HPC Power Increasingly Challenges the Power Grid

SC18: Ninth Annual Workshop for the Energy Efficient HPC Working Group (EE HPC WG)

Dr. Josip Loncaric, LANL

Nov. 12th, 2018

NNSA
National Nuclear Security Administration

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

LA-UR-18-30694

Energy Efficiency Moves Impacts To the Power Grid

The Power Grid Session

11:15 – 12:00

Dallas, TX

- **HPC power profiles are challenging**
  - Measurements in practice
  - Impacts on power grid
  - Step change statistics
- **Why is this happening?**
  - Approaching zenith of CMOS
  - Power models for CMOS
  - Characteristics of HPC applications
- **Outlook**
  - Power challenges growing
  - Work with utilities to mitigate cost impacts

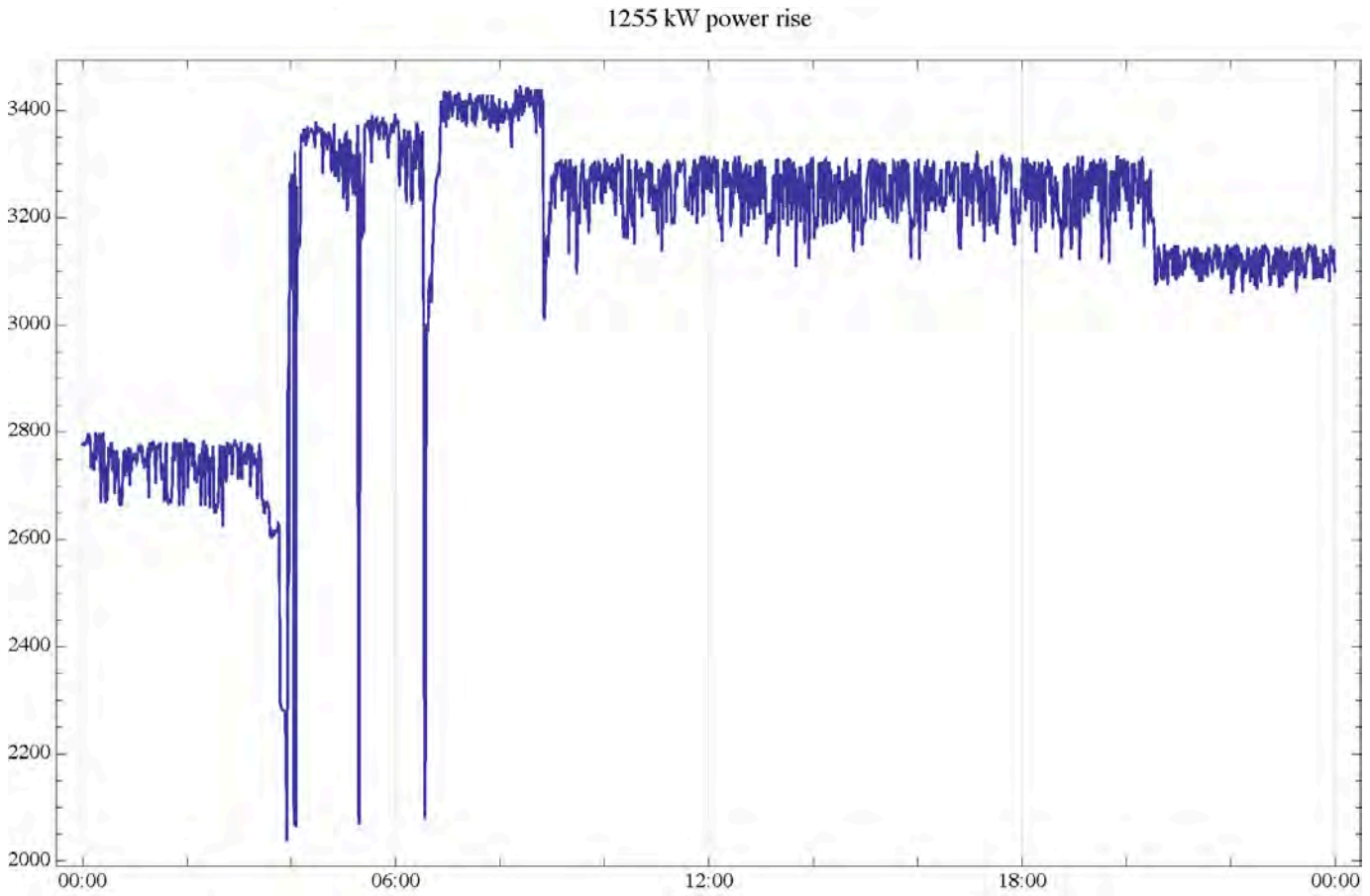
HPC power profiles are challenging

HPC Power Needs & Power Fluctuations Are Growing

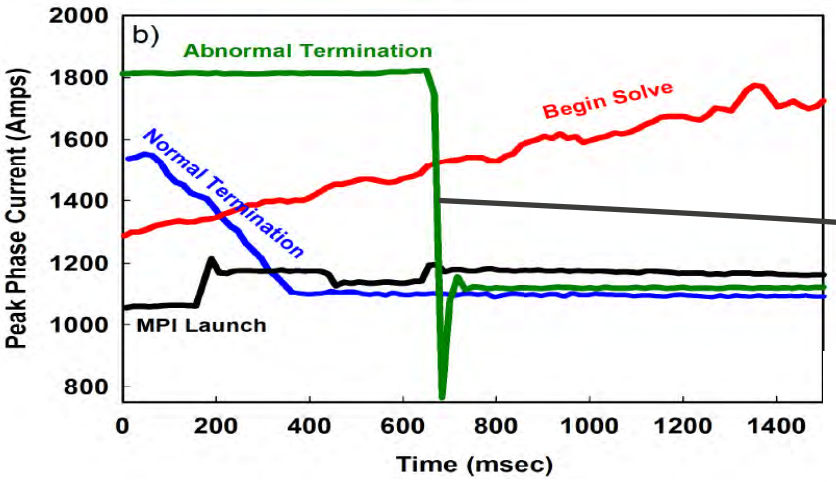
- Power efficiency gains are not keeping up with platform scaling
  - Power requirements are growing, about 6 year doubling time
  - Exascale platforms may need 20-40 MW to operate in 3-5 years
- Power fluctuations are increasing
  - Advanced power management is necessary to reduce power when idle
    - Semiconductor process scaling isn't helping enough
  - Idle power is becoming a small fraction of maximum power
  - HPC workloads often step between idle and max within microseconds
    - Wait until everyone is ready, then everyone starts working hard, or vice versa
  - A 20-40 MW computer could impose fast 15-30 MW power transients
- This isn't what the power grid was designed to do easily: Cost impacts?
  - Rapid power flow changes can cause voltage disturbances if grid is insufficiently stiff. Similar characteristics to a grid fault.

LANL Measured HPC Power Transients in 2012

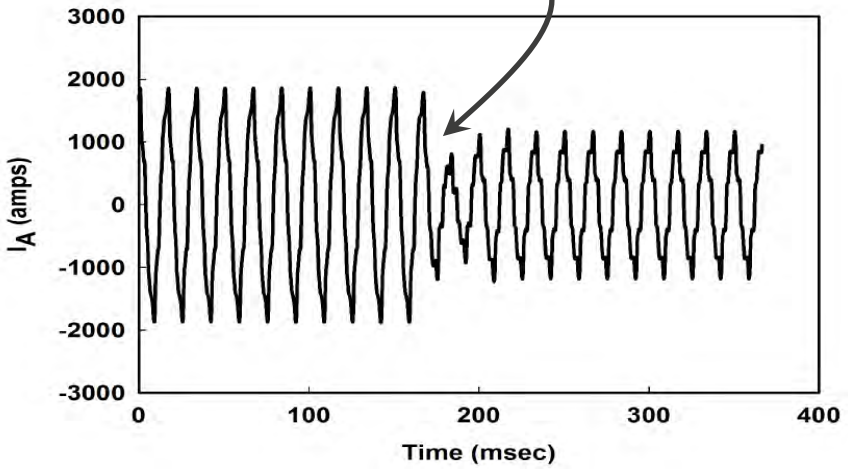
- Power transients on Cielo, a 4 MW platform
- Using facility monitoring at 1 minute intervals
- Actually, these power transients are *much* faster
- Since 2012, many HPC sites have reported similar behaviors



HPL Power Transient Testing at LANL, 2012



- HPL phases use different power
  - Normal power transitions ~0.1s
- Abnormal termination drops to idle power instantly, <1 AC cycle



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

2012: 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

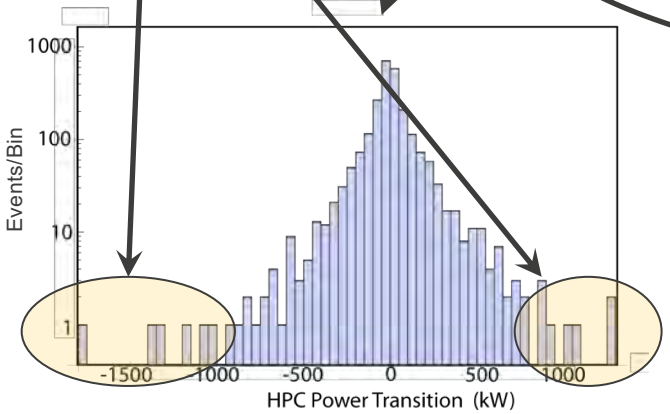
A new class of potentially disruptive grid transients emerging

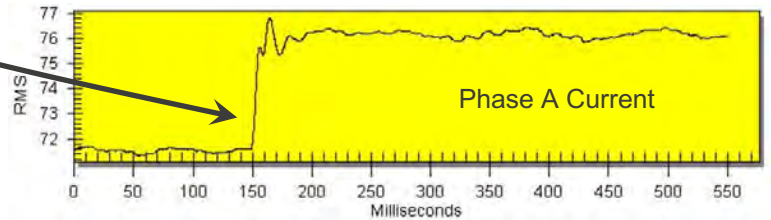
Large—the entire platform (10's MW)

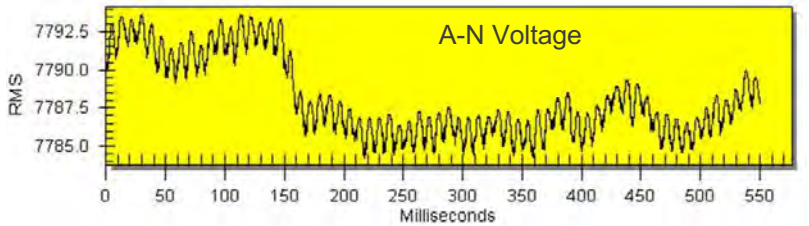
Fast–about one AC cycle (~15 msec)

2012 LANL platform experienced ~ "full-machine" transients daily

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

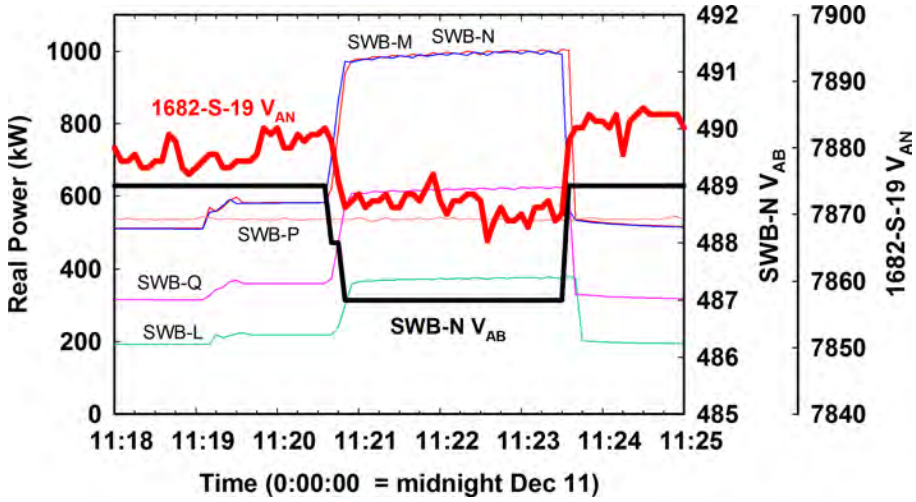




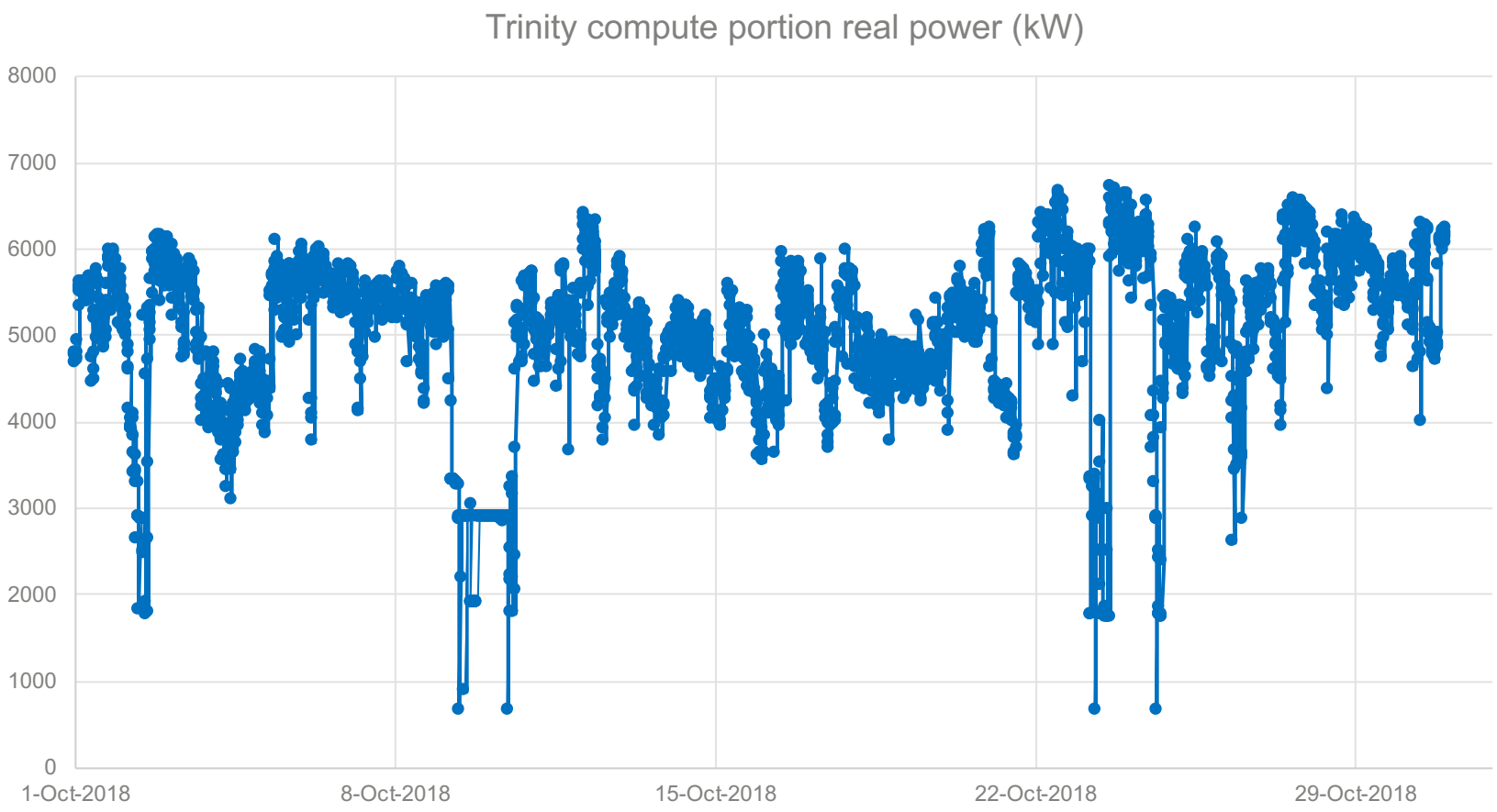


Transient Voltage Disturbances Propagate Upstream

- 2012 HPL test on Cielo
- Red: 13.8 kV line-neutral voltage
- Black: 480 V line-line voltage
- Other: power
- Full machine 1.5 MW transient:
  - 480 V sag: 0.4%
  - 13.8 kV sag: 0.2%
- Extrapolating to 15 MW transient: 4% and 2% respectively
- As expected by National Electric Code feeder design:
  - Up to 5% sag at full load



Recent Example: Trinity Power, October 2018



Why is this happening?

Power Use in CMOS

- CMOS is designed to use power only when active
  - Plus leakage current, significant since about 2004
- Power model (DARPA Exascale Final Report, 2008):
  - *ActivePower = Capacitance * Clock * V^2*
- Options to reduce power:
  - Voltage: Already down to about 1 V, further reduction difficult
    - Near-threshold voltage regime requires Si process change, lower clock rate
  - Clock: Not a path to faster computing
  - Capacitance: Possible with semiconductor process scaling
    - But very costly: Progress is slowing down already, slow transition 14nm → 10nm, most foundries delaying transition to 7nm: Zenith of CMOS technology is near
  - Advanced power management: Turn off what's not in use, or DVFS
    - Implies potential for greater power transients (same max ⇄ lower idle)

Generic HPC Applications Are Bulk Synchronous

- HPC applications assign work to processors, wait until all is done, repeat
  - Example: HPL panel factorization steps
  - Power reaches max while all are busy, drops while waiting for stragglers
- HPC applications have phases of execution
  - Example: Compute, store results, compute
  - Example: Compute something, then compute something else
  - Different phases use different power
- Transitions between phases are very fast
  - Processes can synchronize in microseconds
  - DC power at a CPU is impacted immediately, in nanoseconds
  - AC power is impacted within 1 AC cycle (filtered by P/S capacitors)

Outlook

HPC Challenges To the Power Grid Will Grow

- More computing requires more power
- Zenith of CMOS: Geometric scaling has already slowed down
  - Progress by other means: Advanced power management
- More energy efficiency increases magnitude of transients
- HPC applications have phases required by physics and algorithms
  - Compute phases can't be de-synchronized enough to help with transients
- Energy efficiency is pushing problems downstream to the power grid
  - Creating big problems can increase energy costs
- Even slow power changes can be costly
  - Power blocks are traded by people on hourly timescales, not by AC cycle
  - Some contracts have power swings built into the cost formula
- We need to work with utilities to prepare before costly surprises pop up

Abstract

- HPC applications quickly transition between phases of computation demanding different power levels. As a result, HPC platforms present highly variable loads to the power grid, pushing the problem upstream to utilities. This problem is inherent in the nature of HPC applications, and it is growing because platforms are getting larger and because energy efficient technologies are reducing idle power. Multi-MW step changes are common today, growing to 10's of MW for Exascale platform. Working with utilities to minimize energy cost impacts is the recommended path forward.