70 YEARS OF CREATING TOMORROW

Los Alamos NATIONAL LABORATORY

Power and Cooling Transients: Requirements of HPC Workloads

Josip Loncaric
LANL, HPC-DO

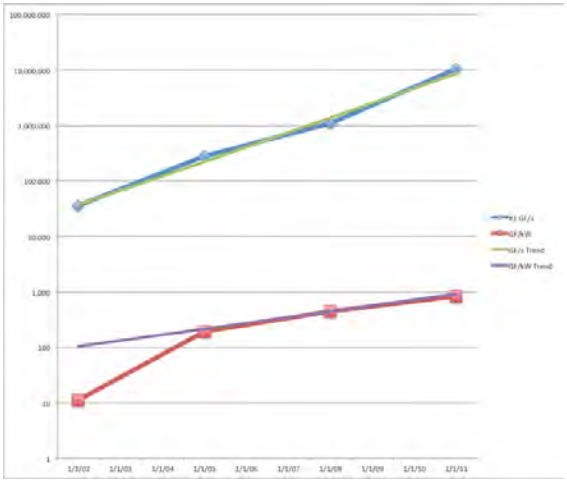
SC13, Nov. 17th, 2013

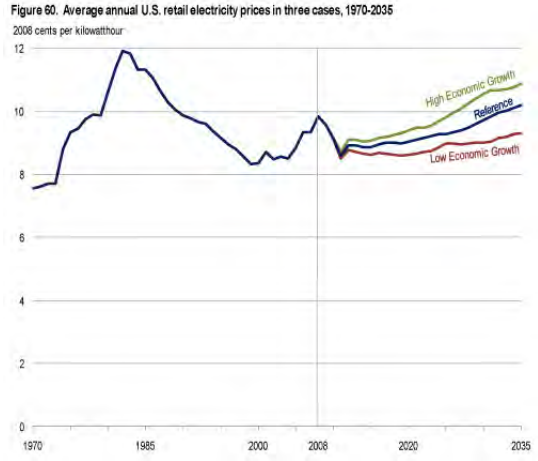
LA-UR-13-25997

Outline

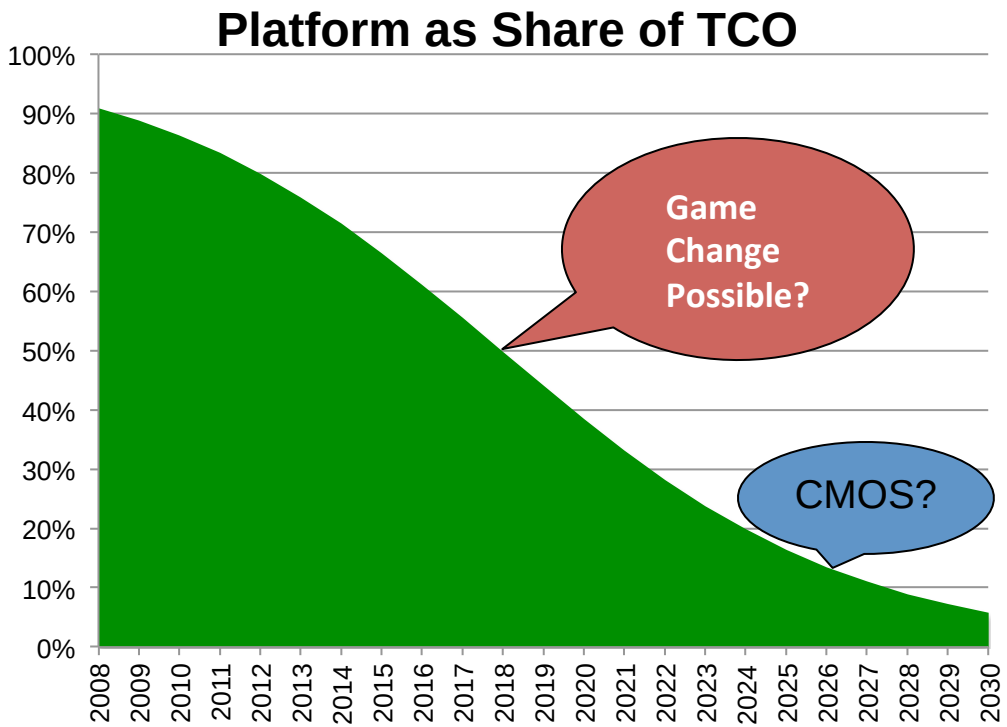
- Power characteristics of HPC workloads
  - What to expect, in theory
  - Long term statistics, as measured
  - Short term power transients, as measured
  - Projections to future
  - Preparing for this future
- Summary: Dynamic behaviors must be considered

Extrapolating Performance & Power Trends





Moore's law & Dennard scaling extrapolated



Sources:
(1) Top500 Nov. 2011 list
(2) U.S. Energy Information Administration, Annual Energy Outlook 2010

Power Patterns Determine Power Cost

- It's about power cost = Demand charge + energy charge
  - Energy charge: Favors energy efficiency
    - Industry is reducing this problem by making power demand variable
    - This can't be ignored at high power levels
    - Energy prices vary, include spot market energy purchases
  - Demand charge: Favors steady power delivery
    - Variable power can double normal demand charge
    - Industry is exacerbating this problem at the chip level
- Power costs could ***increase*** by saving energy!
- Additionally, saving energy could reduce throughput
  - Reducing voltage and frequency on the application critical path extends runtime
  - Additional platform costs must be justified by power cost savings
  - Remember TCO = CapEx + OpEx
  - It's about efficient delivery of value per $, not just saving kWh.

Industrial Electricity Rate Chart Example

NEW MEXICO
PUBLIC REGULATION
COMMISSION
FILED
2011 AUG 11 PM 4 03

PUBLIC SERVICE COMPANY OF NEW MEXICO
ELECTRIC SERVICES

8TH REVISED RATE NO. 30B
CANCELING 7TH REVISED RATE NO. 30B

LARGE SERVICE FOR MANUFACTURING
FOR SERVICE ≥ 30,000 KW MINIMUM AT
DISTRIBUTION VOLTAGE

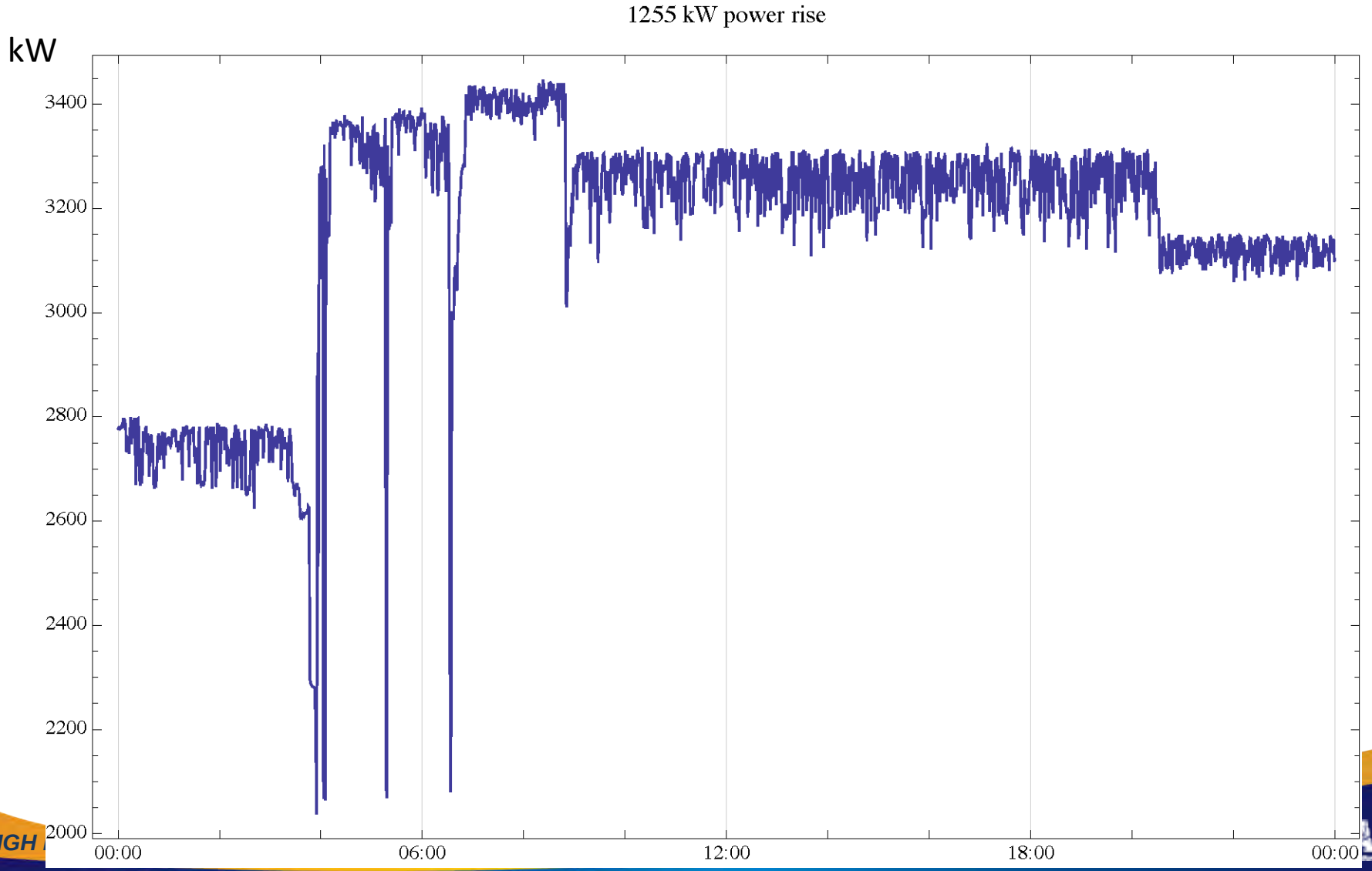
Page 1 of 4

APPLICABILITY: The rates on this schedule are available to any retail manufacturing customer who contracts for a definite capacity commensurate with customer's normal requirements but in no case less than 30,000 kW of capacity, who has a load factor of at least 80%, and takes service at PNM's primary distribution voltage.

"Where highly fluctuating or intermittent loads which are impractical to determine properly (such as welding machine, electric furnaces, hoists, elevators, X-rays, and the like) are in operation by the customer, the Company reserves the right to determine the billing demand by increasing the 15-minute measured maximum demand and kVAR by an amount equal to 65 percent of the nameplate rated kVA capacity of the fluctuating equipment in operation by the customer."

This includes large HPC platforms ➔ demand charge could approximately double.

Extremes: Power Rise 1.255 MW



Daily 1+ MW Transients Growing to 10+ MW

Large HPC platforms consume large amounts of electrical power

Many HPC applications have global synchronization points

Energy efficiency improved via reduction of CPU idle power

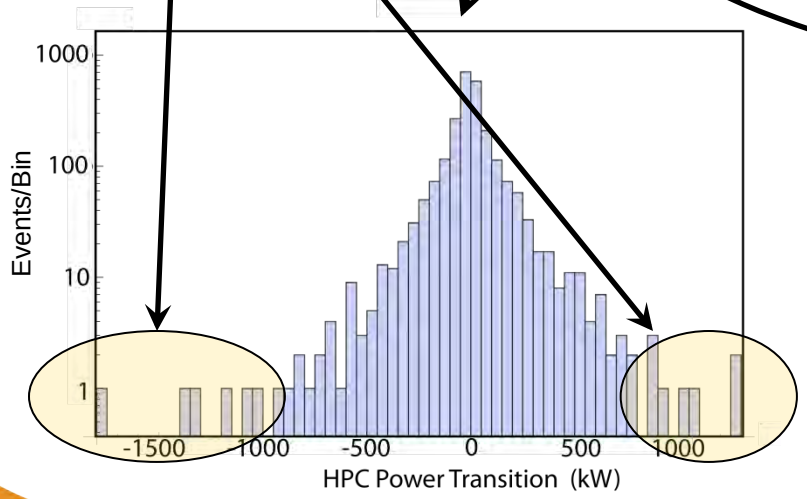
A new class of potentially disruptive grid transients emerging

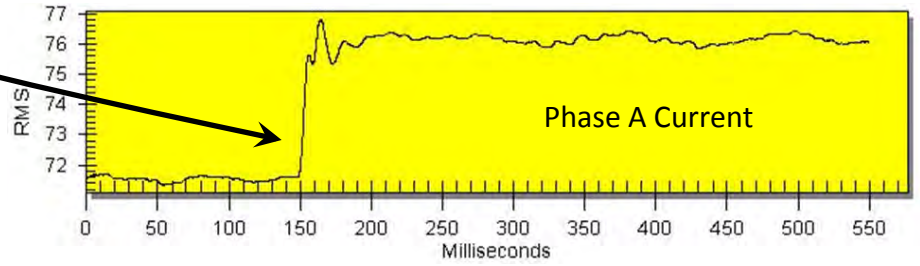
Large—the entire platform (10's MW)

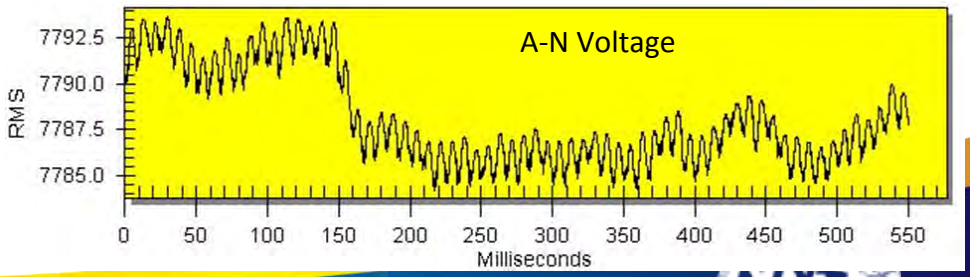
Fast–about one AC cycle (~15 msec)

Current LANL platform experiences ~ "full-machine" transients daily

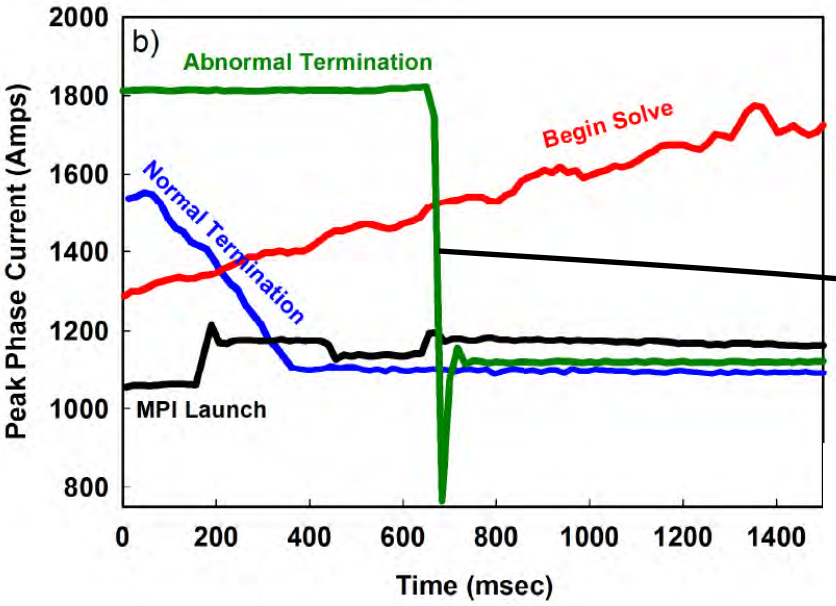
A ~100 kW, single-cycle transition on LANL HPC captured on utility meters at a LANL substation





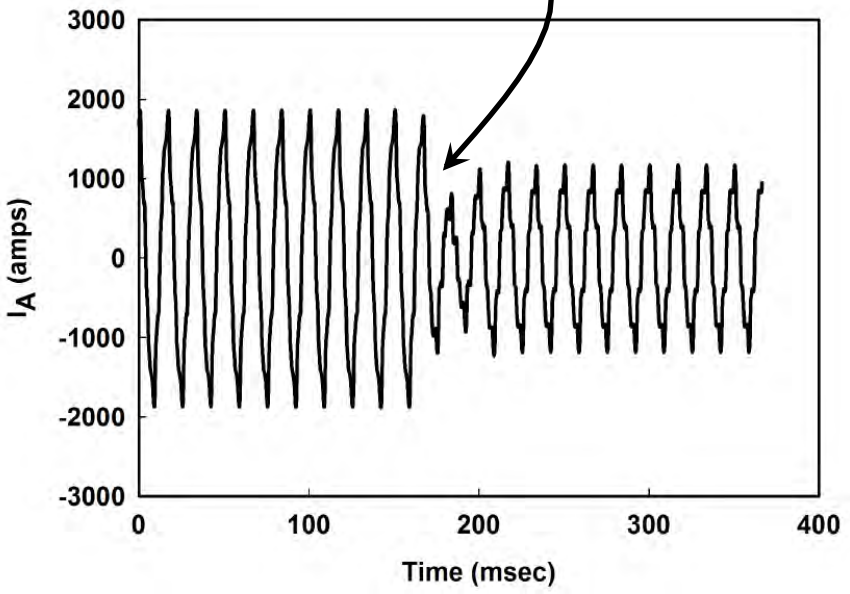


Linpack Power Transient Testing at LANL



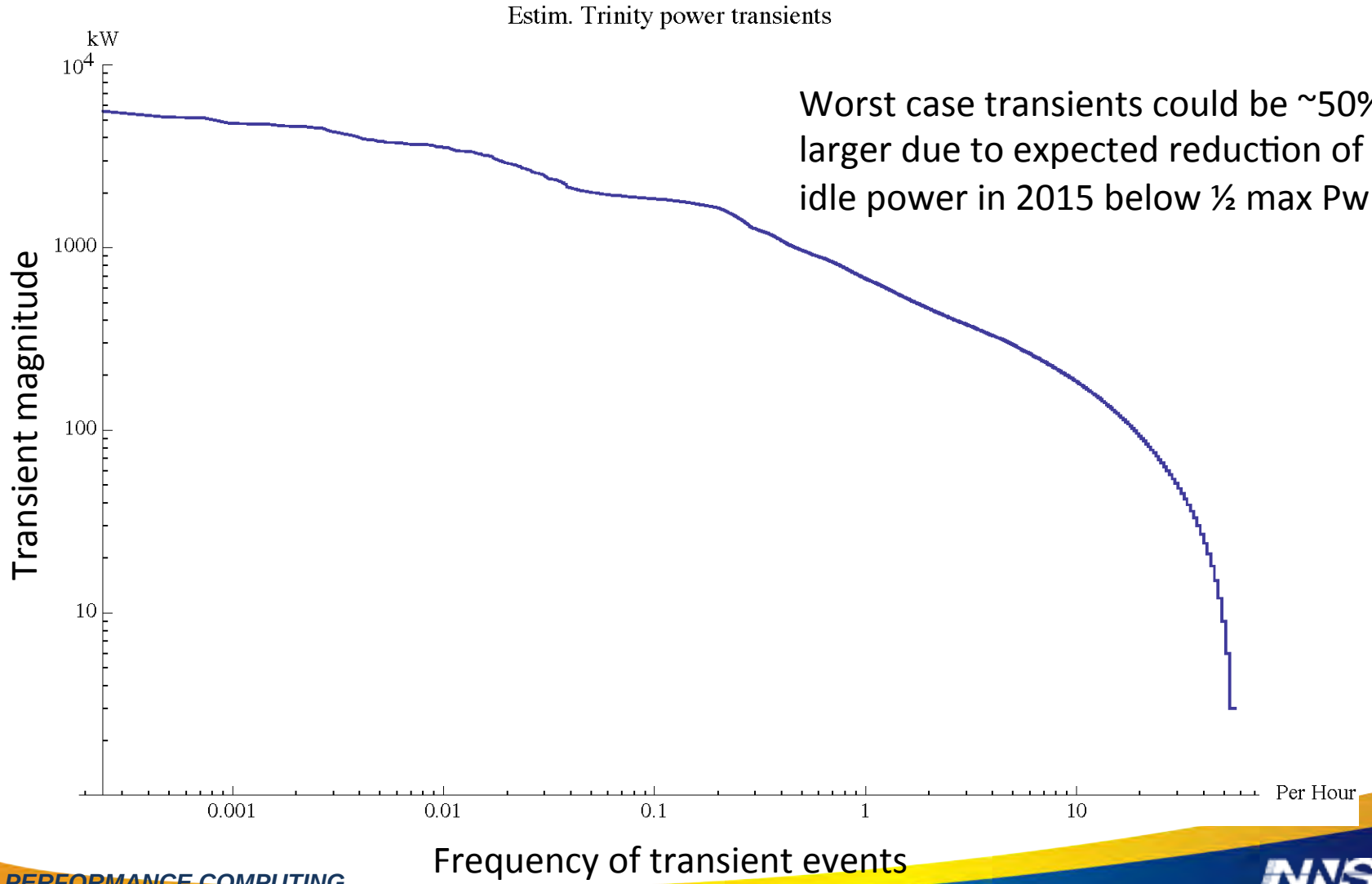
Projections for LANL platforms:

- Today: Transients not a concern
- 2015: Transients noticeable, still within limits
- >2015: Transients likely need to be mitigated (depending on MW growth)



Joint work with Scott Backhaus, Cornell Wright, and Maura Miller at LANL

2015: Extrapolating to 3x Power Platform



Worst case transients could be ~50% larger due to expected reduction of idle power in 2015 below ½ max Pwr

Conclusions

- HPC workload power requirements fluctuate with characteristic step changes arising from phases of computation
- Large power transients at 10 MW level may increase costs of electricity, even if energy use is reduced
- Our utility wants predictable power demand for the next 24+ hours, as hourly averages within ±1 MW tolerance
- Dynamics of variable power must be considered in facility design for large HPC platforms using 10's of MW
- All of the above can be addressed by proper design

Abstract

Power requirements of normal HPC workloads are not steady, but show step changes associated with phases of computation. These changes can be very rapid, as fast as a single AC cycle. Moreover, as chips seek to reduce idle power, the potential for power fluctuations is increasing. Saving energy by converting steady power to fluctuating power may actually increase power costs.