CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Cray Hybrid XC30 Installation– Facilities Level Overview

CUG 2014

Colin McMurtrie
Ladina Gilly
Tiziano Belotti
Swiss National Supercomputing Centre (CSCS)

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Overview

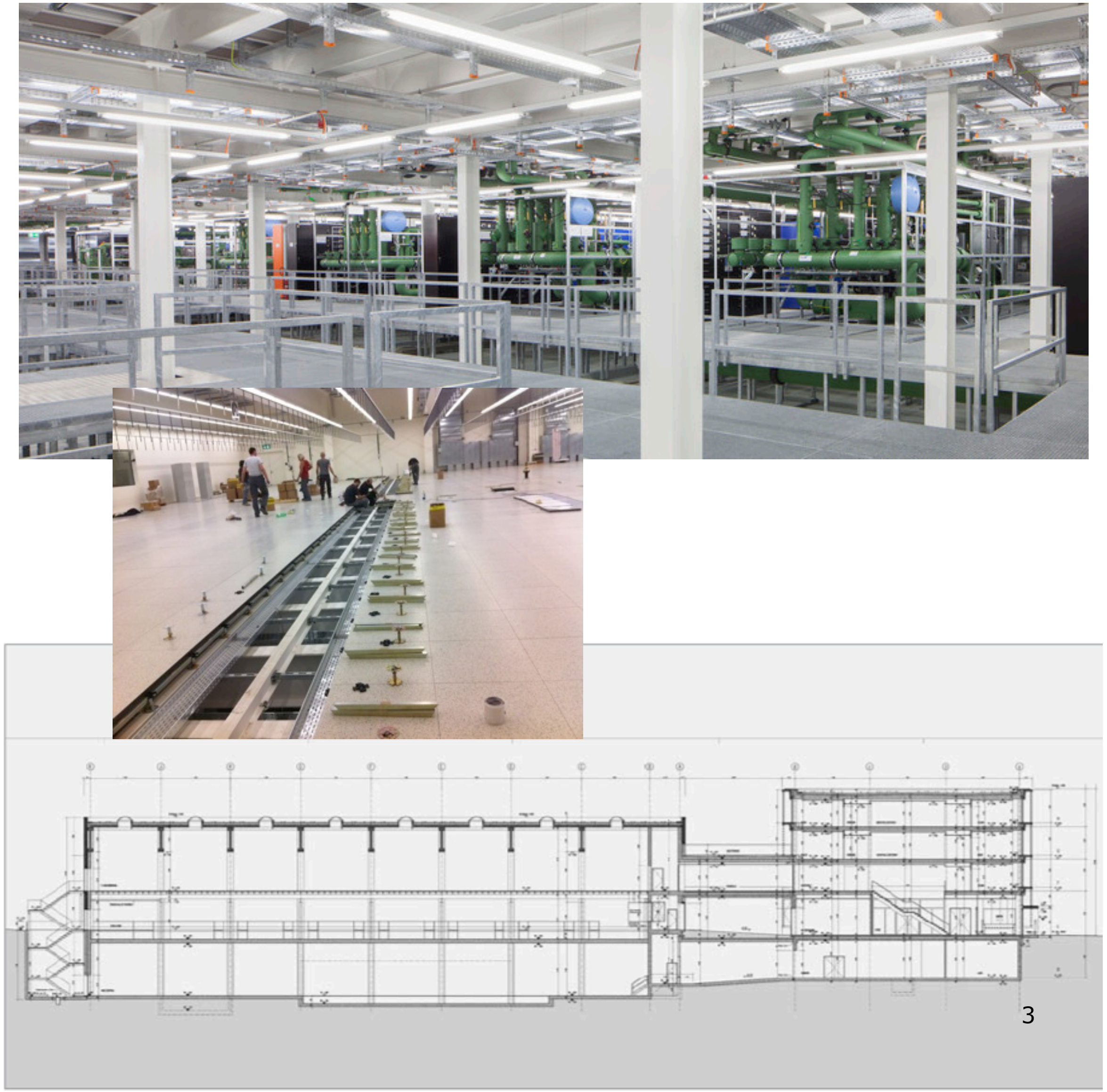
1. **CSCS Data Centre Overview**
    - Overview of Cooling Loops
    - Ambient Machine Room Environment
2. **Hybrid XC30 from a Facilities Point of View**
    - Cabinet Cooling
    - Electrical Supply & Control Elements
    - System Layout
    - Blade and Rack Layout
3. **Design of the Facilities Infrastructure for the System**
    - Secondary Cooling Loop Design
    - Electrical Distribution
4. **Installation Experience**
    - Pre-installation Data
    - Early On-site Testing and Facilities Changes
    - Main System Bring-up
    - Benefits of Secondary Loop Heat Exchangers
    - Monitoring of System Environmental Data

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

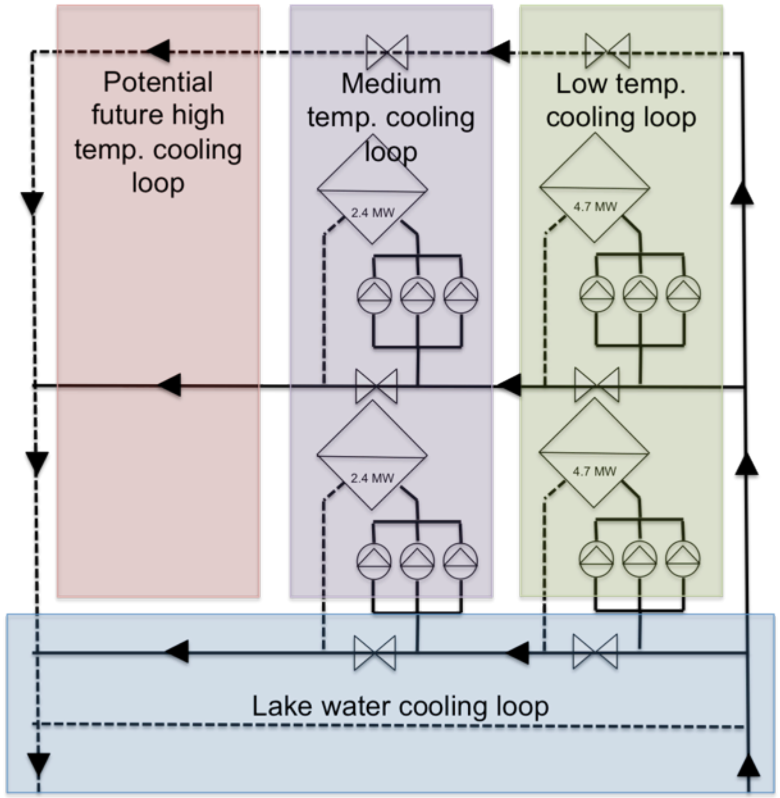
CSCS Data Centre Overview

- Water and electricity enter the building on the underground floor
- Installation Deck is where all secondary power and water distribution is done
- System-dedicated secondary cooling loops
- 5.5m deck height
- I-beam structure supports the machine room floor
- PDUs, Building Mgmt equipment, pumps, walkways above floor level to prevent damage in the event of a water leak
- Machine room is a contiguous space devoid of pillars



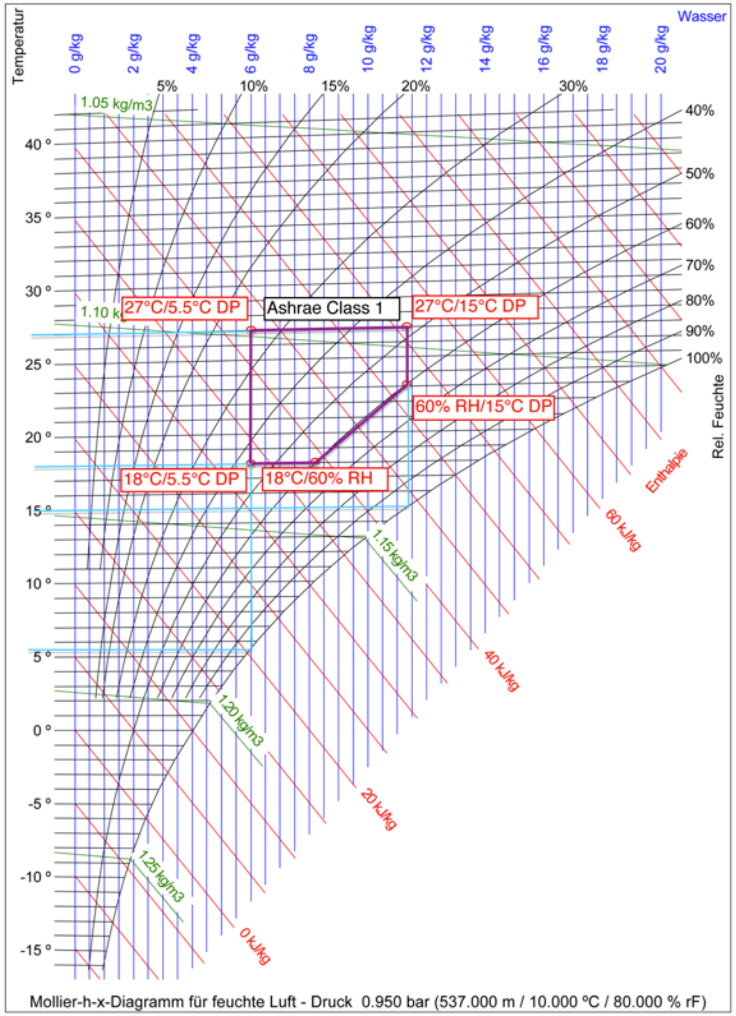
Lake Water and Cooling Loop Overview

- Water pumped from the lake at 7°C, year round
- Primary cooling loops exchange heat with lake water via large heat exchangers
- Primary loops are in series
- Low-temperature (TTN) loop cooled first
  - 9°C inlet, 17°C outlet
  - Bypass can be used if outlet >17°C
- Medium-temperature (MTN) loop cooled second
  - Receives lake water from the TTN outlet
- Lake water return controlled to not exceed 25°C
- N+1 Redundancy in MTN and TTN loops
- There is facility to
  - Add another MTN and TTN loop
  - Add a high-temp. loop (after MTN)
- All system-dedicated loops are attached to the TTN or MTN loops



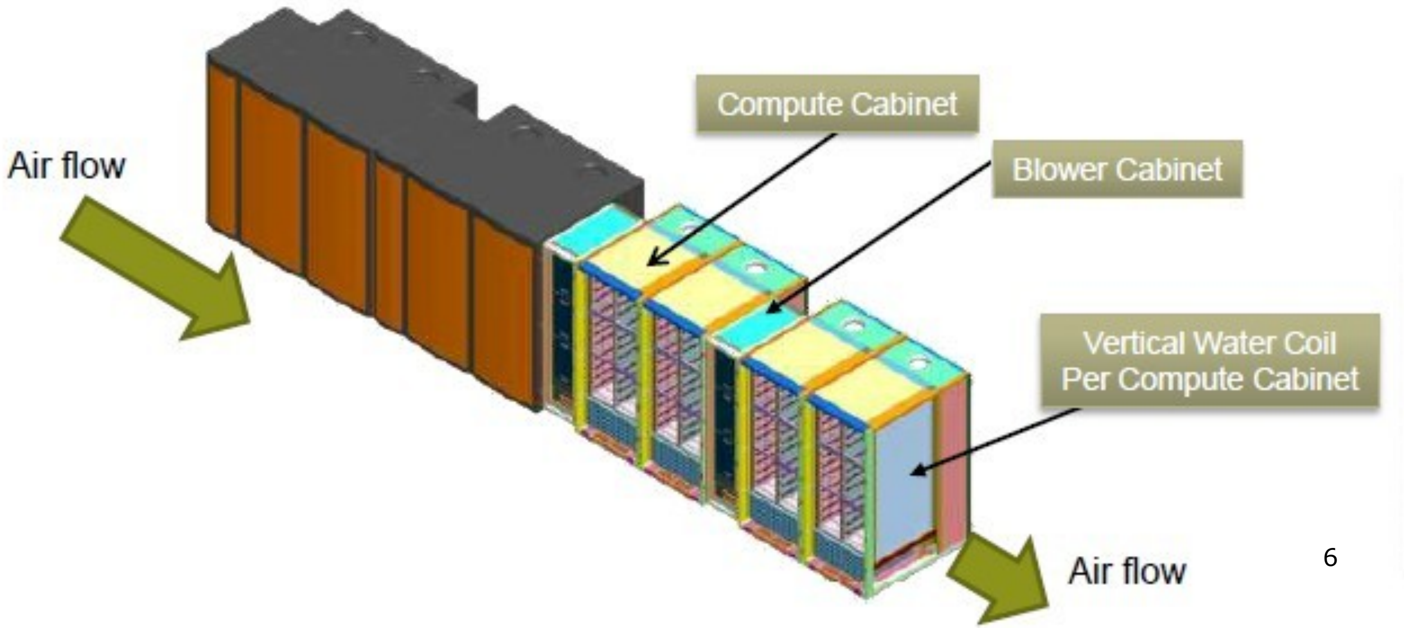
Ambient Machine Room Environment

- ➢ Not cooled in the traditional sense
  - Not plenum cooled
- ➢ The environment is controlled
- ➢ Complies with the ASHRAE Revised Class 1&2 Operating Range
- ➢ Temperature can vary
  - 18°C to 27°C
- ➢ Humidity can vary
  - Dewpoint:5.5°C to 15°C
  - We keep it below 11°C



XC30 Cabinet Cooling

- Down-row air stream cools system components
- Each cabinet has an air-to-water heat exchanger on the downstream side
- Two 2” water pipe connections per cabinet
- Blower cabinets contain 6 large fans (5.5kW at 100%)
    - At the start and (optionally) at the end of the row
    - After every 2nd cabinet within the row
    - N.B. No heat exchanger in the blower cabinets (only fans)
- Optional preconditioner available for the start of each row (conditions air for the first cabinet)



CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

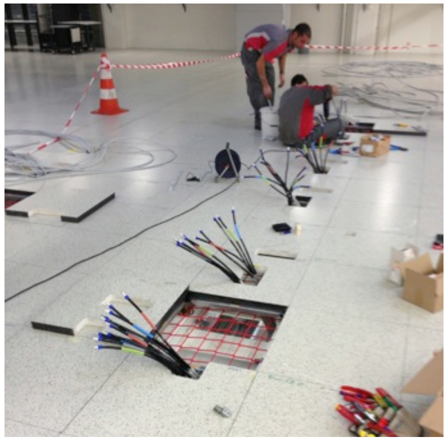
ETH zürich

XC30 Electrical Supply and Control Elements

- Each cabinet requires 2 three-phase supplies
- For 50Hz these supplies are:
  - 400/230VAC, 125Amp
  - WYE+Neutral+Ground
- Total of 10 conductors per cabinet

Control Elements

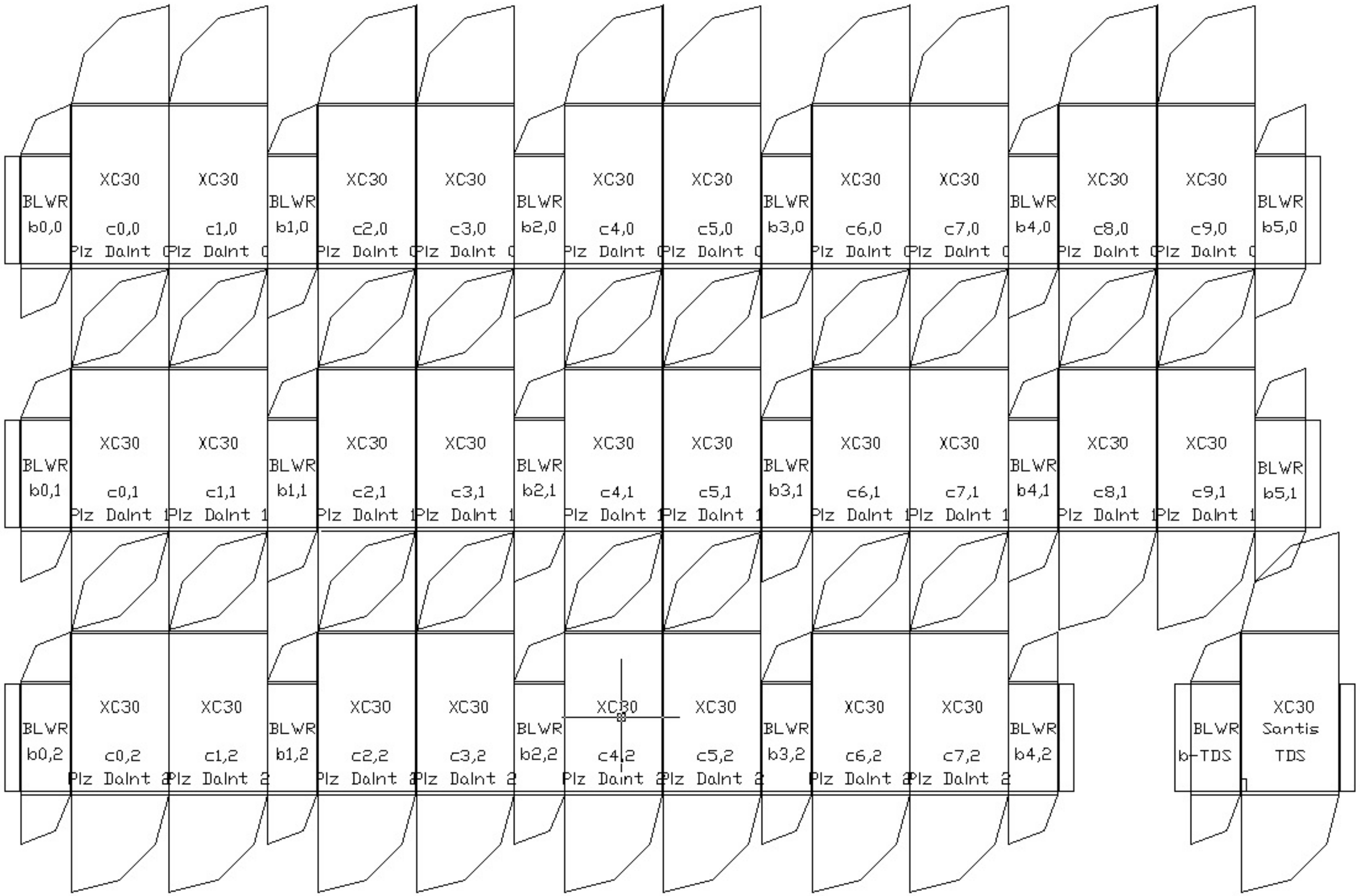
- Variable position water valve with electrical actuator
- Cabinet sensors (temperature etc; 9 total)
  - Downstream of each cabinet heat exchanger
  - 1 at the front, middle and rear of the cabinet
  - Vertically in the middle of each of the 3 chassis
- On-blade component temperature sensors
  - Monitored as part of the PMDB



CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

System Layout



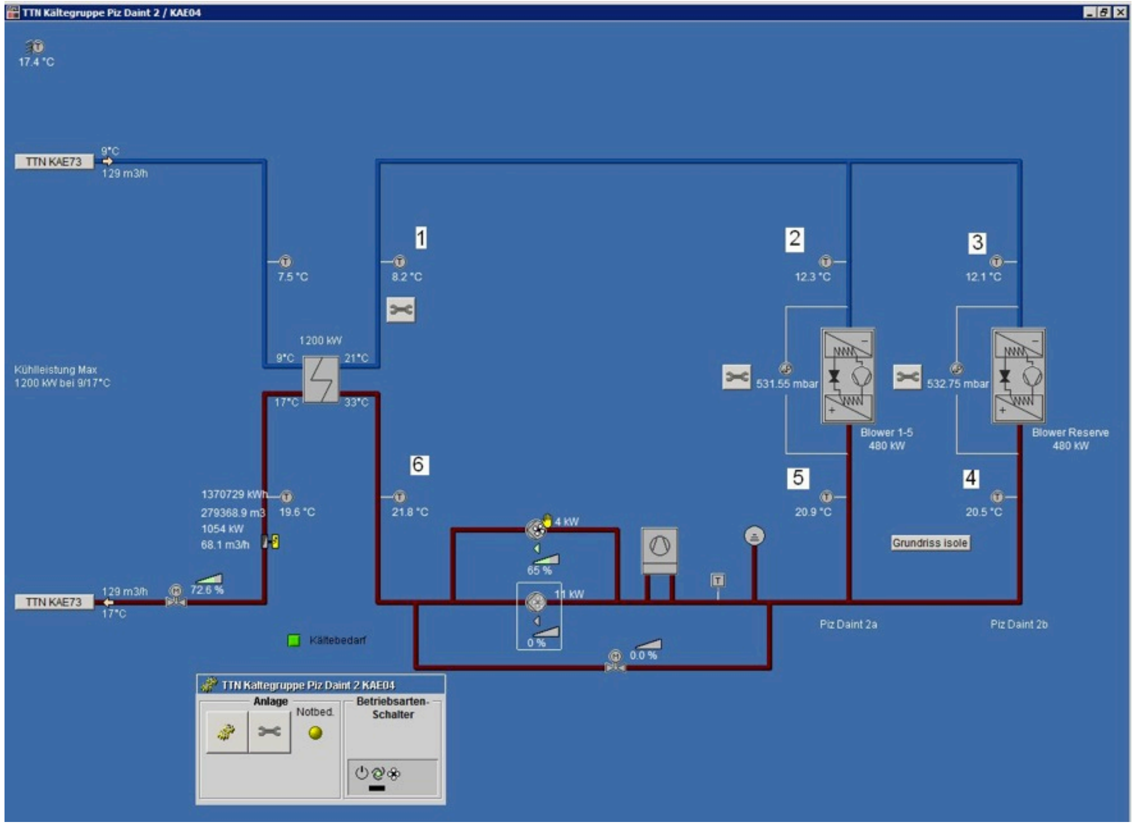
Front

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

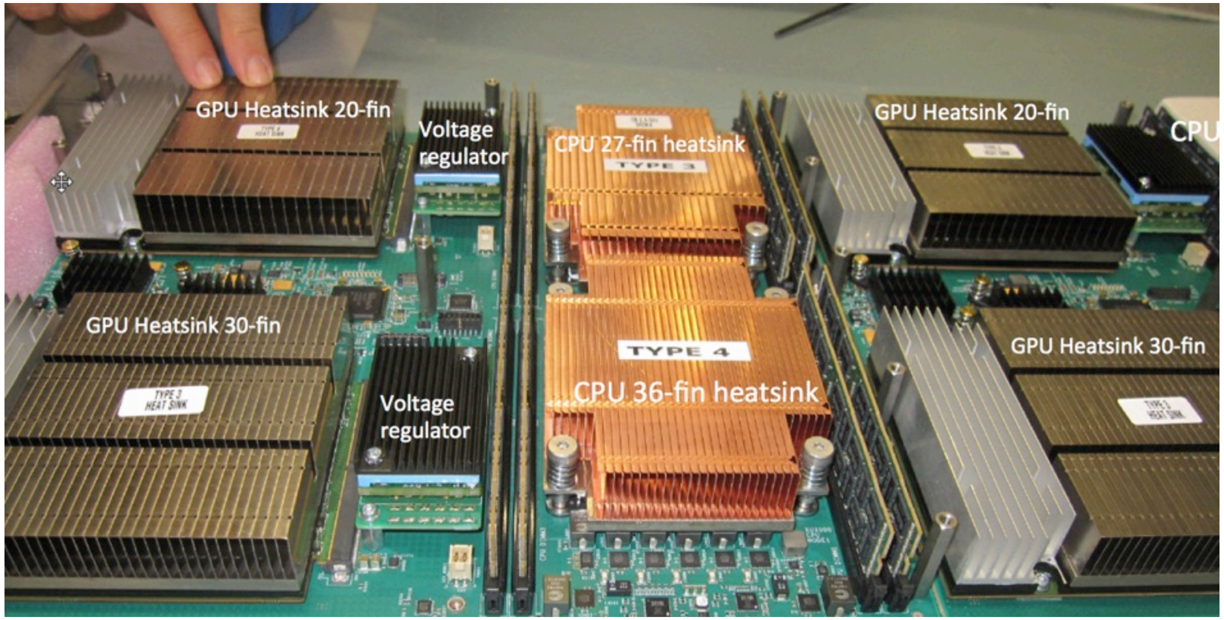
System-dedicated Secondary Cooling Loop Design

- Initial design had 3 system-dedicated loops
  - Front row connected to MTN loop
  - Middle and Back rows connected to TTN loop
- Closed loop control system
  - Varies pump speed to keep supply temp. within a specified range
  - Also monitors DeltaP across cabinet groups

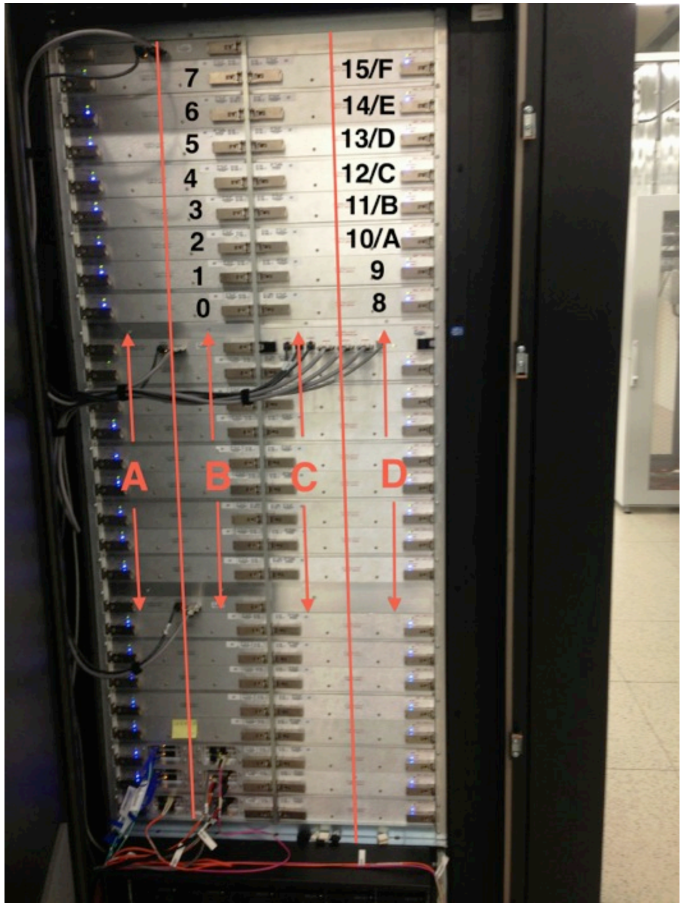


Blade and Rack Layout

- 3 chassis per cabinet
- 16 blades per chassis
    - Lefthand blades (Slots 0 to 7)
    - Righthand blades (Slots 8 to 15)
- Left- and Right-hand blades have different heat sinks
    - Gives rise to 4 distinct regions in the cabinet
    - Locations A to D (more on this later)

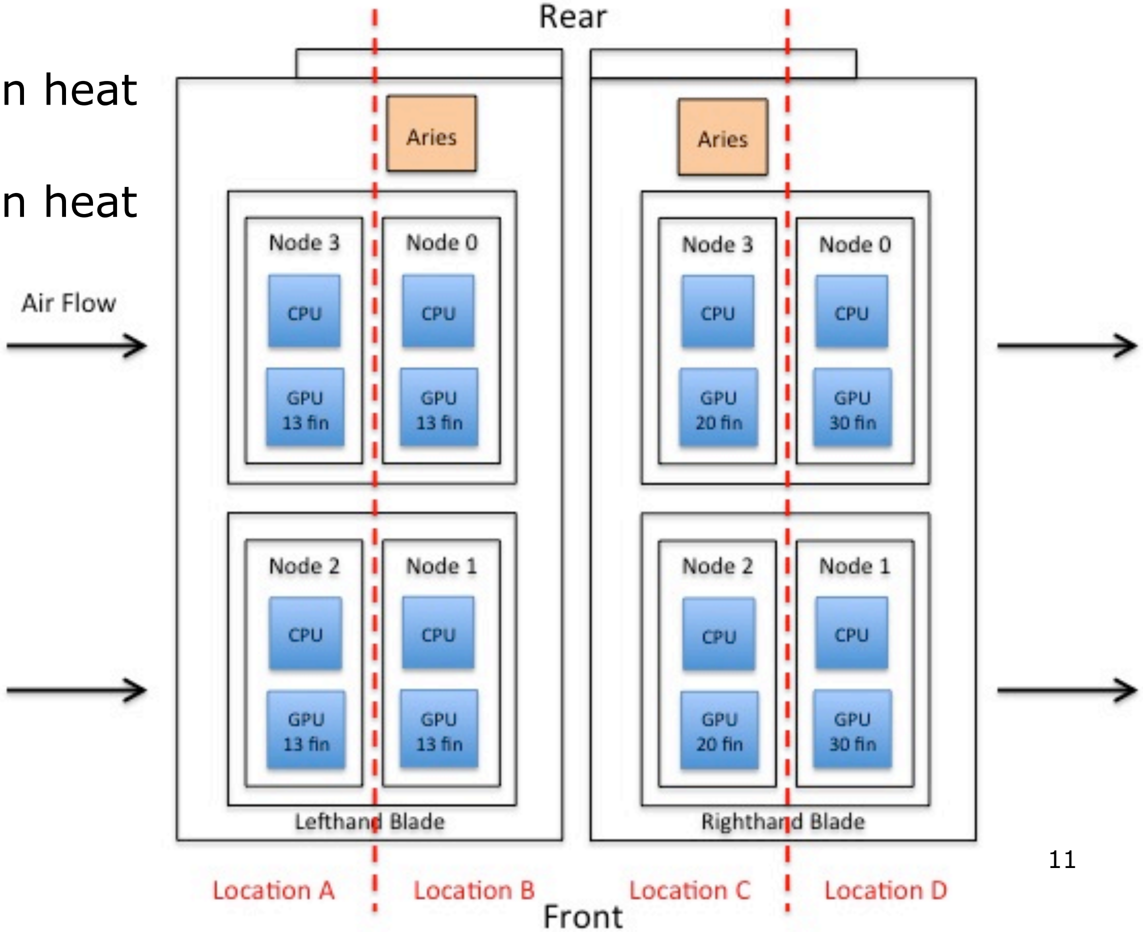


Righthand Blade



Blade Layout (cont.)

- Lefthand blades (Slots 0 to 7)
  - Nodes 0 to 4 have 13-fin heat sink on GPU SXM
- Righthand blades (Slots 8 to 15)
  - Nodes 0 and 1 have 30-fin heat sink on GPU SXM
  - Nodes 2 and 3 have 20-fin heat sink on the GPU SXM



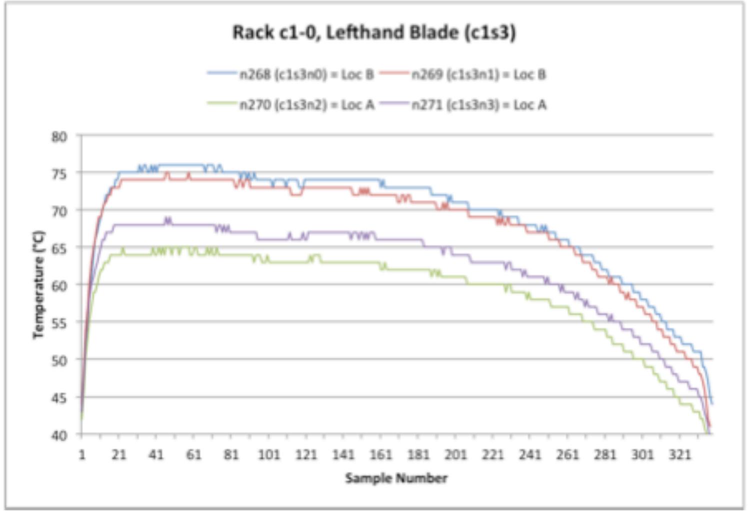
Early System Data

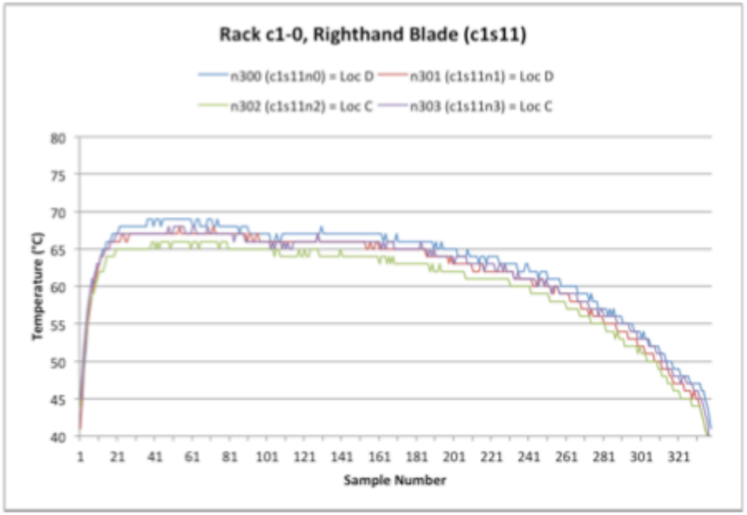
- Three cabinet *clogin85* test system at Cray showed that GPUs in Location ‘B’ ran hotter than GPUs in the other locations
- Cray determined that on average GPUs in Location ‘B’ ran 10C hotter
- Nvidia advised that there would also be some statistical variation in the temperature profile of the GPUs
- => Hot running GPU in the hot Location ‘B’ = thermal capping
    - Thermal capping halves the clock frequency of the GPU
        - Drops to 365MHz
    - This has a big effect on GPU floating-point performance
    - Consequently has a devastating effect on HPL performance

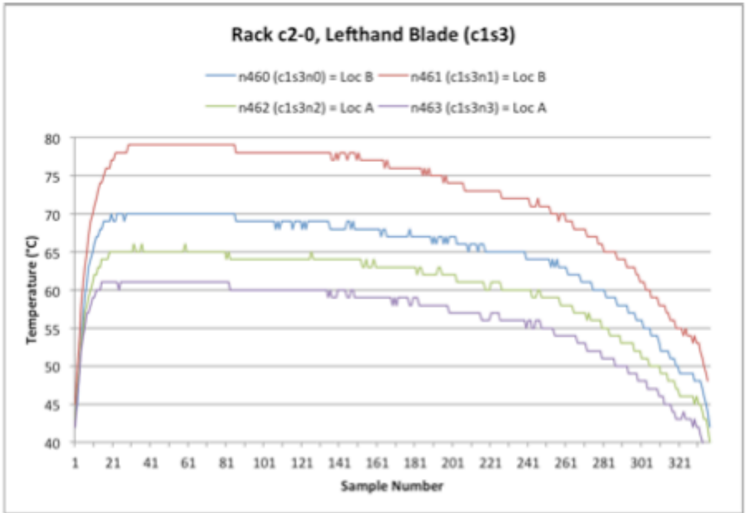
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

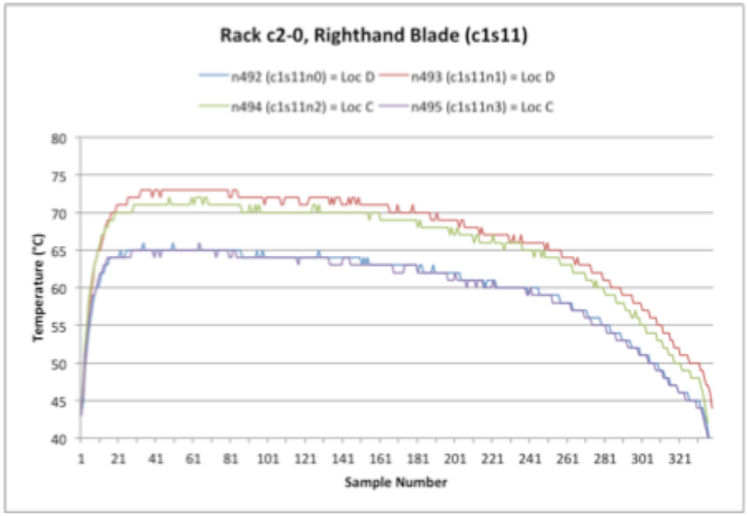
ETH zürich

Early System Data (cont.)









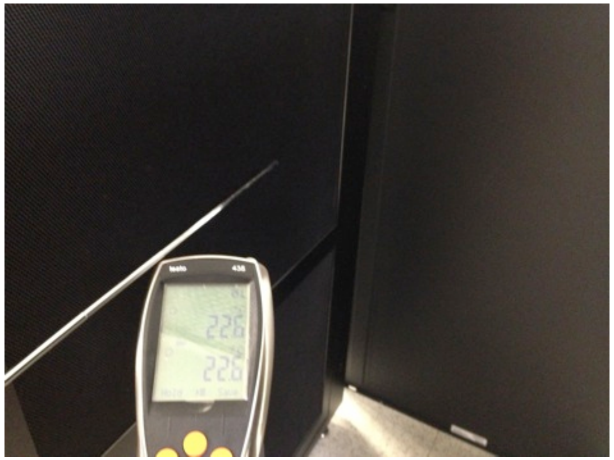
Results courtesy of Gilles Fourestey (CSCS)

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Early System Data (cont.)

- Fully populated TDS (*Santis*) at CSCS showed cabinet exhaust air temperature could not be controlled when running HPL
  - Cray's recommended air temp set-point was 19°C
  - With the temperature of the cooling water it was not possible to get the exhaust air temperature less than 3C above inlet air temperature
  - Also highlighted the potential need for preconditioners
  - Cray implemented an FCO to improve cabinet heat exchanger efficiency but this made little appreciable difference
- => Cooling loop water temperature needed to be lowered



CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Cooling Loop Changes

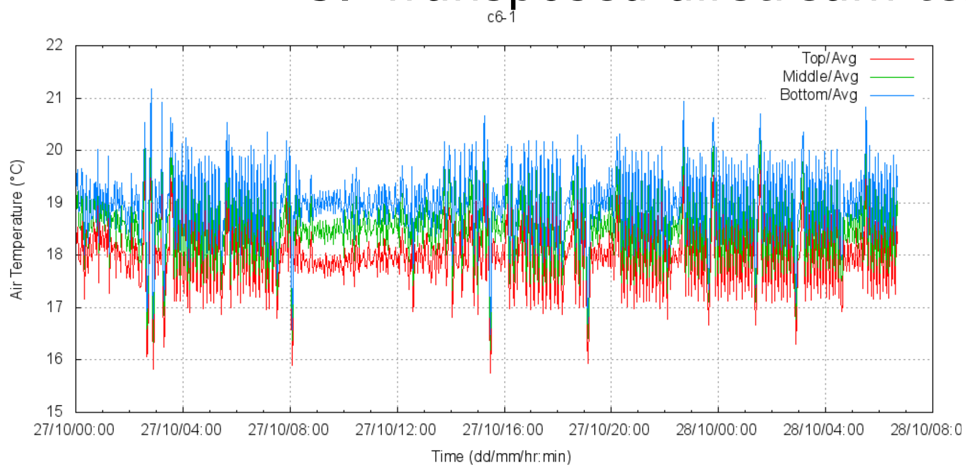
- The **front** row primary-side supply was moved from the MTN to the TTN primary cooling loop
  - Done easily with a small amount of pipework
    - 2 days from design to installation
  - No other changes necessary on this row
    - ✓ Pumps OK
    - ✓ Heat Exchanger OK
- The **middle** & **back** rows were already on TTN supply
  - Heat exchangers had been sized for much higher secondary-side temperature and were too small
  - 4-6 week delivery time didn't fit project timeline
  - => Remove the secondary heat exchangers
  - Done in less than a day
    - ✓ Included manual cross-connect valve
  - No other changes needed
    - ✓ Pumps OK
    - ✓ Control system OK (minor tweaking necessary)

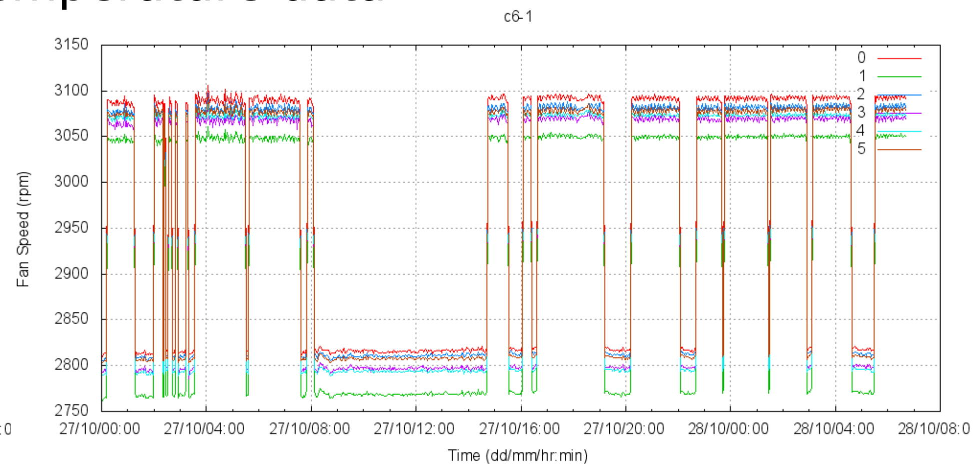
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Main System Bring-up

- ➢ Despite the cooling loop changes the installation remained on schedule
- ➢ Main system bring-up went smoothly
  - 19°C airstream set-point was achievable
- ➢ Monitoring of ccsysd cabinet environmental data showed some anomalies
  1. Cabinet Controller crashes and reboots
  2. Uncontrolled water valve position after CC crash
  3. Erratic water valve position
  4. Unreliable and erroneous cabinet inlet and outlet water temperatures
  5. Transposed airstream temperature data

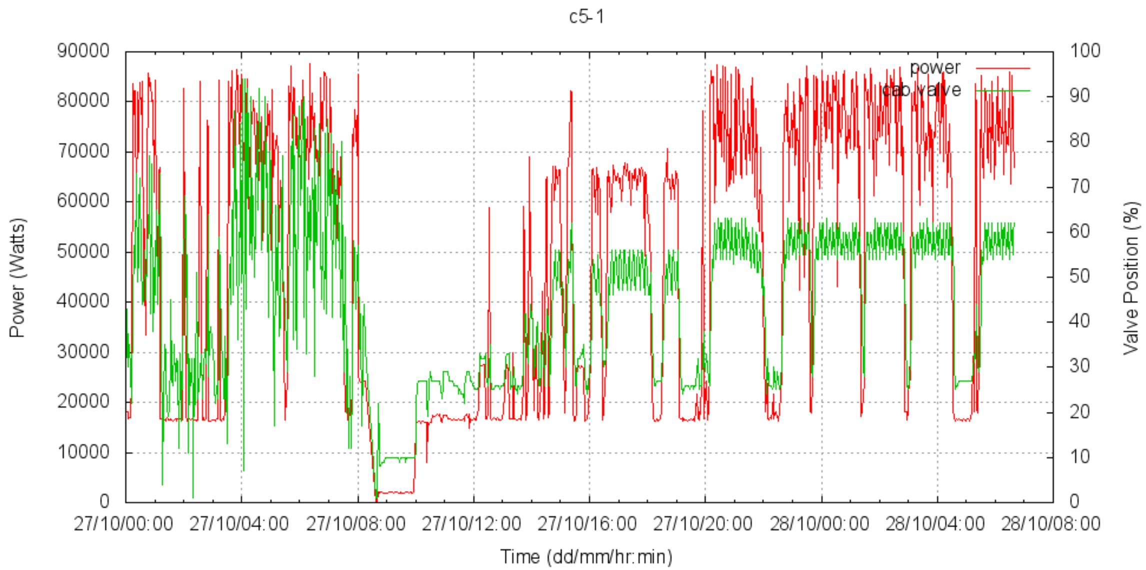




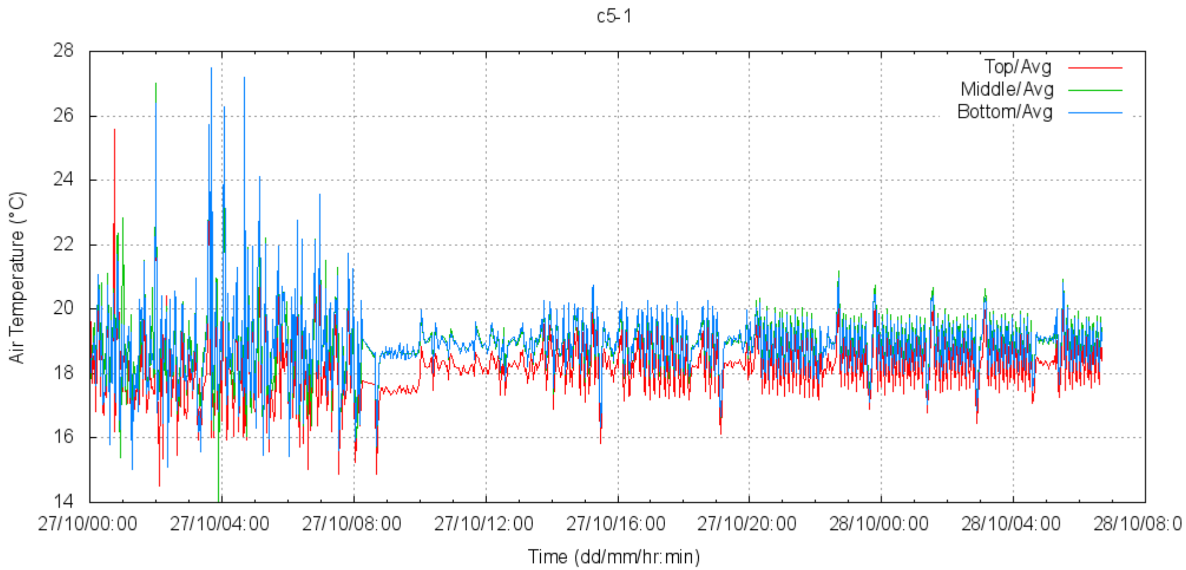
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Erratic Water Valve Position

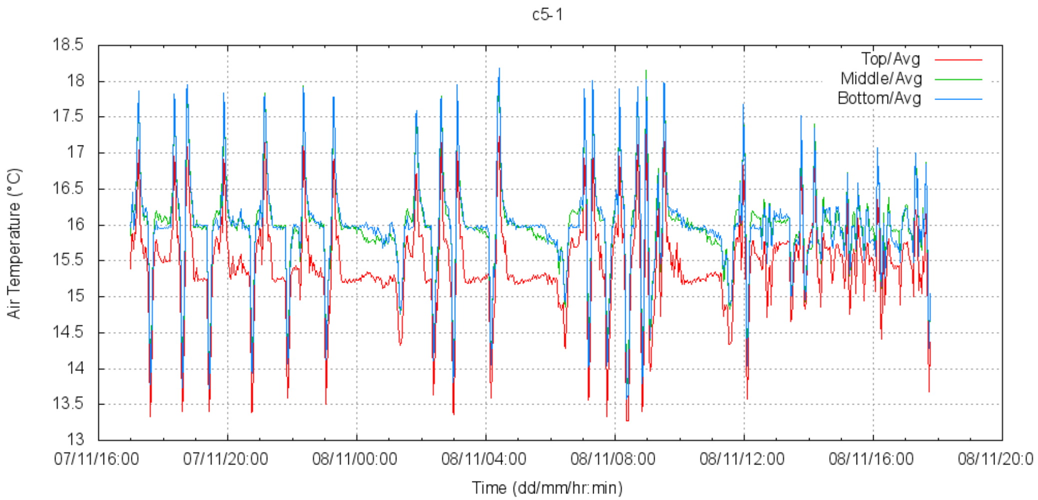


- ➢ Problem tracked to voltage overload on Cabinet Control Board (CCB)
- ➢ Due to sprung-loaded water valve actuators



HPL Data

- Data from early HPL runs showed GPUs (especially in Location 'B') thermal capping
- Decision was taken to lower the airstream set-point to 16°C
  - Substantial difference from Cray's initial design data
- Not a problem because the cooling infrastructure was able to accommodate the change easily following the cooling loop changes
- 14°C was tried as well but the machine seemed very unstable at this temperature
  - High number of HSN-related errors
- 16°C airstream set-point used for Top500 runs

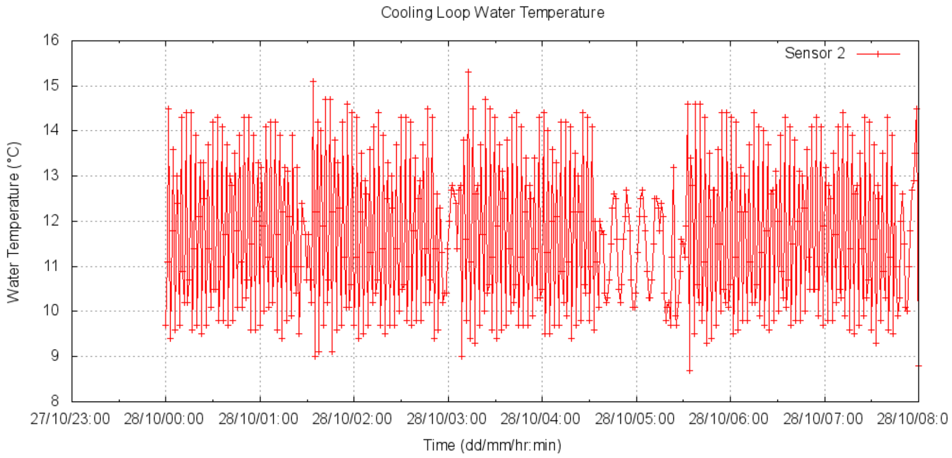


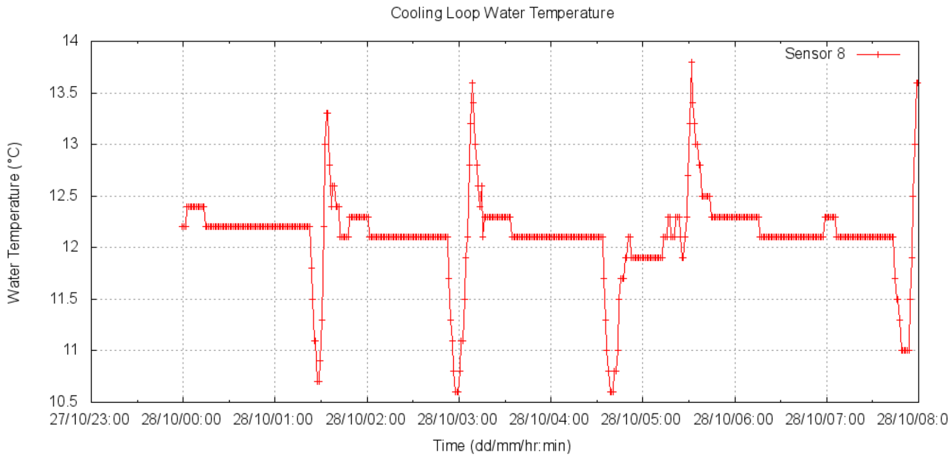
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Benefits of Secondary-loop Heat Exchangers

- With the secondary heat exchangers removed on the cooling loops for the **middle** and **rows** there was the opportunity to assess the differences to the loop on the **front** row
- Decoupling the cooling loops (by means of a heat exchanger) provide an optimal configuration
  - Each loop can control the water temperature independently
  - Control system is not continuously fluctuating around the control variable set-point



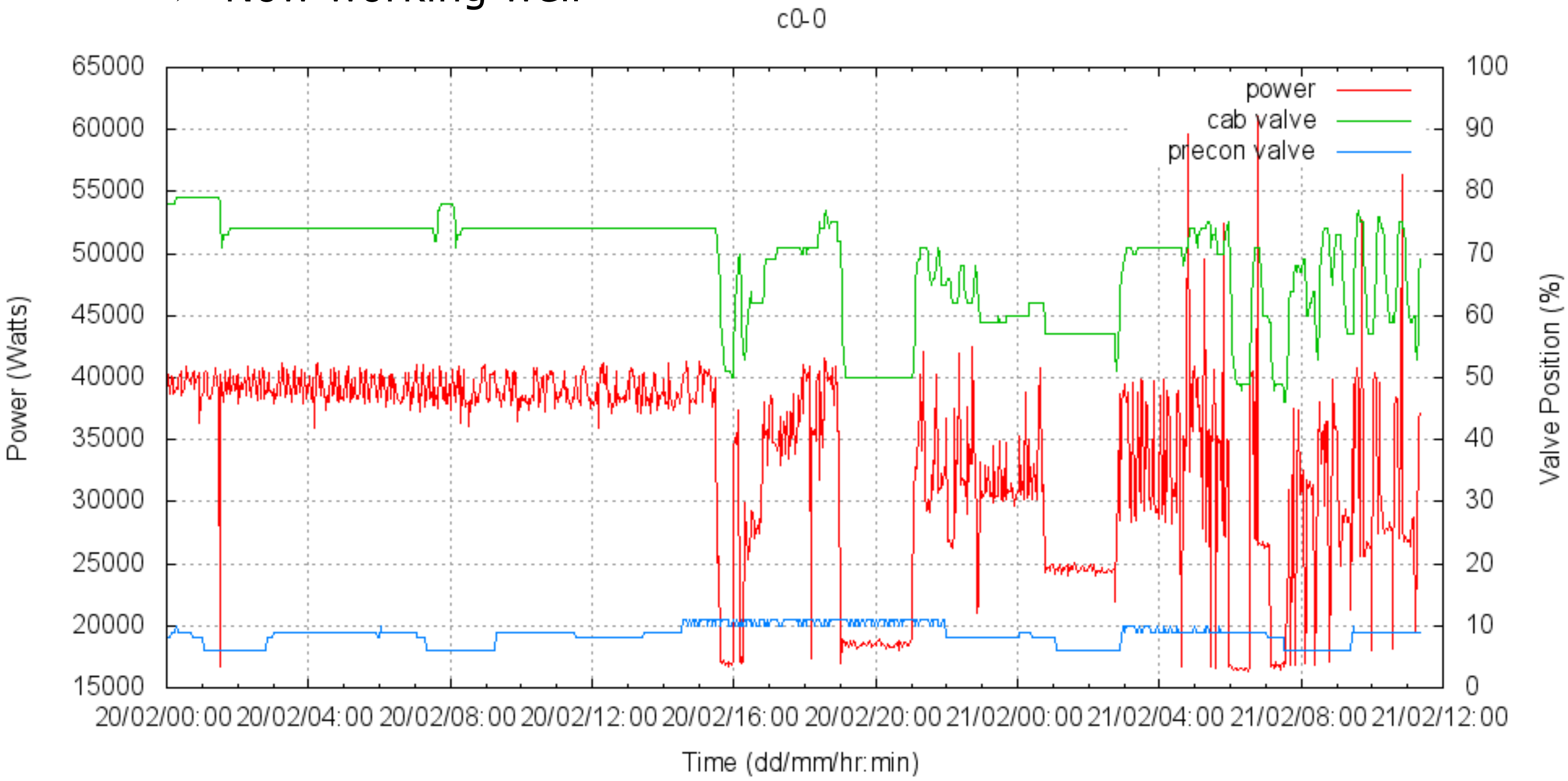


Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Preconditioners

- Installed as part February maintenance (1 day)
- Some initial teething problems with CCB voltage overload
  - Same as before: Sprung-loaded water actuators the culprit
- Now working well



CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Q&A

Colin McMurtrie: colin@cscs.ch
Ladina Gilly: lgilly@cscs.ch
Tiziano Belotti: belotti@cscs.ch