Hybrid Warm Water Direct Cooling Solution Implementation in CS300-LC

Roger Smith
Mississippi State University
Giridhar Chukkapalli
Cray, Inc.

Safe Harbor Statement

This presentation may contain forward-looking statements that are based on our current expectations. Forward looking statements may include statements about our financial guidance and expected operating results, our opportunities and future potential, our product development and new product introduction plans, our ability to expand and penetrate our addressable markets and other statements that are not historical facts. These statements are only predictions and actual results may materially vary from those projected. Please refer to Cray's documents filed with the SEC from time to time concerning factors that could affect the Company and these forward-looking statements.

Legal Disclaimer

Information in this document is provided in connection with Cray Inc. products. No license, express or implied, to any intellectual property rights is granted by this document.

Cray Inc. may make changes to specifications and product descriptions at any time, without notice.

All products, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Cray hardware and software products may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Cray uses codenames internally to identify products that are in development and not yet publically announced for release. Customers and other third parties are not authorized by Cray Inc. to use codenames in advertising, promotion or marketing and any use of Cray Inc. internal codenames is at the sole risk of the user.

Performance tests and ratings are measured using specific systems and/or components and reflect the approximate performance of Cray Inc. products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

The following are trademarks of Cray Inc. and are registered in the United States and other countries: CRAY and design, SONEXION, URIKA, and YARCDATA. The following are trademarks of Cray Inc.: ACE, APPRENTICE2, CHAPEL, CLUSTER CONNECT, CRAYPAT, CRAYPORT, ECOPHLEX, LIBSCI, NODEKARE, THREADSTORM. The following system family marks, and associated model number marks, are trademarks of Cray Inc.: CS, CX, XC, XE, XK, XMT, and XT. The registered trademark LINUX is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis. Other trademarks used in this document are the property of their respective owners.

Agenda

- The Problem
- Many ways to attack The Problem
- Cray's Solution
- Benefits of Warm Water Cooling
- Real-world analysis of CS300-LC performance at MSU

The Problem

Power & Cooling

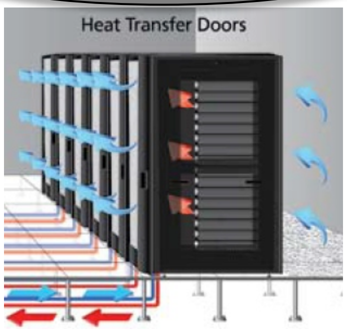
- Power density of processors rising
- System density increasing
- Power now a major design constraint
- Reaching limits in air cooling

Many ways to attack the problem

Any Server Solutions

Server Specific Solutions

Heat Transfer / Rear Doors



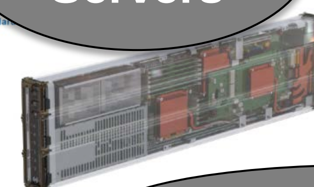
Immersion Racks

CRAY CS300-LC

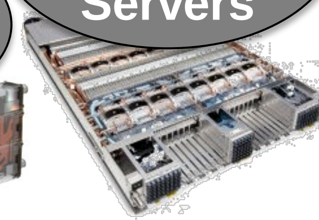
Cold Water Solutions

Warm Water Solutions

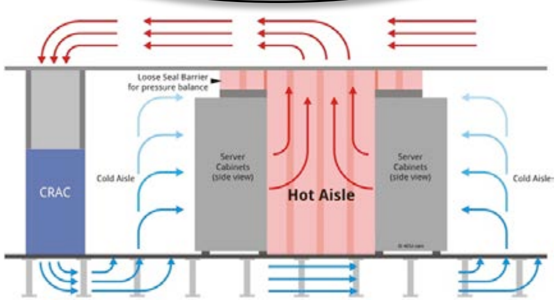
Immersion Servers



IBM /Bull D2C Servers



State-of-the-art Air Cooled Data Center Technology



Sealed Racks

In-Row Coolers

Direct Touch LC

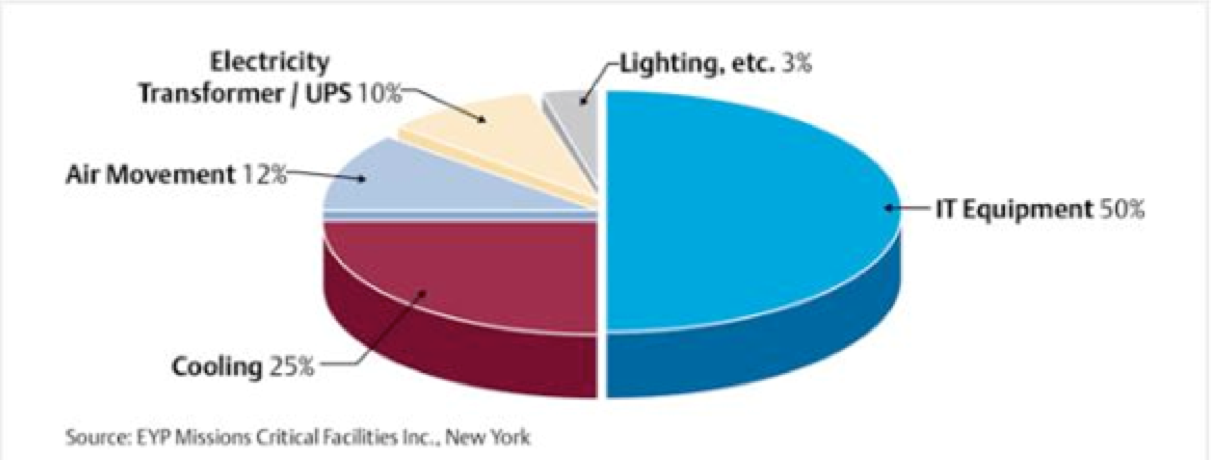
But not all methods are equal…

Cold water solutions address a small part of the problem:

- Reduce energy needed to move air in data center – good
- Same energy for chillers – still using chilled water
- Same energy to move air through servers – still air cooled

Warm water solutions address the whole problem:

- Eliminate chiller use on liquid cooled load
- Reduce energy needed to move air in data center
- Reduce energy needed to move air through servers

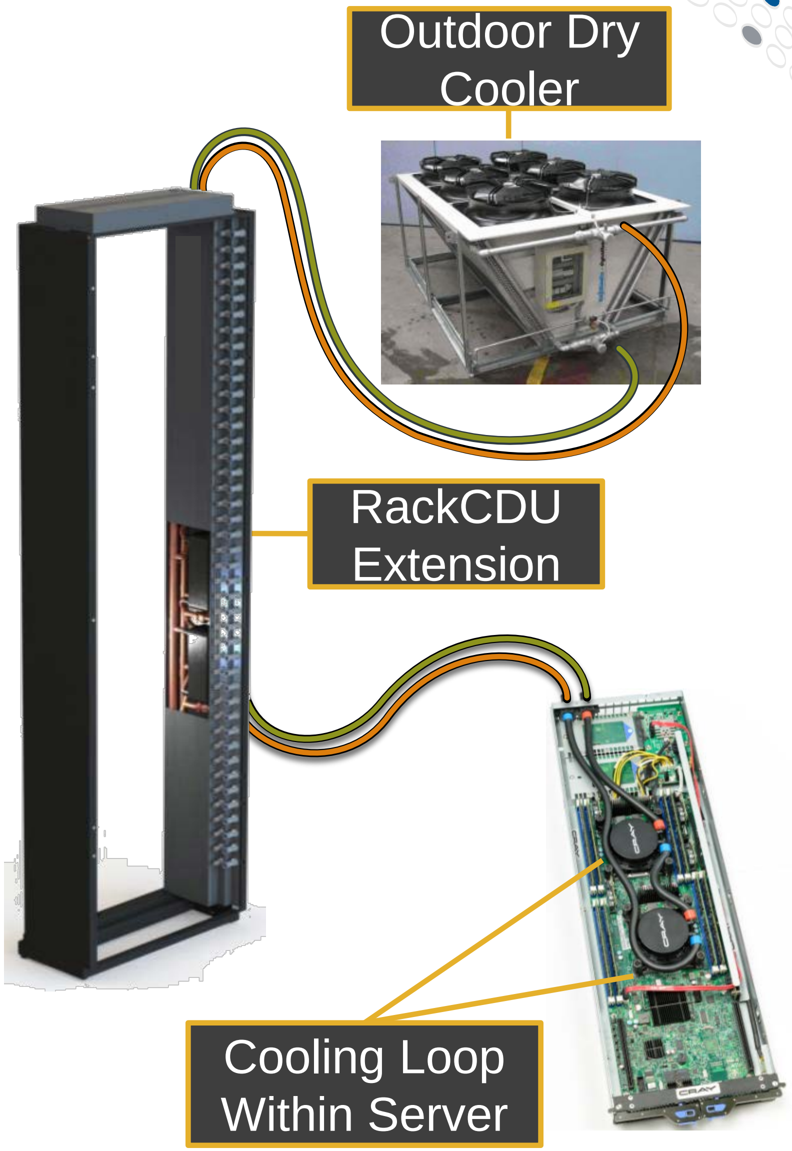


Cray CS300-LC Solution

- **CS300-LC is a Warm Water, Direct-to-Chip Liquid Cooling System that:**
  - Reduces data center cooling costs
  - Enables data center waste heat recovery
  - Support for higher TDP CPUs & Co-Processors in small form factors
- **Partial Liquid Cooling Solution – processors, memories & co-processors**
- **Complete alarm & monitoring system with automated emergency shut down**

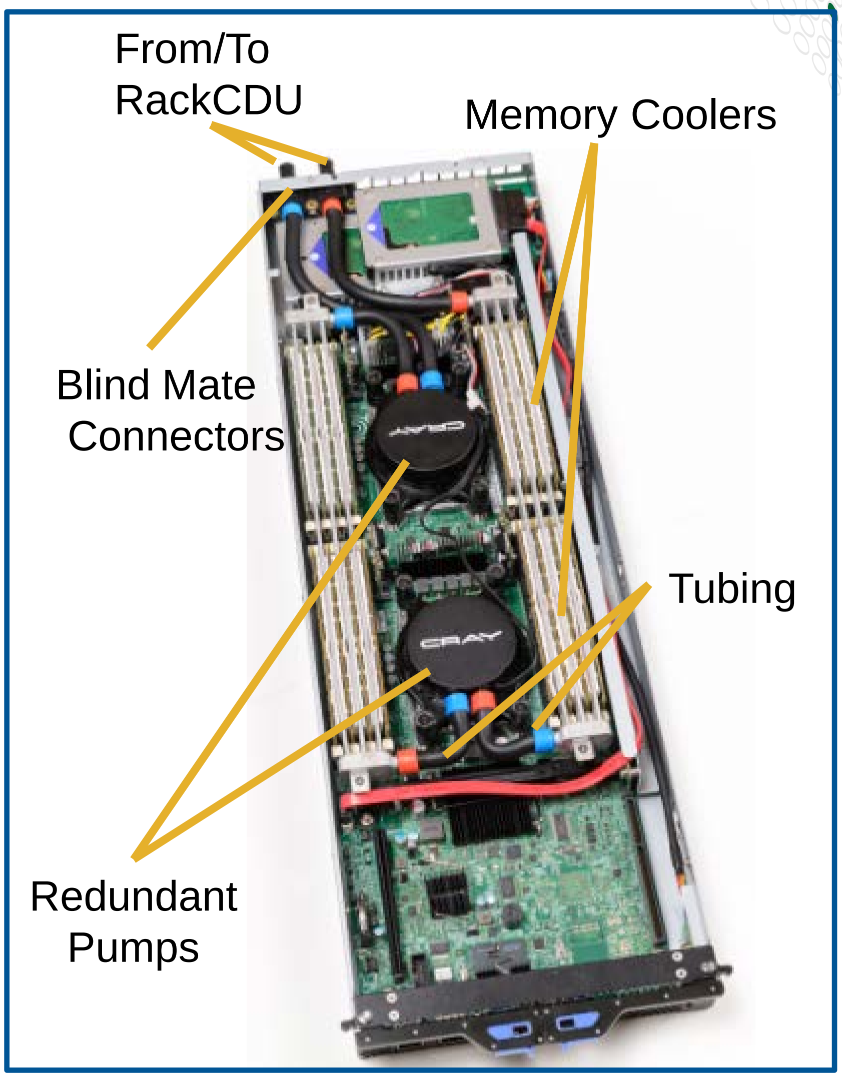
How It Works

- **Three Key Elements in the System:**
  - Outdoor Dry Cooler / Cooling Tower
  - Integrated RackCDU with L2L HEXs
  - Cooling Loop within Blade server
- **Air Cooling is used to cool components that are not liquid cooled**
- **Sever Cooling Loop is a drop in replacement for CPU & GPU air coolers**
- **Memory Coolers insert between standard DIMMS**
- **RackCDU separates Facilities and Blade server Cooling Liquid at the Rack.**
- **Facilities Liquid Cooled with "Free" Ambient Outdoor Air, No Chilling Required**
  - Dry Coolers, Cooling Towers, or Waste Heat Recycling used to take heat from facilities liquid



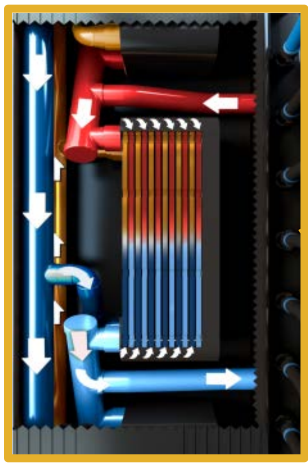
How it Works - Server

- **Server Loops are delivered fully tested, filled and ready to install.**
  - IT staff never has to handle liquid
  - Low Pressure , Factory Sealed Design Eliminates Risk of Leaks
- **Pump / Cold Plates Install Like Standard Air Coolers.**
  - Pumps/Cold Plate Units Replace Air Heat Sinks and circulate cooling liquid
- **Dual In-series Pumps Provide Redundancy**
- **Support for high TDP Intel® Xeon® processor & Xeon Phi™ coprocessor**



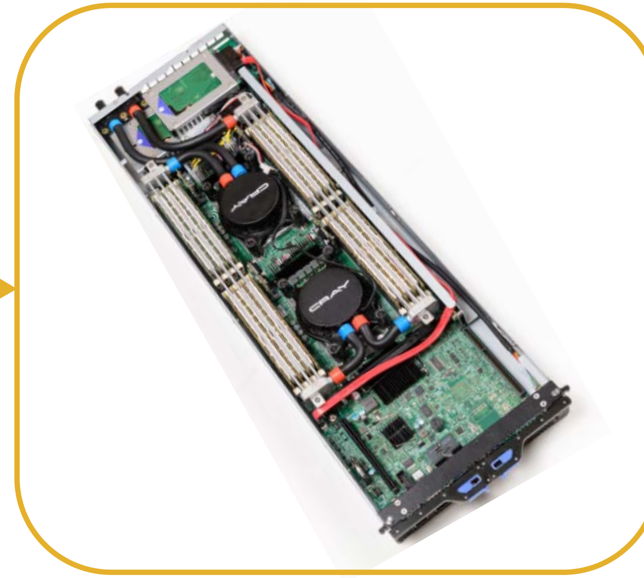
How it Works

Warm water from Facilities dry cooler or cooling tower enters RackCDU, hotter water returns.



Liquid-to-liquid HEXs exchange heat between facilities liquid loop and server liquid loops. Facilities and server liquids are kept separate and never mix.

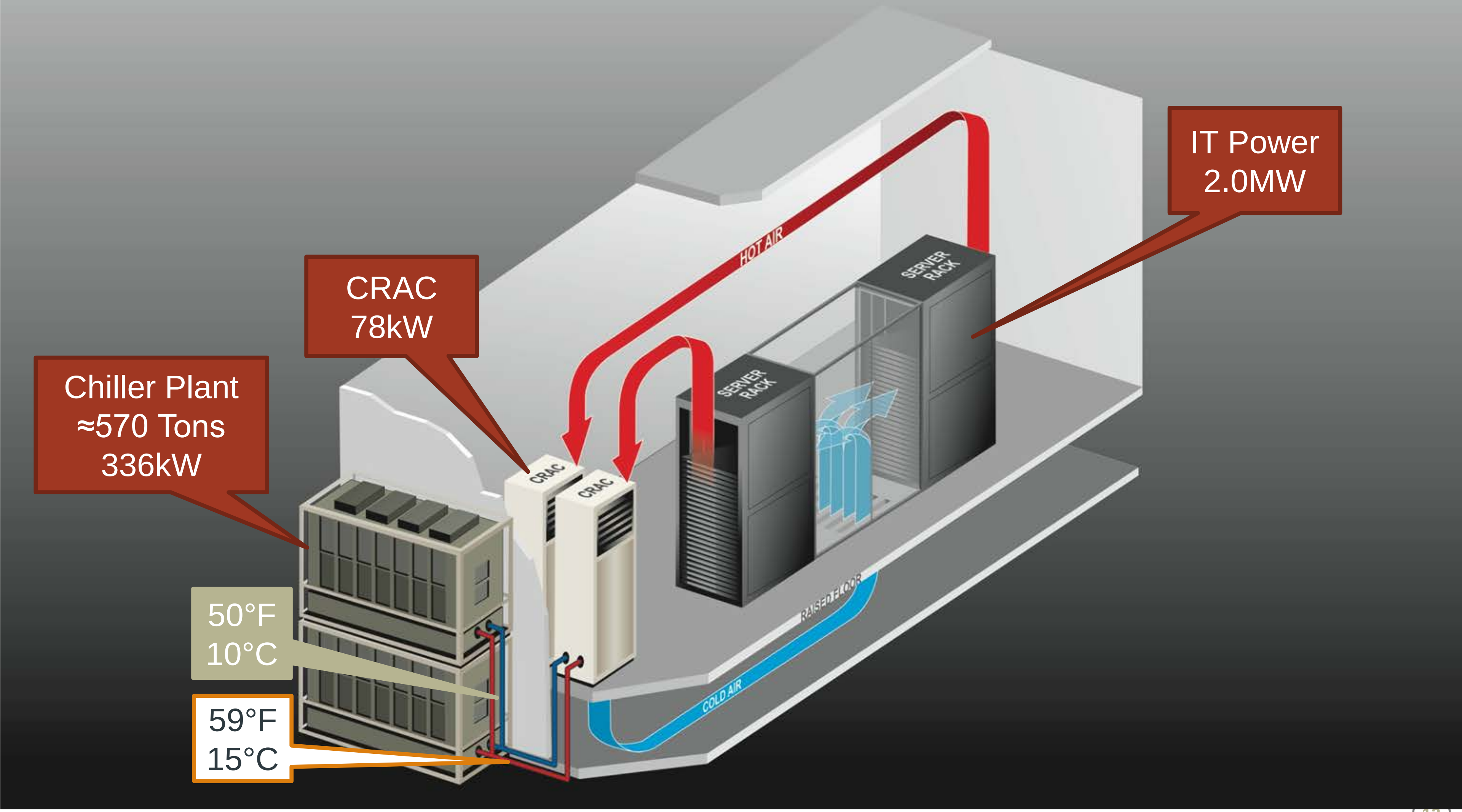
Tubes move cooling liquid to and from RackCDU to servers.



Pump/cold plate units atop CPUs (or GPUs) circulate liquid through blades and RackCDU, collecting heat and returning to RackCDU for exchange with facilities liquid.

Benefits of Warm-Water Cooling Traditional Air Cooled Datacenter

Total Energy: 2.414MW



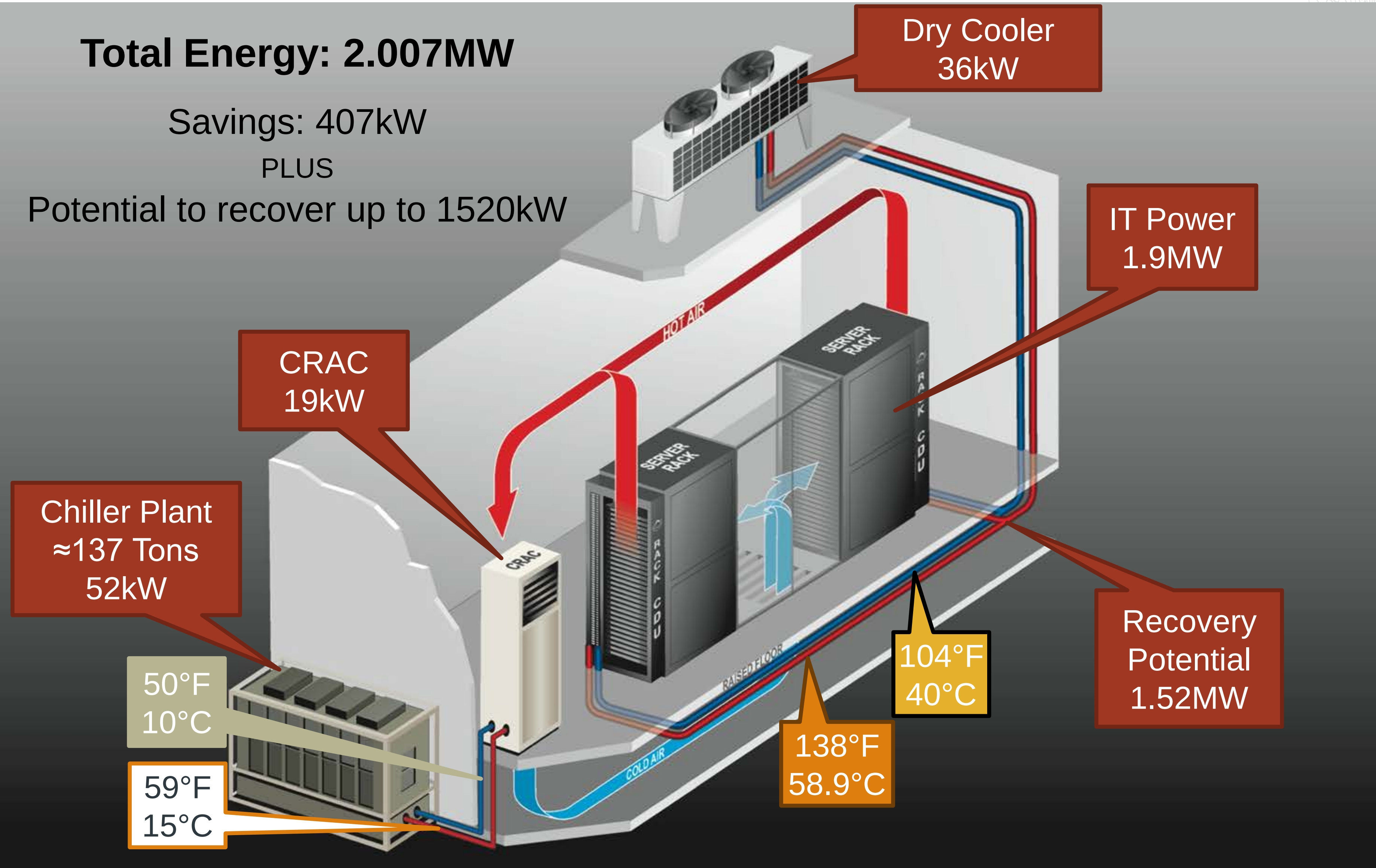
Benefits of Warm-Water Cooling
Warm-water Cooled Datacenter

Total Energy: 2.007MW

Savings: 407kW

PLUS

Potential to recover up to 1520kW



Summary

- Power & Cooling is a major consideration in any HPC system purchase
- Silicon power density and overall rack power density will continue to rise
- Direct liquid cooling is a real viable option for commodity-based clusters
- Warm-water, direct liquid cooling solution can have significant impact to both CAPEX and OPEX

Now, from the MSU perspective

Why water cooling?

- In 2010 we were already struggling to air cool existing systems with CRAC units.
- Began planning for additional system.
- Did not have capacity to air cool new system.
- Installed new computer with chilled rear doors using excess building chilled water capacity.
- This has worked very well. The system actually produces more cooling than the heat load it generates.

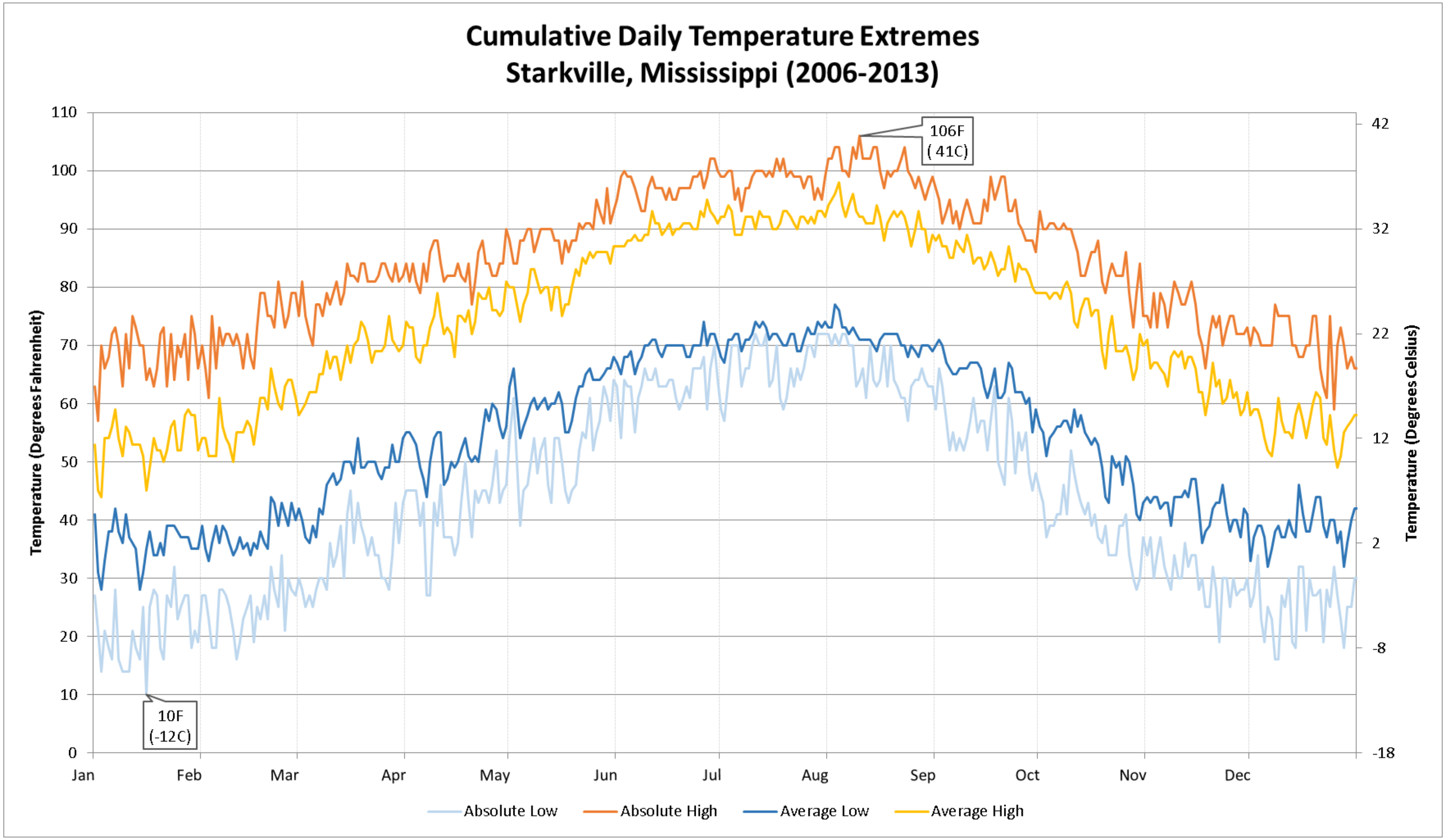
Fast forward

- In 2013, we began planning our next computer.
- Liked water cooling solution, but did not have enough additional capacity in building chiller plant to support an additional large system.
- Additional water cooling would require a new chiller dedicated to the computers.
- Facilities upgrades and computer are funded from the same source. Money spent on a new chiller would mean a smaller computer.
- Researchers want cycles, not chillers.

Warm water cooling?!

- We had previously seen warm water cooled systems, but didn't really understand them as a complete system.
  - Time to do some homework!
  - Water cools processors and memory directly. These are normally very warm, so water doesn't have to be cold to have enough differential temperature to cool them sufficiently.
  - The closer the water is to the heat source, the more efficiently it will cool.
    - ✓ CRAC < chilled rear doors < direct cooling
  - Can be done as a free cooling solution if designed carefully
  - Free cooling is great in northern latitudes, what about Mississippi? Can we really cool it without a chiller?

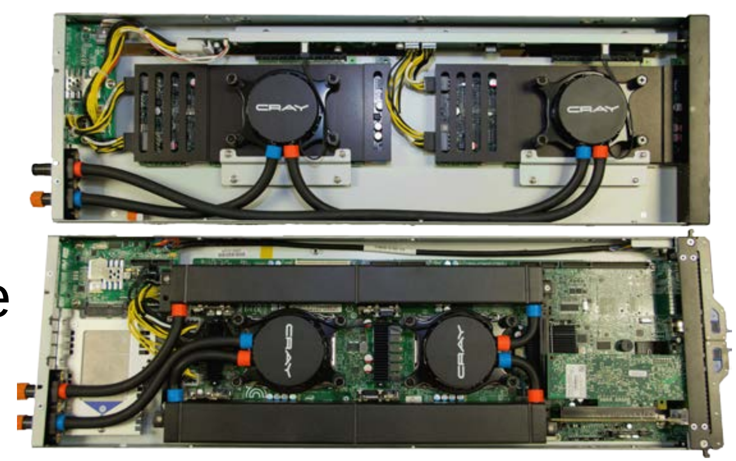
Well, maybe.



MISSISSIPPI STATE UNIVERSITY

Cray CS300-LC

- Direct, warm-water cooled
  - Input water temperature up to 40C (104F)
  - Only system on the market that could water cool the processor, memory, and Intel Xeon Phi (or NVIDIA GPU)
  - Secondary water loop with low pressure and low flow into each node
  - Each CPU (and each Xeon Phi) has its own water pump, so built in redundancy.
  - Lots of water sensors with a nice monitoring and alert capability

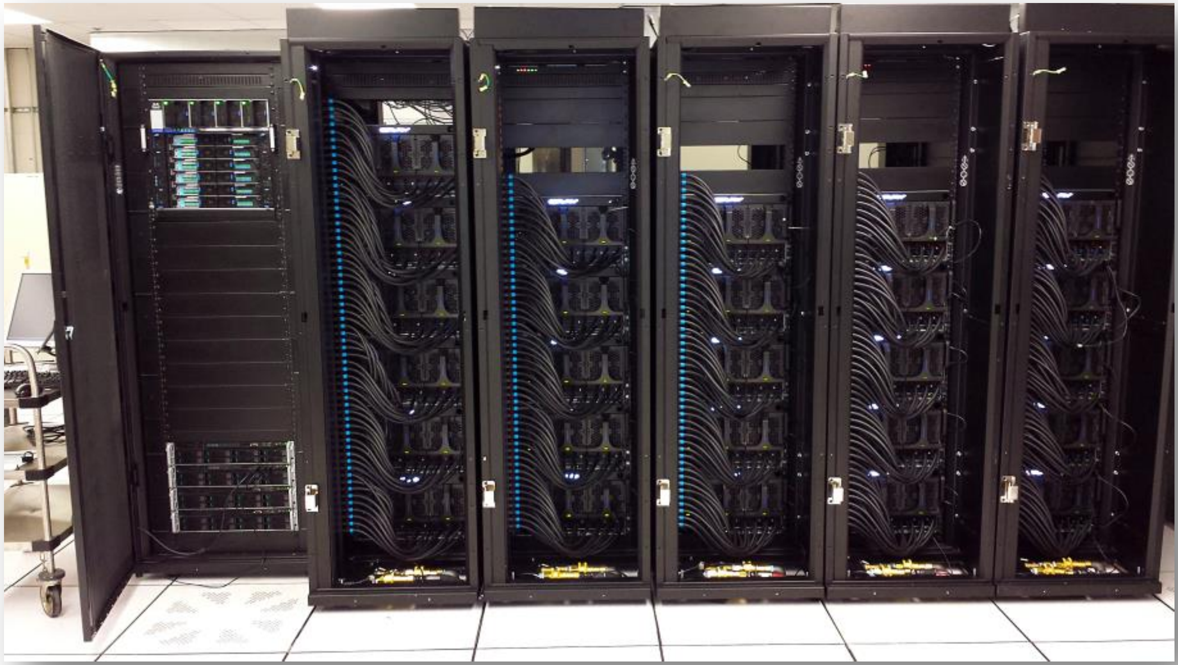


Shadow configuration

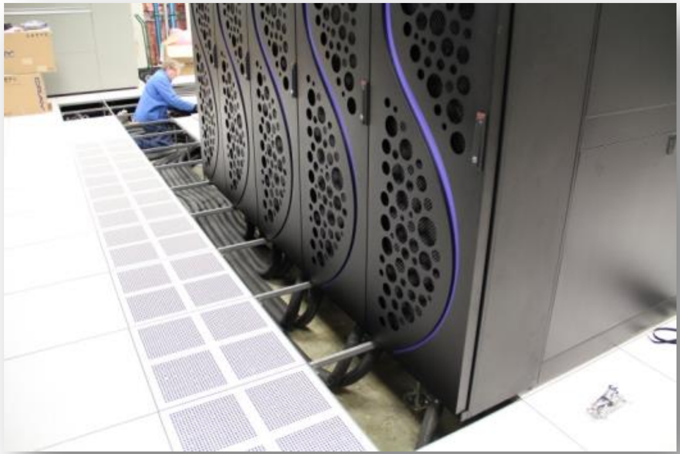
- 128 compute nodes
  - Two Intel E5-2680 v2 “Ivy Bridge” processors,
    - 10 cores (2.8 GHz)
  - 64 GB memory (DDR3-1866)
  - 80GB SSD drive
  - Mellanox ConnectX-3 FDR InfiniBand (56Gb/s)
  - Two Intel Xeon Phi 5110P coprocessors
    - 60 cores (1.053 GHz)
    - 8 GB GDDR5 memory
- Two redundant management nodes
- Three login/development nodes
- Fully non-blocking Mellanox FDR InfiniBand network
- Compute node peak performance: 316 TFLOPS
  - Achieved 80.45% efficiency on Linpack (254.328 TFLOPS)

Shadow configuration

- Five sub-racks per rack in four racks, six sub-racks in one rack.
- Login and management nodes in air-cooled rack with InfiniBand switch
- Facility water fed from below
- (Photo taken during testing at Cray facility in Chippewa Falls, WI)



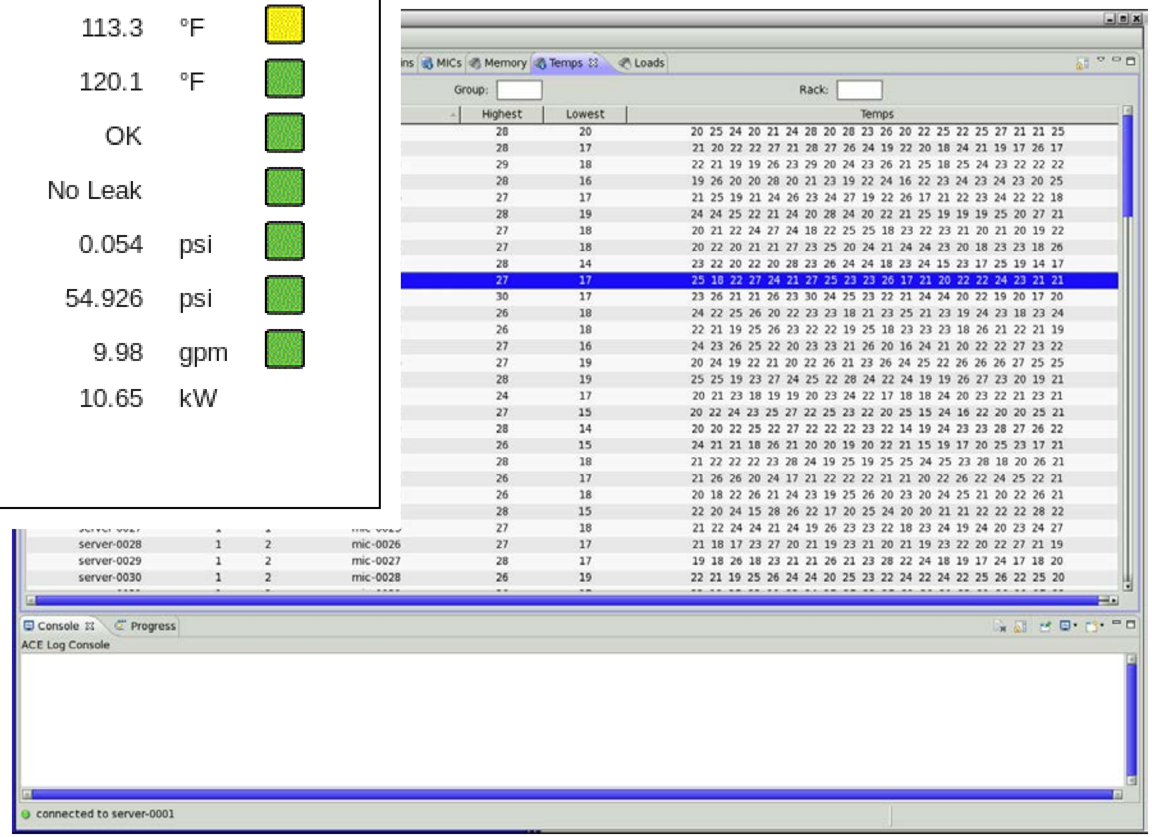
Our cooling system



Monitoring tools

SENSOR OVERVIEW: Rack5

Description:	Value:		
Facility water temperature SUPPLY:	107.0	°F	
Facility water temperature RETURN:	113.0	°F	
Server liquid temperature SUPPLY:	113.3	°F	
Server liquid temperature RETURN:	120.1	°F	
RackCDU™ liquid level:	OK		
RackCDU™ leak detection:	No Leak		
RackCDU™ pressure:	0.054	psi	
Facility pressure:	54.926	psi	
Facility water flow:	9.98	gpm	
Heat Load: (60 sec)	10.65	kW	

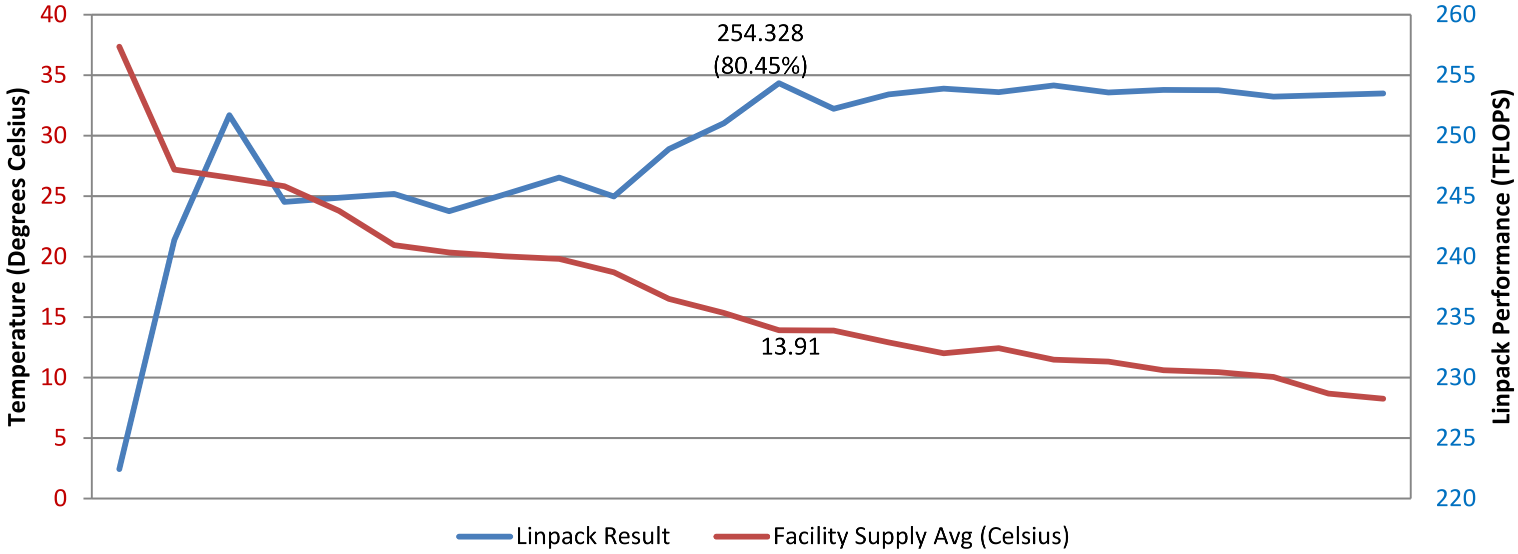


Testing Parameters

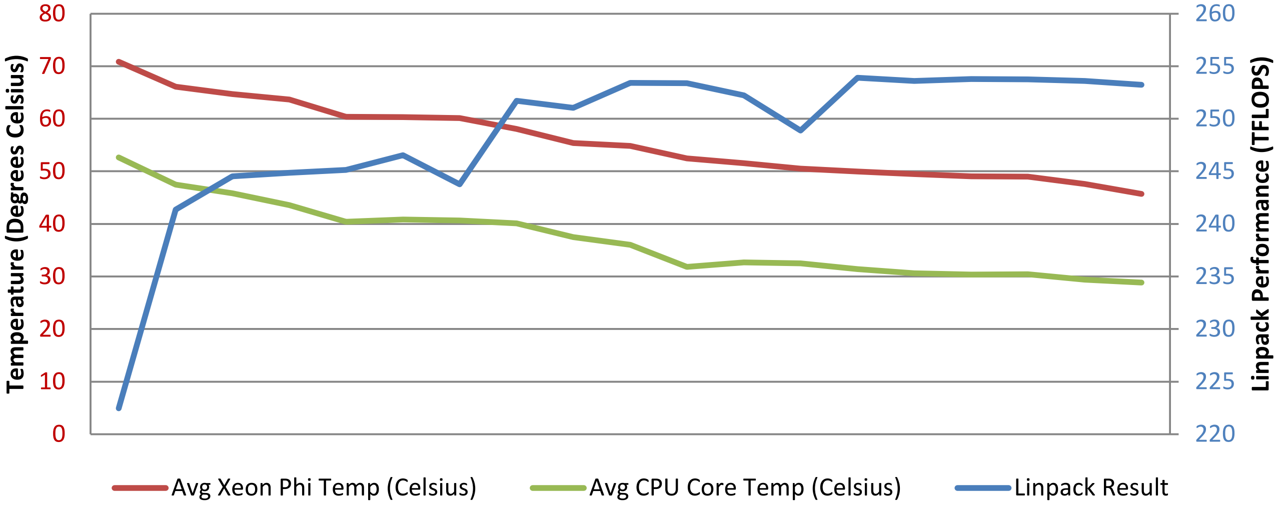
- All data collected during Linpack benchmark runs with system at 100% load and averaged across the duration of the run.
- Facility input fluid flow rate was constant around 586 ml/s (9.3 gpm) except as noted.
- Datacenter input air temperature was 16°C.
- Input fluid temperatures were artificially increased in some cases by shutting down fans in dry cooler

Facility Supply Temp vs. Linpack Performance

- Maximum performance (80.45% efficiency) achieved with input fluid temperature of 13.9°C.
  - No improvement in performance with lower temps.
- Performance dropped ~4% with input temps up to 27°C.
- As input temps reached 40°C, performance dropped to around 70% efficiency



Core Temperatures vs. Linpack Performance



- Best computational performance when CPUs at 21°C and Xeon Phis at 54°C (average)
- Lower temperatures did not improve performance
- Performance remained at 95% of peak with temperatures of 47°C / 65°C

Xeon Phi Frequency Throttling

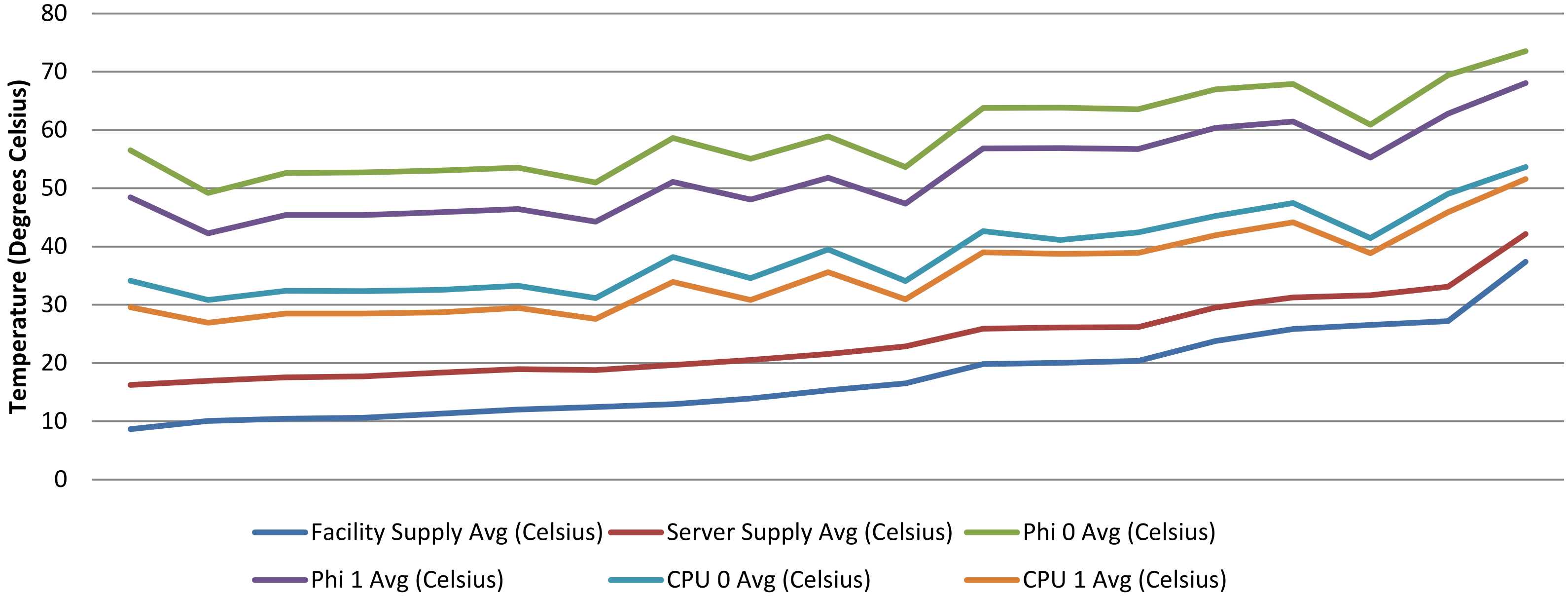
- Xeon Phi 5110P has a TDP of 225W, but may draw up to 245W under load.
  - At 236W for more than 300ms, it throttles ~100MHz (to about 947MHz)
  - At 245W for more than 50ms, it issues PROCHOT_N signal and reduces frequency to minimum values (~820MHz)
  - Linpack and other computationally intense programs may cause this condition.
  - Increased power consumption typically caused by increased chip leakage current at higher temps.

Xeon Phi: Air cooled vs. Water cooled

- Individual node Linpack testing
  - Air cooled login node with same processors, memory and Xeon Phis as compute nodes
  - Input fluid temp: 25°C, input air temp: 16°C
  - Ran 10 consecutive single-node Linpack runs on two air-cooled and two liquid cooled systems. Averaged all results.
    - ✓ Some frequency throttling was observed in air-cooled systems, but not in liquid cooled system

System type	Linpack (TFLOPS)	Xeon Phi Avg Temp (C)	Xeon Phi Max Temp (C)
Air cooled	1.82	72.75	85
Water cooled	2.01	62.5	79

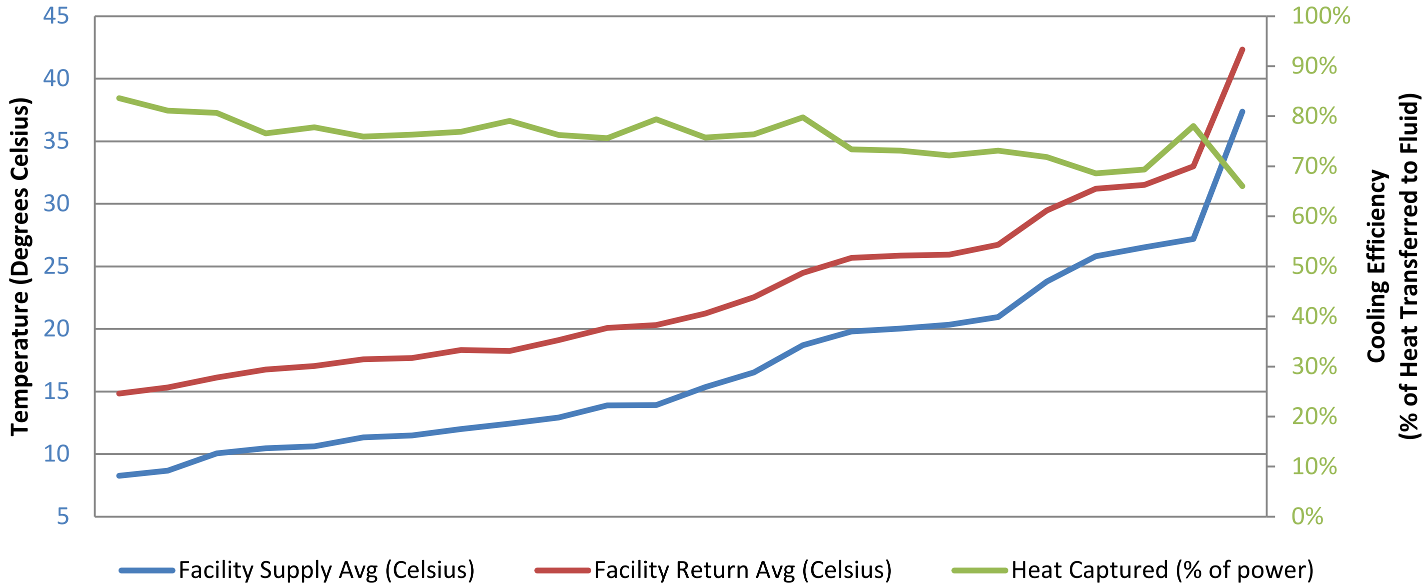
Core Temps vs. Supply Fluid Temps



- With average facility fluid temps of 37°C, internal (secondary) fluid loop temperatures were 42°C
  - CPU core temperatures averaged 53°C
  - Coprocessor core temperatures average 71°C

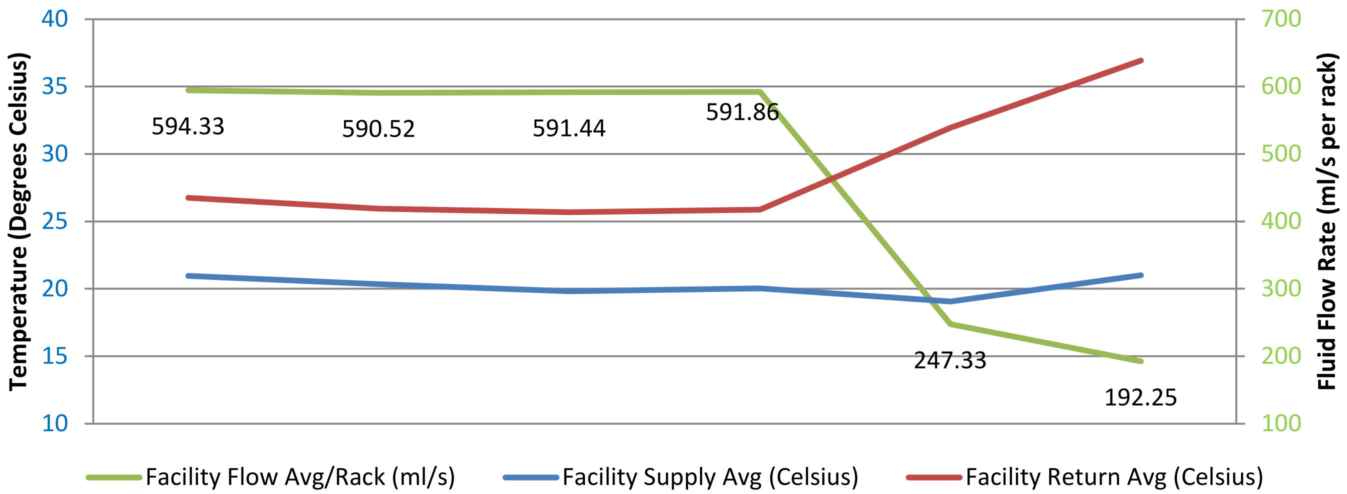
Input Fluid Temperature Effects on Cooling Efficiency

- Cooling efficiency = heat captured by fluid / total power consumed
- Cooling efficiency is better with cooler fluid
  - 84% efficient with >10°C input fluid
  - 66% efficient with >37°C input fluid
  - 76% efficient across all test cases.
- Pretty good across a wide range of facility input temperatures



MISSISSIPPI STATE UNIVERSITY

Effect of flow rate reduction



- Reducing flow of facility fluid to racks increases ΔT of input temperature vs. output temperature
  - Input fluid temperature was 19-21°C during tests
  - At ~586 ml/s (9.3 gpm), cooling efficiency averaged 73% with a ΔT of 6°C
  - At 247 ml/s (3.9 gpm), cooling efficiency averaged 68% with a ΔT of 13°C
  - At 192 ml/s (3.0 gpm), cooling efficiency averaged 64% with a ΔT of 16°C
- Higher differential temperatures are useful for waste heat recovery systems, lowering PUE

Conclusions

- Clear correlation between facility input fluid temperature and processor/coprocessor efficiencies due to throttling.
  - This appears more related to power consumption than directly to cooling (although there is a correlation to power consumption based on core temperature)
  - Tests seem to indicate that water cooled coprocessors are outperforming air cooled versions.
- Even at very high input fluid temperatures (40°C), core temperatures all remained well within recommended temperature parameters.
- With 70%-80% of heat eliminated to water, which is then free-cooled, operational costs are much lower than previous cooling techniques implemented at MSU.
- Free cooling a CS300-LC is possible in warm, humid climates such as in Mississippi with proper design and planning

Questions?