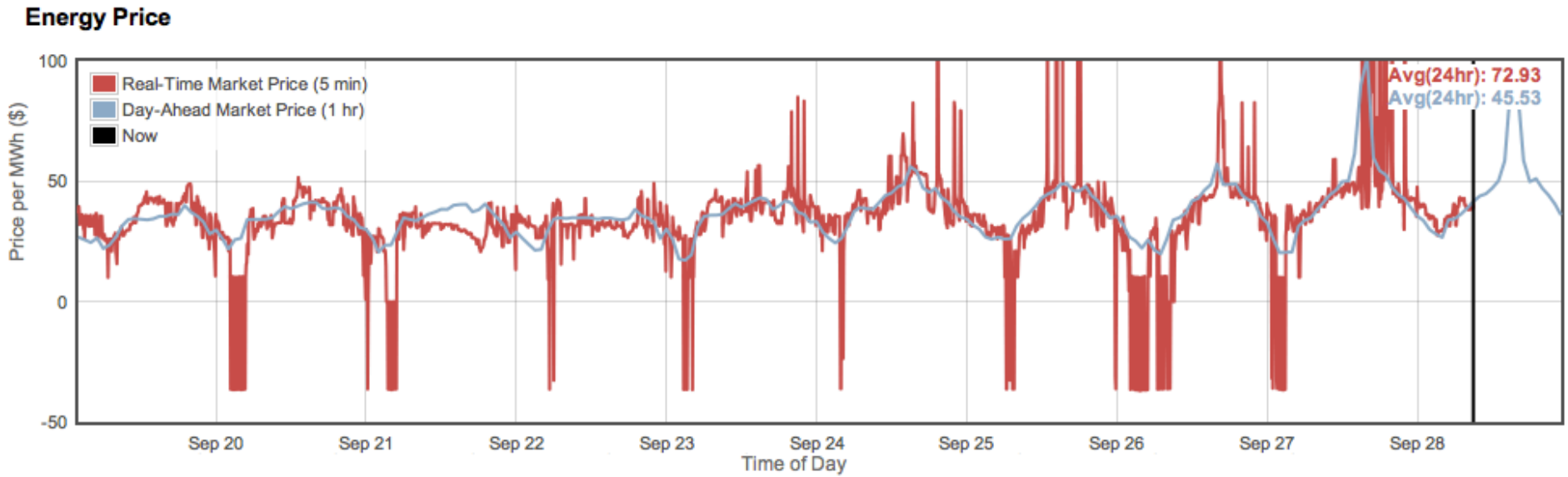
Load-following Supply (?)

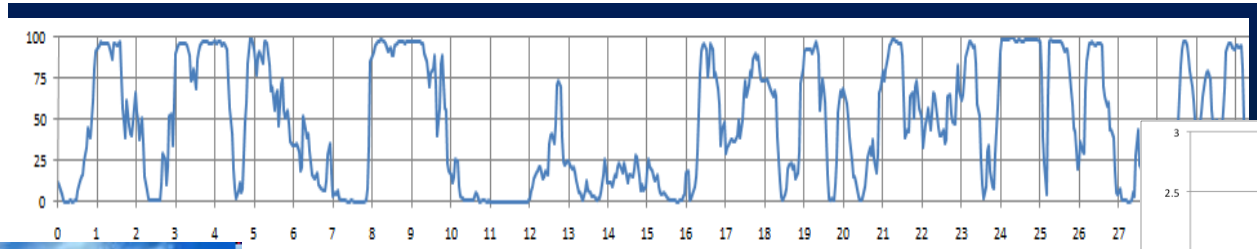
Growing proportion of renewables leads to higher price volatility. October 2008 to March 2010:
>**90 hours** with negative prices; highest price reached: +€500/MWh, lowest -€500/MWh

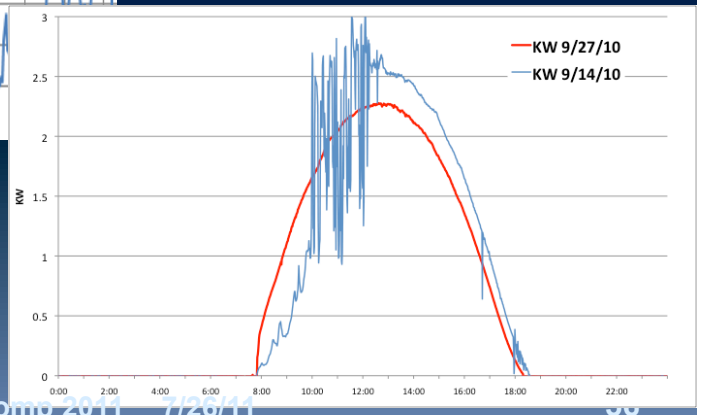


Source: EEX spot prices.

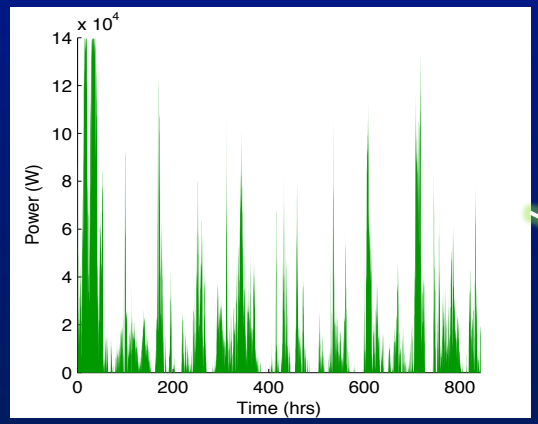
… and @ CA


Energy Price
Real-Time Market Price (5 min)
Day-Ahead Market Price (1 hr)
Now
Avg(24hr): 72.93
Avg(24hr): 45.53
Price per MWh ($)
100
50
0
-50
Sep 20
Sep 21
Sep 22
Sep 23
Sep 24
Sep 25
Sep 26
Sep 27
Sep 28
Time of Day



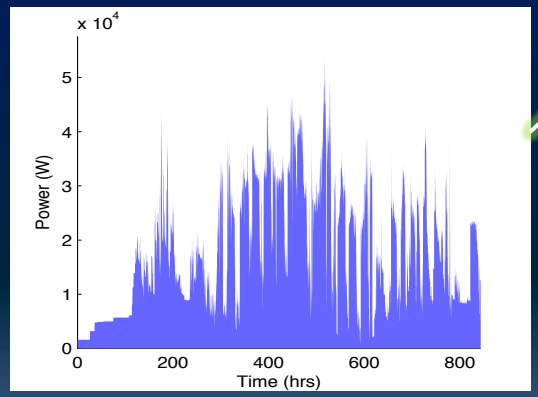

KW 9/27/10
KW 9/14/10
KW

Renewable Integration



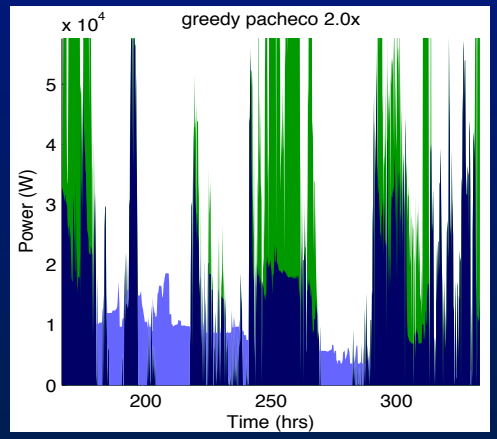
Pacheco wind farm

Non-dispatchable, variable supply



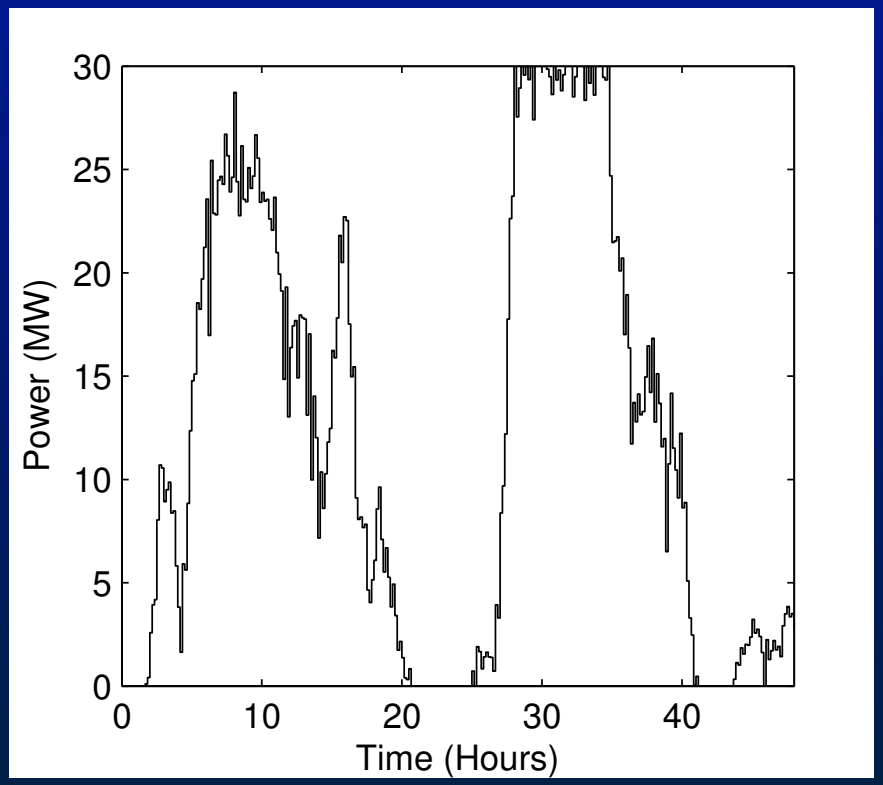
Scientific computing cluster

Power proportional, grid-aware loads

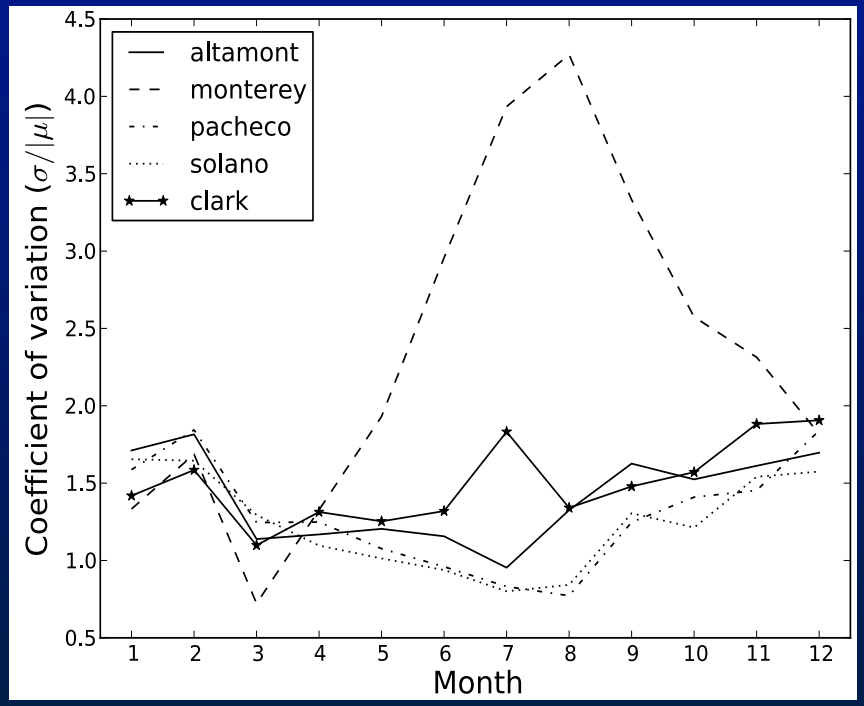


NREL Western Wind and Solar Integration Study Dataset
http://wind.nrel.gov/Web_nrel/

Wind

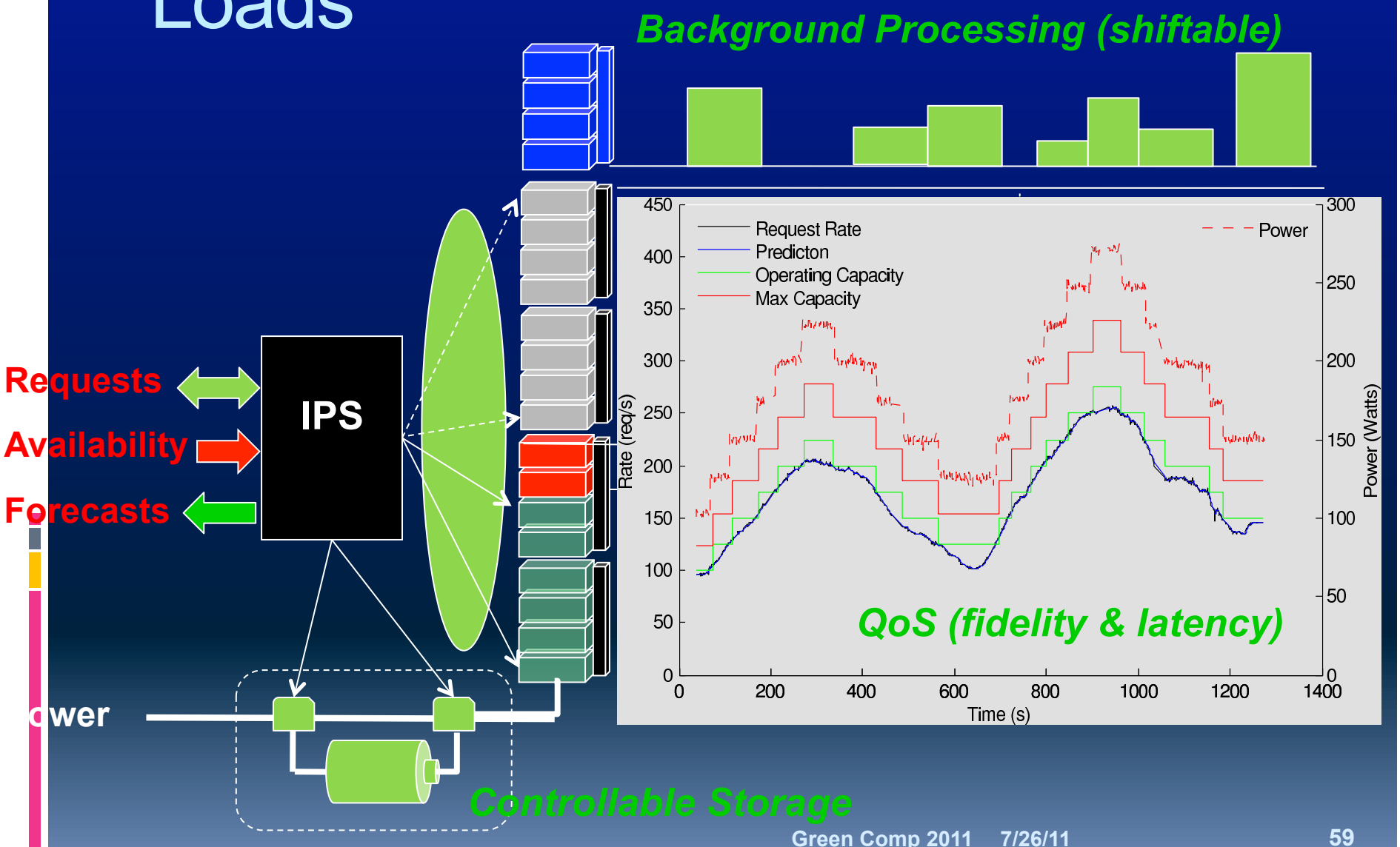


Wind power over 48 hours from a wind farm in Monterrey County, California.

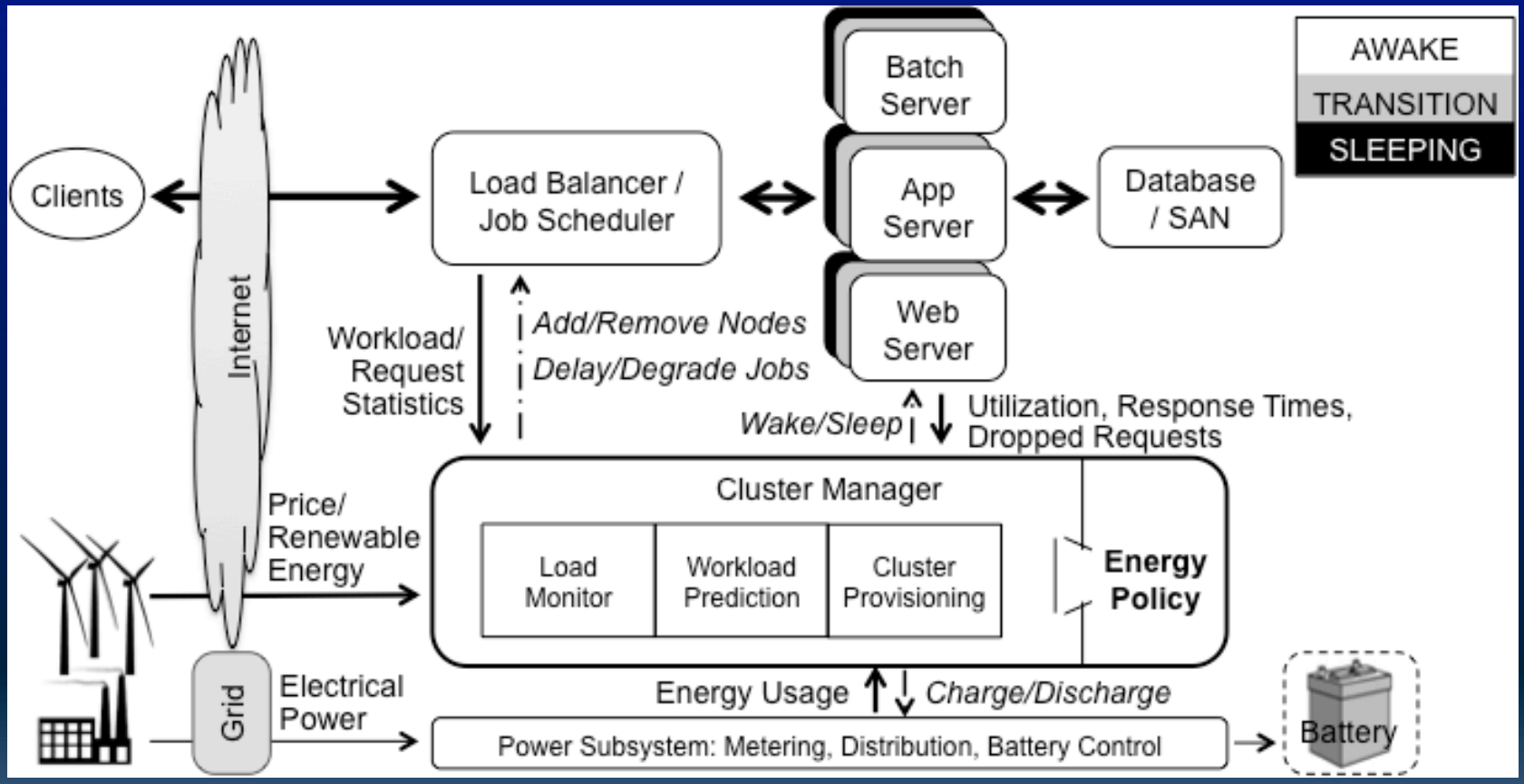


Variation in wind power for month long intervals at multiple wind farms.

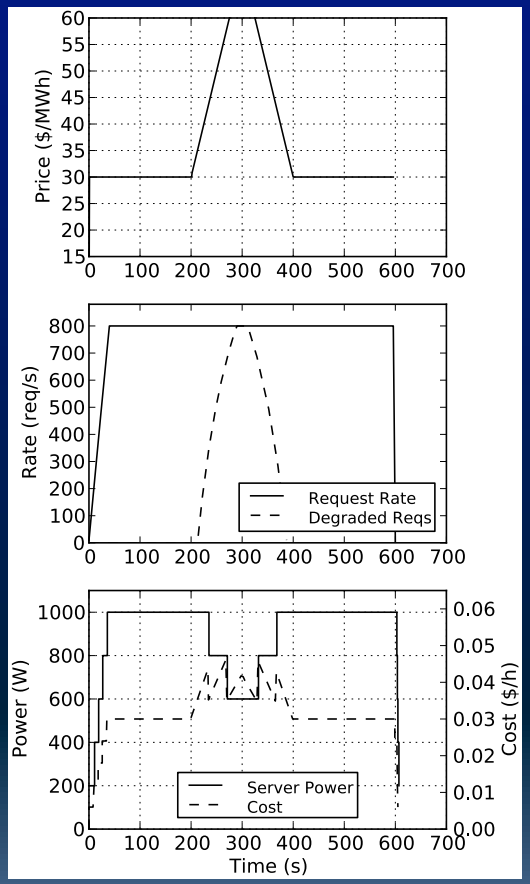
Supply-Following Computational Loads

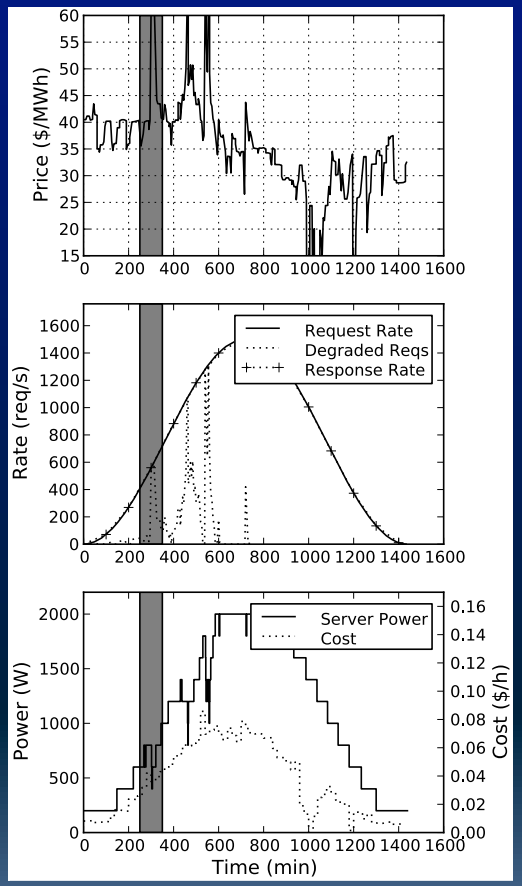


Energy-Agile Cluster Architecture



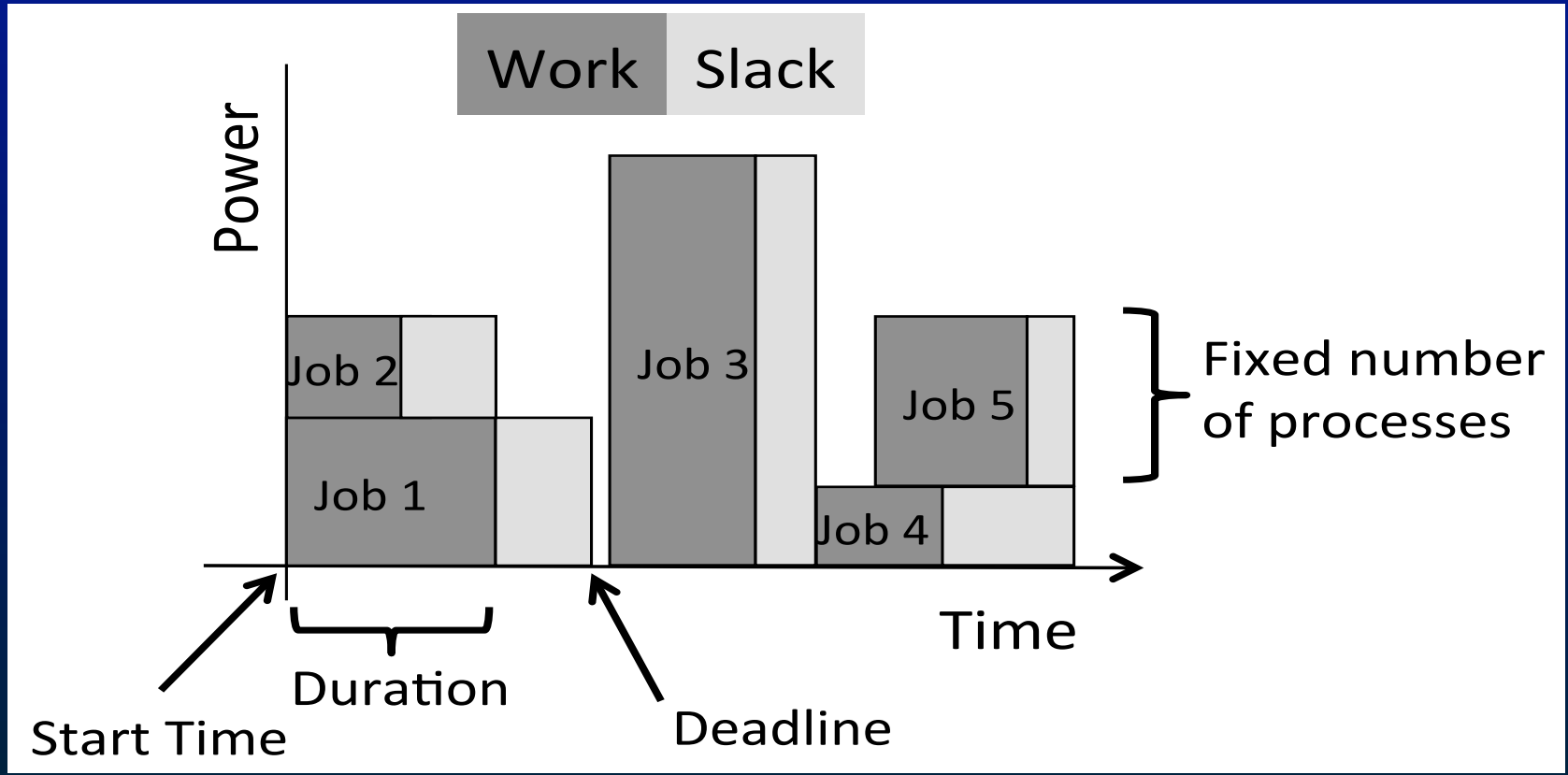
Grid-Aware Interactive Web Service



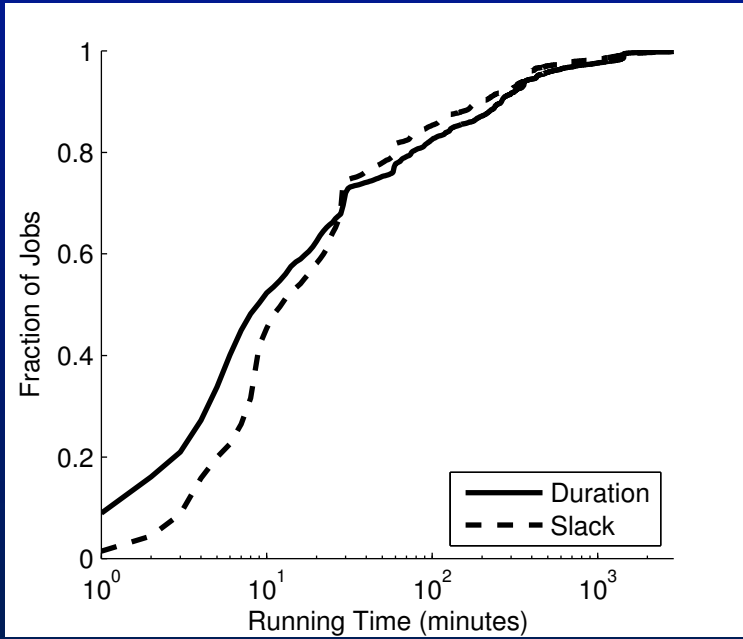


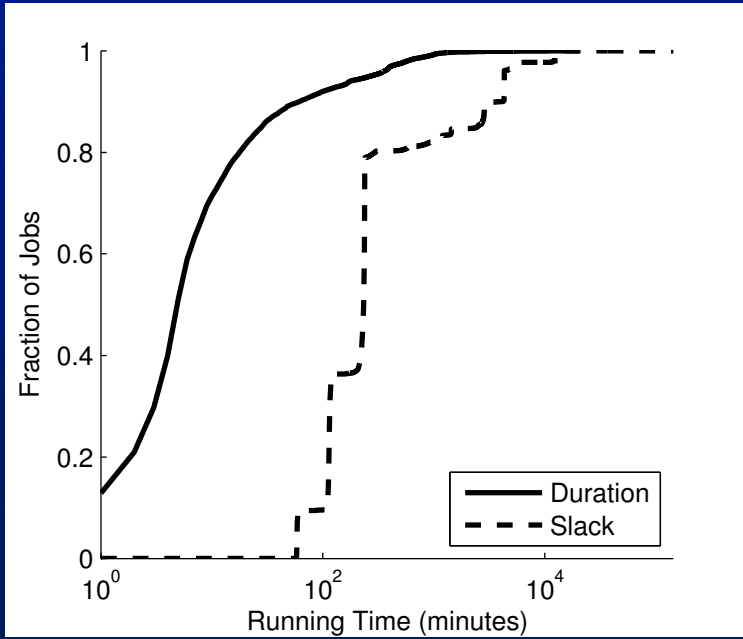
- **example goal:** maintain delivery cost despite Elect. price spikes
- **method:** distill content based on price & let PP workload mgr govern demand

Slack in Compute Load



slack = max run time – job duration

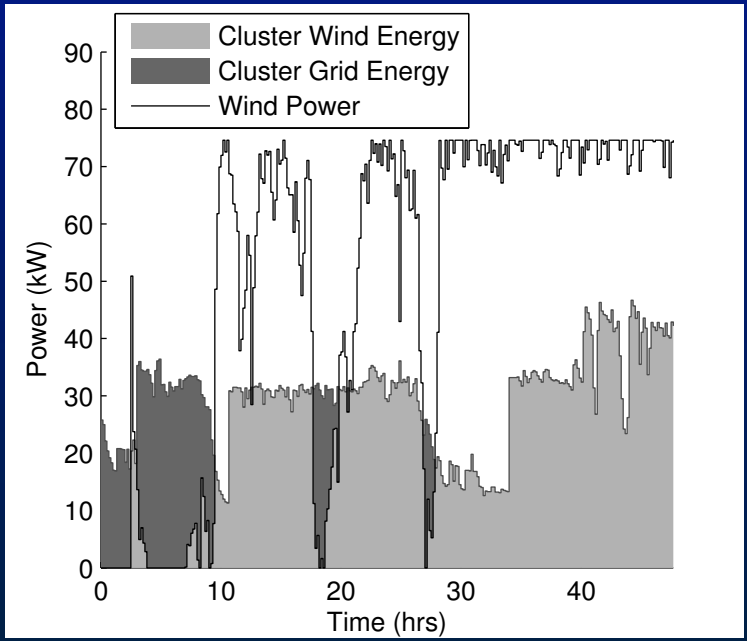




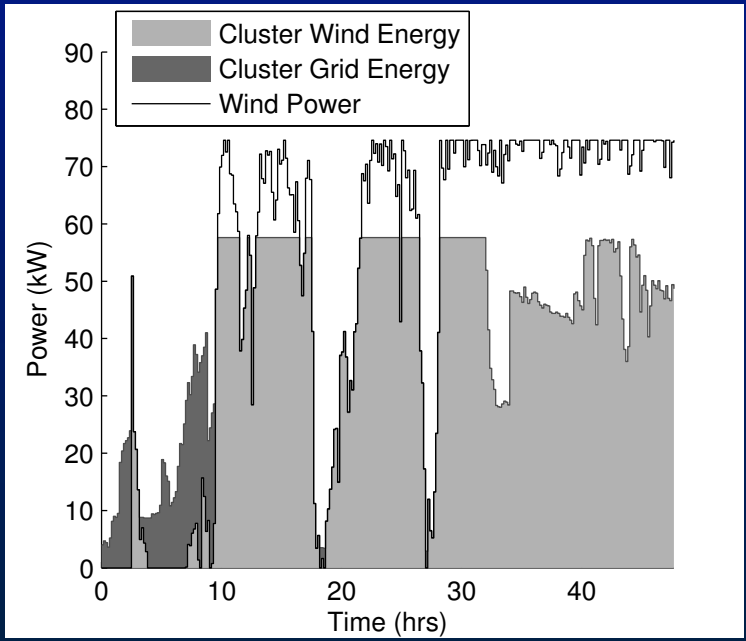
Cluster: NERSC Franklin
Average duration: 98 min
Average slack: 68 min

Cluster: EECS PSI
Average duration: 55 min
Average slack: 17 hours

Grid-Aware Scheduler



Run immediately, grid-oblivious scheduler



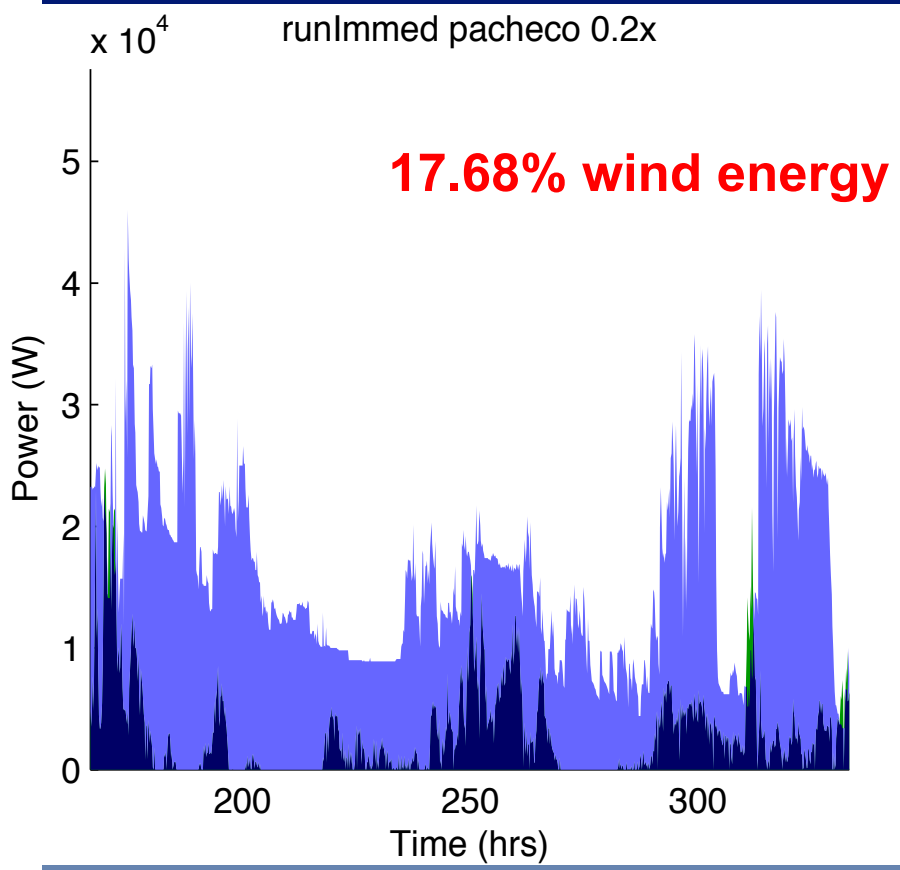
Greedy, grid-aware scheduler

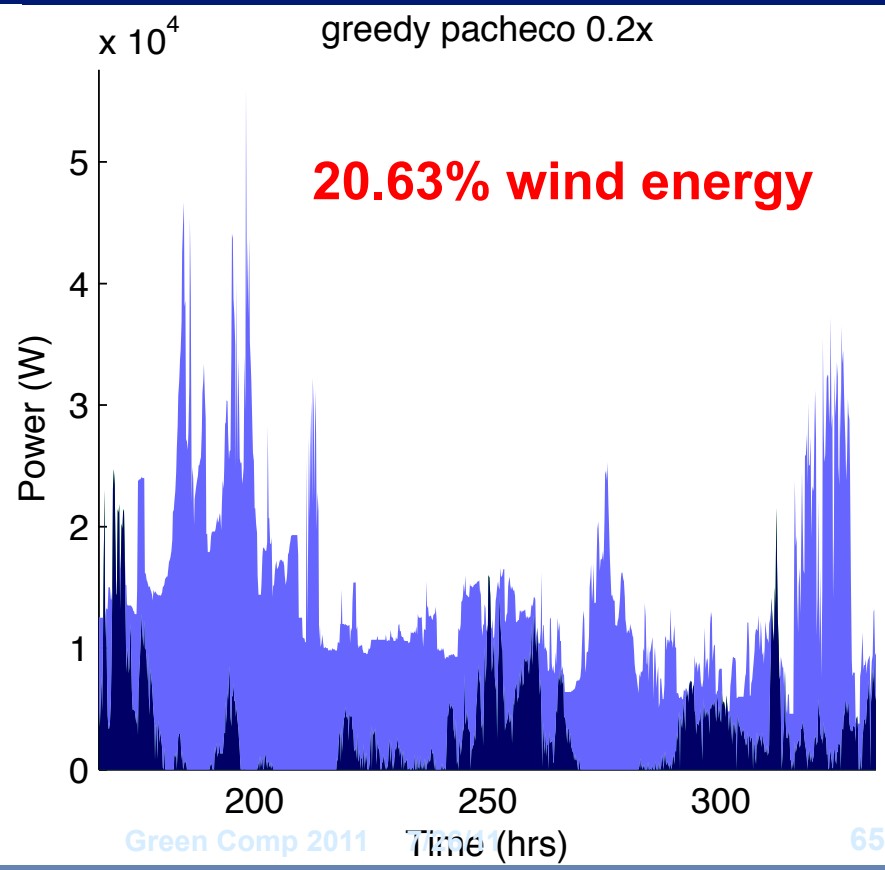
Low Renewable Penetration

Most of the available wind is used

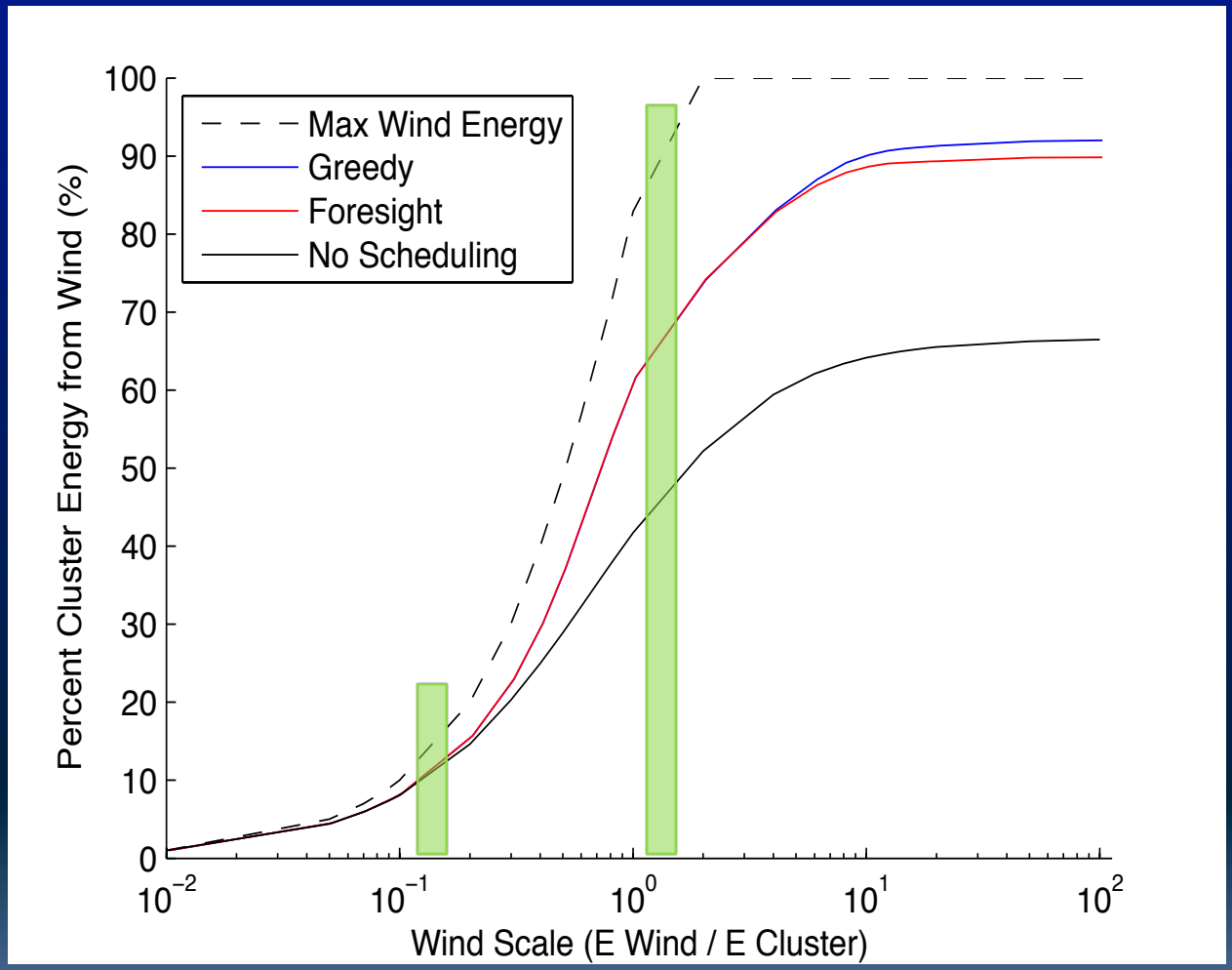
All algorithms perform comparably

Unused Wind Energy
Cluster Grid Energy
Cluster Wind Energy



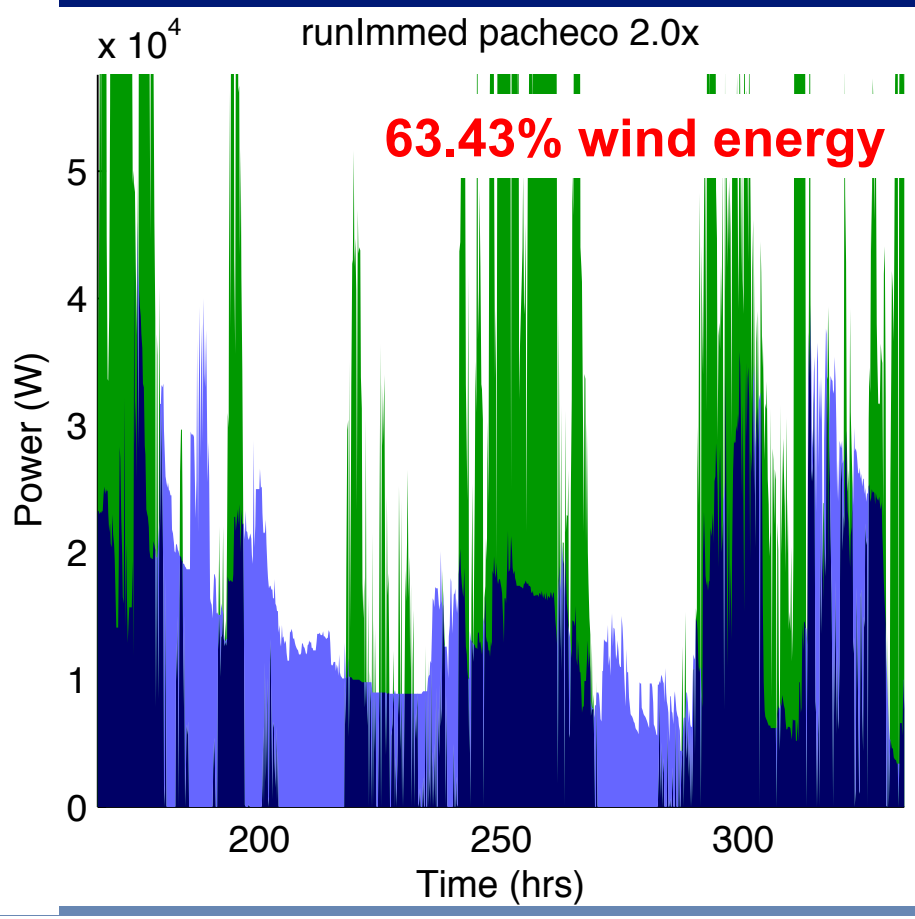


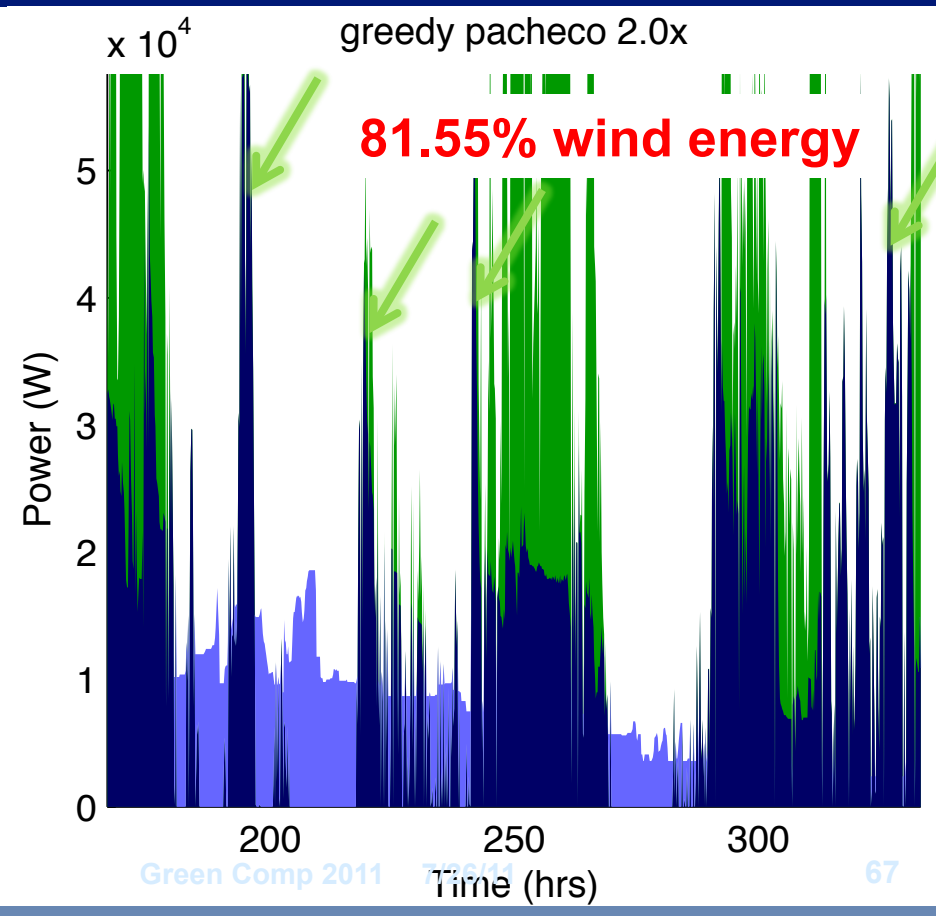
Wind Scale Impact



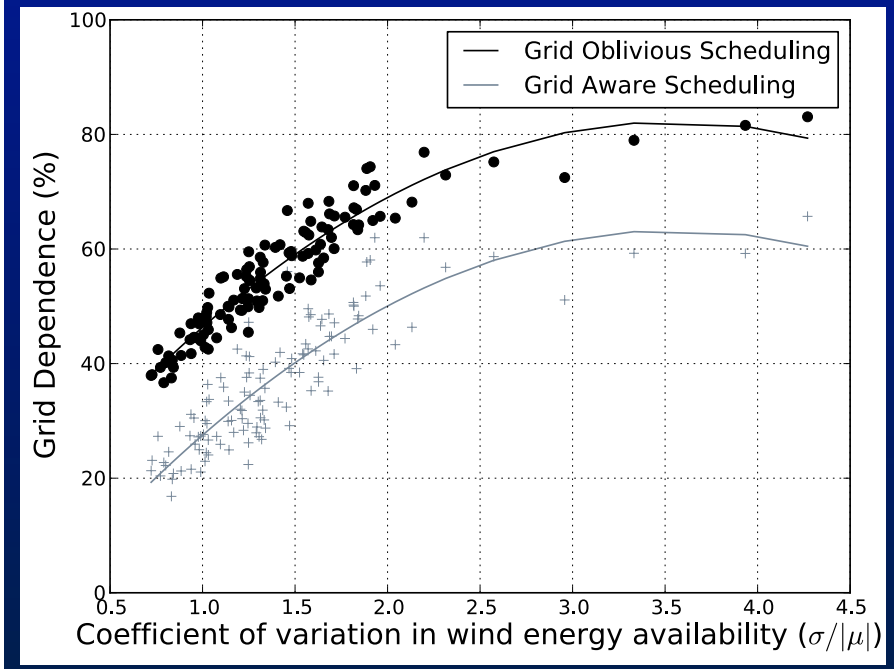
Gusts and Bursts

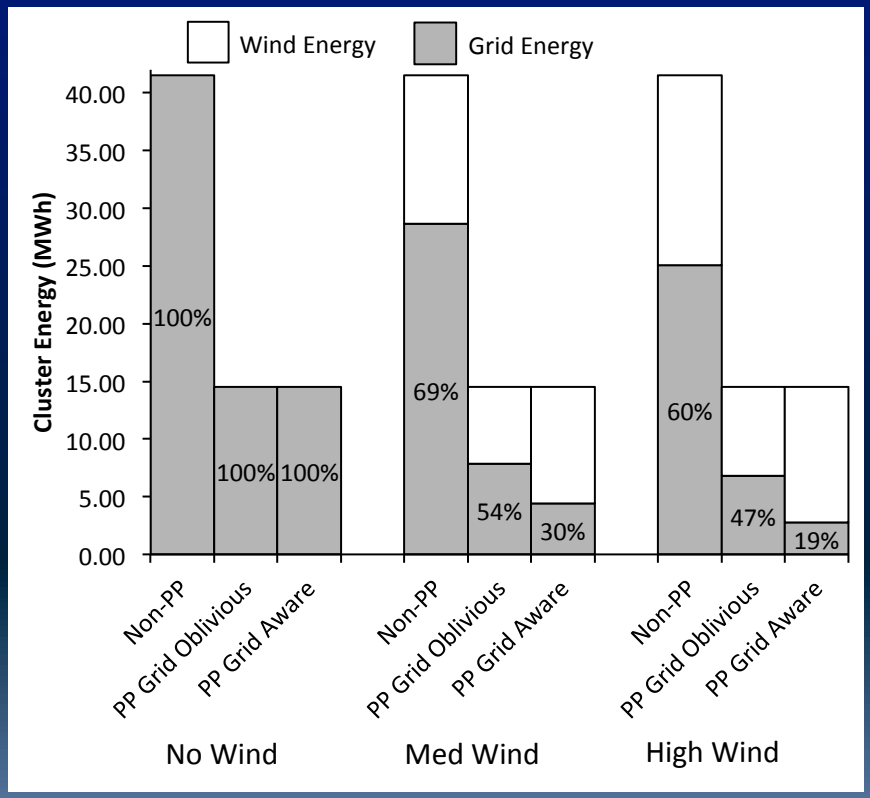
Unused Wind Energy
Cluster Grid Energy
Cluster Wind Energy



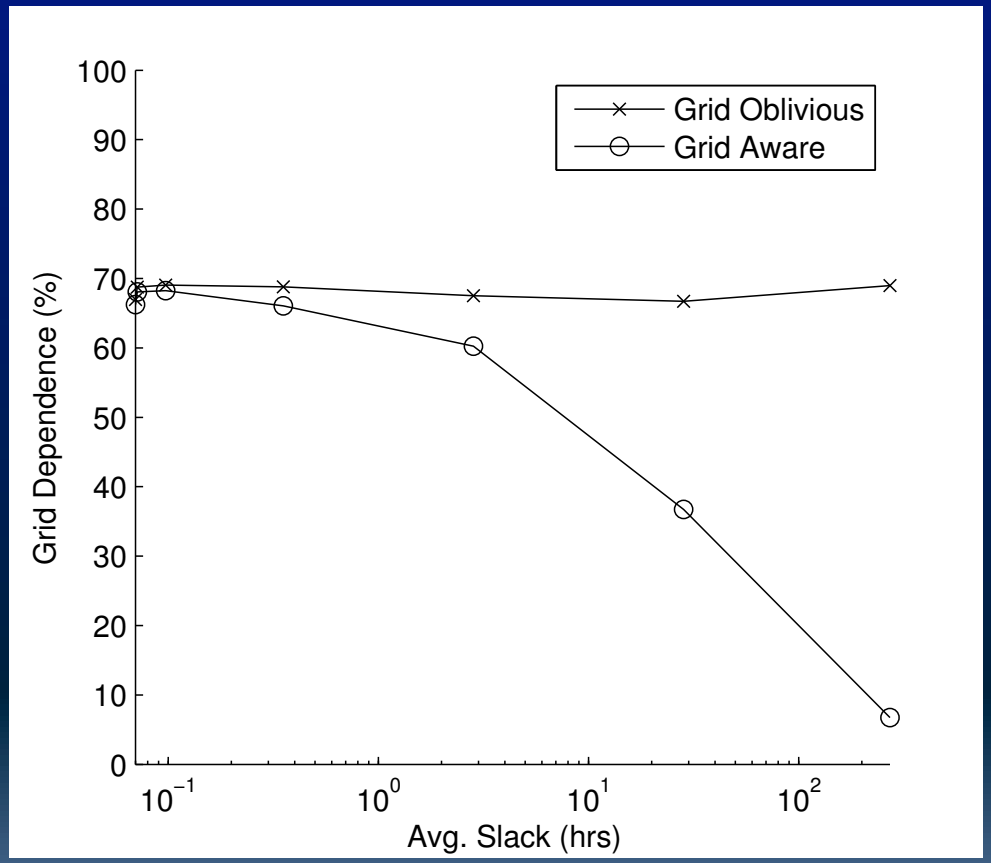


Wind Characteristics

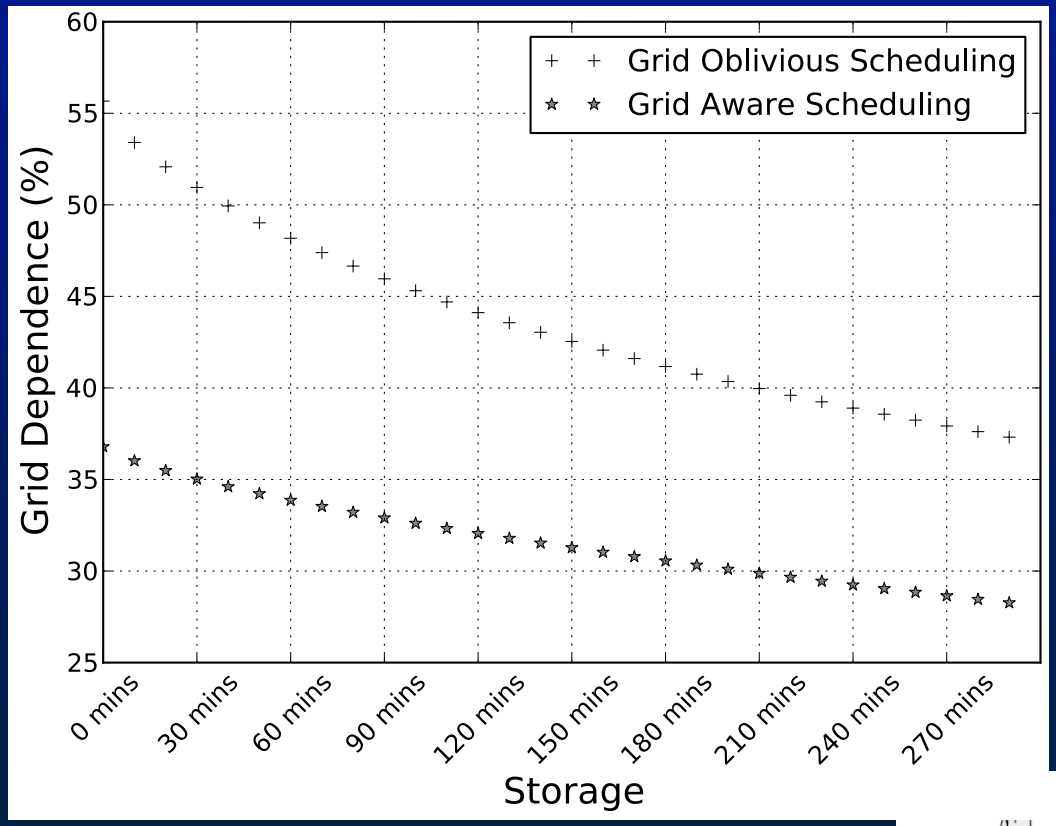


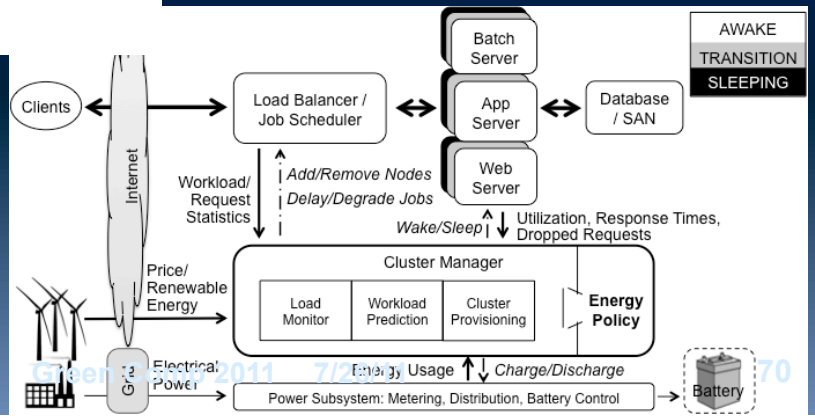


Job Characteristics



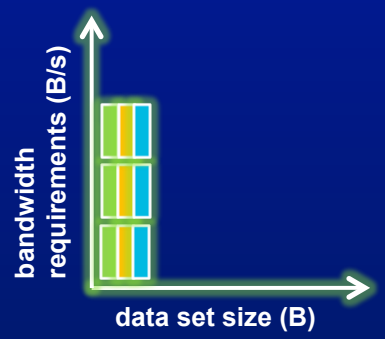
Scheduling vs Energy Storage

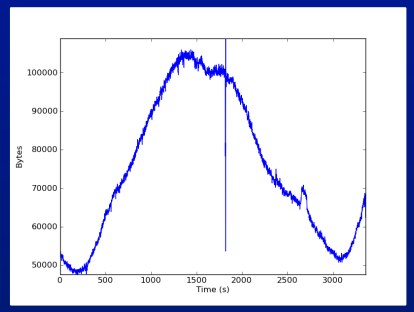




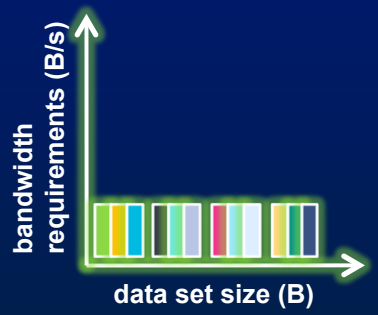
Data Storage?

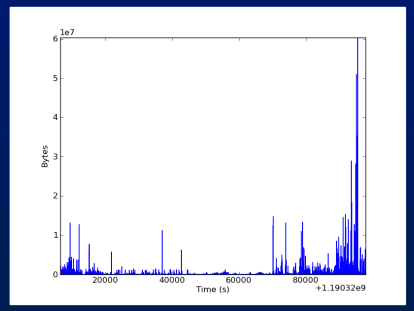
Why Replication?



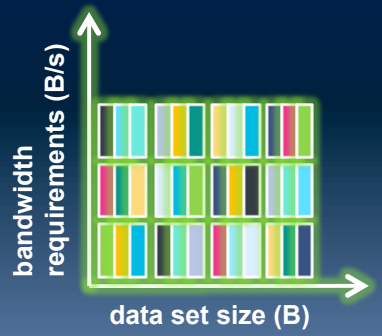


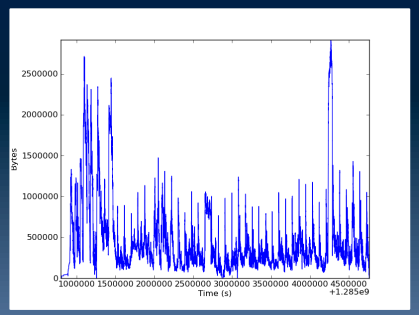
web farms (Wikipedia)





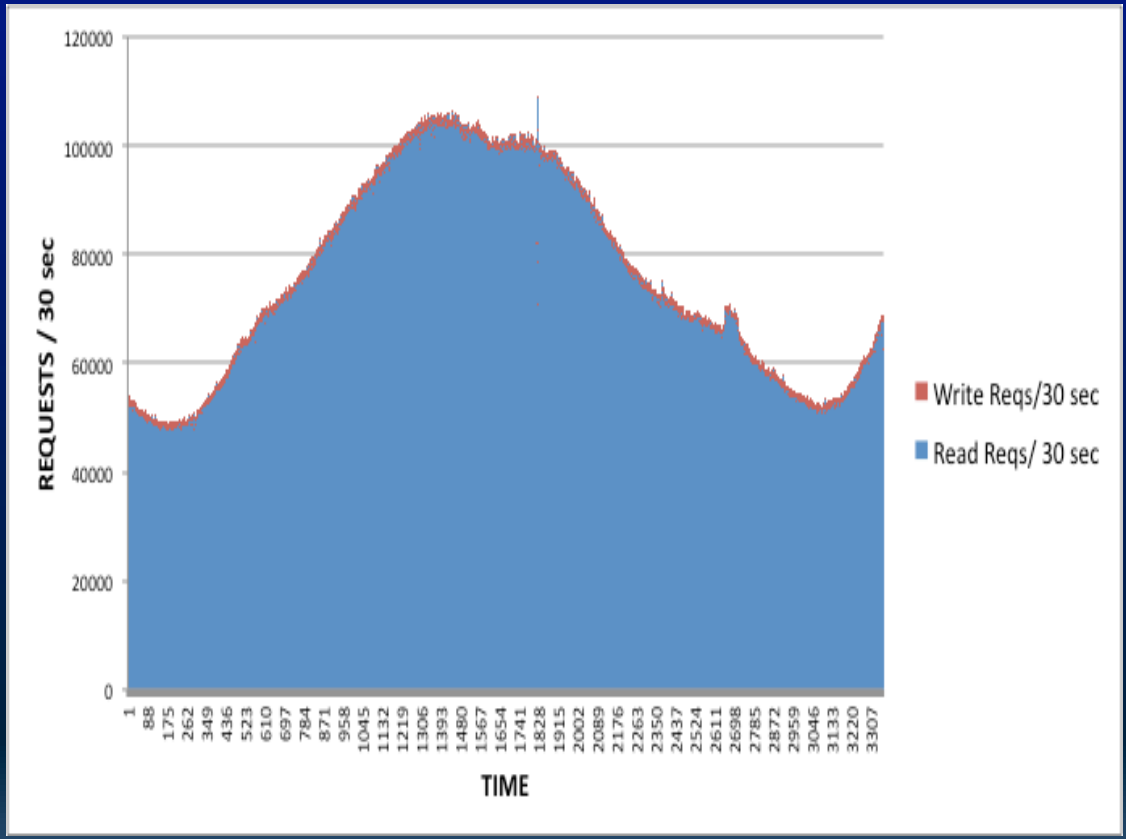
NAS / NFS (NetApp), disk arrays





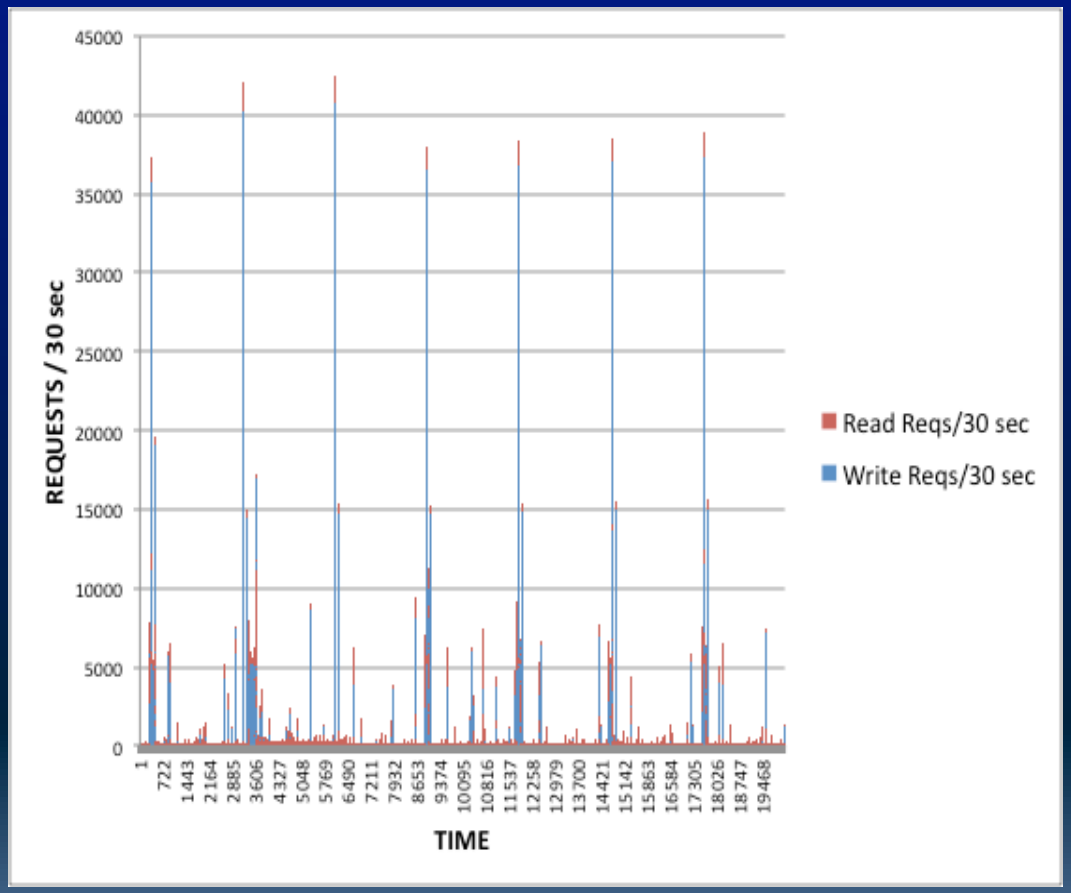
data analytics, DFS (Hadoop)

Wikipedia 1-day trace



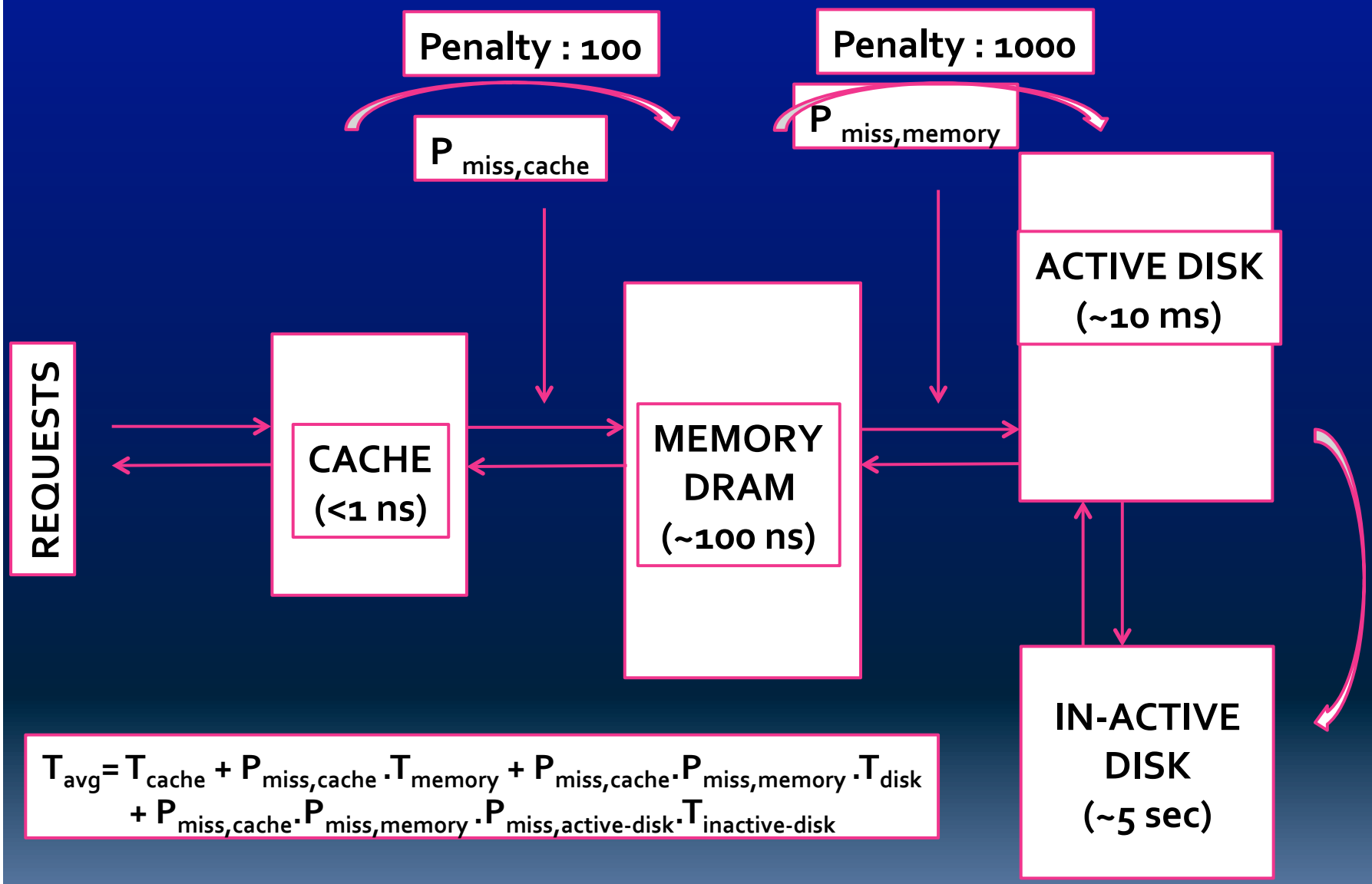
- **Replicated thrice**
- **Partitioned over 5 disks**

MSR-Cambridge Web/SQL server 7 day trace



- **Not replicated**
- **4 volumes**
- **10 disks each**

Sleepy Hierarchy



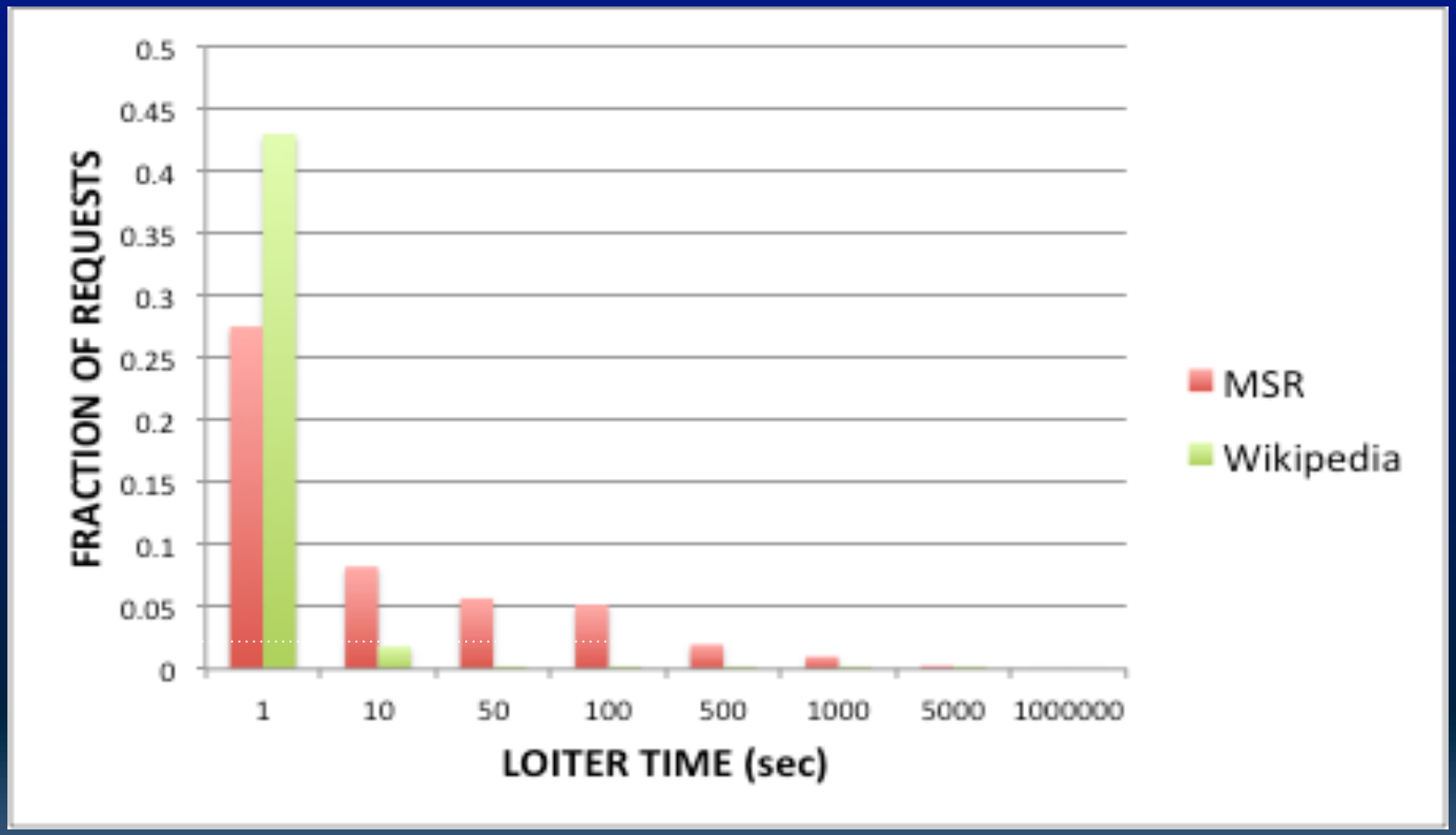
$$T_{avg} = T_{cache} + P_{miss,cache} . T_{memory} + P_{miss,cache} . P_{miss,memory} . T_{disk} + P_{miss,cache} . P_{miss,memory} . P_{miss,active\text{-}disk} . T_{inactive\text{-}disk}$$

Techniques

1. Put disks to sleep after a loiter time , and wake-up on demand
2. Caching
3. Prefetching disk blocks
4. Dynamic object creation/Intelligent layout

- $T_{avg} \approx T_{active\text{-}disk} + P_{miss,active\text{-}disk} \cdot T_{non\text{-}active\text{-}disk}$

Active-disk misses



Stages of Energy Effectiveness

- Waste Not
  - Do Nothing Well !!!
- Power Proportionality
  - Peak Performance : Power => Safety
  - Optimize Partial Load - from nothing to peakl
- Sculpting
  - Identify the energy *slack* and utilize it
- Negotiated Grid / Load / Human Interaction
  - Plan, Forecast, Negotiate, Manage

Why we don't do it today…

- … didn't buy 'em to turn 'em off
- … only a small fraction of the cost of the data center is the utility bill
- … rather raise the utilization by co-locating instances
- … in rare cases may see some performance degradation
- … might not wake up reliably (wake-on-lan)
- … a sleeping machine might die on wake-up
- … don't get paid to serve to do nothing.

The Common Case

Doing Nothing

- "Perhaps the most important and pervasive principle of computer design is to make the common case fast: In making a design trade-off, favor the frequent case over the infrequent case."

green

– John L. Hennessy and David A. Patterson, Computer Architecture – a Quantitative Approach

☑ Do Nothing Well

Thanks