Monitoring and Management of Data Centers in the Internet-of-Things Era

Prof. David Atienza,
Scientific Director of EcoCloud Center and
Head of Embedded Systems Lab. (ESL), EPFL

David.Atienza@epfl.ch

https://ecocloud.epfl.ch/

6th Intelligent Maintenance Conference 2022 (IMC 2022), Lausanne, Sep. 6-7, 2022

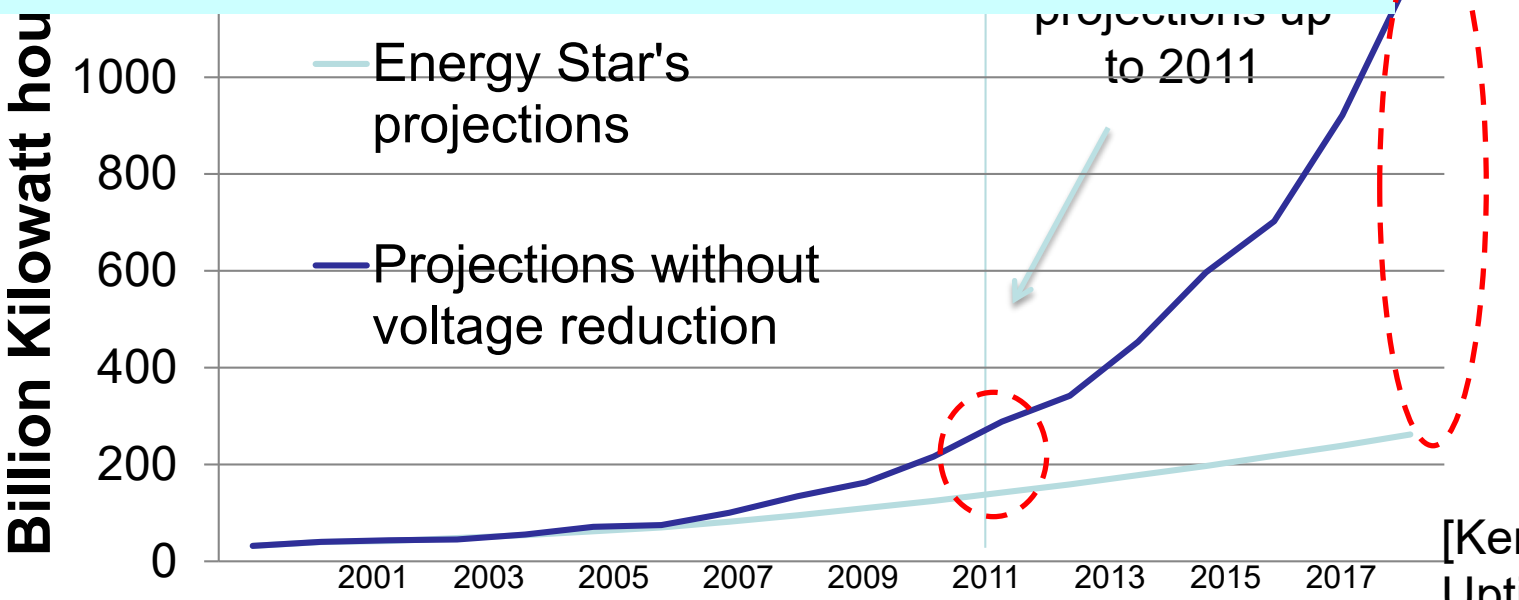
Current Information and Communication Technology (IT) Ecosystem

- ICT ecosystem encompasses:
  - Multi-level computing: edge, switching centers and cloud
  - Core part of computing at hyperscale data centers (DCs)

Why an energy problem? Power Density in Computing Systems: "Economic Meltdown of Moore's Law"

- 45 years of IT industry & Moore's law being discontinued
  - 2x more transistors every 2 years, quadratic energy reduction (voltage)
  - Voltages have leveled, ITRS projections in 2000 for voltage levels in 2011 were 40% lower…

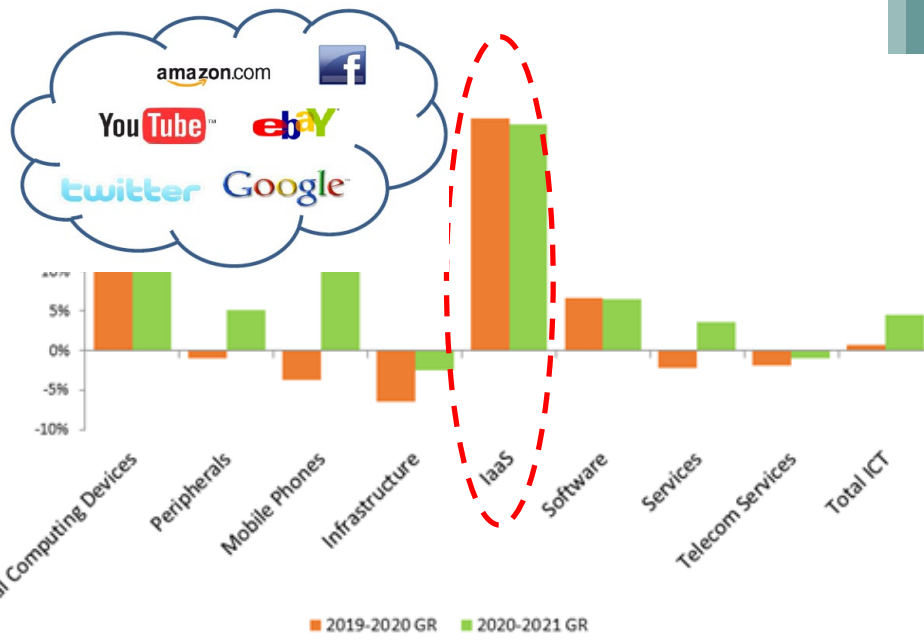
Dramatic increase in energy usage every new IT generation (80% more than expected today!)



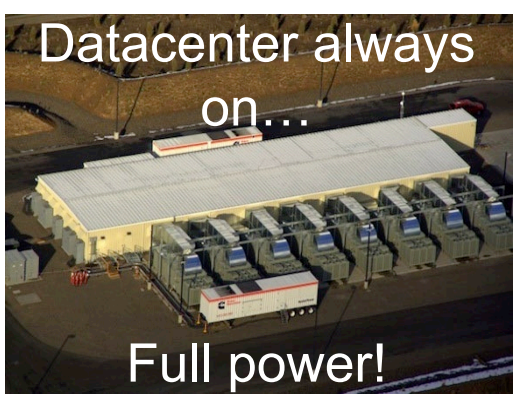
[Kenneth Brill, Uptime Institute]

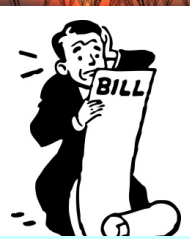
Ineffective Data Centers: Large Energy Expenditure… And Not Even Well Used!

- Cloud computing is… Growing always!
  - Amazon: $1 billion profits annually
  - Infrastructure as a Service (IaaS): Data Centers (DCs) = essence of Cloud

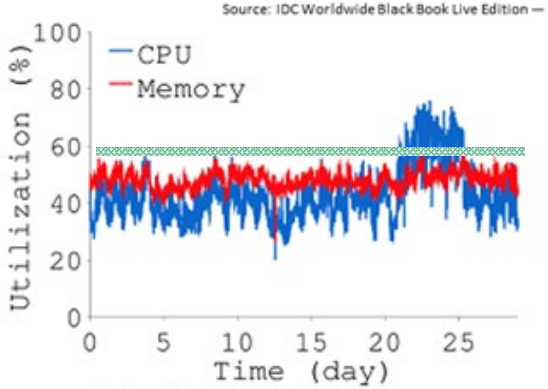


- Datacenters electricity can reach 8% worldwide electricity by 2030 [IDC]

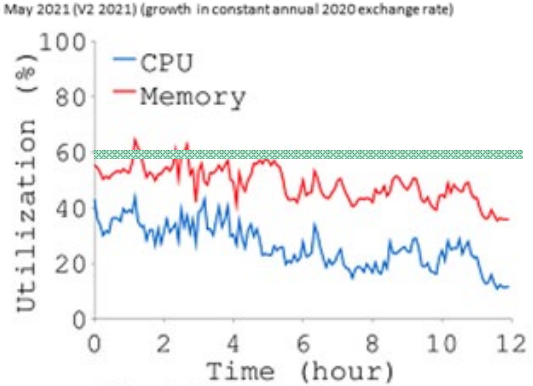




Source: IDC Worldwide Black Book Live Edition — May 2021 (V2 2021) (growth in constant annual 2020 exchange rate)



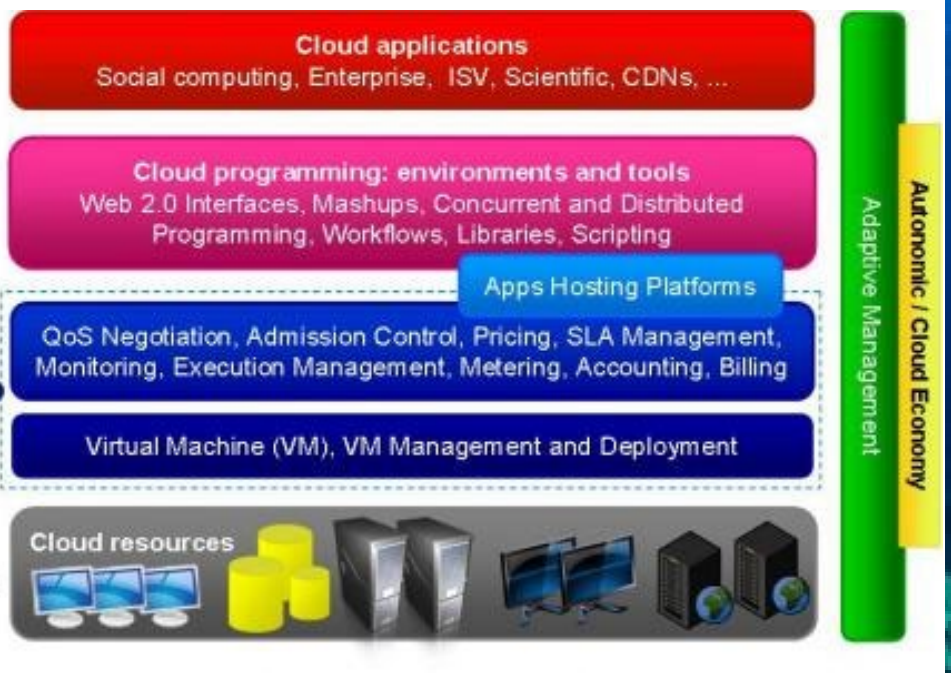
(a) Google Cluster

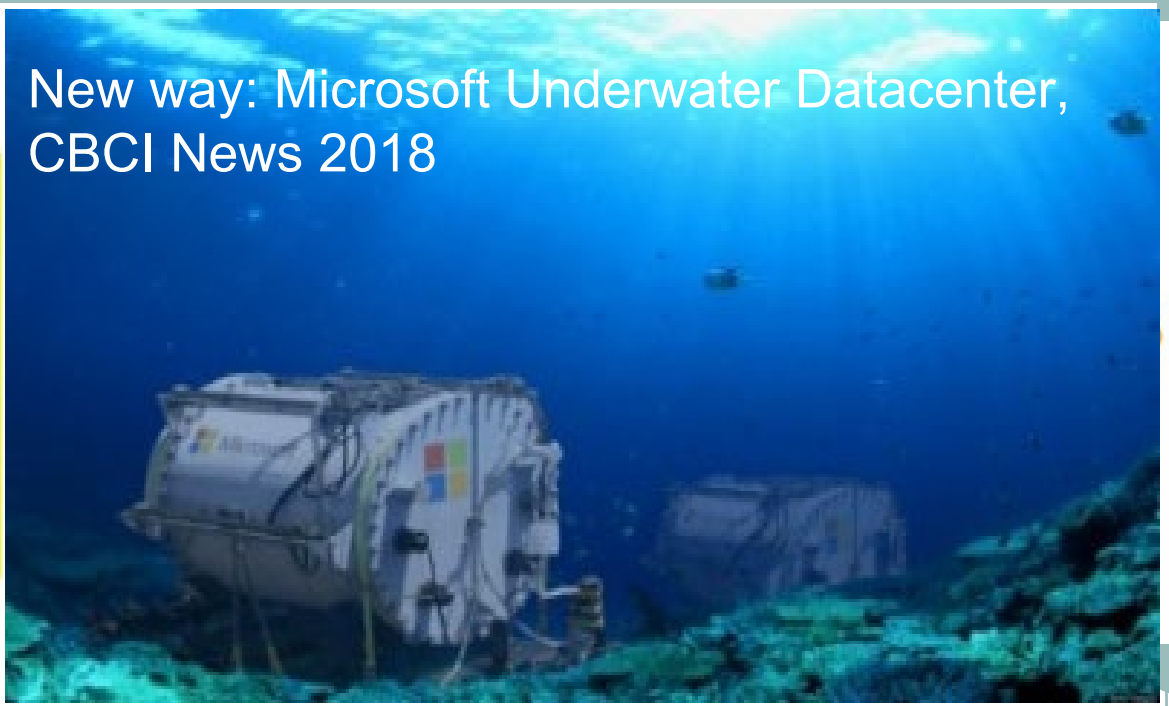


(b) Alibaba Cluster

But typically used at 60-65% of capacity!

Current DC Design Is Not Optimal

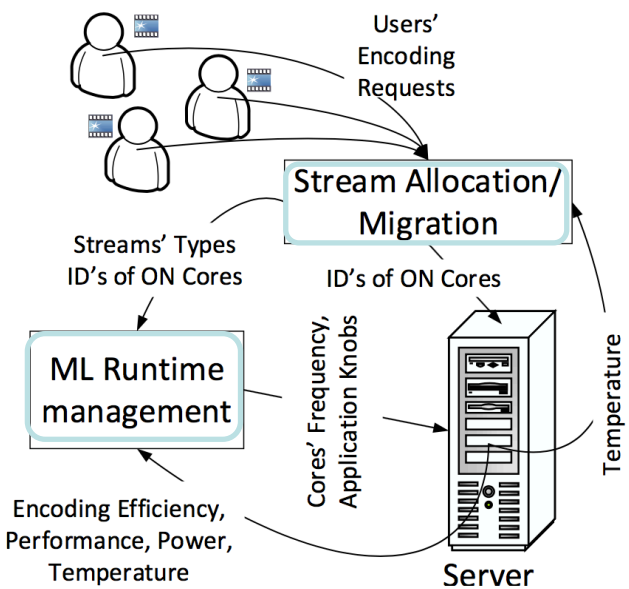




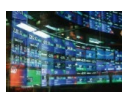
- Segregation of tasks: **multi-scale** modeling problem (components choice, applications, etc.)
  - Not clear communication of knowledge from one layer to another
  - IT equipment and cooling (over)provisioned separately to increase resilience
- Current DCs are **energy inefficient**
  - Power-cooling need as much energy as IT equipment… and thrown away
  - **For an 18 MW datacenter : Up to US$ 2.5M wasted per year!**

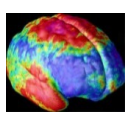
YINS: Thermal-Aware Design for Next Generation Energy-Efficient Data Centers

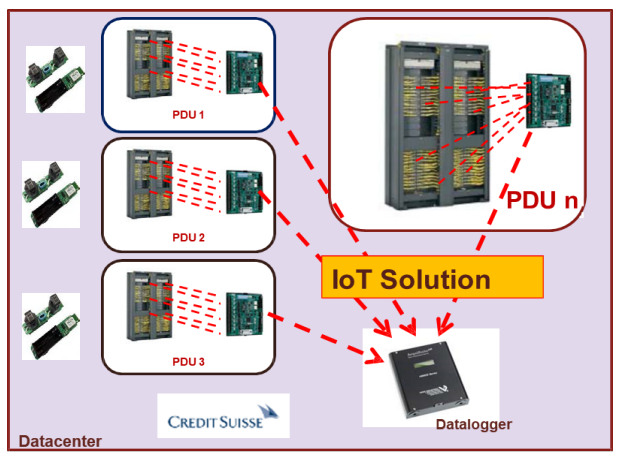
1. Use of Internet of Things (IoT) to monitor computing infrastructure: PMSM
2. Thermal/Power/Faults maps: 3D-ICE open-source multi-scale compact simulator
3. Two-step optimization approach
   a) TheSPoT: Thermal Stress-Aware Power and Thermal Management
   b) MAMUT: ML-based runtime management using multi-agent reinforcement learning (RL)



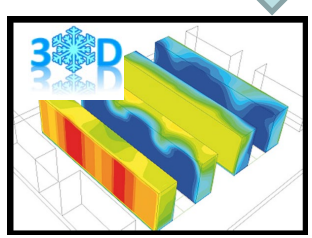
Phase 1) Monitoring Arrival of Applications

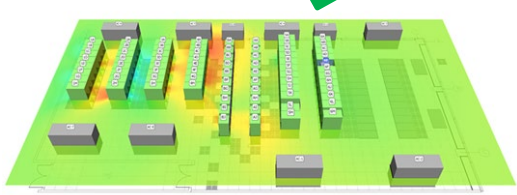

Drugs Simulat.


Medical Analytics



Phase 2) Power and Temp. Monitoring Tool





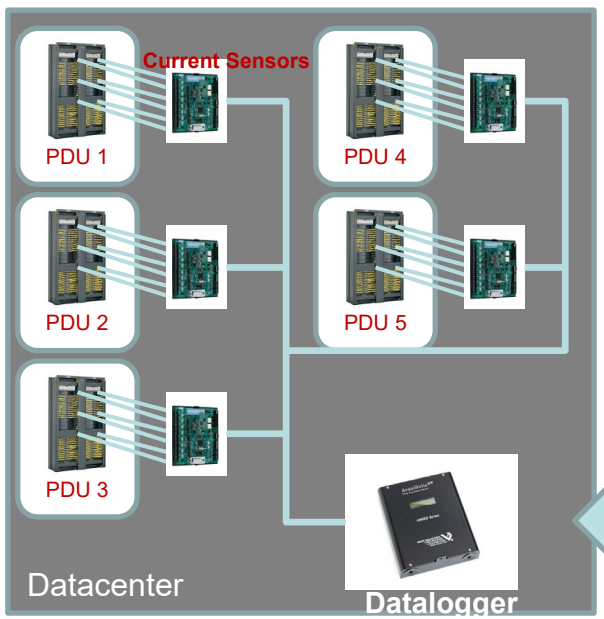
Phase 3) Machine Learning Based Run-Time Management

ETH zürich
EATON
CREDIT SUISSE

Power Monitor System and Management (PMSM)

- Real-time monitoring and automatic alerts for 3 DCs (>7,000 servers)
  - Servers monitoring using on-board sensors
  - Rack: extra IoT sensors for temp., humid. and vibrations

- System-level data center modeling/maintenance
  - Historical servers power consumption analysis: daily, weekly, monthly, …
  - Faults detection: rack circuit breaker failure, power overload, “zombie” servers,..



In cooperation with CREDIT SUISSE

See info at: http://esl.epfl.ch/page-57400-en.html

PMSM: Data collector and Analysis server

PMSM: User Management Interface

Network

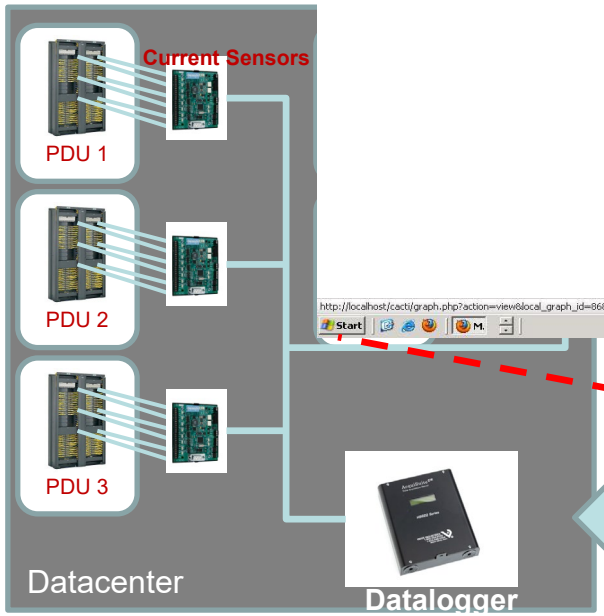
Power Monitor System and Management (PMSM)

- Real-tim
  in place
  - Server
  - Rack:
- System-l
  - Histori
  - Faults

Graphical energy analysis of selected racks in a datacenter

Energy evolution in time (up to 1 year) for specific PDU

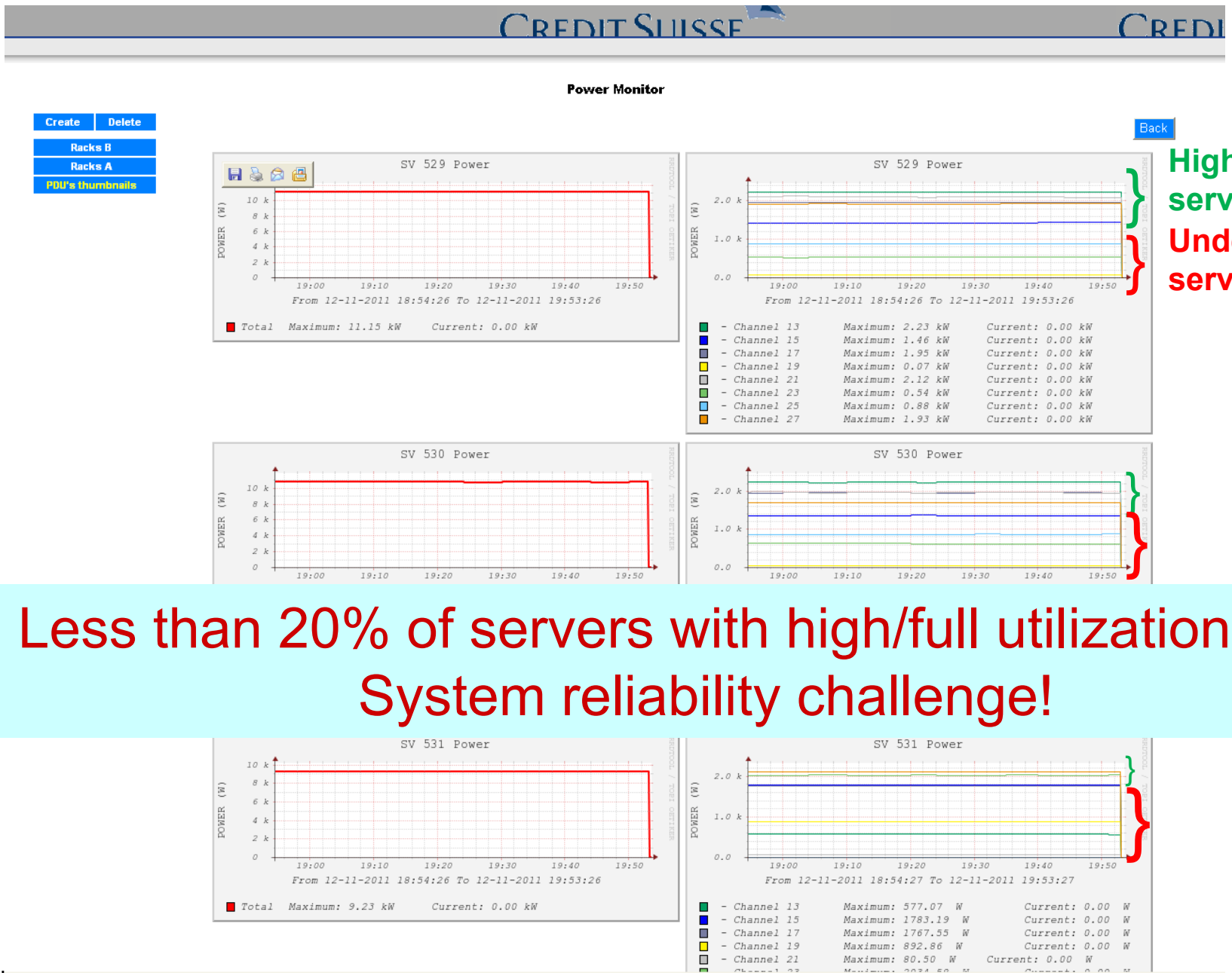
Power variation in time for in each PDU of selected rack



Network

PMSM: User Management Interface

Example Case Study: Detailed Energy Use View for Different Racks in one Credit Suisse DC



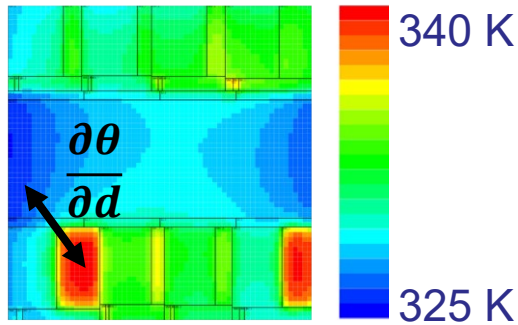
Highly Utilized servers

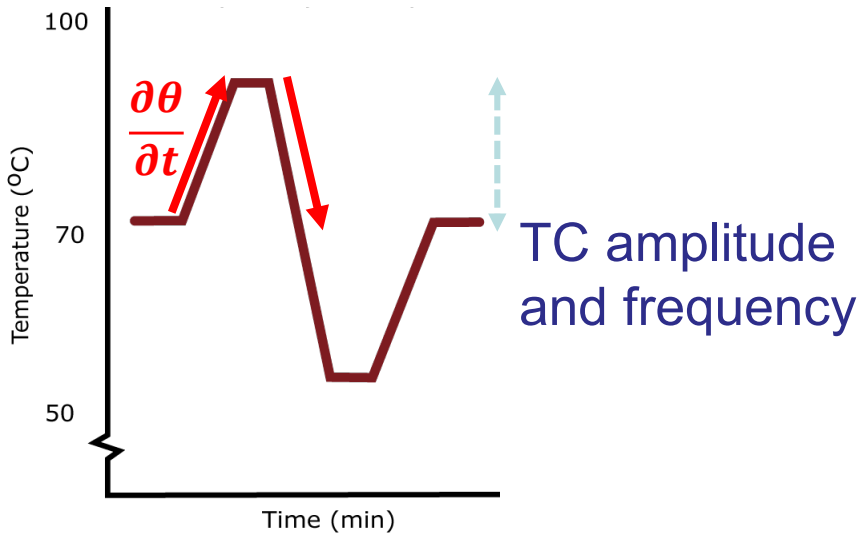
Under-utilized servers

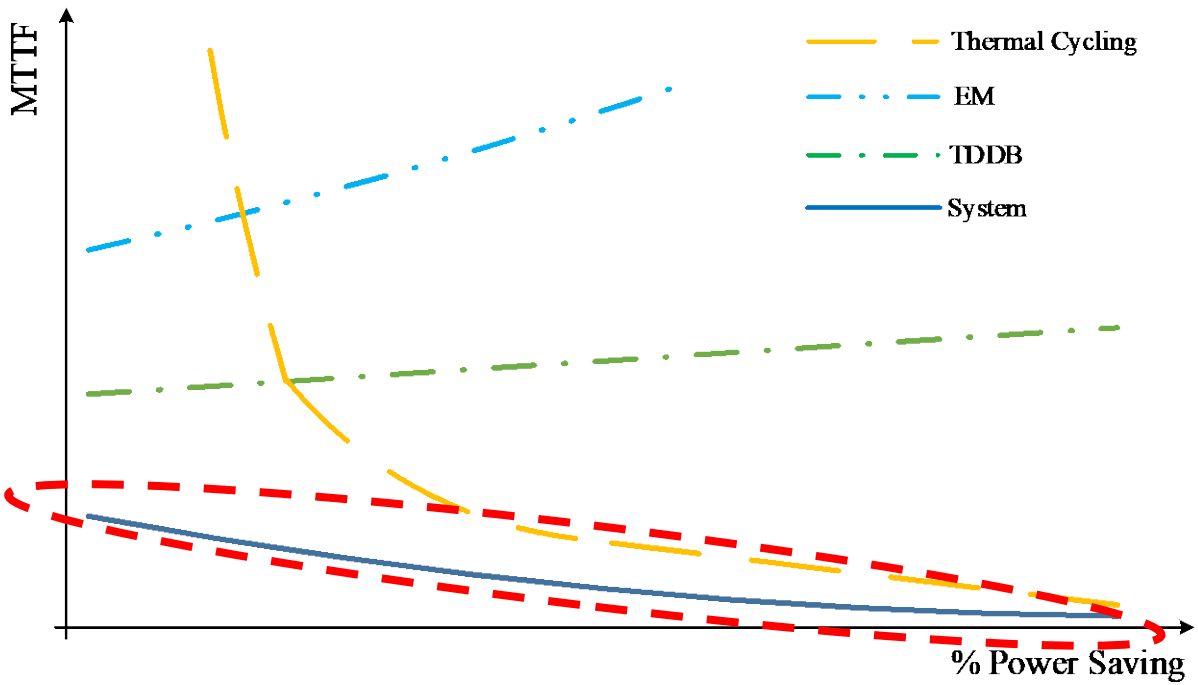
Less than 20% of servers with high/full utilization, why?
System reliability challenge!

Mean Time To Failure (MTTF) Factors: Reliability Analysis for Servers

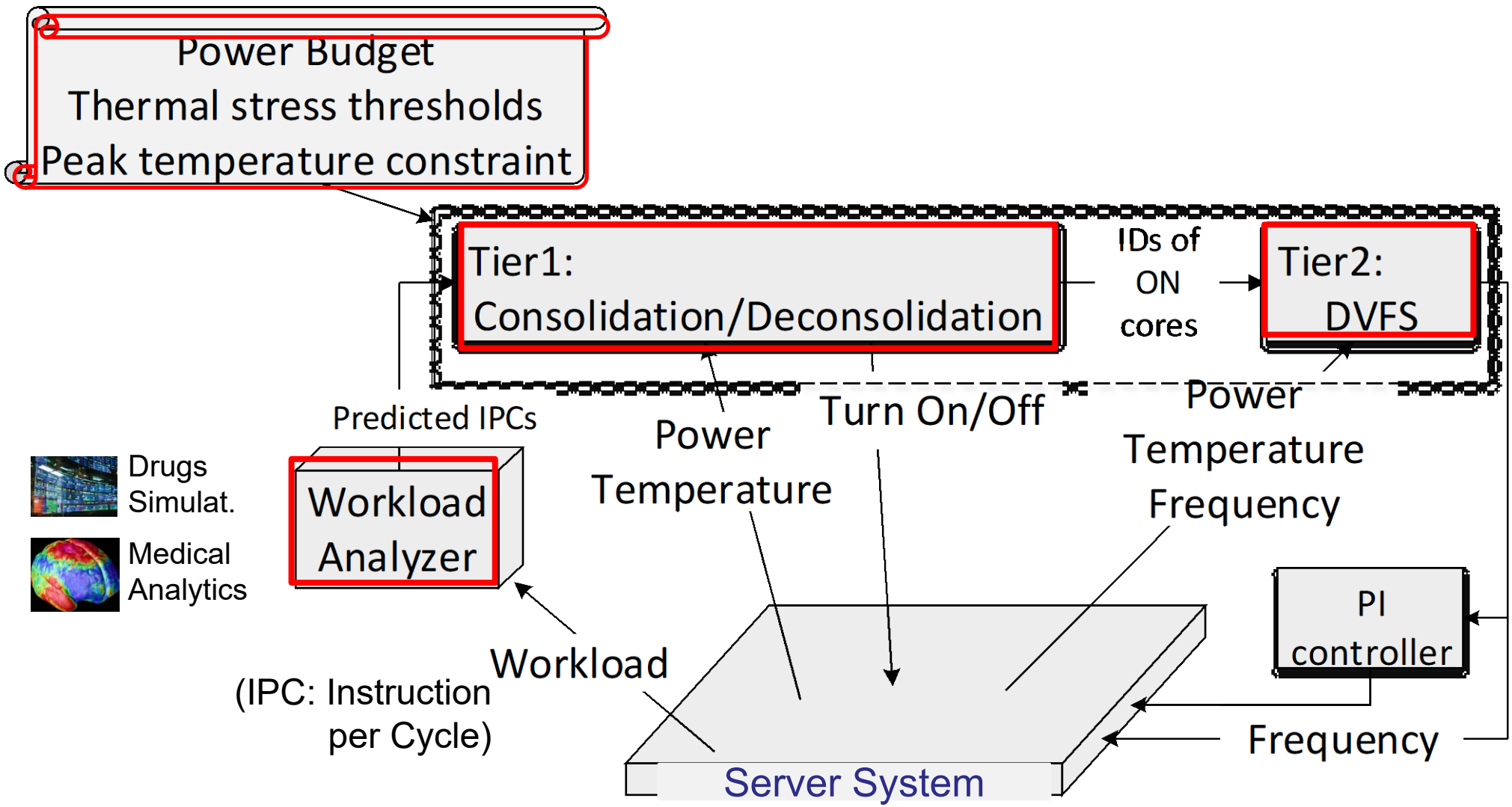
- Traditionally hot spots (peak/average) temperature in CMOS-based electronics
  - Mechanisms: Time-Dependent Dielectric Breakdown (TDDB), Electromigration (EM), etc.
- New Mechanisms in servers: Thermal Stress
  - Spatial Thermal Gradient (STG)
  - Temporal Thermal Gradient (TTG)
  - Thermal Cycling (TC)





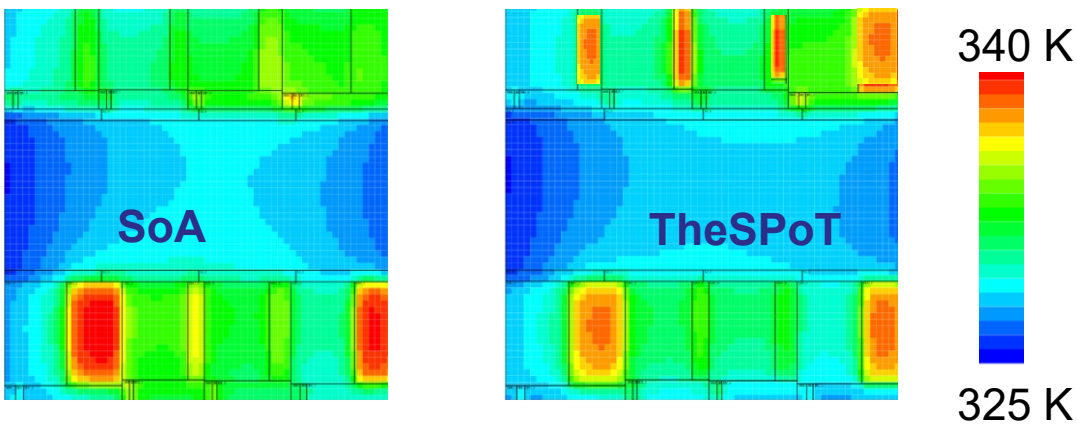


TheSPoT: Thermal Stress-Aware Power and Thermal Management



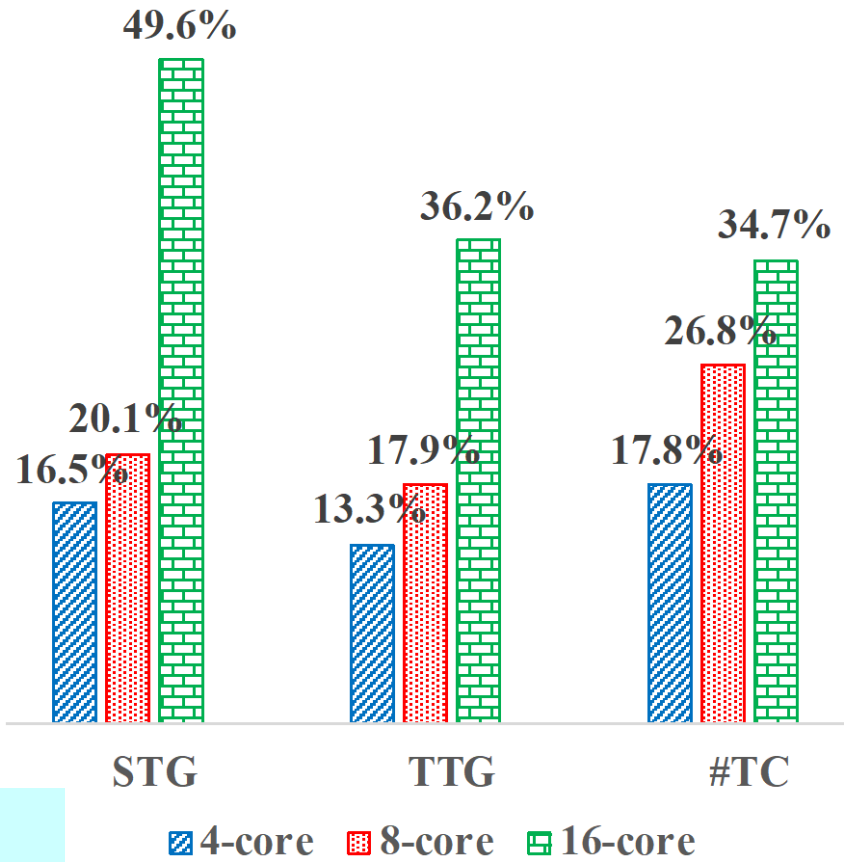
Results: TheSPot vs State-of-the-Art (SoA)

- Slightly higher total power in server
  - Thermal gradient less than 8ºC



8-core server thermal modeling

Improvement (%) achieved compared to SoA



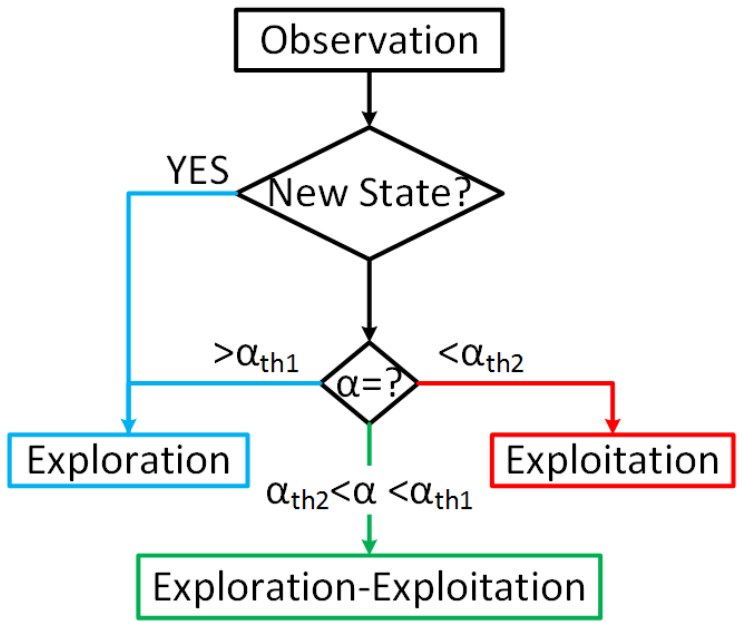
MTTF improved by 49%!
But how do we improve utilization?

MAMUT: Multi-Agent Reinforcement Learning for Efficient Real-Time Multi-User Tuning

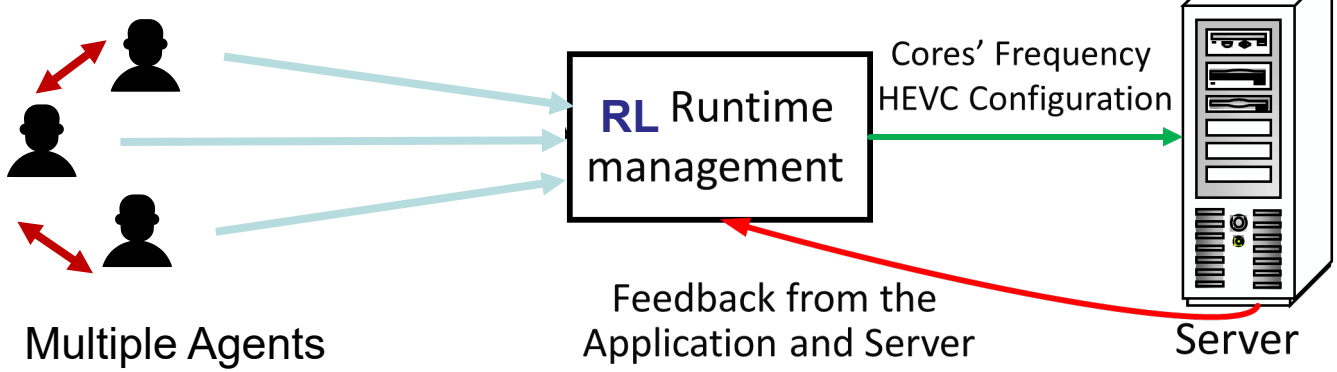
- Actions (a):
  - Application: QoS, CU
  - Processor: Frequency
- States (s):
  - Application: PSNR, Bitrate, Op/s
  - Processor: Power, Temperature
- Formulation: $\alpha_t(s_t, a_t) \stackrel{\text{def}}{=} \text{Learning Rate}$

$$\alpha_t(s_t, a_t) \propto \frac{1}{N(s_t, a_t)}$$

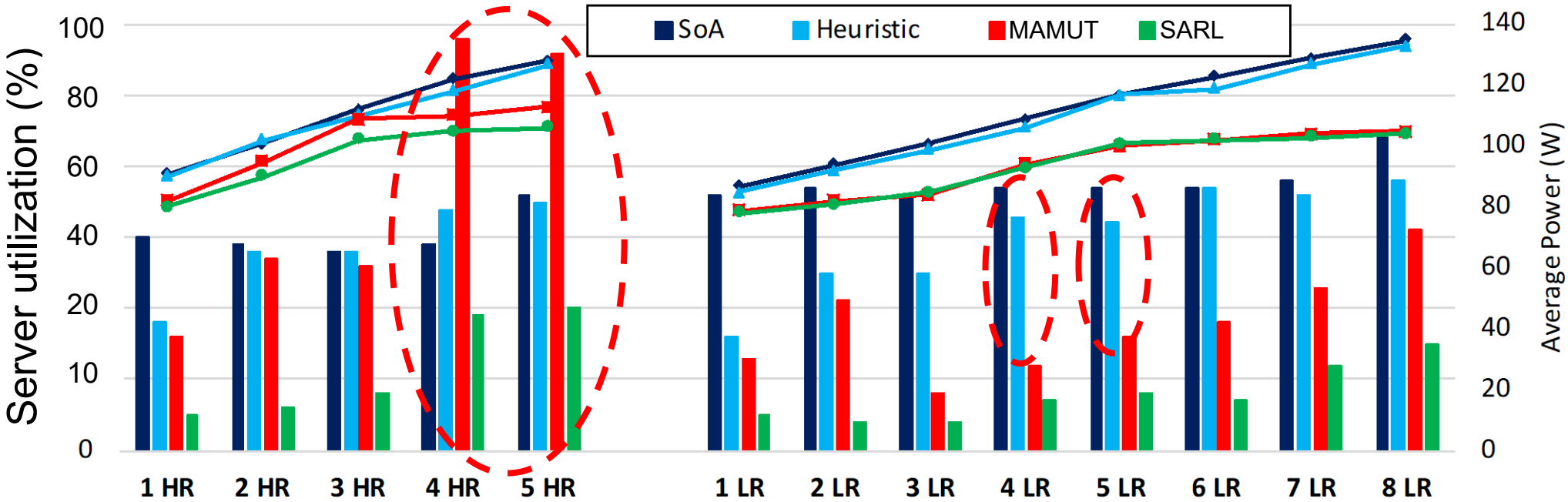
Number visits



- 3 agents:
  - DVFS,
  - Quality Proc. (QP),
  - Number threads



Results: MAMUT vs. SoA



Server utilization reaching 95-97% with more types of Heterogeneous Racks (HR) to manage

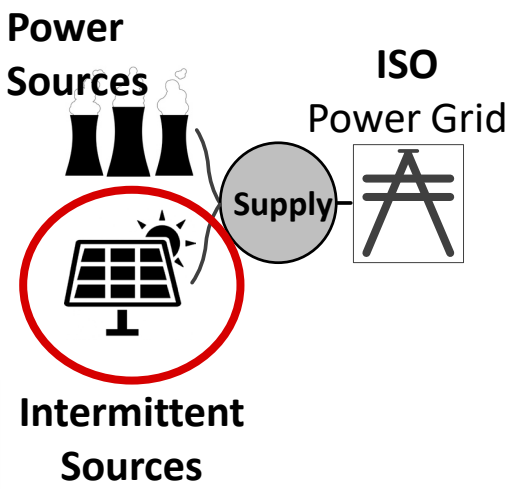
40-50% less power per server than SoA
(but with much higher resilience thanks to TheSpot!)

Conclusion

- "Green" (energy efficient) datacenters are key for ICT
  - Drastic increase of energy use on servers and cooling side
  - Sub-optimal design to prevent degraded performance/reliability
- Future: energy-reliability centric design = system-level
  - Multi-level modeling (utilization, energy, reliability, etc.)
  - Cross-layer optimizations
- IoT and Machine Learning to the rescue
  - PMSM: new sensors to understand DC operation and failures
  - TheSPoT: Thermal stress-aware power and thermal management to increase MTTF in servers (**up to 49% improvement**)
  - MARL: Multi-agent RL for efficient DC use (**up to 97% server use**)

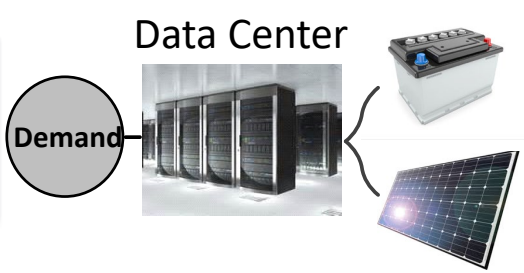
Future Challenges (1): Emerging Power Markets

- Integrating renewable energy on supply-side
  - EU Union targets to integrates over 20% share of renewables by 2025 [Bhringer 2019]



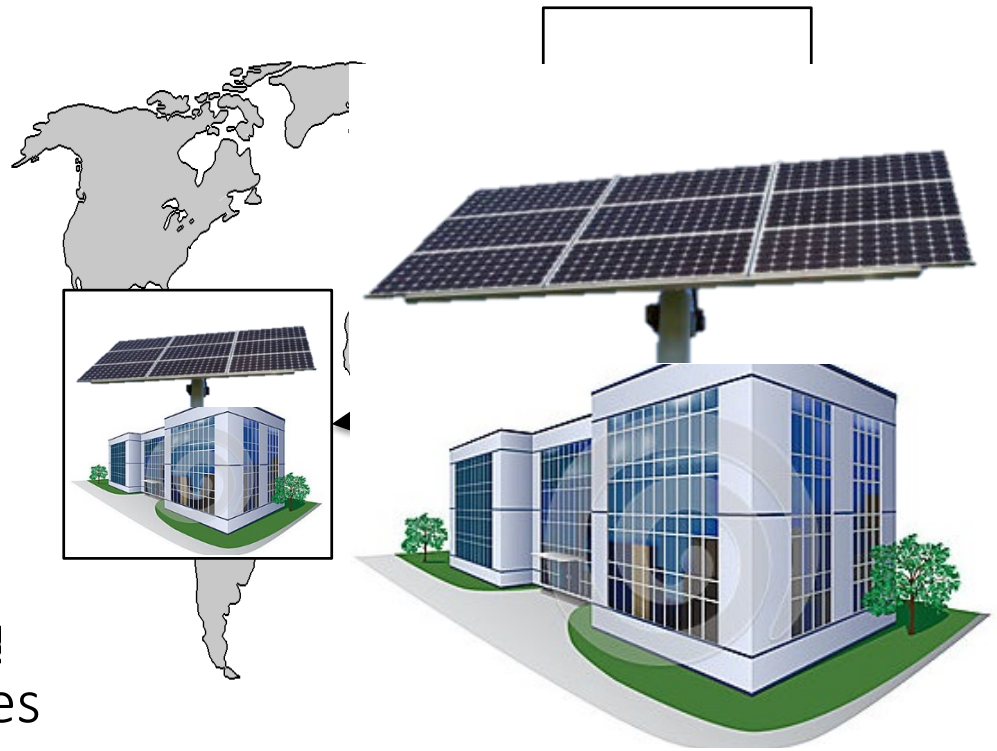
Challenge: high uncertainty to match supply and demand in the power grid (volatility and intermittency of renewables)

Goal: use demand-side capacity reserves, and dynamically adapt DCs power use when SLA allows it



Future Challenges (2): Modern Geo-Distributed and Green DCs

- **Geo-distributed DCs**
  - Multiple DCs in different locations connected through network
  - **Causes:** increasing users' demands, data protection policies (e.g., GDPR)
- **And they must be green DCs!**
  - Coupled with renewable energy sources in each different places
  - Different peaks for renewables



Conclusion

- "Green" (energy efficient) datacenters are key for ICT
  - Drastic increase of energy use on servers a
  - Sub-optimal design to prevent degraded pe
- Future: energy-reliability centric design
  - Multi-level modeling (utilization, energy, rel
  - Cros
- IoT and Machine Learning to the rescu
  - PMSM: new sensors to understand
  - TheSPoT: Thermal stress-aware po to increase MTTF in servers (**up to**
  - MARL: Multi-agent RL for efficient D

Included in Eaton's Intelligence Platform for data centers since 2021

Home › Green Tech › Power Saving

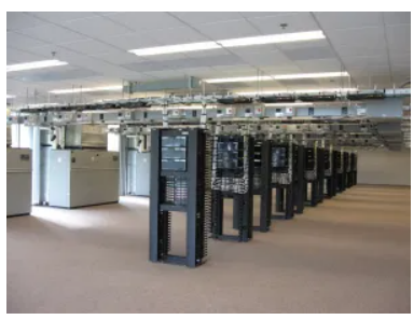
GREEN TECH POWER SAVING

Credit Suisse Zurich Datacenter Saves Up to 50% Electricity Thanks to Innovative Management System

By **Ovidiu Sandru** Modified date: May 30, 2017 120 1

Like 0

Share



An innovative method of saving the energy consume by data centers has been invented by EPFL scientists and applied for Credit Suisse in Switzerland. The solution, developed at the **Embedded Systems Laboratory at EPFL**, will save up to 50% of the energy currently used.

The **Power Monitor System and Management** uses a set of (probably Hall) sensors connected to the server racks' main power cables and measures the current passing through. Then, the consumption is logged and sent to a software feedback system that adjusts the load on each server.

"Two servers running at 40% of their capacity each, consume much more than only one at 80%," said David Atienza, ESL director.

The system has already been implemented on the racks of some 5,200 servers in **Credit Suisse**'s Zurich **datacenter**. The institution had been planning such a "virtualization" approach for about six years.

Thank You!
Questions?

david.atienza@epfl.ch

Acknowledgements:

Nano-Tera.ch Swiss Engineering Programme

Swiss National Science Foundation

European Commission

European Research Council

Key References and Bibliography (1/2)

- **Energy- and sustainability-aware design and scheduling of datacenters and Cloud computing systems**
  - D. Huang, A. Pahlevan, et al, “COCKTAIL: Multi-Core Co-Optimization Framework with Proactive Reliability Management”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (T-CAD)*, February 2022.
  - A. Iranfar, M. Zapater, and D. Atienza, “Multi-Agent Reinforcement Learning for Hyperparameter Optimization of Convolutional Neural Networks”, IEEE T-CAD, December 2021.
  - A. Pahlevan, M. Zapater, A. K. Coskun, and D. Atienza, “ECOGreen: Electricity Cost Optimization for Green Datacenters in Emerging Power Markets”, *IEEE Transactions on Sustainable Computing (T-SUSC),* April/June 2021.
  - W. Simon, Y. Qureshi, et al., “BLADE: An in-Cache Computing Architecture for Edge Devices”, *IEEE Transactions on Computers (TC)*, February 2020.
  - K. Haghshenas, A. Pahlevan, M. Zapater, S. Mohammadi, D. Atienza, “MAGNETIC: Multi-Agent Machine Learning-Based Approach for Energy Efficient Dynamic Consolidation in Data Centers”, *IEEE TSC*, July 2019.
  - L. Costero, A. Iranfar, M. Zapater, et al., “MAMUT: Multi-Agent Reinforcement Learning for Efficient Real-Time Multi-User Video Transcoding”, *Proc. of DATE*, March 2019.
  - A. Iranfar, M. Zapater, D. Atienza, “Machine Learning-Based Quality-Aware Power and Thermal Management of Multistream HEVC Encoding on Multicore Servers”, *IEEE TPDS,* October 2018.
  - A. Pahlevan, X. Qu, M. Zapater, D. Atienza: “Integrating Heuristic and Machine-Learning Methods for Efficient Virtual Machine Allocation in Data Centers”, *IEEE T-CAD*, August 2018.
  - K. Kanoun, et al., “Big-Data Streaming Applications Scheduling Based on Staged Multi-Armed Bandits”, *IEEE TC*, Dec. 2016.
  - A. Pahlevan, P. Garcia, et al., “Exploiting CPU-Load and Data Correlations in Multi-Objective VM Placement for Geo-Distributed Data Centers”, *Proc. of DATE*, March 2016.
  - A. Pahlevan, J. Picorel, et al., “Towards Near-Threshold Server Processors”, *Proc. of DATE*, March 2016.

Key References and Bibliography (2/2)

- **2D/3D Thermal modeling and simulation for many-core SoCs**
  - Y. M. Qureshi, et al., “Gem5-X: A Many-Core Heterogeneous Simulation Platform for Architectural Exploration and Optimization”, *ACM Transactions on Architecture and Code Optimization (TACO)*, December 2021. (https://esl.epfl.ch/research/2d-3d-system-on-chip/gem5-x/)
  - F, Terraneo, A. Leva, et al., “3D-ICE 3.0: efficient nonlinear MPSoC thermal simulation with pluggable heat sink models”, *IEEE T-CAD*, December 2021, (http://esl.epfl.ch/3d-ice.html).
  - J. Ranieri, A. Vincenzi, et al., “Near-Optimal Thermal Monitoring Framework for Many-Core Systems-on-Chip”, *IEEE Trans. of Computers (TC)*, Nov. 2015.
  - S. Raghav, C. Pinto, et al., “GPU Acceleration for simulating massively parallel many-core platforms”, *IEEE TPDS*, Vol. 26, Nr. 5, pp. 1336-1349, May 2015.
  - A. Sridhar, et al., “A Semi-Analytical Thermal Modeling Framework for Liquid-Cooled ICs”, *IEEE T-CAD,* Aug. 2014.
  - A. Sridhar, et al., “Neural Network-Based Thermal Simulation of Integrated Circuits on GPUs”, *IEEE T-CAD*, Jan. 2012
- **Microfluidic based power-cooling delivery for 2D/3D MPSoCs**
  - H. Najibi, J. Hunter, et al., “Enabling Optimal Power Generation of Flow Cell Arrays in 3D MPSoC with On-Chip Switched Capacitor Converters”, *IEEE ISVLSI*, 2020.
  - A. Andreev, A. Sridhar, M. M. Sabry, M. Zapater, P. Ruch, B. Michel, and D. Atienza, “PowerCool: Simulation of Cooling and Powering of 3D MPSoCs with Integrated Flow Cell Arrays,” *IEEE Trans. on Computers*, 2018.
  - A. Seuret, et al., “Design of a two-phase gravity-driven micro-scale thermosyphon cooling system for high-performance computing data centers”, *Proc. of ITHERM*, June 2018.
  - A. Sridhar, M. Sabry, P. Ruch, D. Atienza, B. Michel, “PowerCool: Simulation of Integrated Microfluidic Power Generation in Bright Silicon MPSoCs”, *Proc. of ICCAD*, November 2014.