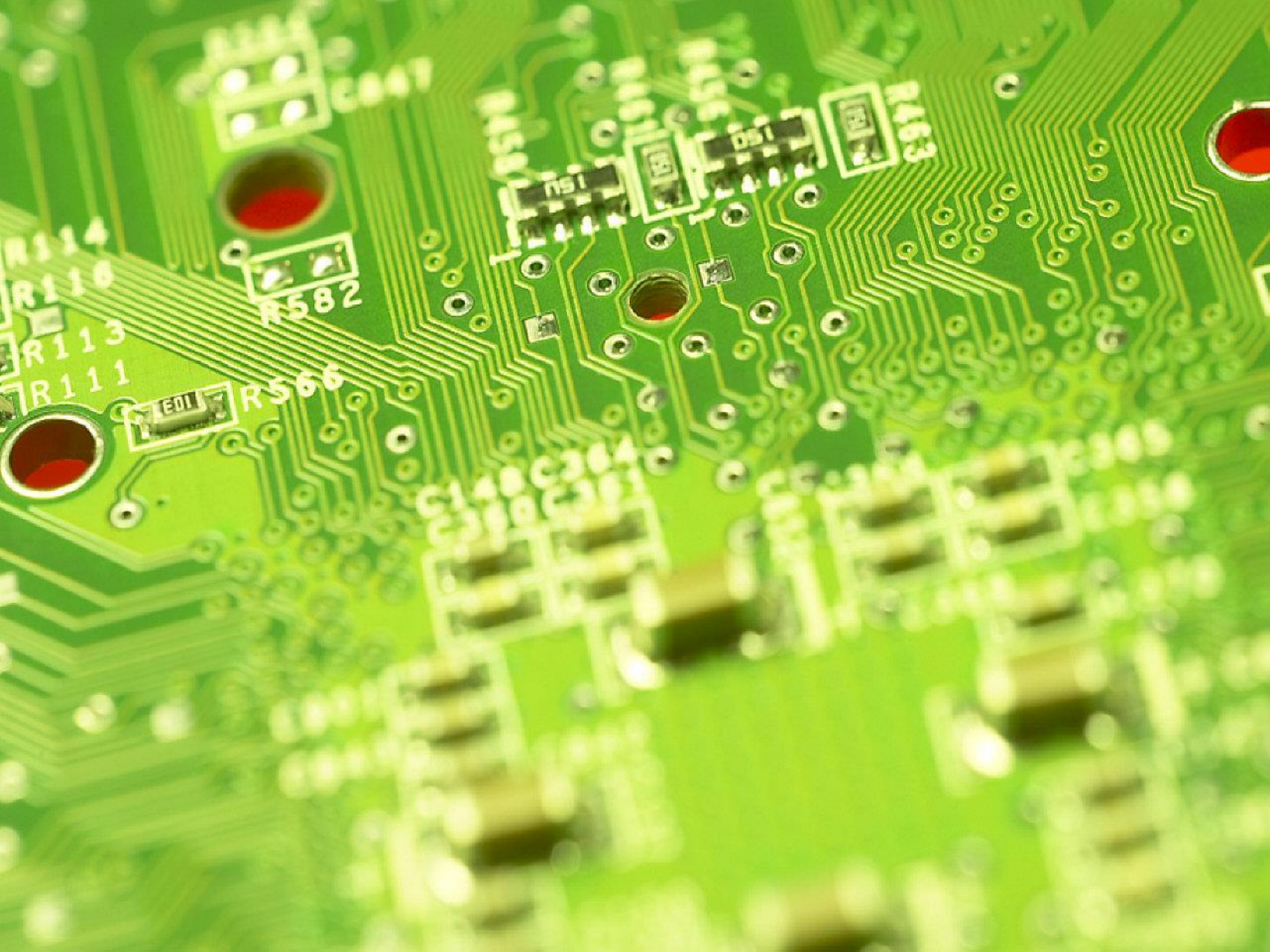

R114
R116
R113
R111
R582
R566
D51
R463

11th Charm++ Workshop:

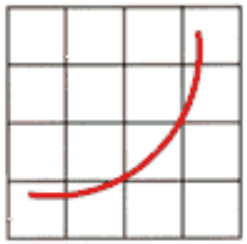
Power-performance modeling, analyses and challenges

Kirk W. Cameron
Computer Science
Virginia Tech

This material is based upon work supported by the National Science Foundation under Grant No. 0910784 and 0905187.

My Green HPC Upbringings

- Over $6M related federal funding (since ‘04) (NSF, DOE, SBIR, IBM, Intel, and others)
- EPA Energy Star for servers (since ‘05)
- SPECPower Founding Member (since ‘05)
- Co-founder Green500 (since ‘06)
- Green IT Columnist (*IEEE Computer*)
- CEO and Founder, MiserWare Inc. (since ‘07)

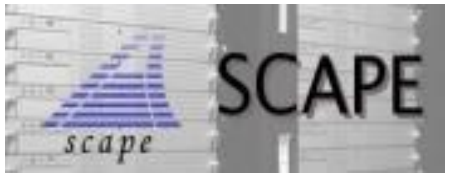


spec

MiserWare

Saving Energy, Saving Money, Saving the World.

THE GREEN 500



The way we were (circa 2003)

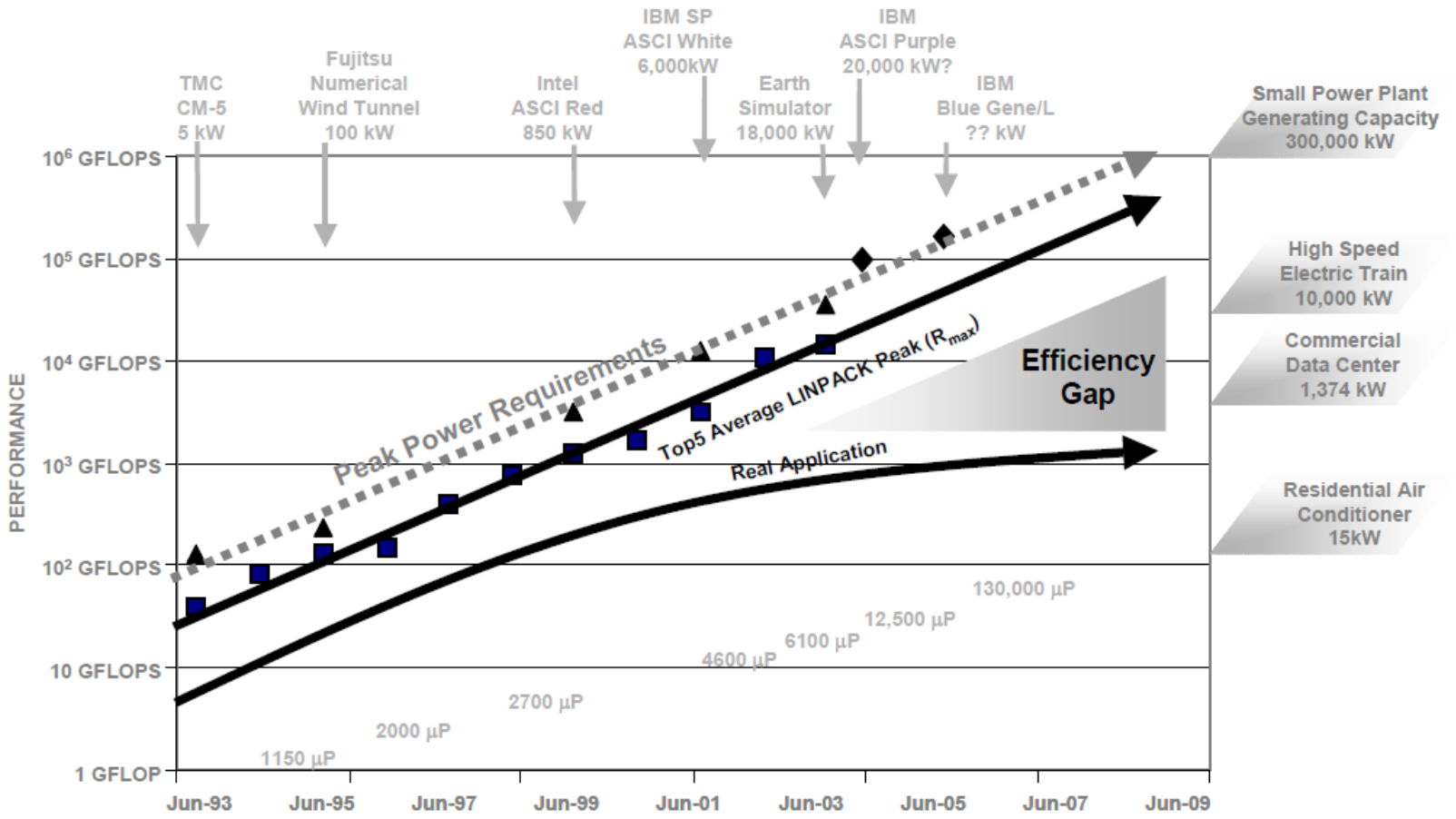


Fig. 1 Power-performance trends in the supercomputer industry. The computational demands of scientific applications have led to exponential increases in peak system performance (shown as average of peak LINPACK measurements), system power consumption (shown for several supercomputers), and

Source: CAREER: High-performance, Power-aware Computing
K. Cameron, NSF CCF-0347683, 3/1/04-2/28/09)

Getting there…

From 2007-2012…

[6x ↑ Flops/watt]

[~2.5x ↑ power consumption]

Projections for 2012-2019…

[2100 to ~15,000 MFlops/Watt]

[66 kW for 1 Petaflop System]

[66 MW for 1 Exaflop System}

[Need 50,000 Mflops/Watt for 1 Exaflop @ 20 MW by 2019!!!]

Conclusion: We need help.

What do we need…?

Insight

Where does energy go?

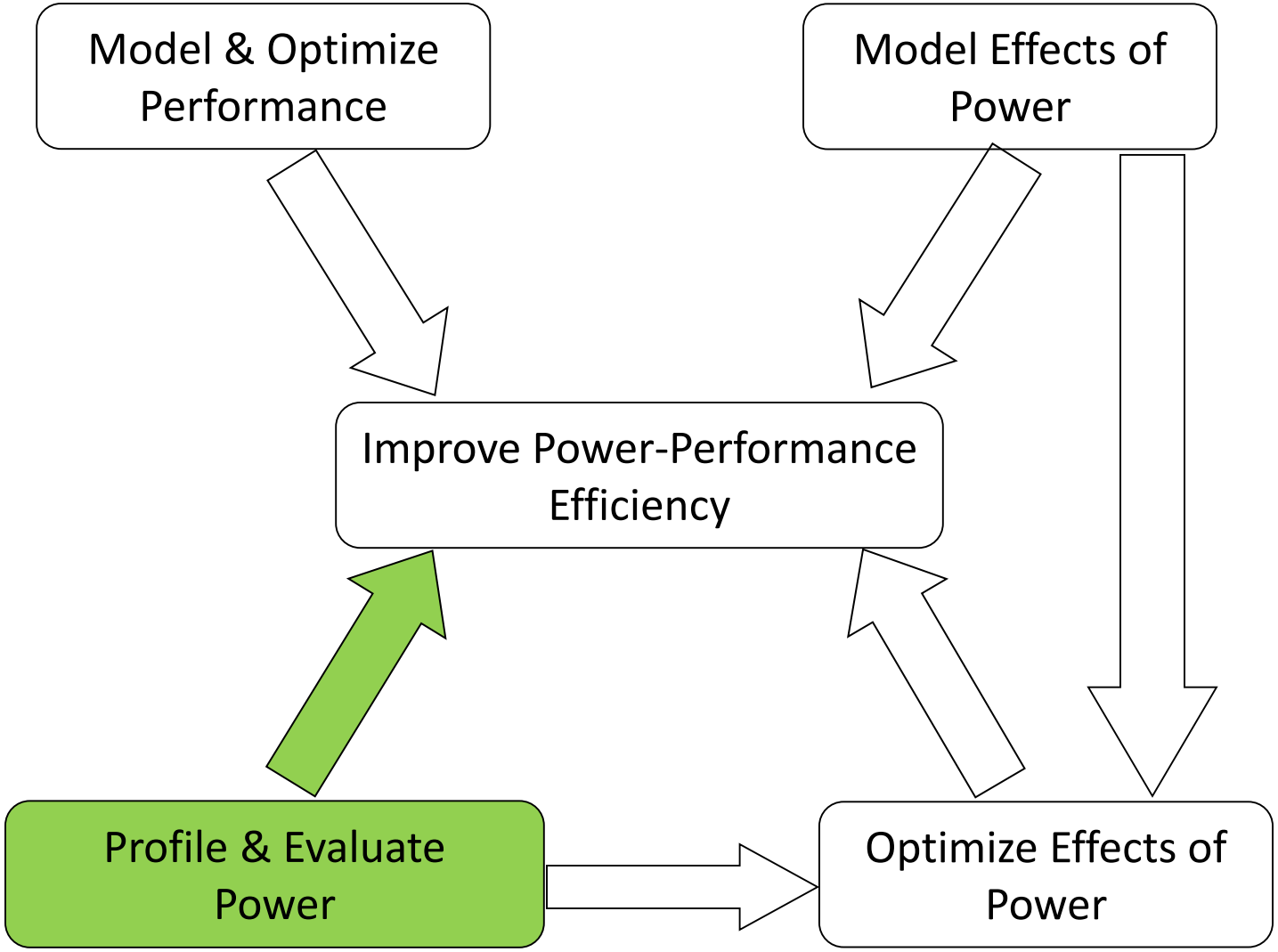
Understanding

How does energy scale?

Action

What can we do?

Power-Performance Efficiency,



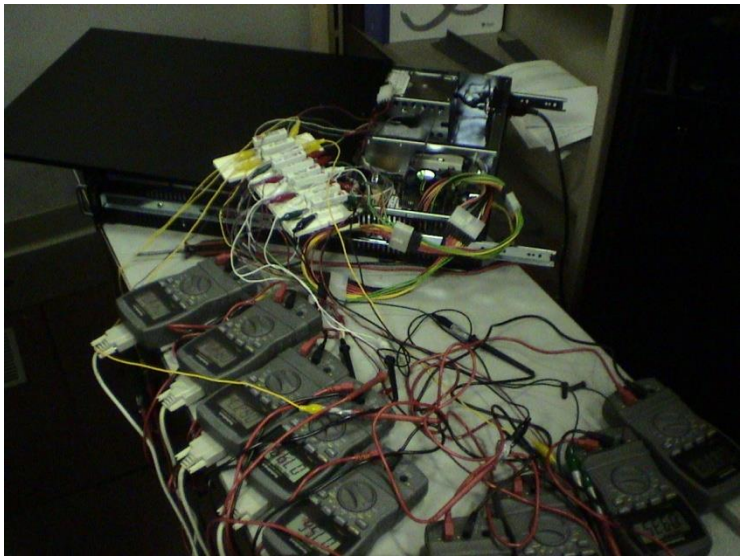
[SC04], [SC05], [IPDPS 2005], [IJHPCA 2009], [TPDS 2010]

How can we…help you…help us…

“You can only manage what you can measure.”

Peter Drucker, writer

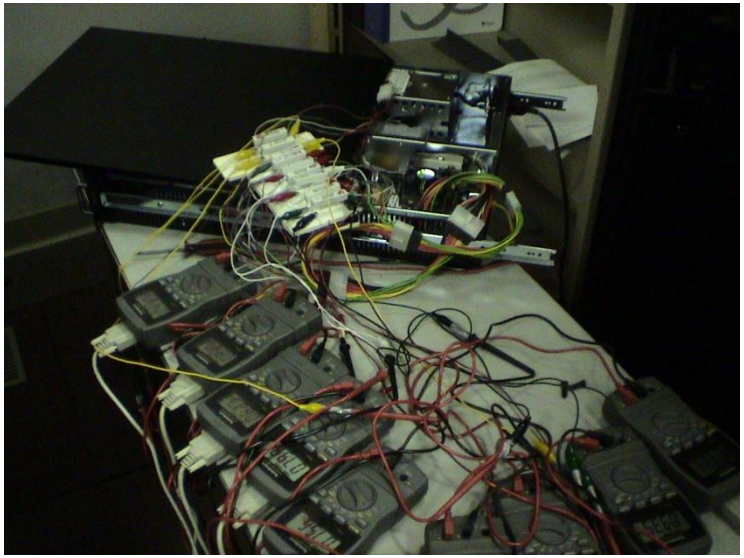
Measuring power is “tough”



What is PowerPack?

[IEEE Computer 38(11) 2005, TPDS 21(5) 2010, http://scape.cs.vt.edu/software/]

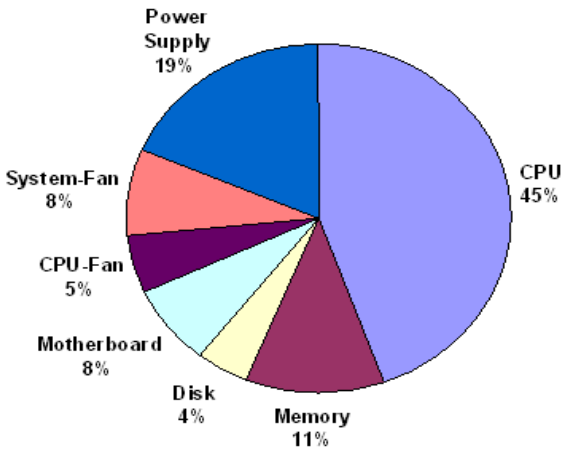
- Modularized measurement software
- HW sensors (component, room, etc.)
- Fine-grain API (function-level)
- Analytics



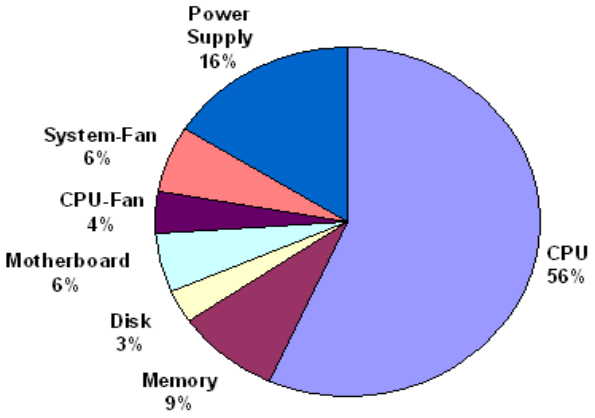
```
If node .eq. root then
        call pmeter_init (xmhost,xmport)
        call pmeter_log (pmlog,NEW_LOG)
endif

<CODE SEGMENT>

If node .eq. root then
        call pmeter_start_session(pm_label)
endif

<CODE SEGMENT>

If node .eq. root then
        call pmeter_pause()
        call pmeter_log(pmlog,CLOSE_LOG)
        call pmeter_finalize()
endif
```

SystemG Supercomputer

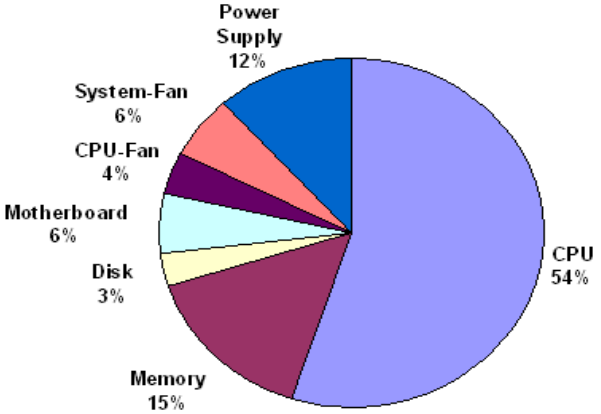
Power Profiles – Single Node



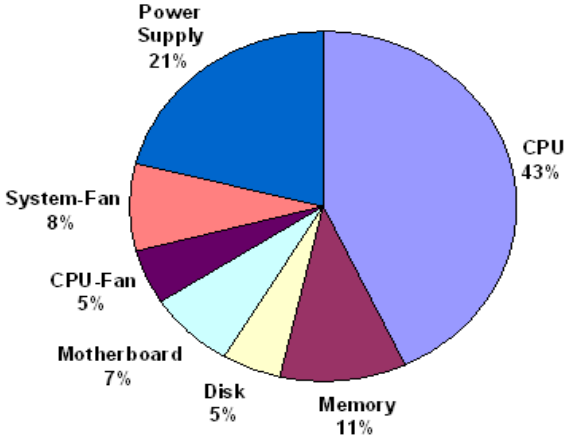
(a) Power distribution for **system idle**: system power 152.5 Watts



(b) Power distribution for **164.gzip**: system power 206.5 Watts



(c) Power distribution for **171.swim** : system power 209.2 Watts

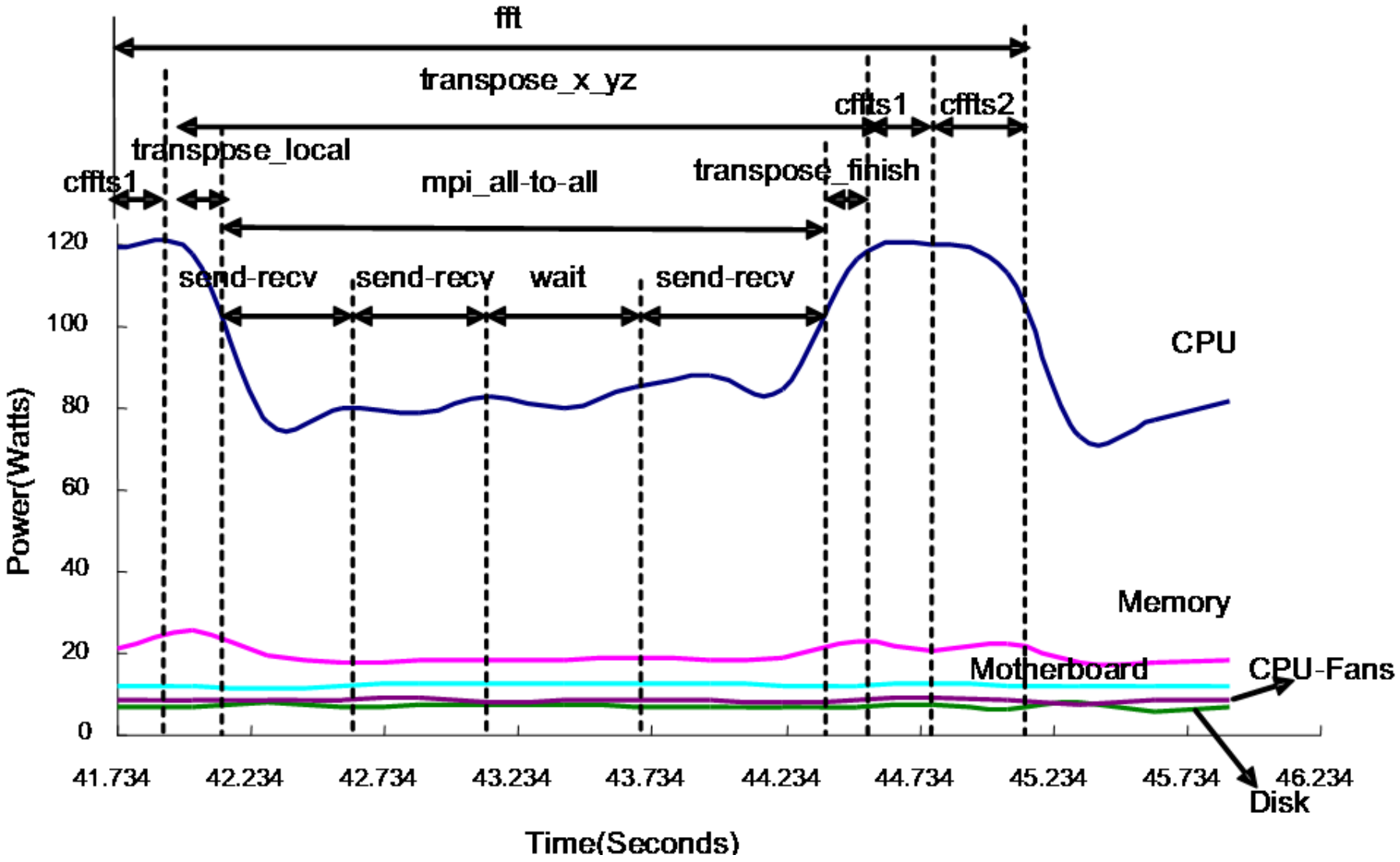


(d) Power distribution for **cp**: system power 165.2 Watts

Fig. 5. Power distribution for a single node under different workloads: (a) zero workload (system is in idle state); (b) CPU bounded workload; (c) memory bounded workload; (d) disk bounded workload.

PowerPack Function-level Profiling

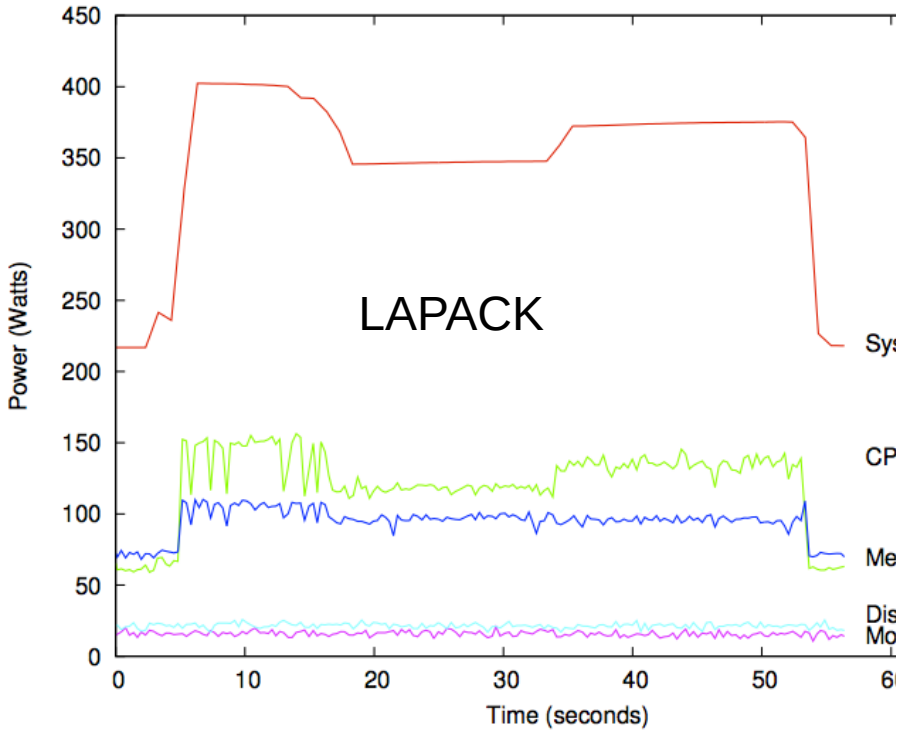
[IEEE Computer 38(11) 2005, TPDS 21(5) 2010, http://scape.cs.vt.edu/software/]

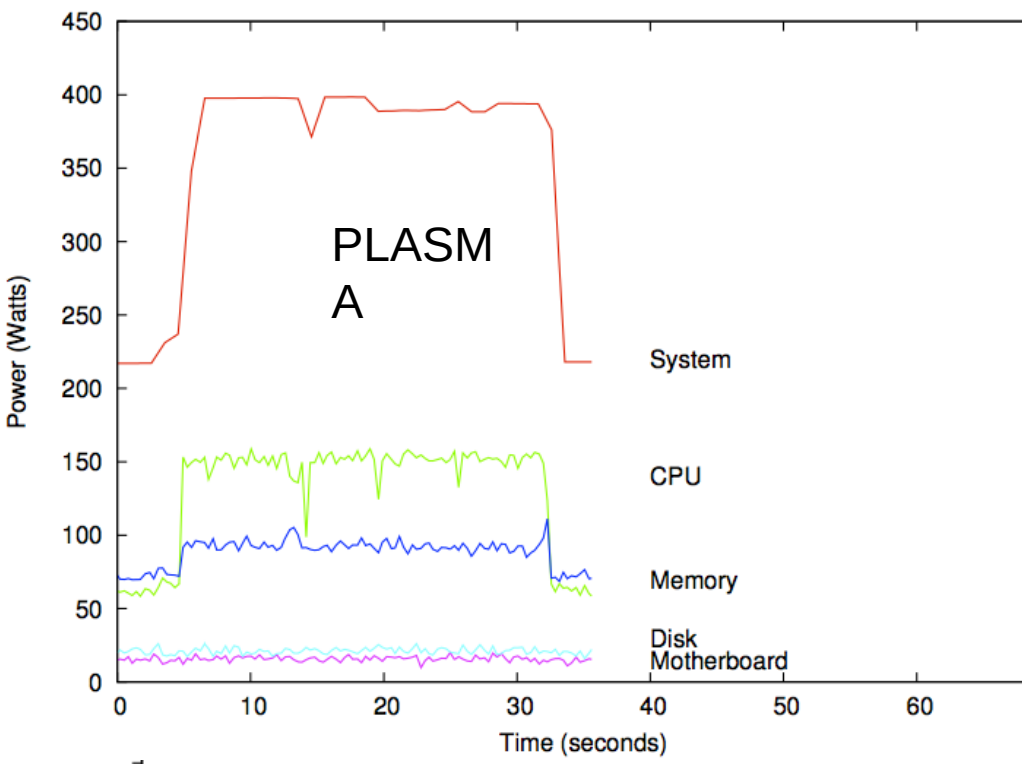


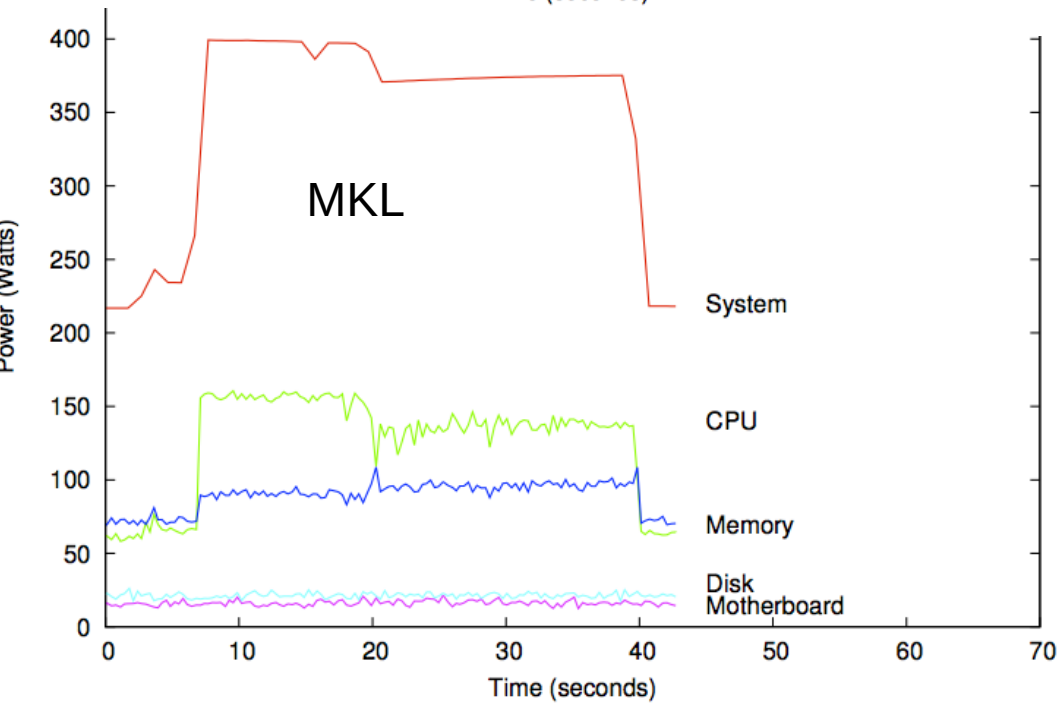
Who uses PowerPack? SystemG?

- Texas A&M (Taylor et al)
- UTenn-Knoxville (Moore, Dongarra, et al)
- Oxford University
- Lawrence Livermore National Lab
- Pacific Northwest National Lab
- Oak Ridge National Lab
- University of Florida
- KAUST (Saudi Arabia)
- University of Madrid (Spain)
- UC Berkeley

...and many others







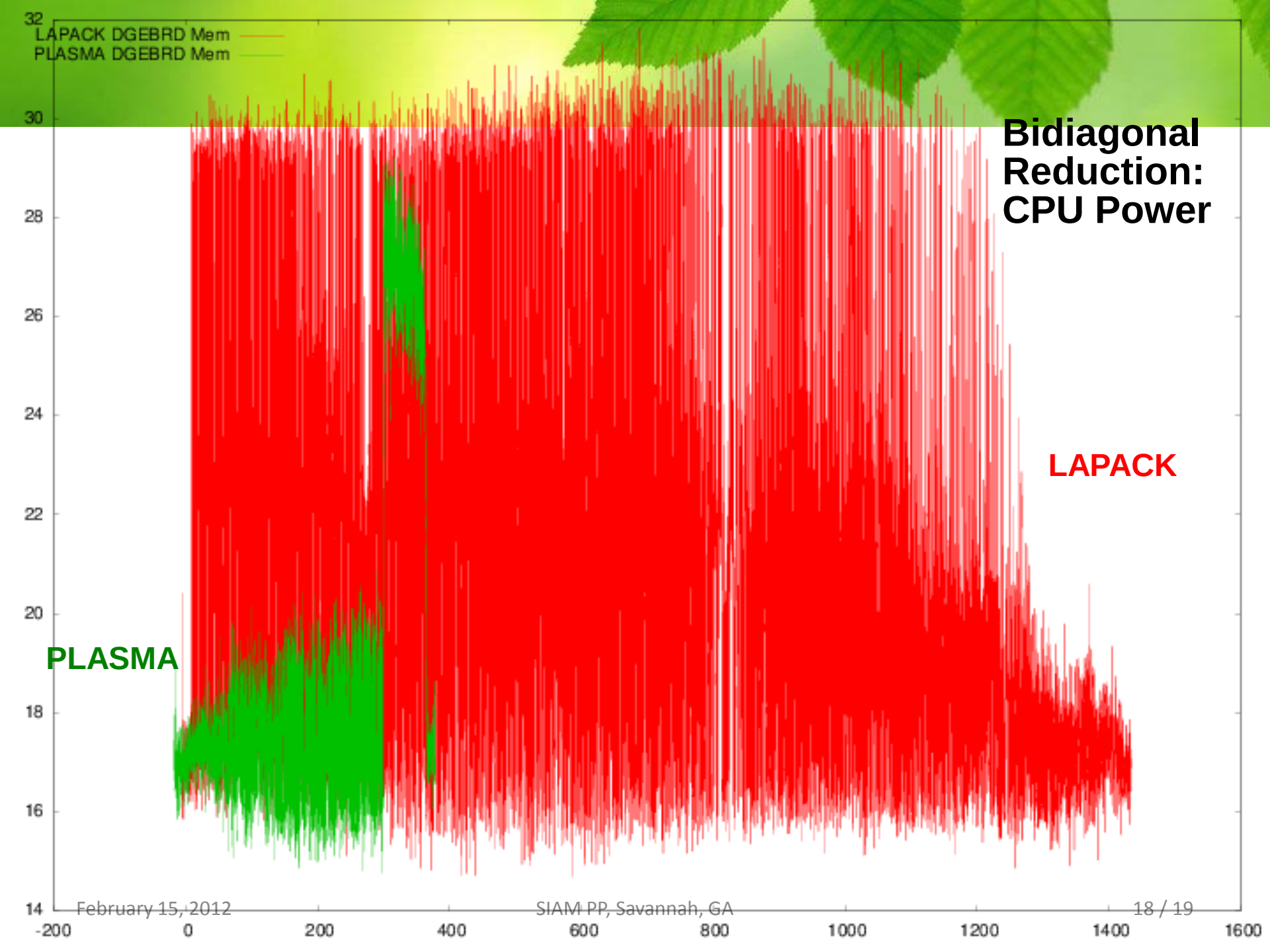
Power consumption over time

Matrix inverse

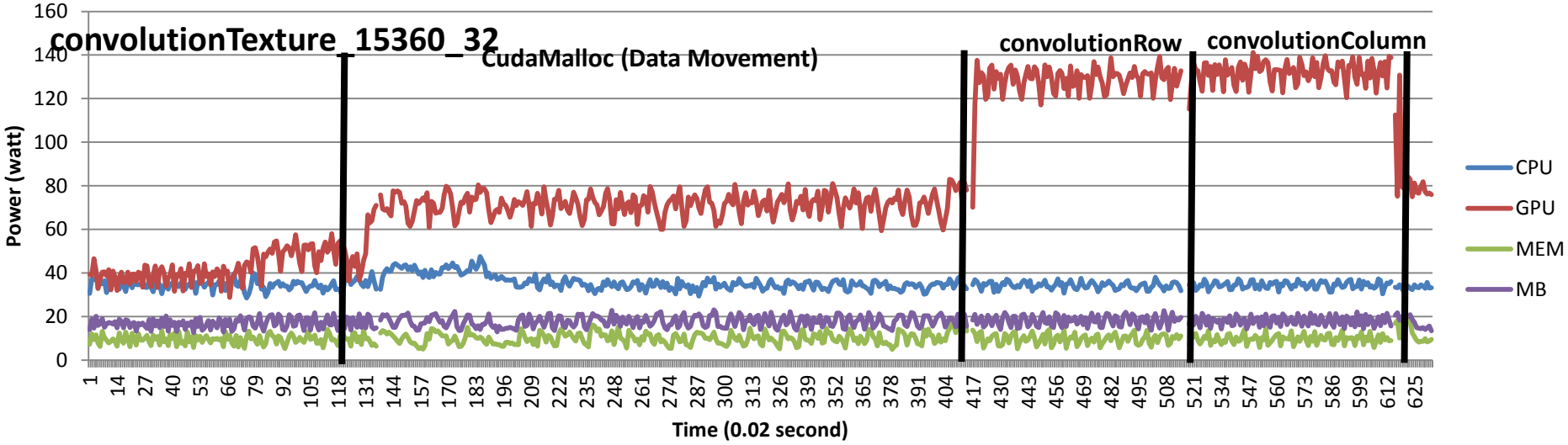
Sources:

Piotr Luszczek Hatem Ltaief

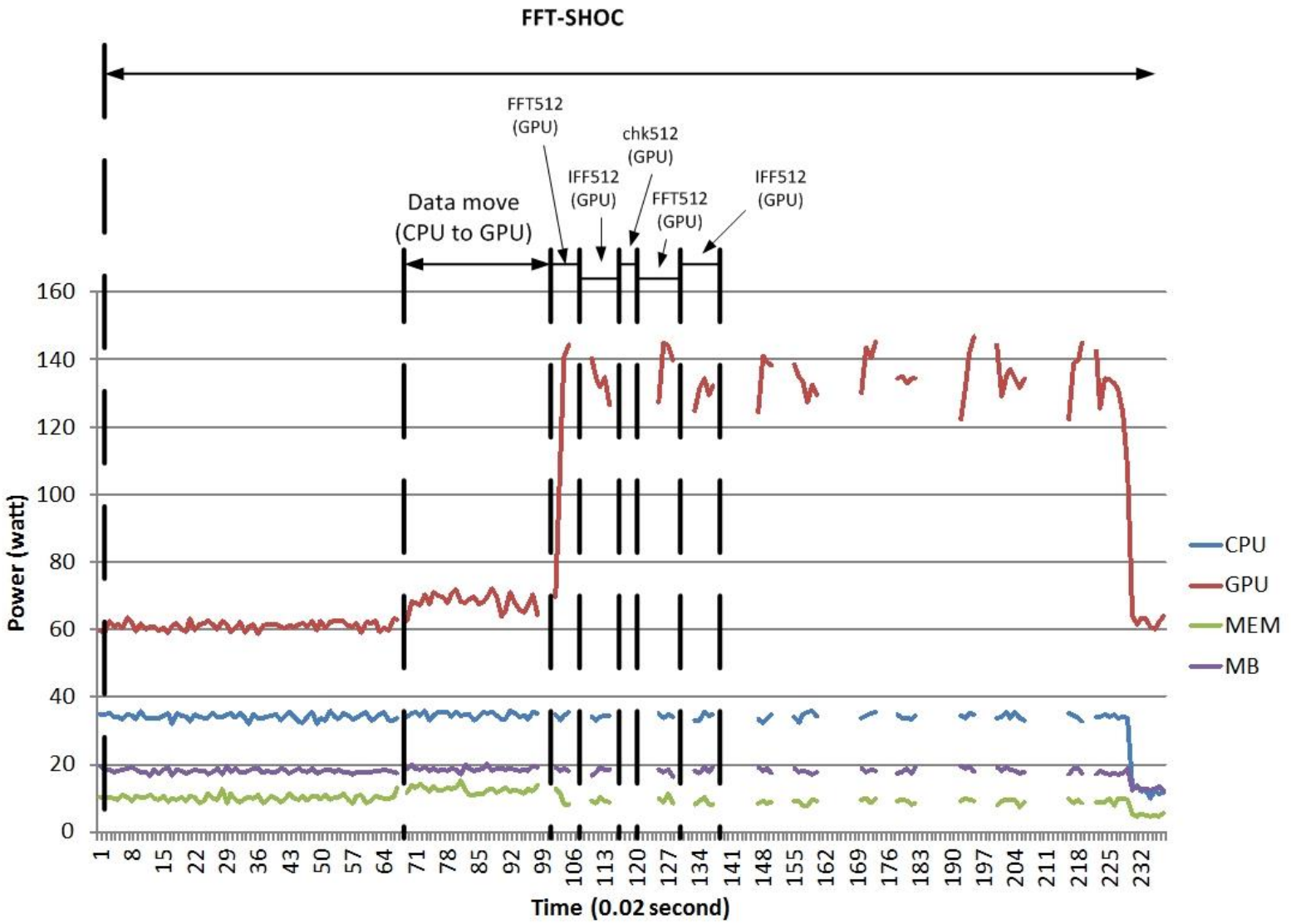
Bidiagonal Reduction: CPU Power



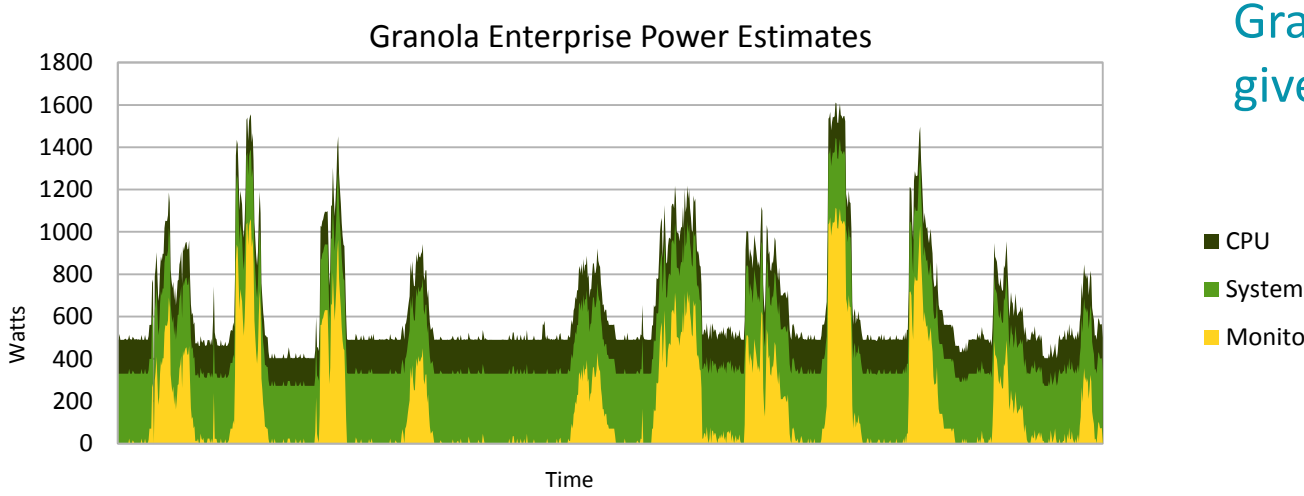
PowerPack 4.0 (accelerator support)



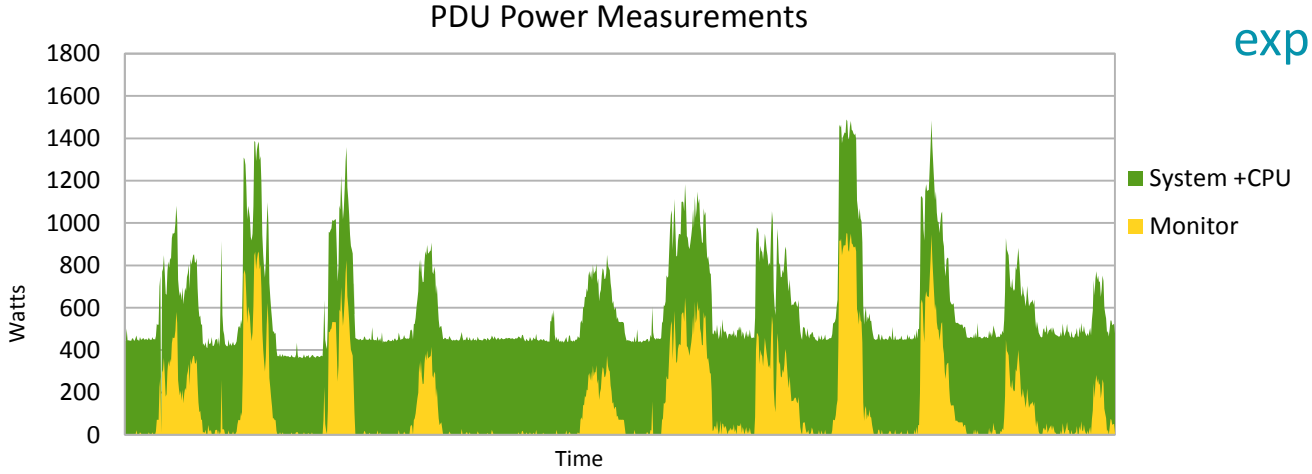
PowerPack 4.0 (API+accelerator)



Commercial grade measurement…

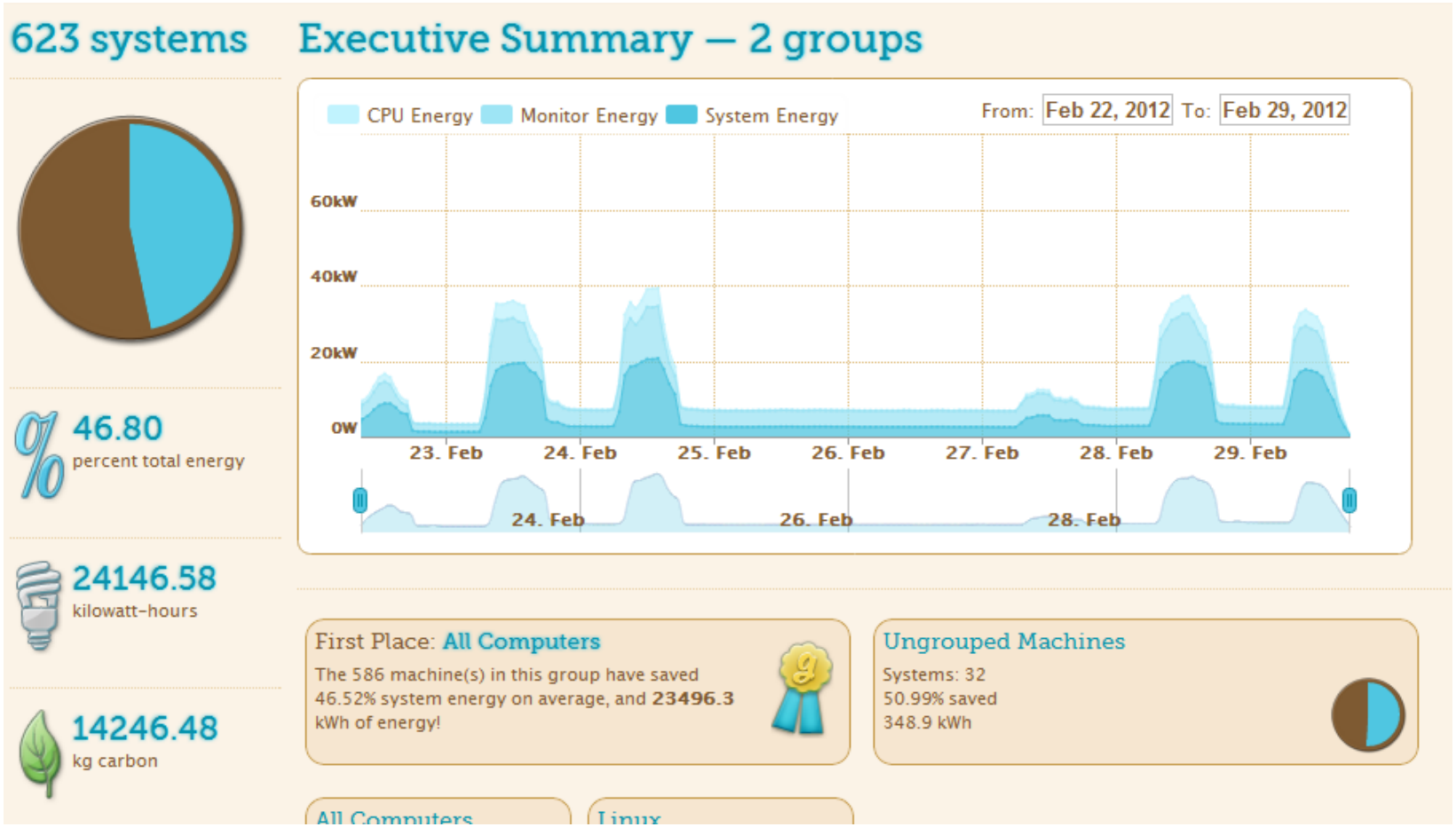


Granola software gives more detail…



…same accuracy as expensive hardware

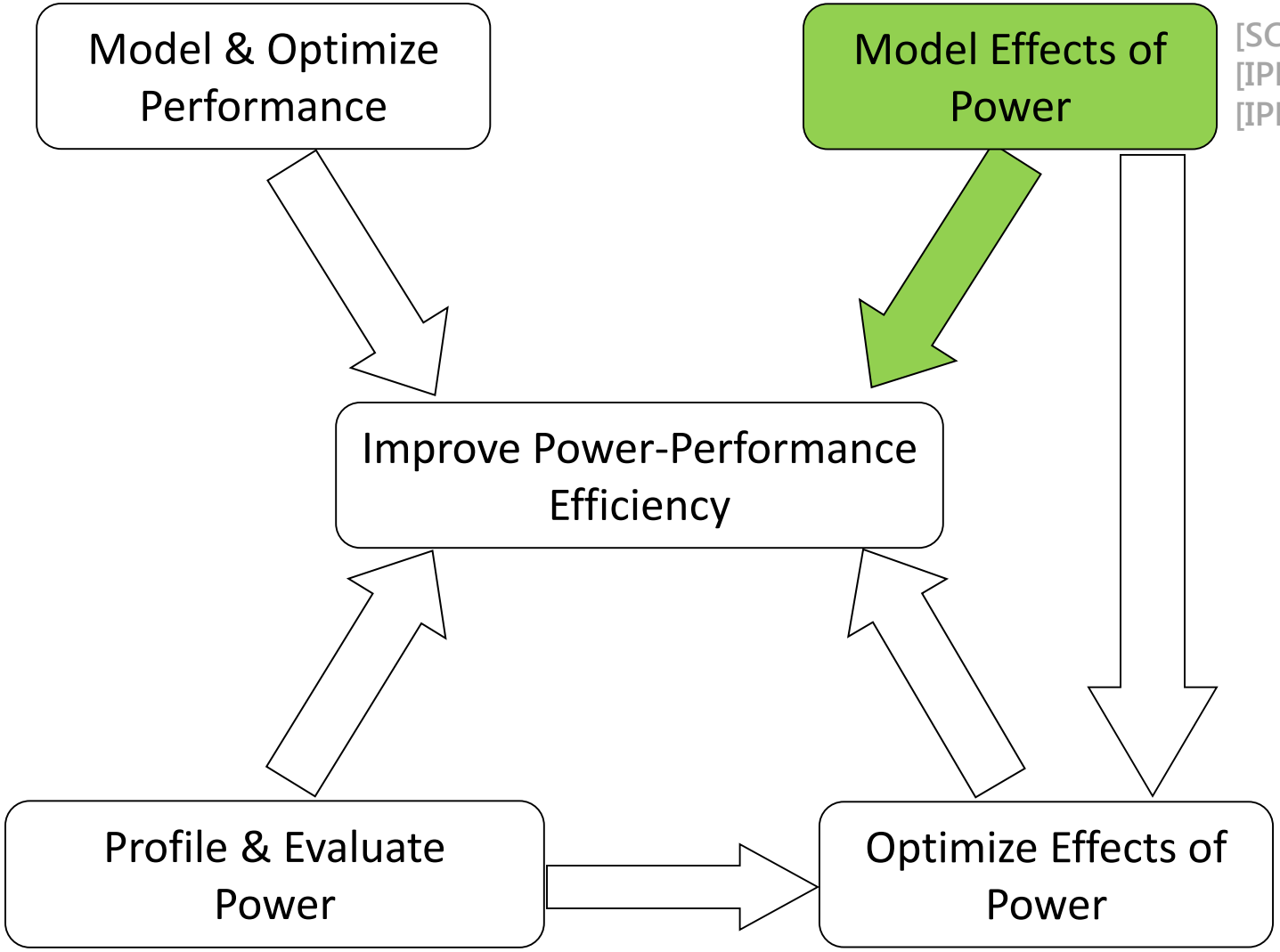
Granola Enterprise (Freeware)



“To know is to understand.”

Aristotle

Power-Performance Efficiency



[SC 2004], [SC 2005], [IPDPS 2011], [IPDPS 2013]

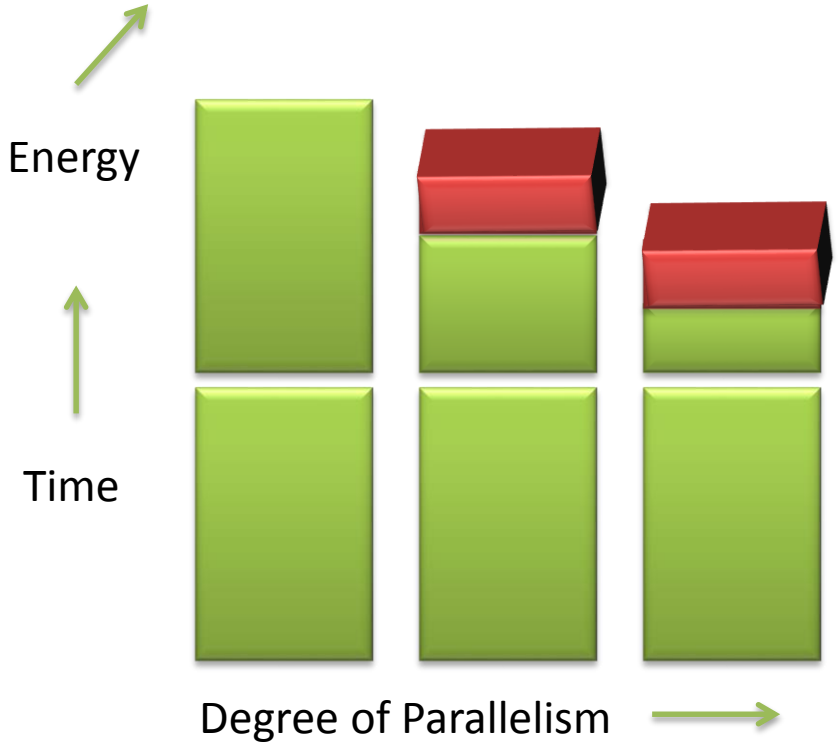
Early Green HPC questions…

- What happens to energy at scale?
- How can we scale energy/perf efficiently?

Amdahl's Law (for energy?)

- Classical speedup
  - Amdahl's law for 1 enhancement (parallelism)

$$S_N(w) = \frac{T_1(w)}{T_N(w)} = \left[(1 - FE) + \frac{FE}{SE} \right]^{-1}$$



Time ~ energy. Right?

So we only get energy savings by reducing time. Right?

Then why does PM (e.g. DVFS) save energy? And sometimes without affecting time?

Amdahl = no overhead

But, overhead is the key to savings energy without loss!

Power-Aware Speedup

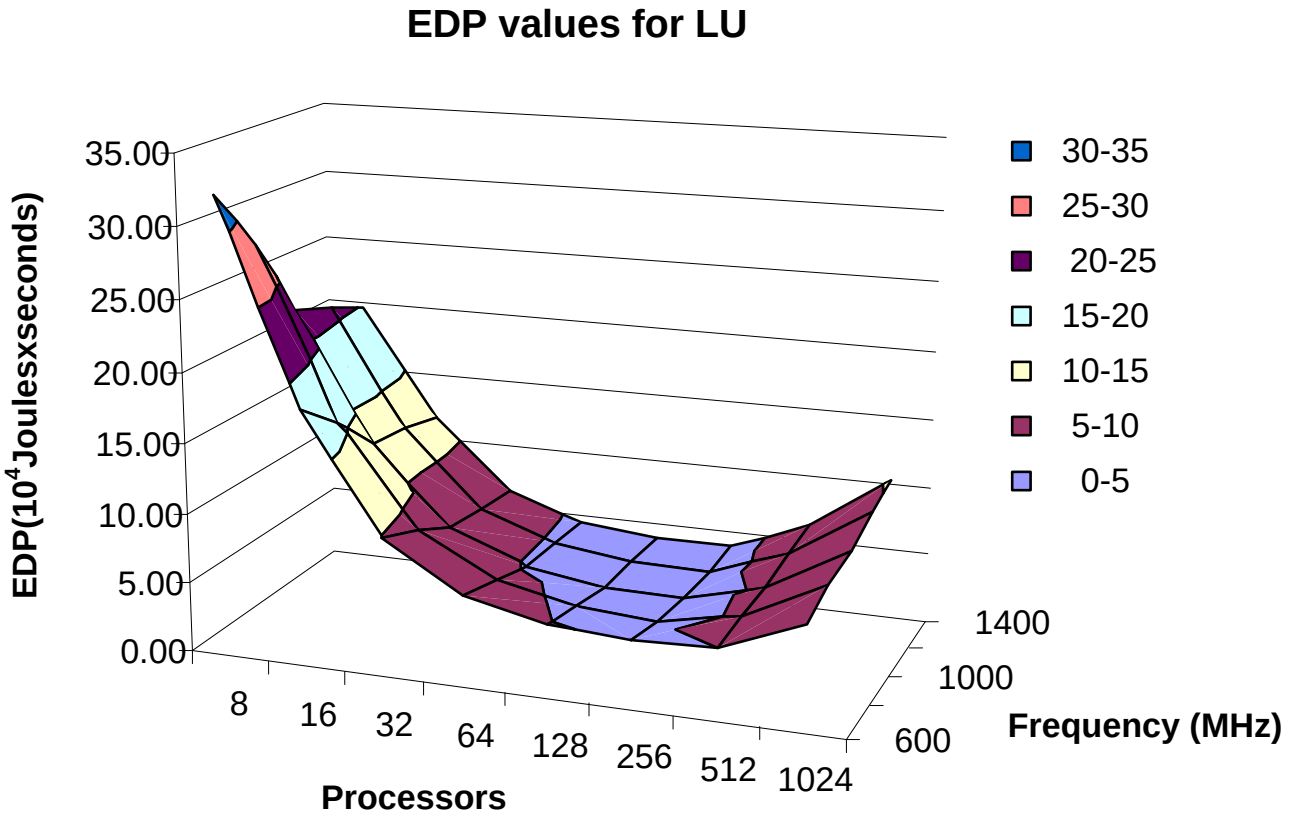
[IPDPS 2007]

- **Definition**
  - Speedup

$$S_N(w, f) = \frac{T_1(w, f_0)}{T_N(w, f) + O(w, f)}$$

  - *w:* workload
  - *N:* number of nodes
  - *f*: the clock frequency and f_0 is the base value
  - $T_1(w, f_0)$: sequential execution time at base frequency f_0
  - $T_N(w, f)$: parallel execution time at *N* processors at frequency *f*

Bounding Efficiency at Scale



- Energy/performance optimal system configuration
  - # processors: 256
  - CPU frequency: 1200MHz

Early Green HPC questions…

- What happens to energy at scale?
- How can we scale efficiently?

Iso-energy-efficiency

Grama et al: performance efficiency can be held constant if we increase both number of processors and problem size simultaneously.

Algorithm + Scale → fixed performance

Iso-energy-efficiency

Algorithm + Scale + Power Modes → (power, performance)

- Requires accurate performance model
- Requires accurate power model
- Must be accurate, useful, usable

Iso-energy-efficiency Derivation

[IPDPS 2011],[IPDPS 2013]

General form of our Iso-energy-efficiency model:

$$EE = \frac{E_1}{E_p} = \frac{E_1}{E_1 + E_o} = \frac{1}{1 + {E_o}/{E_1}}$$

EE *: system-wide energy efficiency*

E_1 *(baseline):* total energy consumption of sequential execution on one processor

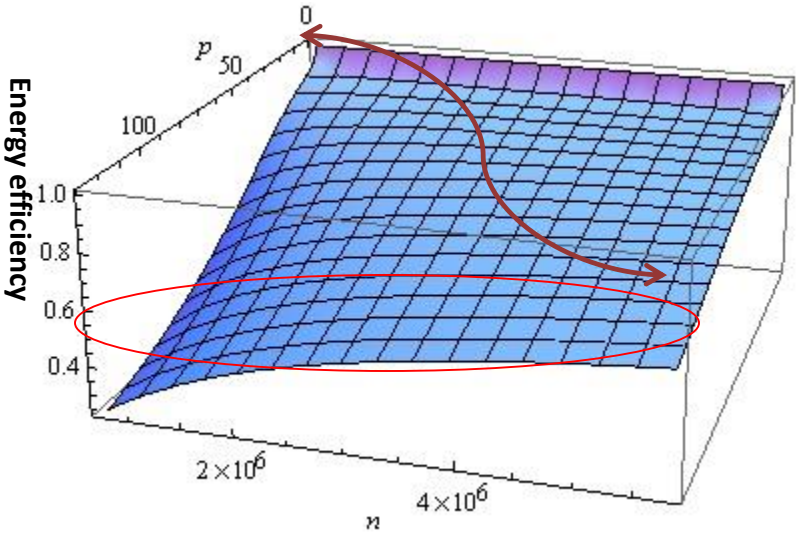
E_p : the total energy consumption of parallel execution for a given application on *p* parallel processors

E_o : the additional energy overhead required for parallel execution and running extra system components

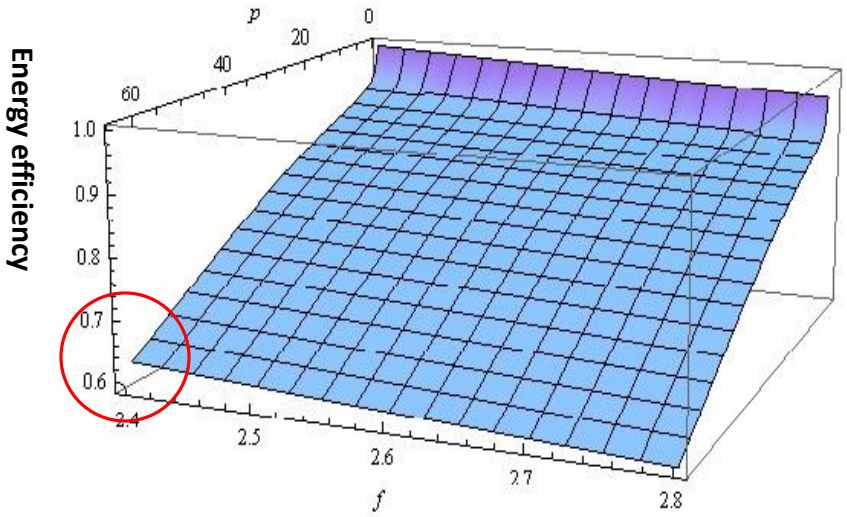
Maintaining Efficiency in 3-D FFT

$$EE_{FFT} = \frac{1}{1 + \frac{6.87log_2p - 1.75flog_2p + p(p-1)f(\frac{11500}{n} + \frac{0.376}{4^{log_2p-2}})}{163 + 22.7f}}$$

FT's system-wide energy efficiency with p and n as variables



FT's system-wide energy efficiency with p and f as variables



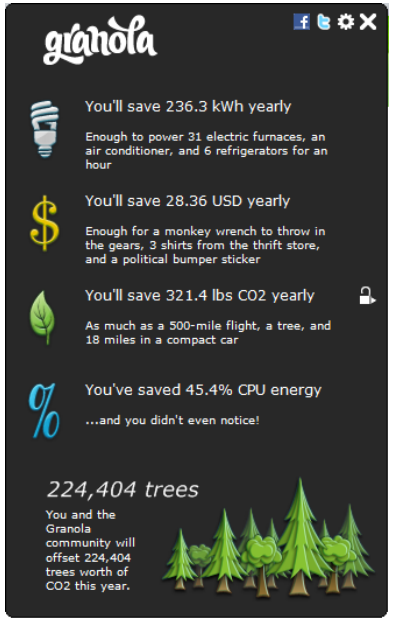
- *Problem size scaling effective in maintaining overall system energy*
- *CPU frequency scaling: only slightly improves EE*
- *But, the effects of CPU clock frequency on on-chip workload diminish while scaling up system size.*

Commercial grade management…

Granola (http://grano.la)

- Launched Earth Day 2010
- Free home version
- 350K+ Downloads so far…
- 165+ Countries
- Uses: laptops, PCs, servers
- *Performance Guarantees*

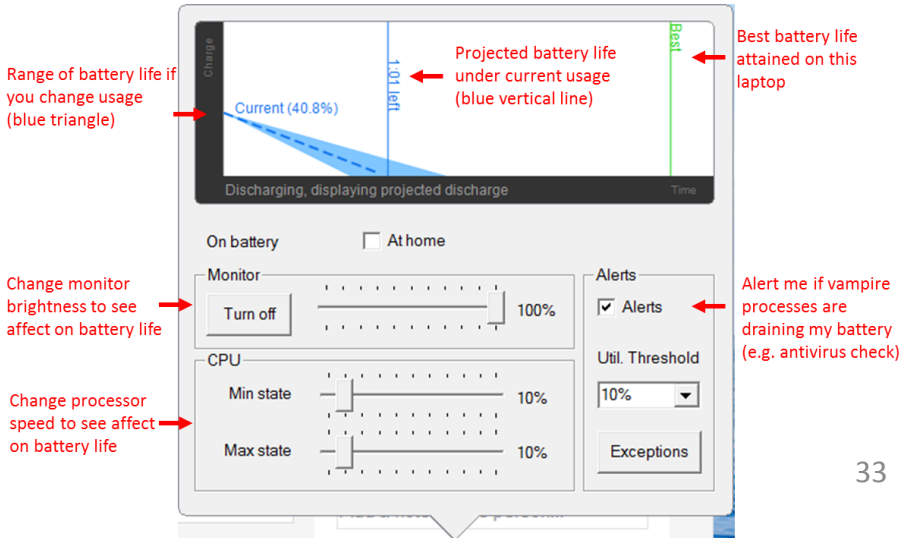
Patents: [USPTO: #13/061,565] [UK: #GB2476606B]



Fatbatt (http://fatbatt.com)

- Launched March 2013
- Free ad-version

Display when running on battery.



Where do we go from here?

We need lots of help.

Disruptive vs. Incremental.

Silver bullet is unlikely.

Commodity matters.

Markets matter.

Tools matter.

Wanted: Major catastrophe.

Custom system is likely the only answer by 2019. Energy wall?

“Victory” is inevitable when you change the game.

Thank you.