Hewlett Packard Enterprise

lrz

CUG2024 Diverse Universe PERTH AUSTRALIA

FIRST ANALYSIS ON COOLING TEMPERATURE IMPACTS ON MI250X EXASCALE NODES (HPE CRAY EX235A)

Dr. Torsten Wilde (HPE)
Dr. Michael Ott (LRZ)
Pete Guyan (HPE)

Presenter: Steve Martin (HPE)

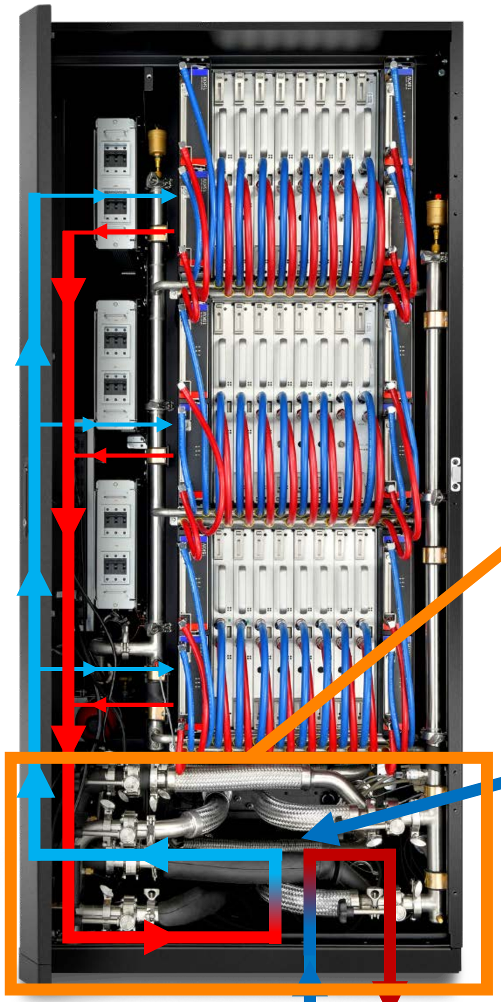
Mai, 2024

THE WORLD SMALLEST HPE CRAY EX AT LRZ

- HPE Cray EX2500
- 100% Direct Liquid Cooled (DLC)
- 4 modified EX235A nodes
- Integrated Switches
- Redundant Cooling Distribution Units (CDUs)

HPE CRAY EX2500

- 100% Direct Liquid Cooled (DLC)
- 24 compute blades
- Integrated Switches
- Redundant Cooling Distribution Units (CDUs)



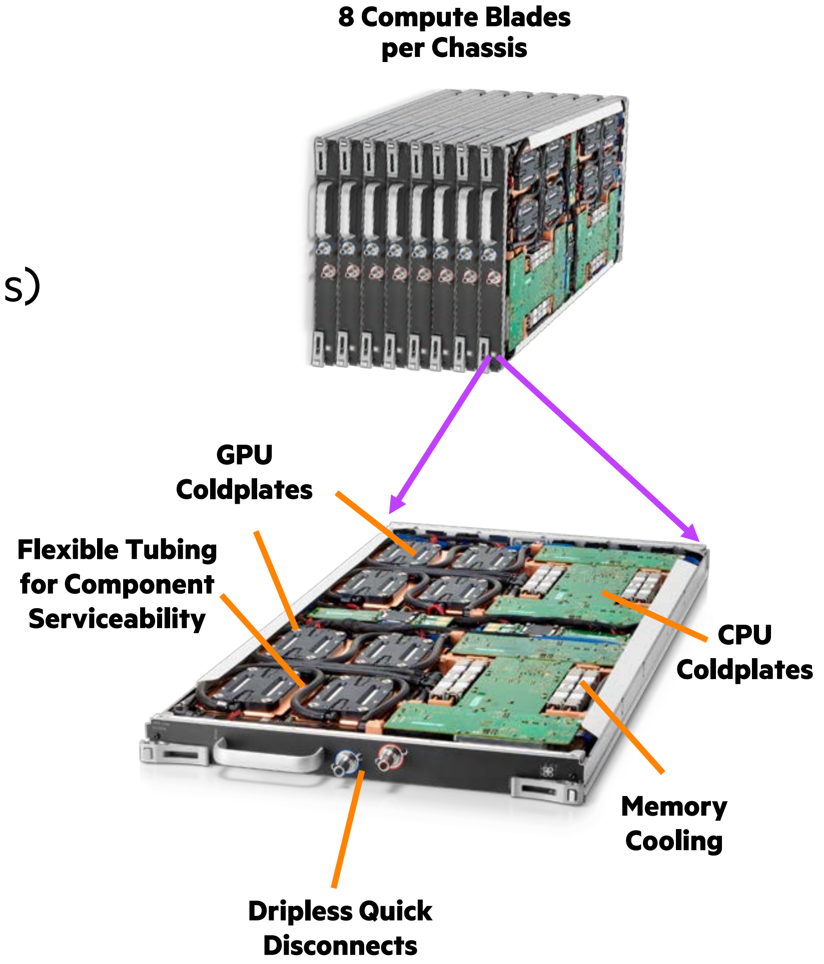
Coolant Distribution Units (CDU)

Primary and Secondary Side fluids do not mix.

Heat is transferred through a plate-to-plate heat exchanger in the CDU.

Primary Side

(Facility Water to CDU)

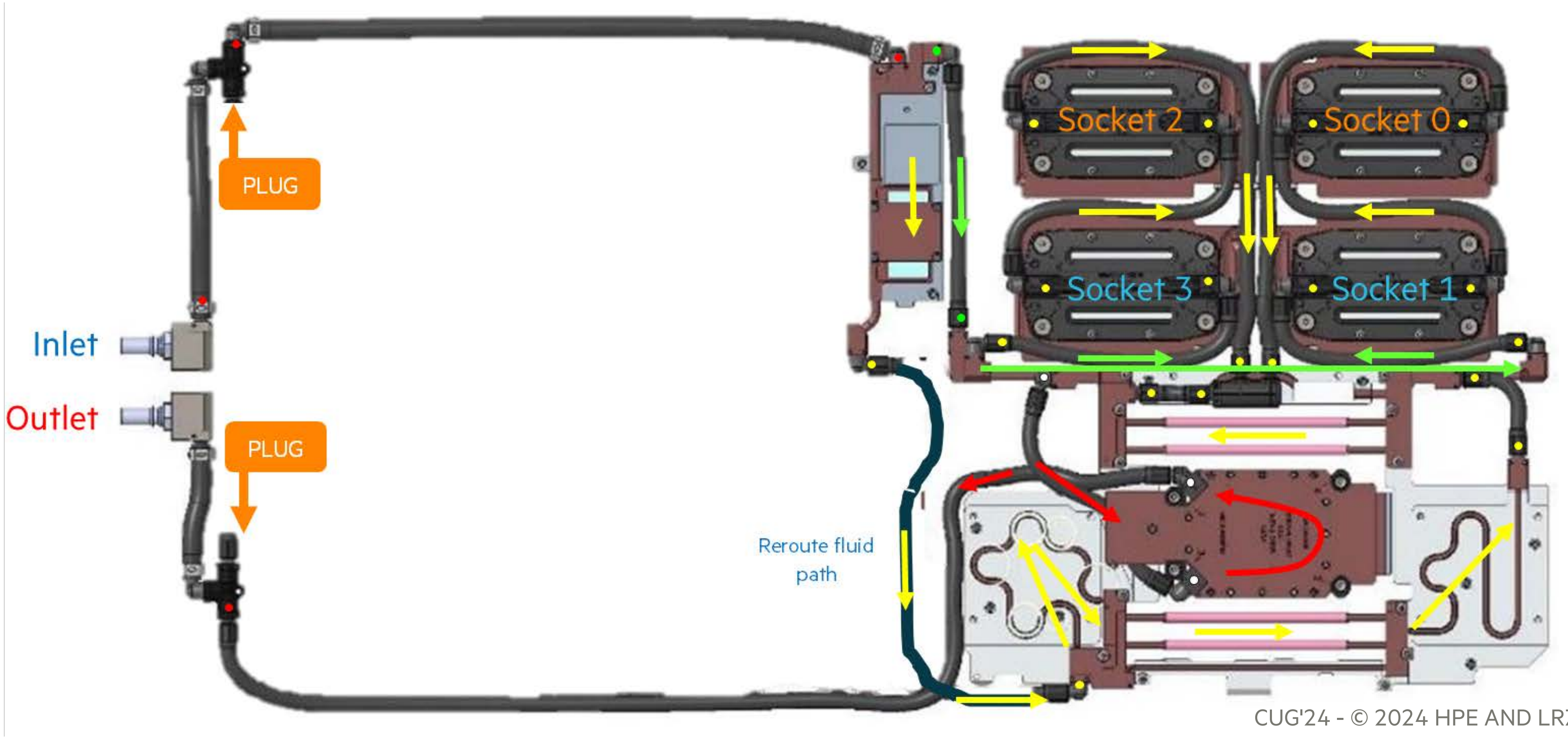


Note: Demonstration system has only one node per blade installed

LRZ CO-DESIGN SYSTEM SINGLE NODE FLUID DISTRIBUTION

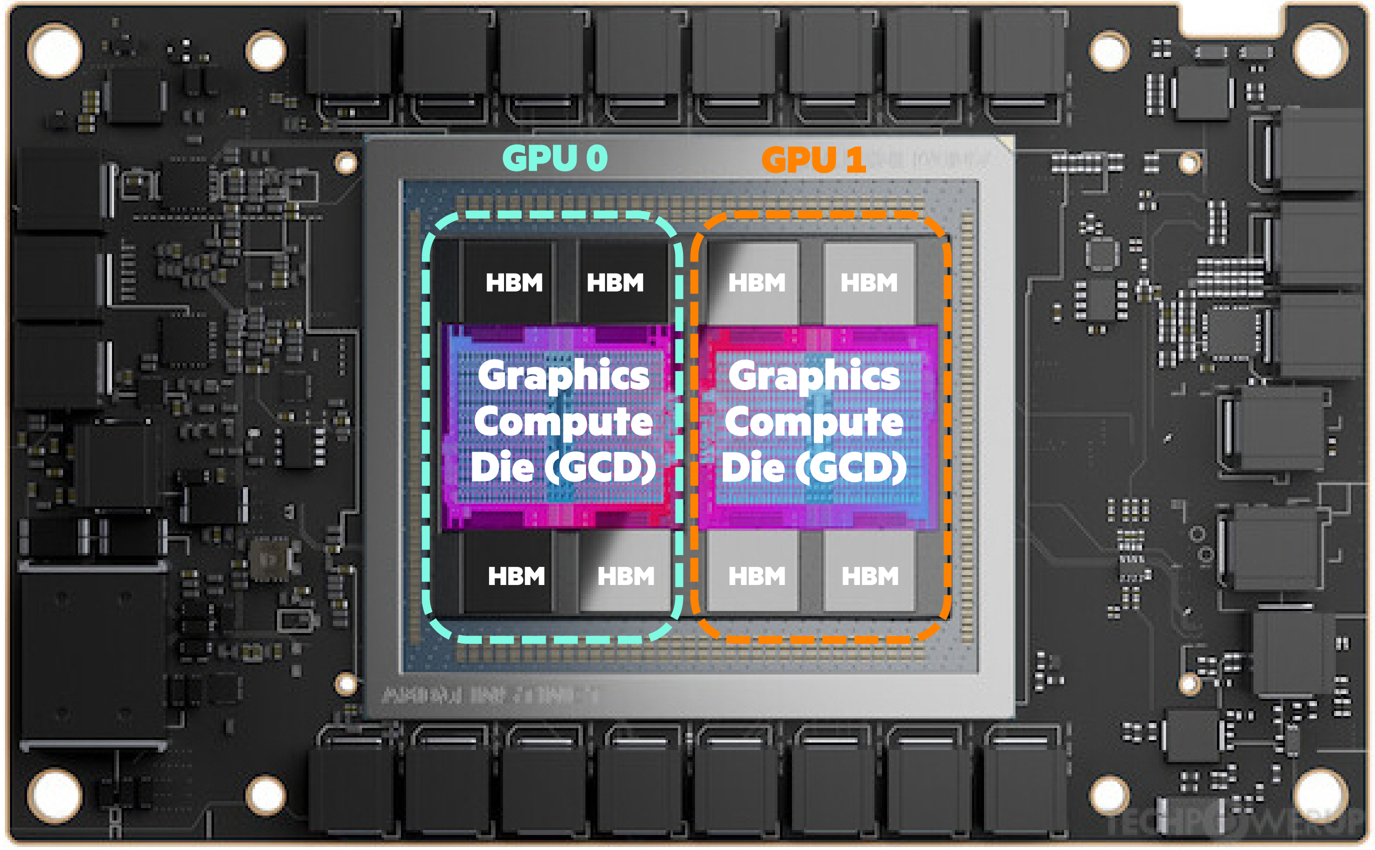
Modified Ex235a Blade – One Node Removed

- Modification to enable higher flow rates for testing expanded thermal operating range
- Preheat from upstream GPU in serial flow causes downstream GPU to reach thermal thresholds first



AMD MI250X ACCELERATOR THERMAL REPORTING

- Two thermal zones, each corresponding to one Graphics Compute Die (GCD) complex
- HBM temperature reported as maximum temperature of the four HBM stacks within one GCD complex
- Die junction temperature is reported



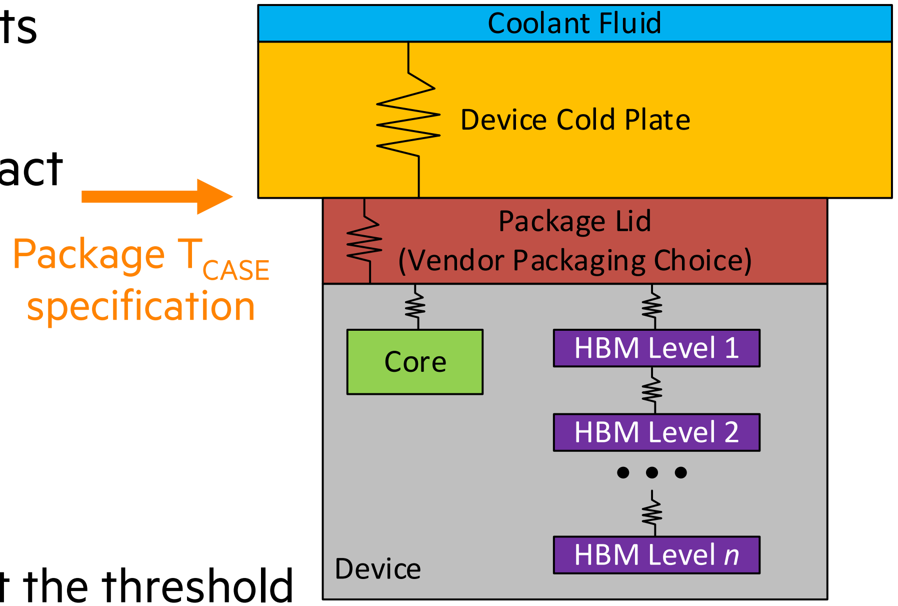
Source: https://www.tweaktown.com/news/82636/amd-instinct-mi250x-aldebaran-mcm-gpu-128gb-hbm2e-memory-500w-power/index.html

COMPONENT COOLING (CONTINUED)

- System monitoring capabilities cannot directly measure TCASE
  - Can measure individual chiplet and/or subsystem die junction temperatures

- Exceeding chiplet (die) thermal thresholds have performance impacts

- Multiple thresholds may exist and have increasing performance impact
  - HBM transition from 1x to 2x memory refresh
  - HBM 2x max temperature is reached -> GPU throttling
  - GPU throttles full package



- Notes
  - Highest temperature HBM die/stack per zone is used to compare against the threshold
  - Lowest die within HBM stack typically has the highest temperature
    - Due to being the furthest away from the package lid (cooling surface)
  - Throttling occurs when first zone reaches threshold

ANALYSIS SETUP

- Benchmarks (HPCG and HPL) are run on all 4 nodes multiple times and results are averaged
  - HPL ~50 runs (together ~1h)
  - HPCG, 1 run with 3000 iterations (~1h)

- Benchmark instrumentation
  - pm_counters and rocm-smi to collect in-band node and device power/energy
  - Engineering tools to collect monitoring data (Redfish)
  - Performance information
  - System log to see hardware events

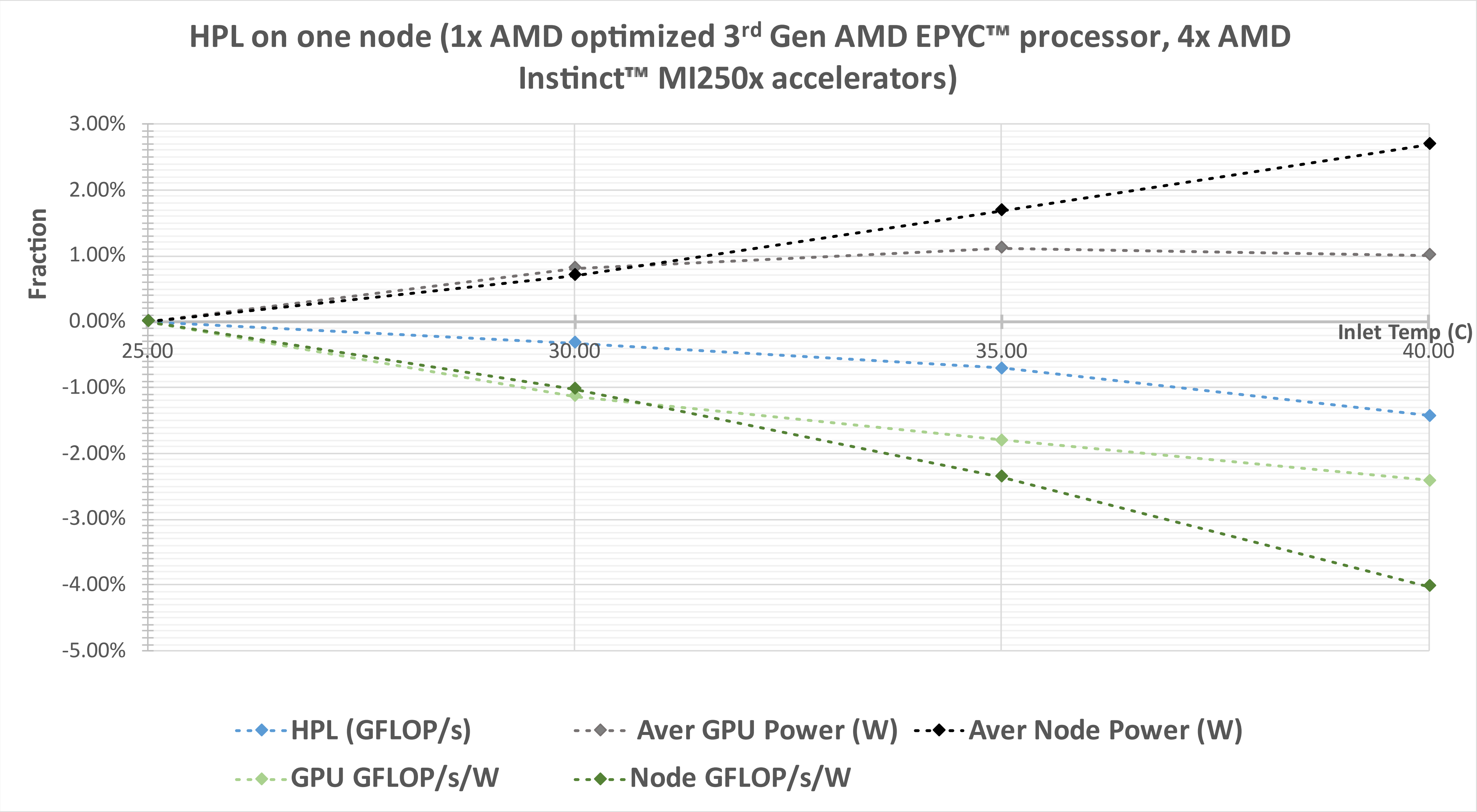
- Jupyter notebook to process and analyze collected data
  - Average results over all 4 compute nodes
  - Generate plots
  - Generate csv for excel analysis

HPL AND HPCG AVERAGE NODE POWER (CONSTANT FLOW RATE)

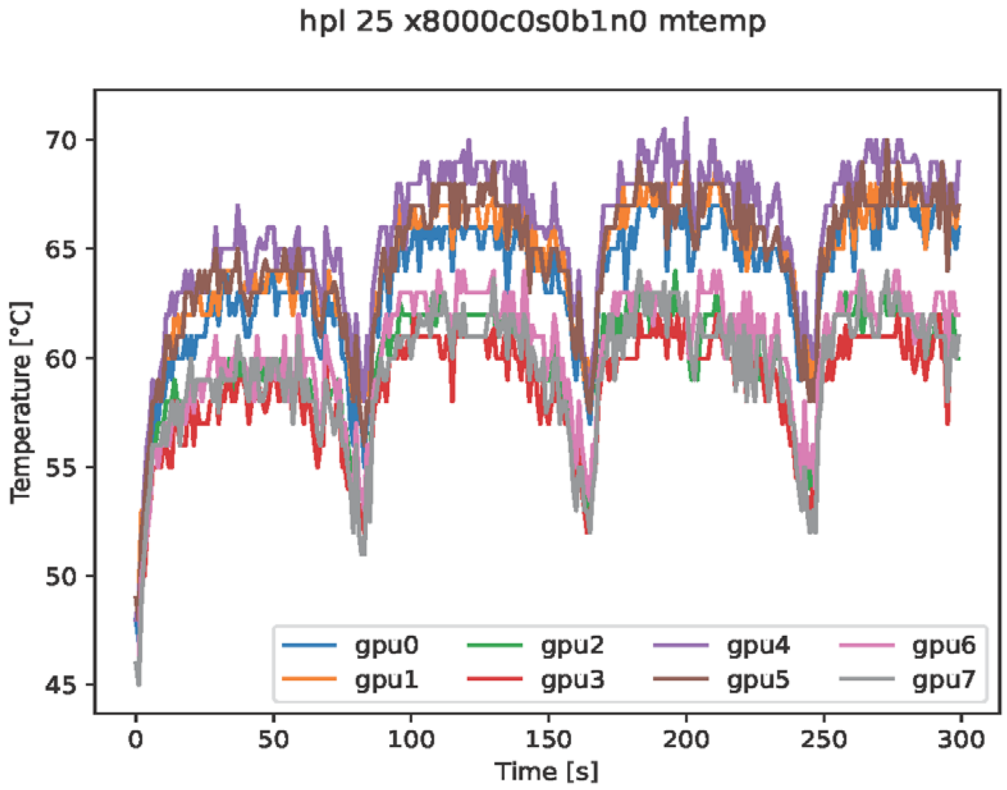
Inlet Temperature to Compute Blades (℃)	**HPL**			**HPCG**		
	Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)	Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)
25	2582	142525 (0%)	55.209 (0%)	2493	1969.35 (0%)	0.790
30	2600	142058 (-0.3%)	54.643(-1.03%)	2533	1966.36 (-0.15%)	0.776 (-1.73%)
35	2625	141513 (-0.7%)	53.906 (-2.36%	2554	1959.08 (-0.52%)	0.767 (-2.89%)
40	2651	140490 (-1.4%)	52.993 (-4.02%)	2559	1901.12 (-3.46%)	0.743 (-5.95%)

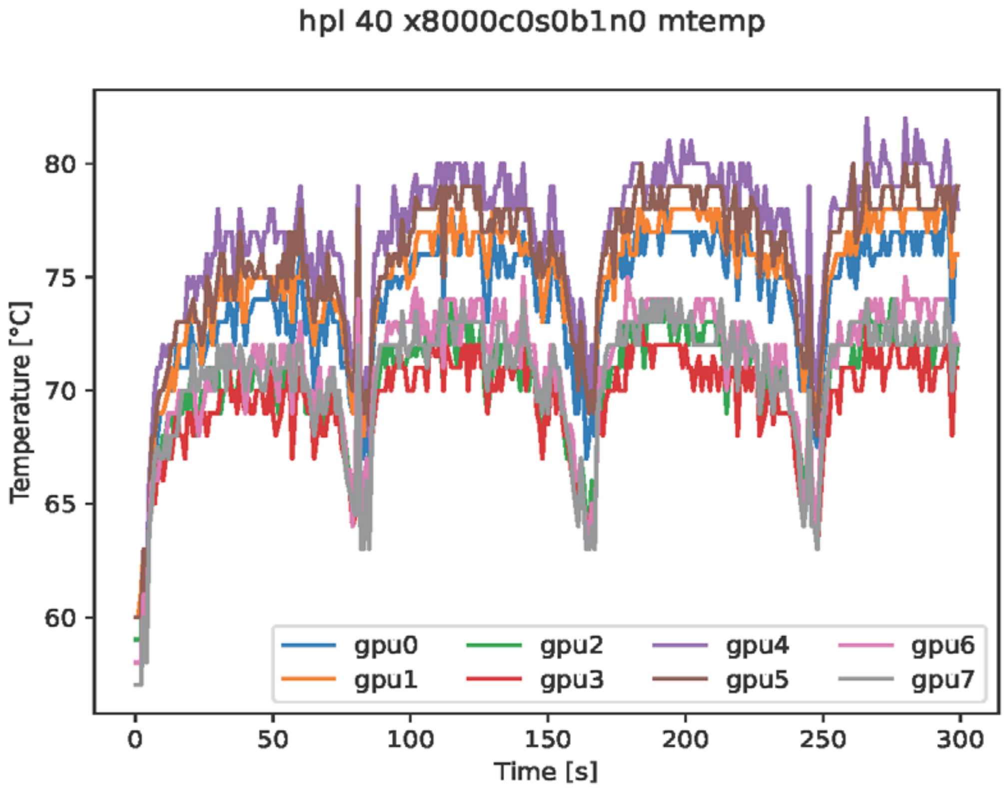
- HPCG average power consumption close to HPL (~3.5% difference)
- Power increase for higher inlet temperatures not uniform
- Performance drop more visible when moving from 35℃ to 40℃

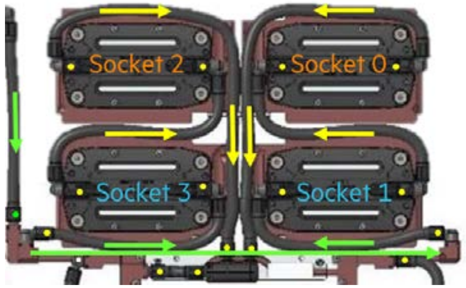
HPL AVERAGE NODE MEASUREMENTS NORMALIZED TO 25℃



HPL NODE LEVEL HBM TEMPERATURE MEASUREMENTS



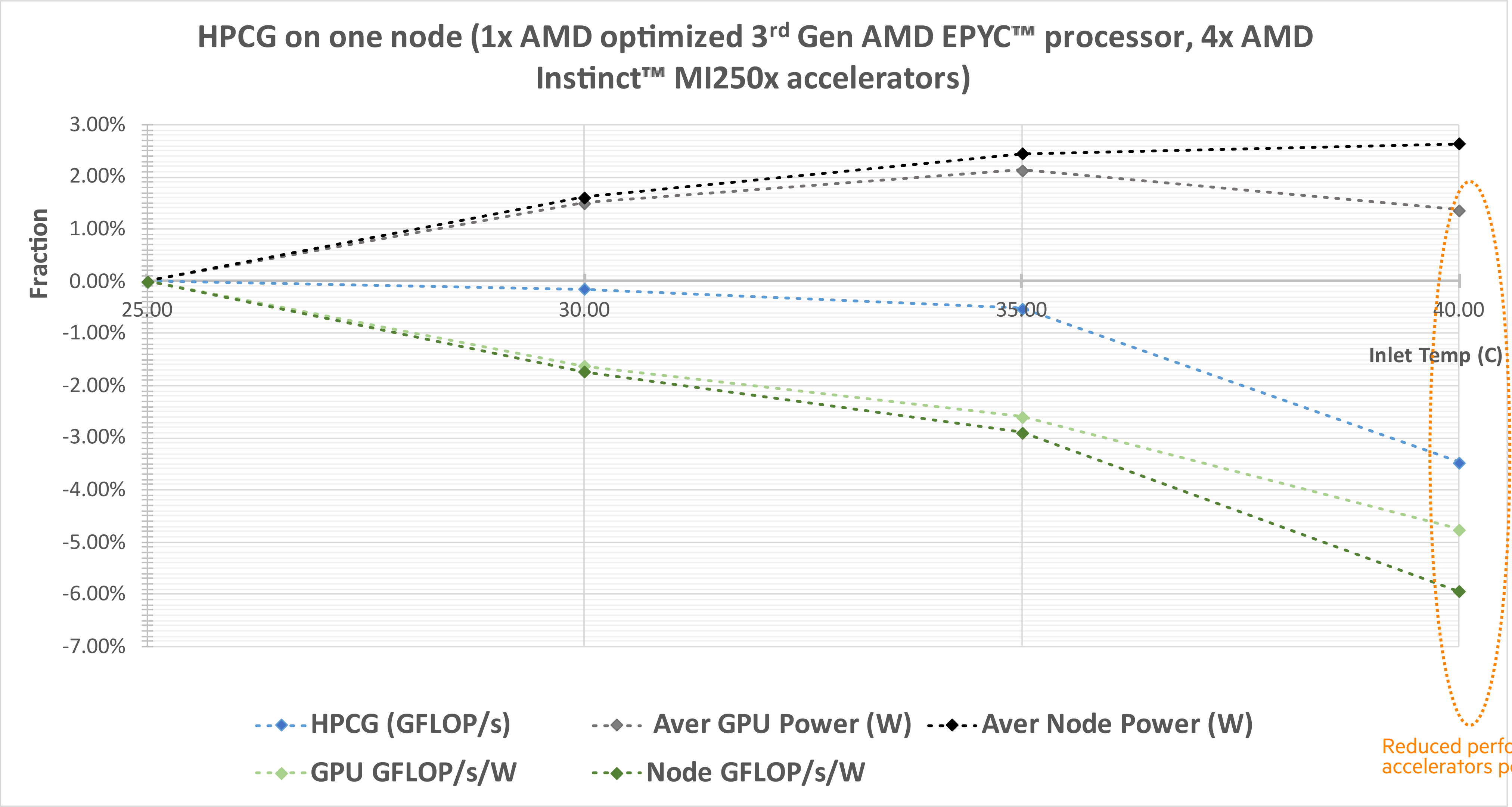




GPUs are labeled per GCD in each socket:

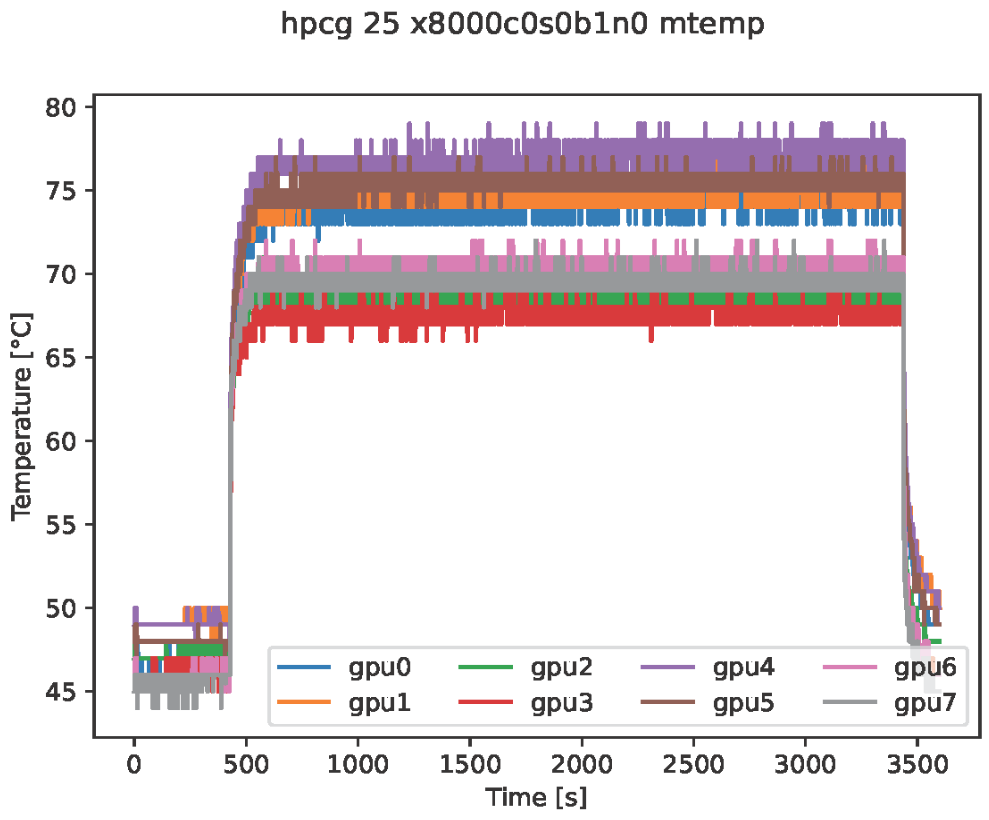
- Socket 0 (gpu0, gpu1)
- Socket 1 (gpu2, gpu3)
- Socket 2 (gpu4, gpu5)
- Socket 3 (gpu6, gpu7)

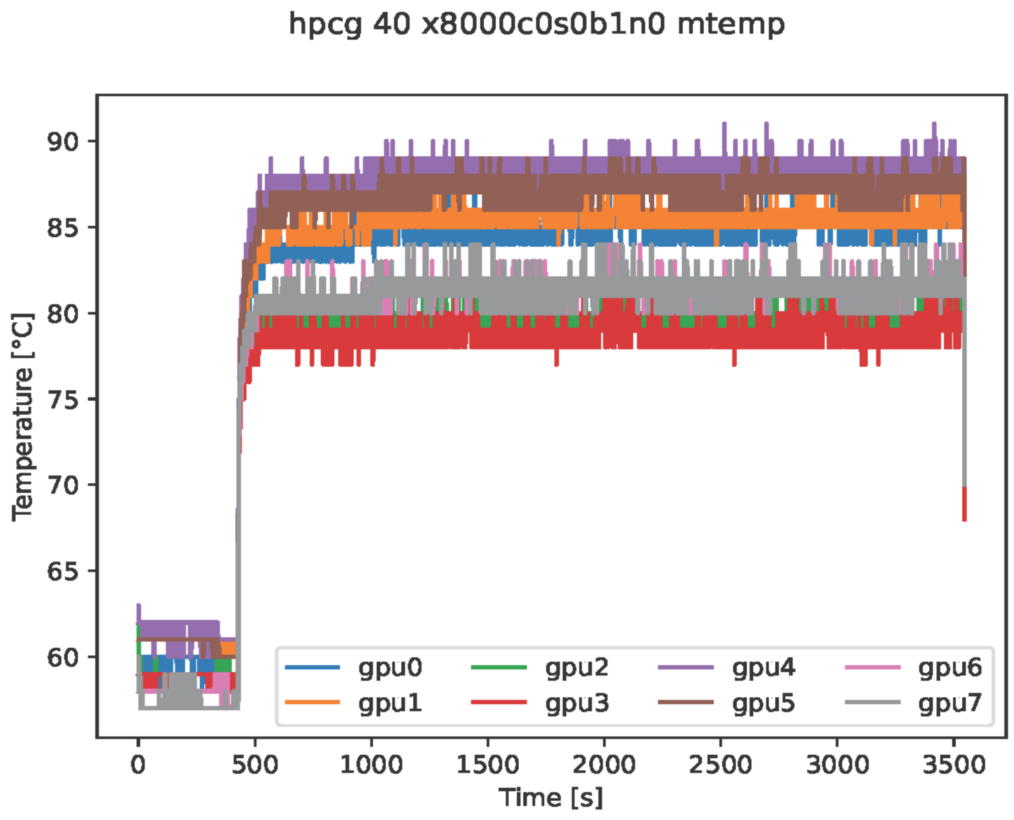
HPCG AVERAGE NODE MEASUREMENTS NORMALIZED TO 25℃



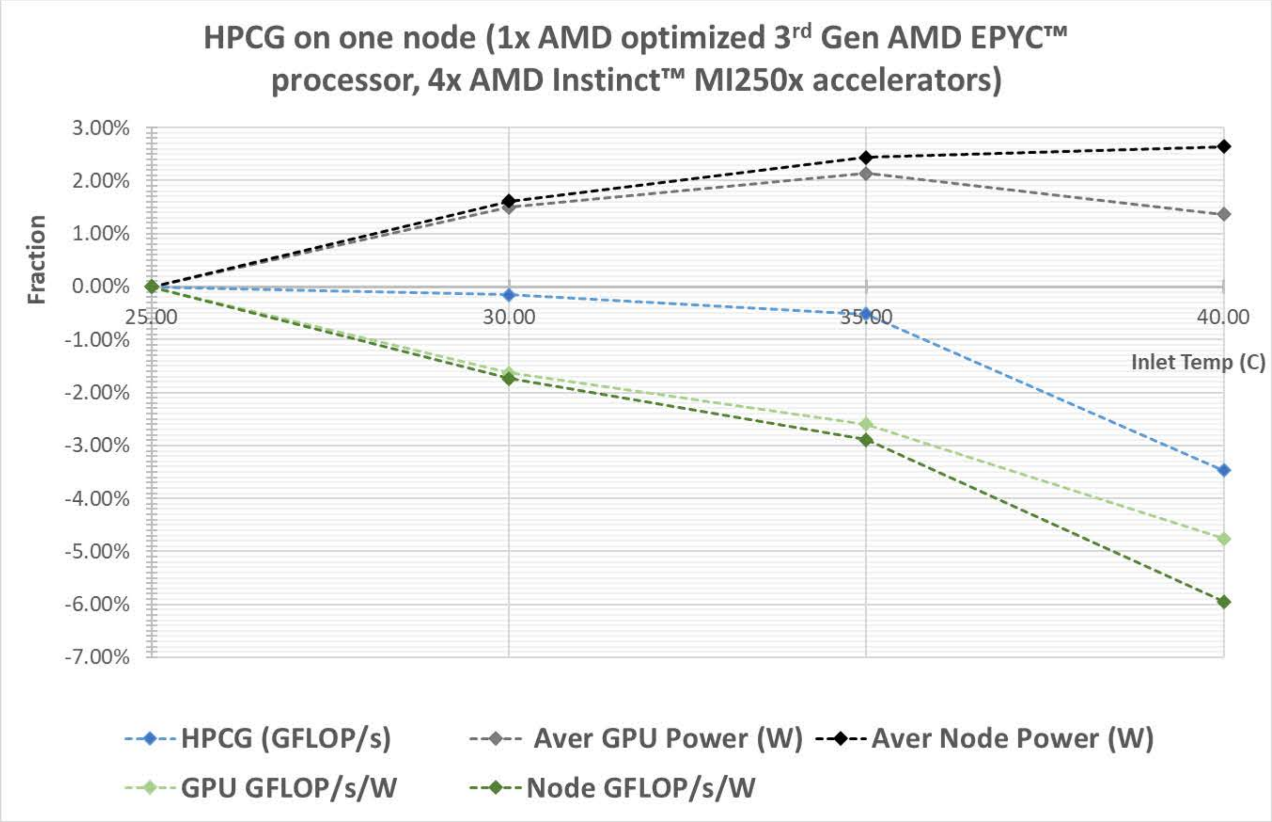
Reduced performance limits accelerators power increase

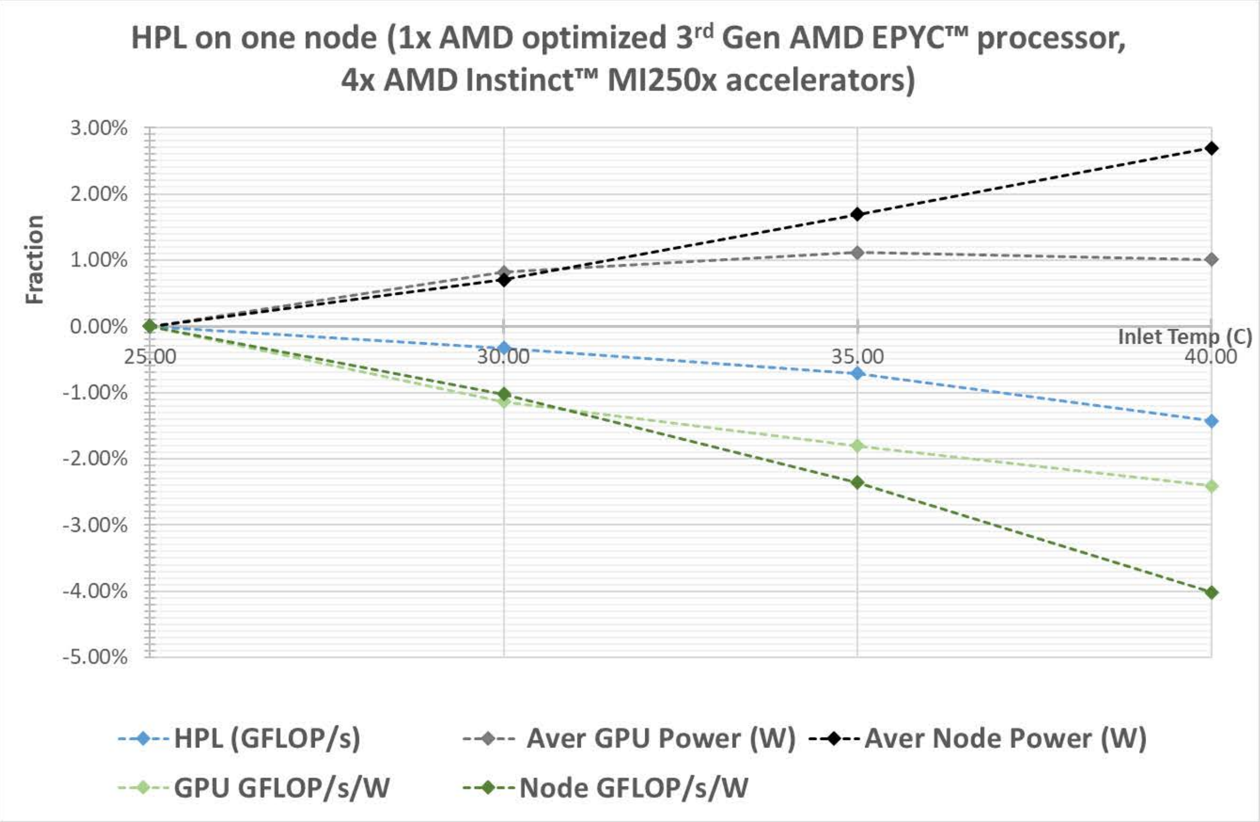
HPCG NODE LEVEL HBM TEMPERATURE MEASUREMENTS





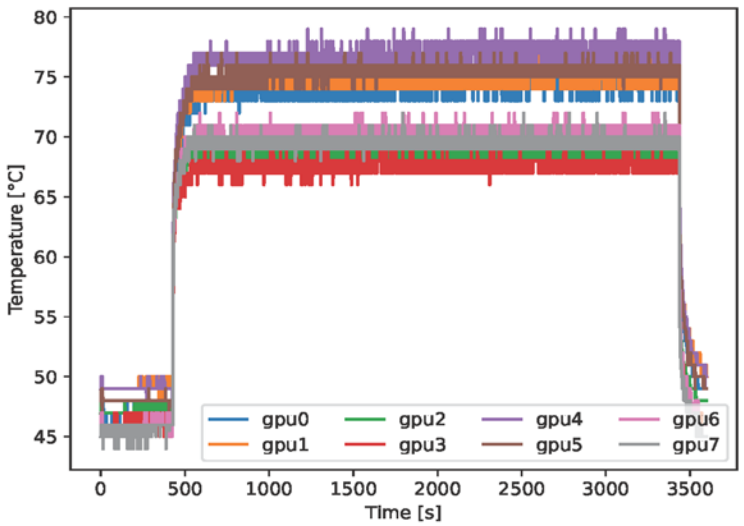
HPL AND HPCG COMPARISON SUMMARY



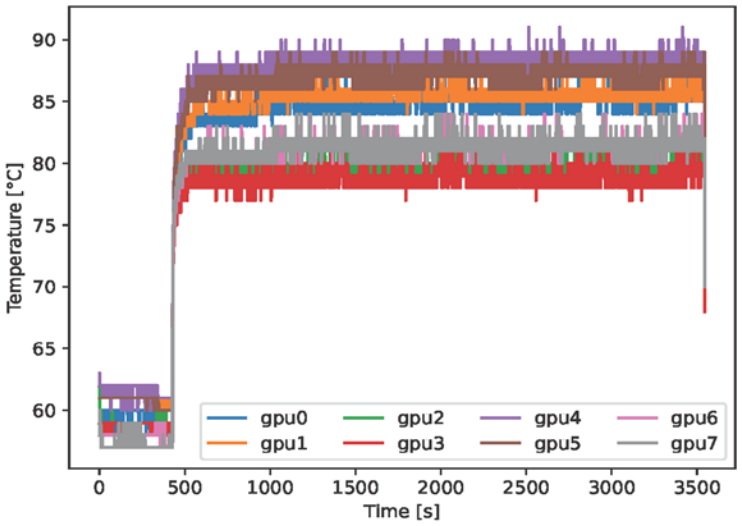


HBM USAGE DEFINES PERFORMANCE IMPACT AT HIGH COOLING TEMPS

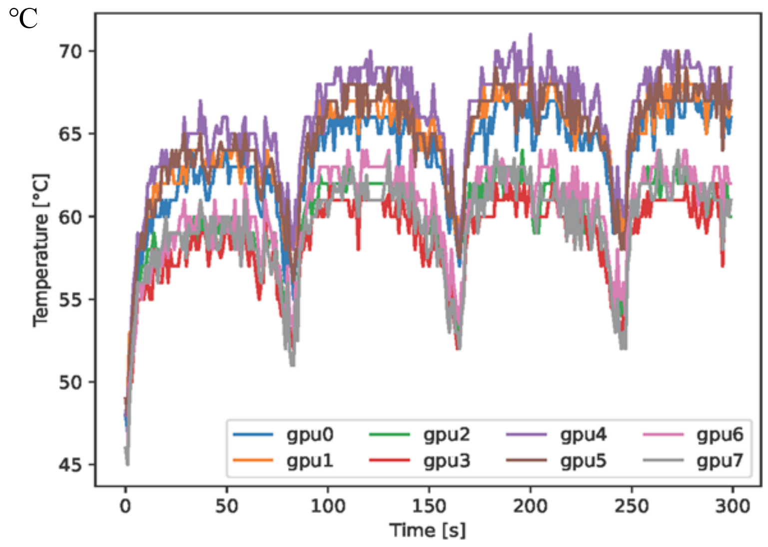
hpcg 25 x8000c0s0b1n0 mtemp



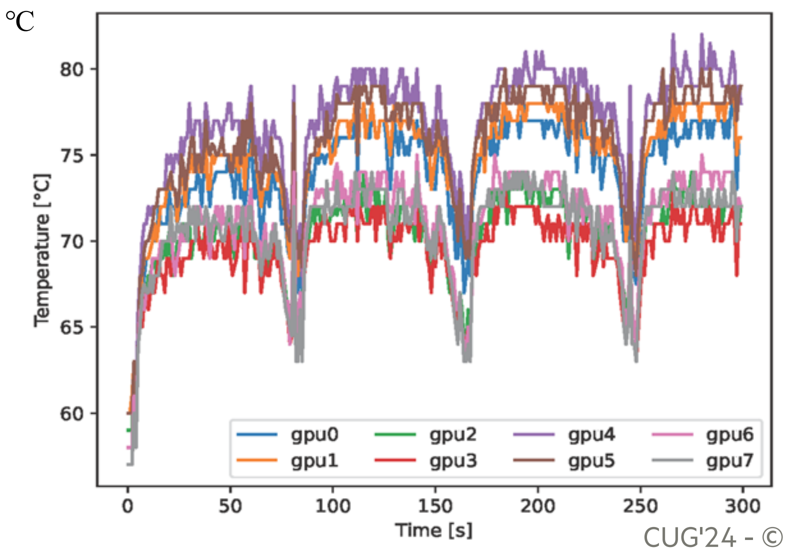
hpcg 40 x8000c0s0b1n0 mtemp



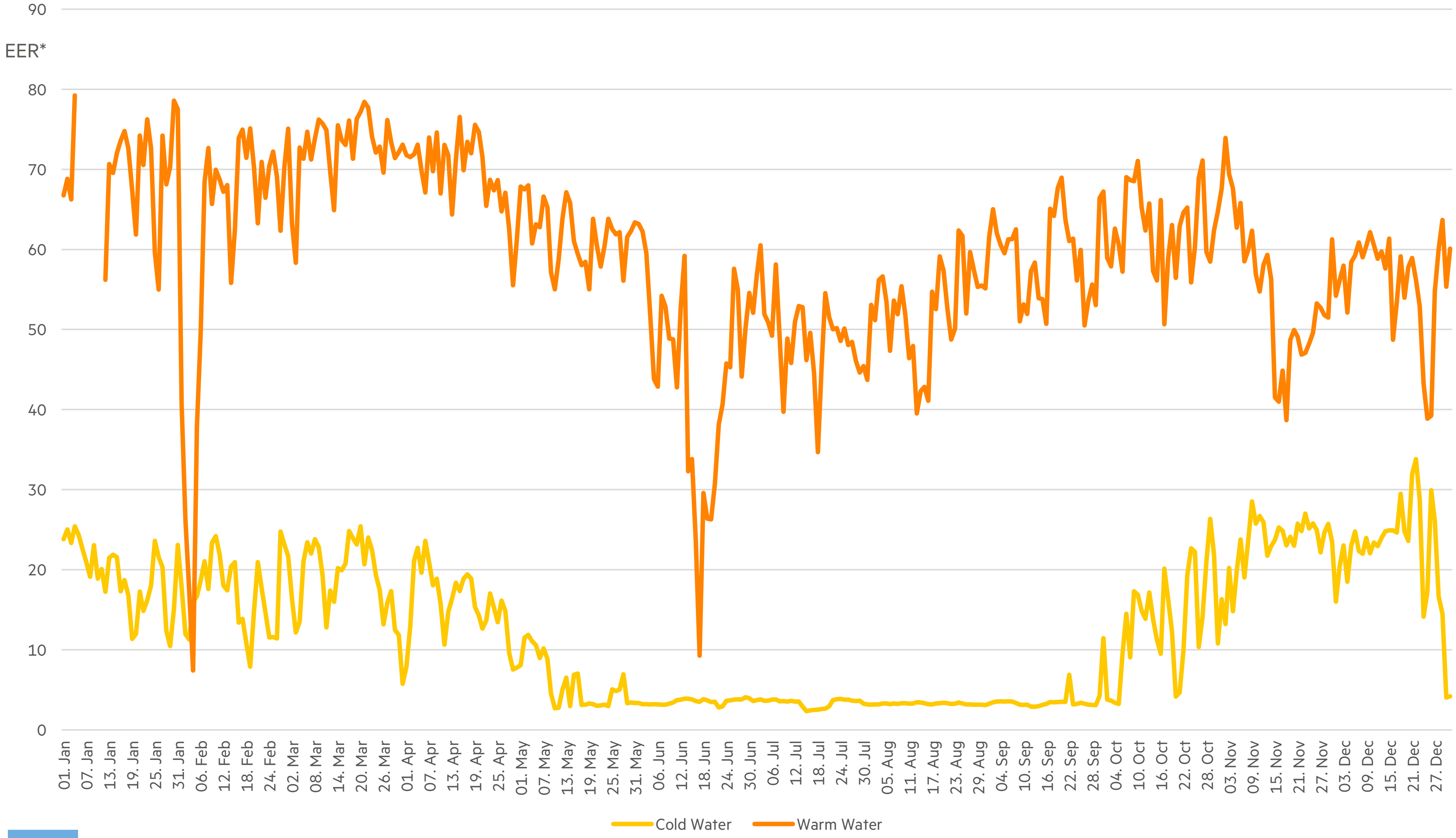
hpl 25 x8000c0s0b1n0 mtemp



hpl 40 x8000c0s0b1n0 mtemp



EER* OF LRZ WARM AND COLD WATER PLANTS



*EER (Energy Efficiency Ratio) = $\frac{Cooling_{generated}}{P_{Cooling}}$

EER = cooling generated per 1W of electrical power consumption of the cooling infrastructure

INCLUDING FACILITY EFFICIENCY – CHILLER & CHILLER FREE

	LRZ CER*
Chiller supported (Cold Water) (20℃ facility cooling loop outlet – 25℃ inlet IT)	**12**
Chiller-less (Warm Water) (40℃ inlet IT)	**59**

Inlet Temperature to Compute Blades (℃)	**HPL (compute only)**			**HPCG (compute only)**		
	Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)	Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)
25	2582	142525 (0%)	**55.209 (0%)**	2493	1969.35 (0%)	**0.790 (0%)**
30	2600	142058 (-0.3%)	**54.643(-1.03%)**	2533	1966.36 (-0.15%)	**0.776 (-1.73%)**
35	2625	141513 (-0.7%)	**53.906 (-2.36%**	2554	1959.08 (-0.52%)	**0.767 (-2.89%)**
40	2651	140490 (-1.4%)	**52.993 (-4.02%)**	2559	1901.12 (-3.46%)	**0.743 (-5.95%)**

*CER (Cooling Efficiency Ratio) = $\frac{Heat\ to\ remove\ over\ a\ year\ (kWh_{th})}{Electical\ Energy\ used\ (kWh_{el})}$

CER = factor that shows how much heat can be removed per 1W of electrical power consumption of the data center cooling infrastructure

INCLUDING FACILITY EFFICIENCY – CHILLER & CHILLER FREE

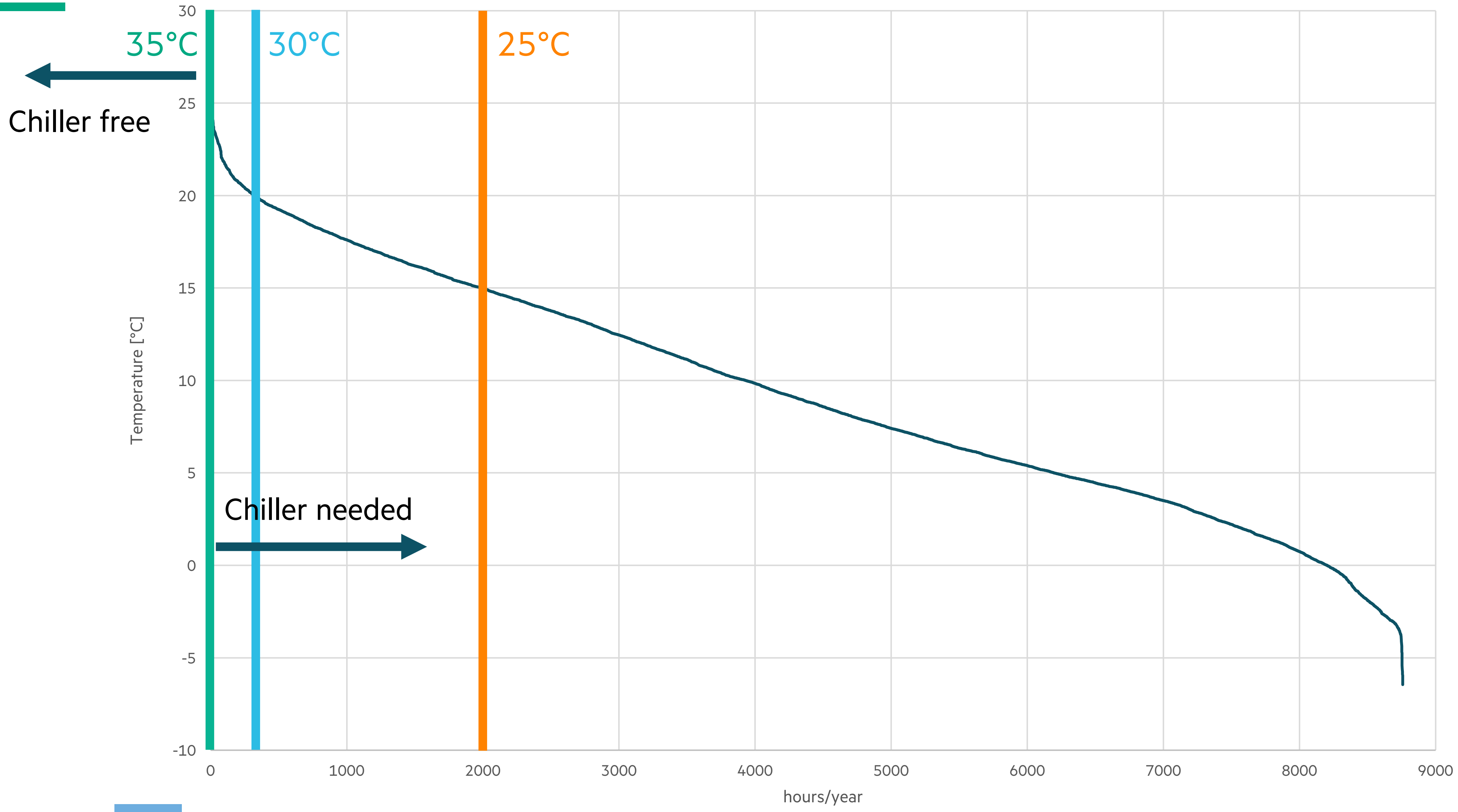
	LRZ CER*
Chiller supported (Cold Water) (20℃ facility cooling loop outlet – 25℃ inlet IT)	12
Chiller-less (Warm Water) (40℃ inlet IT)	59

Inlet Temperature to Compute Blades (℃)	HPL (including facility cooling efficiency)			HPCG (including facility cooling efficiency)		
	Adjusted Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)	**Adjusted** Avg Node Power (W)	Performance (GFlops)	Efficiency (GFlops/W)
25	2797	142525 (0%)	**50.956 (0%)**	2741	1969.35 (0%)	**0.719 (0%)**
40	2696	140490 (-1.4%)	**52.110 (+2.27%)**	2641	1901.12 (-3.46%)	**0.720 (+0.14%)**

*CER (Cooling Efficiency Ratio) = $\frac{Heat\ to\ remove\ over\ a\ year\ (kWh_{th})}{Electical\ Energy\ used\ (kWh_{el})}$

CER = factor that shows how much heat can be removed per 1W of electrical power consumption of the data center cooling infrastructure

WET BULB TEMPERATURES AT LRZ (WITH APPROACH OF 10°C)



SUMMARY

- When **only** looking at node performance/watt:
  - Higher inlet temperatures reduce performance and increase power consumption
  - Workloads using heavily HBM are more impacted by higher cooling temperatures

- To make a sound discission, one **need to include the facility effectiveness**
  - Combined efficiency for chiller-less cooling with high inlet temperature (at 40 °C)
    - Identical to chiller supported (at 25°C)
  - Higher inlet temperatures improve facility effectiveness
    - Best inlet temperature for LRZ demonstration system for chiller less cooling in climate is > 35°C

- **Heat re-use** requirements might shift the most efficient inlet temperature to 40°C or above
  - This will depend on the outlet temperature requirements
  - The effectiveness of a heat pump might need to be included

THANKS

WILDE@HPE.COM
MICHAEL.OTT@LRZ.DE