Operations Lessons Learned and Energy Efficiency Case Studies Panel

Miloš Puzović

STFC Hartree Centre
Daresbury Laboratory
Sci-Tech Daresbury
Cheshire, WA4 4AD, UK

Eighth Annual Workshop for the Energy Efficient HPC Working Group

Hartree Centre
Science & Technology Facilities Council

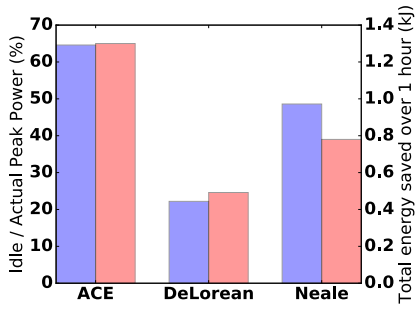
Data Center Birds-Eye View

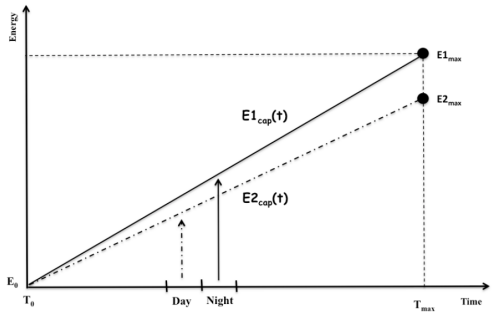
- Wide array of architectures - IBM Power, Intel x86 and ARM
- Different types of cooling - air, water and oil (Neale)
- Three rooms:
  1. Low Density Room - only air cooled
  2. High Density Room - only water cooled
  3. Green Room (experimental hardware) - mix of air and water cooling

Room	Low Density Room (LDR)		High Density Room (HDR)		Green Room	
Systems	*Panther*	32× 16-core IBM Power 8	*Scafell Pike*	846× 32-core Intel Skylake	*DeLorean*	5× Maxeler MPC-X2000 (8× DFE)
		4× NVIDIA K80 GPU each node		840× 64-core Intel KNL		5× 16-core Intel Ivy Bridge
	Paragon	32× 16-core IBM Power 8		128 GB - 24× 1TB/KNL 96GB		128GB RAM
		4× NVIDIA P100 GPU each node	*JADE*	22× NVIDIA DGX-1 DLS	*ACE*	24× 96-core Cavium ARMv8 ThunderX
	IBM TS3500 tape library		*Napier*	360× 24-core Intel Ivy Bridge		
	Centralised Data Store (CDS)			64GB RAM		132GB RAM
	ATOS Bullion	16× 18-core Intel Haswell	*Dawson*	x86-based Data Analytics		
		4TB RAM	*Iden*	84× 24-core Intel Ivy Bridge	*Neale*	120× 16-core Intel Ivy Bridge
	ATOS x86-based XRV system			42× Intel KNC co-processors		
				64GB RAM		132GB RAM

Idleness

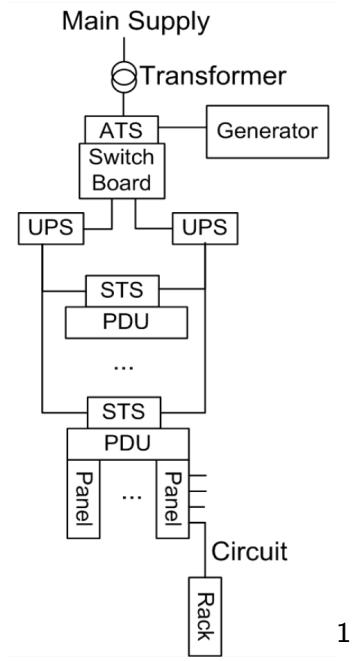
- non-peak power efficiency improvements:
  - power efficiency measured when system is at maximum load
  - **most** of the time systems operate *far away* from peak activity
- ***IDEAL SYSTEM***
  - consumes **NO** power when idle; and
  - power increases proportionally to increasing activity



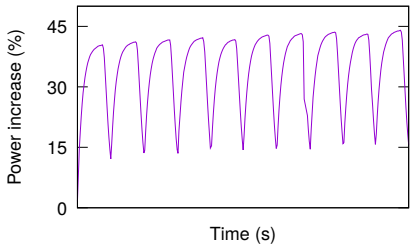


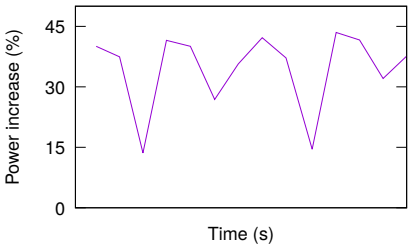
A significant improvements to energy efficiency are possible without any changes to peak power consumption

Workload variability



[1]



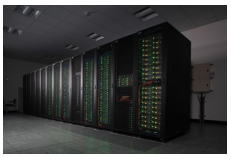


- larger time window has larger power variation
- relative power variation depends on power hierarchy level
- variation also depends on the application

For power policy to be successful power sampling must be at high frequency and at appropriate level in power delivery network

[1] *Fan et al.* 2007. Power provisioning for a warehouse-sized computer. In Proceedings of the 34th annual International Symposium on Computer Architecture

Heterogeneity



- Advantages:
  - dynamic range of power consumption narrower (*stable draw*)
  - small correlation between power consumption peaks
  - actual power peak consumption less than with homogeneous systems
  - wide set of opportunities for optimisation
- Disadvantages:
  - different systems $\implies$ different headaches
  - no unified policies across multiple different policies
  - ***BUT*** there is work towards standardisation (see RedFish and PowerAPI)

Co-locating heterogeneous systems can provide more compute within power budget

Summary

1. significantly noticeable gains in reducing energy consumption are possible by lowering idle power as much as possible (striving to zero with help from system designers)
2. to make the maximum use of power capping and similar technologies it is necessary to understand to characterise power consumption behaviour of workloads and power delivering infrastructure at high frequency sampling rate
3. occupying data centres with heterogeneous systems will improve compute capacity within constrained power budget at the cost of administrating complexity