Thermal Behaviors in Liquid Immersion Cooling under Various Workloads: a Case Study

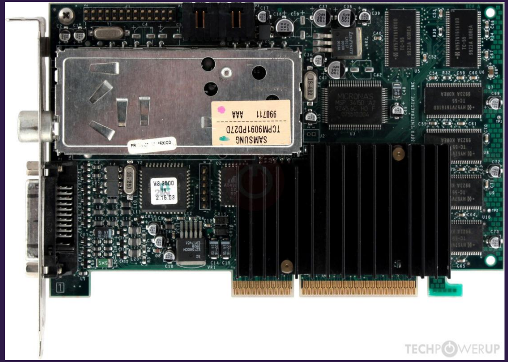
Thomas Randall, Bennett Cooper, Naman Kulshreshtha, Rong Ge

CLEMSON

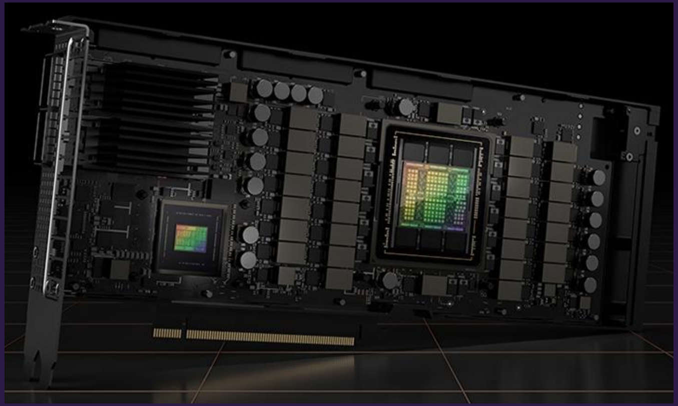
School of COMPUTING Clemson® University

Costs of Ineffective Cooling

- Energy expenditure → heat → performance reduction
  - Lower clock rate
  - System taken offline
  - Hardware damage
- Cooling systems maintain performance
  - Sustain activity
  - Recover faster
- Dense compute == higher performance AND thermal burden
  - GPUs and other specialized chips
  - Towards kW-scale TDP



Voodoo3 3500 TV AGP (1999, 15W TDP)



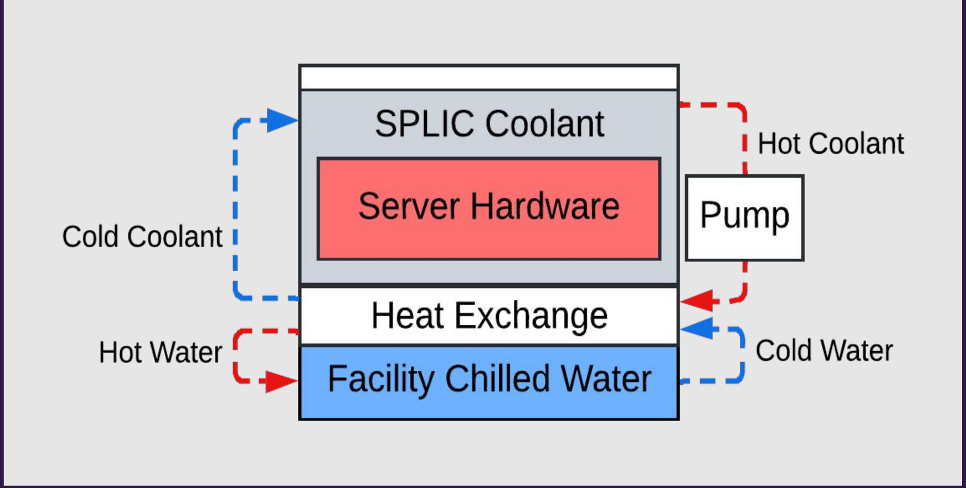
NVIDIA H100 (2024, 350W TDP)

Available Techniques for Effective Cooling

- Heat dissipation via liquid > gas
  - Distilled water possible, but MUST remain pure!
  - Many dielectric fluids similar to water in viscosity and thermal conductivity
- Liquid cooling and evaporators well-established technologies, developing best integration with servers
  - Key future developments
    - Novel fluids
    - Liquid circulation practices
    - Specialized hardware for liquid immersion

Single Phase Liquid Immersion Cooling (SPLIC)

- Previous studies established
  - Foundational effectiveness of liquids over gases
  - Designs and properties for dielectric fluids
  - Placement and number of pumps in tank
  - Impacts on hardware warranties
- Some effects not studied
  - Separate sensors drive cooling systems
    - Delay in system’s reaction
  - Varying activity and physical arrangements



Case Study Contributions

- Understand thermal burdens by component
  - Commodity hardware is compatible, but not maximally effective
  - Potential to share heat between components
  - System cooled equally rather than per-component
- Observe heat dissipation capabilities under load
  - Determine appropriate system responsiveness
  - Other liquid cooling approaches can improve efficiency at high temperatures
- Differentiate heat accumulation based on real workloads
  - Hardware utilization
  - Periodicity
  - Demand

Hardware Listing

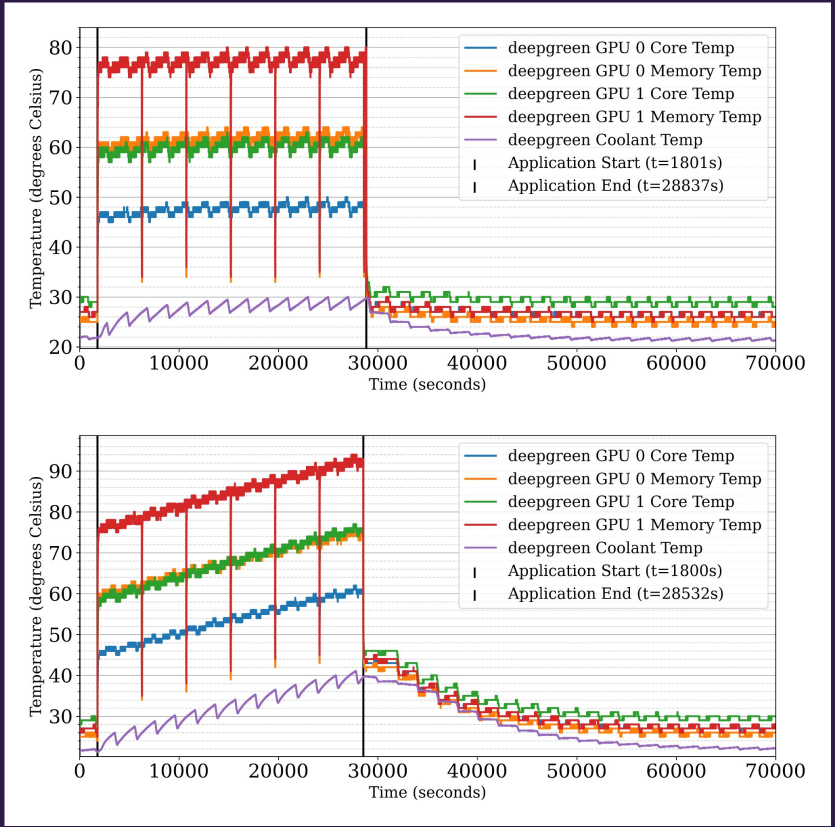
- Cooling Hardware: Submer SmartPod v3
  - White mineral oil dielectric coolant
  - 11C datacenter facility water
- Server Hardware:
  - 3 distinct nodes with two architectures
    - (2 CPU nodes) Intel Xeon E5-2670 v3 2.3 GHz 64 cores
    - (1 GPU node) AMD EPYC 7551P 2.0 GHz 32 cores; two NVIDIA Titan V 12 GB HBM2 GPUs
- Preparation for Immersion:
  - Thermal paste → indium foil
  - Remove fans / blades from power supplies, GPUs, CPUs

Software Benchmarks & Metrics

- GPU-Accelerated Applications
  - Memory-Bottlenecks
    - EMOGI
  - Compute-Bottlenecks
    - CUDA DGEMM
    - MD5 Bruteforcer
- Multi-Node CPU Applications
  - Memory-Bottlenecks
    - NPB DT (Class C)
    - NPB IS (Class D)
  - Compute-Bottlenecks
    - NPB EP (Class E)
    - HPCC HPL

- CPU metrics
  - Core temperature and frequencies
- GPU metrics
  - Core and memory temperature and utilization
  - Power usage and performance state
- NVMe metrics
  - Temperature
- PDU metrics
  - Total system power draw
- SPLIC Tank metrics
  - Coolant and water temperature
  - Power consumption and pump RPM
  - Coolant and water flow rates
  - Effective dissipation and mPUE

Rising Tide Lifts All Boats Together

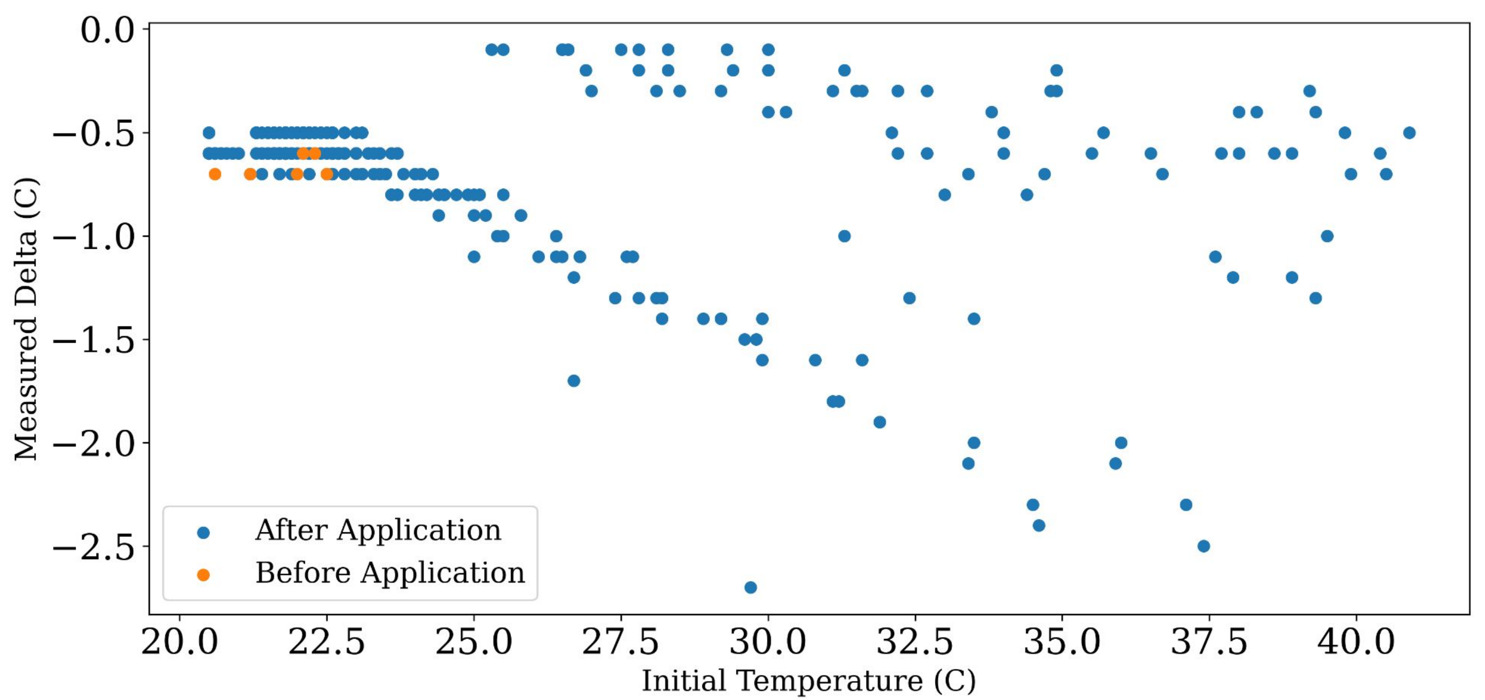


- Top: Normal SPLIC operation
- Bottom: No chilled water until application period ends
- Coolant temperature affects ALL components equally
  - Physical proximity to hot GPUs does not increase average temperature or rate of increase
  - Even within the same server node!

Hardware Component	Idle Temp. Min/Mean /Max (°C)	Interval (s) to Increase Temp by 1°C	Op. Temp. Observed Max (°C)
SPLIC Coolant	19.80 / 21.41 / 22.50	970.99	47.10
deepgreen CPU	14.75 / 17.51 / 19.66	702.20	55.62
deepgreen GPU	24.25 / 26.54 / 28.00	266.83	96.00
deepgreen NVMe	14.00 / 17.91 / 20.00	1068.48	44.00
n01 CPU	17.44 / 21.30 / 24.04	607.33	66.00
n02 CPU	17.67 / 20.81 / 24.26	560.35	69.00

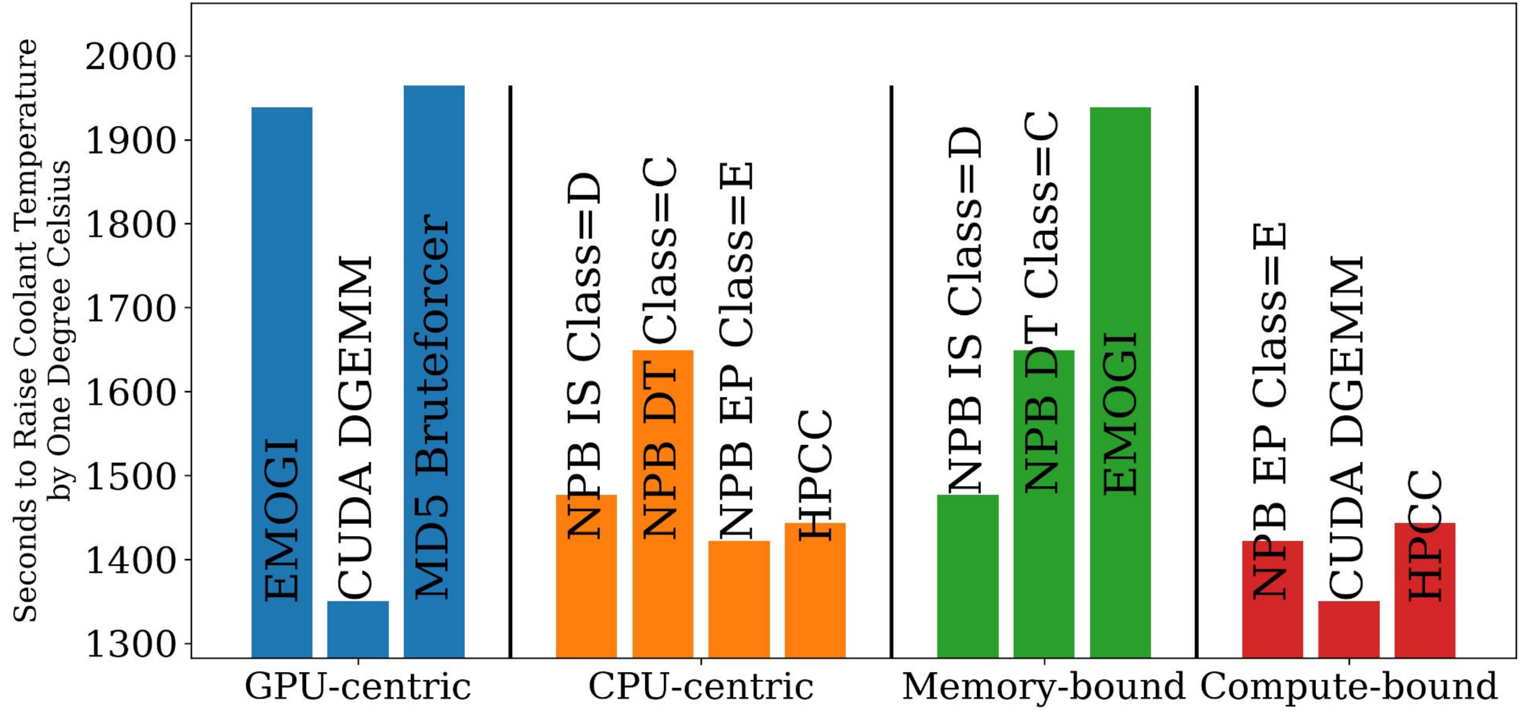
Inconsistent But Improving Efficiency

- Pumps generally manage to lower coolant temperature better with greater differential
- However, many pumping periods end up being *ineffective*, even compared to initial conditions

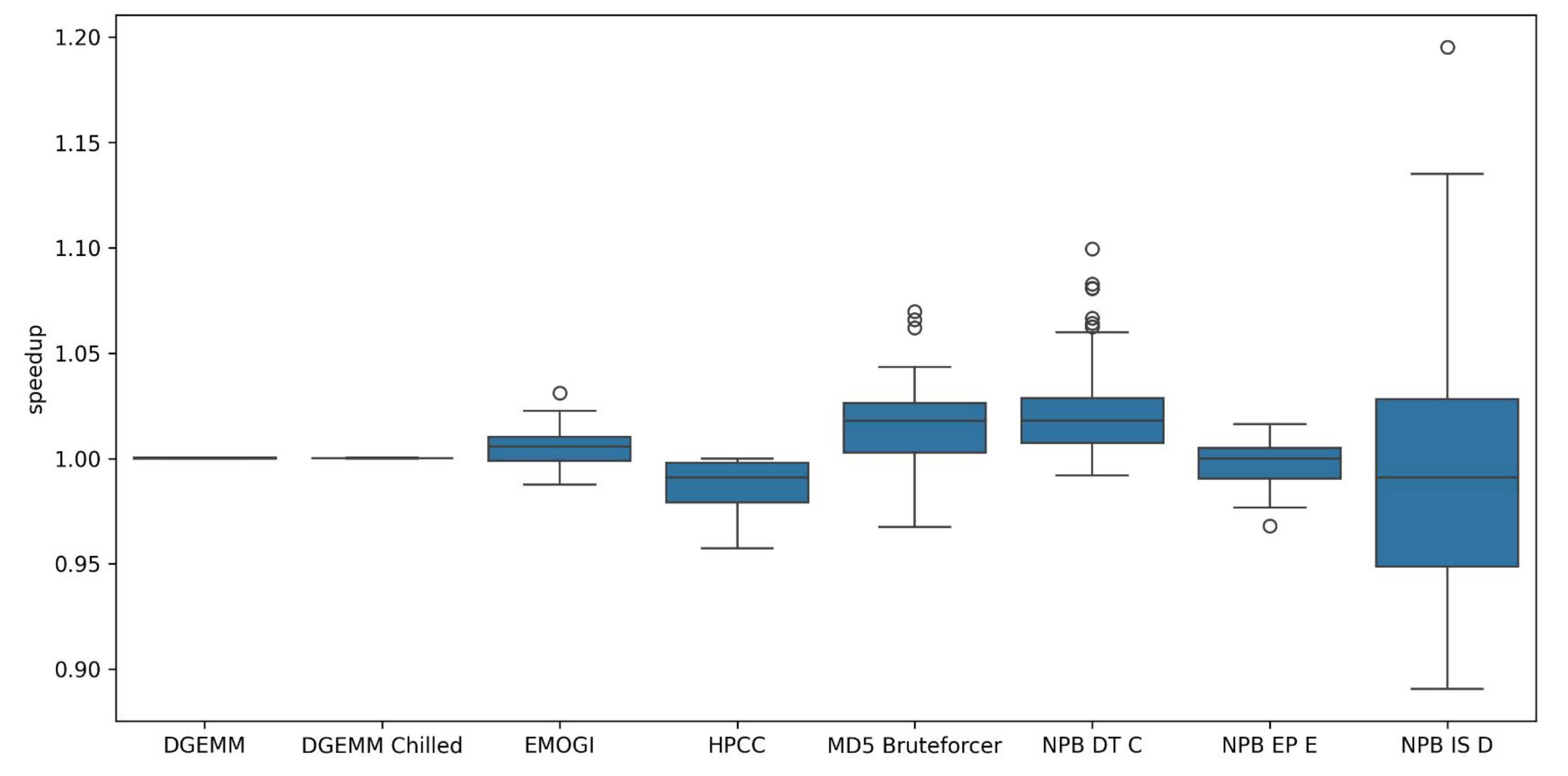


Compute-Utilization More Important than Device

- Lower values = faster heat accumulation



No Significant Performance Deviations



Conclusions

- Coolant normalizes per-component heat increases
  - Circulation semi-sufficient due to excess coolant volume
- Delayed responses require thermal benchmarking
  - Can permit components to throttle at low coolant temperatures
- Compute is primary driver of heat accumulation
- External observations and considerations
  - Preparation and maintenance
    - Irreversible commodity hardware modifications
    - Additional care during maintenance
  - Reliability
    - No degradations in computing performance
    - Plastic components may require periodic replacement
    - Power outage lead to pump failure – requires specialized nontrivial maintenance

Contact: tlranda@clemson.edu

Anticipated dissertation defense in May 2025

Experiment code and data are available online

This research was partially supported by the U.S. National Science Foundation under Grant CNS-CCF-1942182.