Hybrid Warm Water Direct Cooling Solution Implementation in CS300-LC

Roger L. Smith
Mississippi State University
Starkville, MS USA
roger@hpc.msstate.edu

Giridhar Chukkapalli
Cray, Inc.
San Diego, CA
gchukkap@cray.com

Abstract—**In this white paper, the warm water direct liquid cooling solution implementation in the CCS cluster solution is described. First, literature review of a range of direct liquid cooling solutions implemented in various competing vendors and their pros and cons are briefly presented. Design choices of datacenter cooling water distribution mechanisms (single loop vs primary/secondary loops), types of cooling surfaces (sandwiched single milled metal plate vs individual heat sinks) and types of cooling fluids are reviewed. Next, business and technical constraints of the CCS cluster solution with respect to warm water cooling solution are reviewed. The bulk of the paper explains the implementation details of the hybrid direct liquid cooling solution within the CCS constraints of adhering to open standards and sensitivity to system TCO.**

The CS300-LC direct liquid cooling architecture implements a dual loop system isolating the high pressure datacenter loop from a completely sealed secondary IT loop via in-rack heat exchanger. Deionized water-glycol mixture is circulated to blade sub-racks through chassis manifolds and through drip free quick connectors to individual blades. Redundant low-power pumps circulate secondary fluid through CPU, Memory and accelerator heat sinks at very low pressure and flow rates. A remotely accessible PLC device monitors and manages the cooling system at the rack level. System design is such that compute blades are hot swappable and CPU, memory and individual pumps are field replaceable. In the hybrid cooling system, sub-rack level cooling fans run at lower speeds to extract residual heat thereby substantially reducing the fan power and noise.

In the Cray CCS Lab, a prototype CS300-LC system is instrumented to measure temperatures, flow rates of liquid and air at inlets and outlets. Depending on the system configuration (CPU, memory and accelerators in a node), 60% to 85% of the heat is removed to warm water for a wide range of datacenter inlet water temperatures, and the rest of the heat is rejected to air or indirect cooling systems, hence the term hybrid. In addition, in collaboration between Cray Mississippi State University (MSU), additional detailed measurements will be conducted in a customer datacenter environment to understand energy efficiency and PUE.

Additional details of resiliency, remote monitoring and management of the hybrid cooling system are described. The chassis level leak detection mechanism is described. How the CS300-LC cooling system reacts to rapidly changing application workloads and hence cooling loads is reported. The paper is concluded by exploring potential future work of making CS300-LC close to 100% warm water cooled and enhancements to Software tools needed for fine grained, tightly coupled management of the cooling system.

Keywords-CS300-LC, direct coolimg, warm water, Xeon Phi, MIC, performance metrics

I. INTRODUCTION

While Moore's Law scaling of doubling transistor density continues to hold, Dennard scaling of voltage and transistor power scaling ended several years ago. In addition, leakage currents are also gradually growing. As a result, processor chip power requirements have been gradually increasing. As an example server CPU thermal design power (TDP) has grown from 64W to 115W over the last few years and is expected to grow further. Similarly, as memory clock speeds continue to increase, DIMM power consumption is also increasing despite operating voltage reductions. Socket power is also increasing due to packaging innovations and moving additional components like fast memory, and the network interface controller into the package. For example, accelerator sockets can produce greater than 200W/socket. Board power is increasing as the I/O speeds and SerDes speeds increase. Blade, chassis and rack power densities are also growing as the pressure to pack more components per rack increases. Heat loads per standard 19" rack have increased from few kWs per rack to well over 50kW over last few years.

As the heat dissipation densities increase, airflow rates needed to extract the heat effectively need to increase (with a rule of thumb of 1000 CFM per kW) with a corresponding decrease in chilled air temperature. As component densities increase, pressure differences needed to drive the required air flow rates increase as well as the fan power needed to drive the airflow.

Direct liquid cooling can be defined as transferring heat directly to the liquid from the heat dissipating IT components without involving air. Efficiency of direct liquid cooling becomes clear when the key thermal properties, namely heat capacity and thermal conductivity of air vs various liquids (especially water), are observed. Table I lists such a comparison for commonly used heat transfer fluids in the datacenter like Freon, water, etc.

TABLE I. COOLING EFFICIENCIES OF VARIOUS MATERIALS

	Thermal Conductivity (W/m*K)	Heat Capacity (kJ / L*K)	Latent Heat (kJ/kg)
Air	0.024	0.00121	
Freon-Liquid	0.07-0.09	1.285	200
Water	0.582	4.18	2257
Mineral Oil	0.138	1.34	

As can be seen from the table, thermal conductivity of water is approximately 25 times that of air. As a result, to transfer a given amount of heat, the temperature difference between the surface of a processor case and air needs to be 25 times higher than that of the processor case and water. Similarly, the volumetric heat carrying capacity of water is three orders of magnitude higher than that of air, requiring proportionally higher air flow rates compared to water.

A wide range of direct liquid cooling (DLC) solutions have been implemented over the years. Immersive cooling is where heat dissipating IT equipment is completely immersed in a non-conductive fluid like Fluorinert or mineral oil. The advantages of this technique are that 100% of the heat can be captured by the liquid and the power required to transfer the heat can be low. Some of the disadvantages are system serviceability and non-standard datacenter requirements. Most commonly used DLC methods encase or sandwich the system board (PCB) with a cold plate with milled surfaces in direct contact with heat generating components. The heat carrying fluid is circulated through micro-channels in the cold plate. If the cold plate covers 100% of the board and the components, the resulting system is fully direct cooled with the potential exception of the power supply unit (PSU). In partially covered cases, the remaining heat can be extracted by heat pipes that in turn transfer the heat to the IT liquid loop through direct contact heat exchange. Cost, weight, and serviceability complexities are high with 100% direct liquid cooling. Different fluids can be circulated though the cold plate. Water and Freon are the most common fluids. In the case of Freon, latent heat from evaporation is used to transfer heat. In the case of water, either datacenter water can be directly circulated through the cold plate in a single loop implementation or datacenter water can be isolated from the IT loop through a heat exchanger.

II. IMPLEMENTATION

One of the Cray CS300-LC direct liquid cooling implementation criteria is the requirement to use off the shelf technologies as much as possible while still adhering to Cray Cluster solutions serviceability models, in order to achieve the best possible sustained performance per System level TCO and maintaining or improving system reliability, accessibility, and serviceability. Analysis of various direct liquid cooling solutions using the above criteria pointed to restricting direct liquid cooling to the components which generate the most heat, namely, CPUs, accelerators and memory. Depending on the node and system configurations, these three components contribute to 65% to 85% of the total system heat load. This enables Cray to incur the least possible non-recurring engineering (NRE) costs, keeps the system acquisition costs reasonable, maintains serviceability, and enables the use of current and future high TDP CPUs and accelerators. With only 15% to 35% of the heat load from each rack going to the datacenter air cooling, current and next generation denser and higher power racks can be accommodated in more traditional datacenters without requiring an upgrade to the cooling infrastructure (i.e., CRAC/CRAH, chillers, etc.). In addition, server, chassis, and rack level air flow rates (CFM) can be substantially lower to extract the remaining 15% to 35% of the heat. As a result chassis fan power is substantially reduced as well as the system noise levels. In cases where room neutral cooling is required, rear cooling doors can also be employed.

In the following section, the Cray CS300-LC direct liquid cooling set up is briefly described. The CS 300-LC employs a dual loop setup, i.e., the primary data center liquid loop is isolated from the secondary IT loop via a rack level cooling water distribution unit (CDU) heat exchanger. This dual loop setup possesses several advantages over single loop setups. Various components of the IT loop (pipes, couplers, etc.) do not need to be subjected to the high pressures of the datacenter loop. For example, in the MSU setup, the datacenter loop runs at an average pressure of around 2600 mbar (260 kPa or 38 psi) whereas IT loop runs at an average of less than 11 mbar (1.1 kPa or 0.16 psi) pressure. Similarly, the various IT loop components do not need to be subjected to unconditioned datacenter water, which may be conductive, may be corrosive and may contain a high particulate count. As a result, a single loop setup adds to the design complexity, serviceability, and cost of IT loop components.

In addition to the heat exchanging function, the rack level CDU provides onsite and remote monitoring and control functionality. A programmable logic controller (PLC) device on the CDU can monitor and control datacenter loop flow rates and temperatures as well as monitor various IT loop pressures, flow rates and temperatures. It provides remote Ethernet access to the data and is capability of providing alert notifications through SMTP or SNMP protocols. The CDU also has leak detection and control functions built in. The CDU provides individual inlet and out ports to each node in the rack. These ports are connected through pipes to a chassis level manifold which contains drip-free quick disconnects. As a result, server blades can be hot swapped in the field. The secondary IT loop is prefilled with nonconductive deionized water with special additives. Each CDU provides a small 3.6 liter (about 1 gallon) reservoir for the unlikely event of any small leaks or evaporation. In fact, the entire IT loop, including the reservoir, only contains about 8.5 liters (about 2.25gallons) of water in each rack. With such little water in the loop, even a catastrophic failure could cause little to no damage to the facility. Customers should never need to touch the IT loop liquid during a 5 year operation of the system. As a result of the low IT side pressure, economical polypropylene pipes can be used.

The air heat exchanger blocks on the CPU and accelerator sockets in a traditional air cooled setup are replaced by special liquid cooling blocks. Micro-channels are cut through the liquid cooling block connecting the water supply and outlet to the block. A small 2W electric pump is embedded in the cavity of each CPU/accelerator cooling block. Node memory DIMM modules are sandwiched in liquid jackets containing IT loop inlets and outlets. The components of a single node are connected in series from the CDU outlet to the CDU inlet. Since the heat capacity of water is orders of magnitude higher than that of air, downstream components in the IT loop only see a marginal increase in the inlet temperatures as shown in the results section. Since at least two pumps are in tandem in an IT loop, liquid continues to circulate through the node in the case of a pump failure. In addition, a failed pump motor can be field replaced without touching the IT loop liquid. In the case of a leak in the IT loop, loop pressure sensors, as well as physical leak detectors inside the rack itself, will provide notification and automatically take steps via the PLC to protect the system from damage.

III. MSU System Setup

The Mississippi State University CS300-LC system contains both Intel Xeon E5 v2 CPUs and Xeon Phi coprocessors. Each two-socket node has 64 GB of RAM, two ten-core E5-2860 v2 processors, and two Xeon Phi 5110P coprocessors. All processors, coprocessors, and memory are direct water cooled. Five nodes are installed in 6RU sub-racks with either five or six sub-racks in each standard 19” rack. There are a total of 128 compute nodes distributed in five racks, plus one rack for InfiniBand and Ethernet switches, management nodes, and login nodes. The nodes are connected with Mellanox FDR InfiniBand through a single 216 port director-class switch. There are two Ethernet networks for management and datacenter communications. The sub-rack chassis are powered using standard 208V power distribution units (PDUs). Chassis management modules, called iSCBs, monitor and control various aspects of the nodes and sub-racks. Cray ACE scalable system management and provisioning software logs various environmental data points, including sub-rack fan speed and power consumption for the nodes and sub-racks.

The facility cooling subsystem consists of a pumping station, a dry cooler, and a small heat exchanger tied into the main building chilled water loop. This heat exchanger is bypassed most of the year, but is equipped with a temperature sensor and automatic valves that will allow the computer’s main water loop to exhaust heat into the building chilled water system in the event that the computer’s input water temperature exceeds the maximum recommended temperature of 40°C. Once the supply temperature returns below this maximum temperature, the heat exchanger is automatically bypassed and the system returns to a free cooling mode. The computer’s main facility fluid loop is a 25% glycol mix. The dry cooler is a simple radiator with six fans. Two fans run continuously at constant speed while the other four are variable speed and only run when the fluid temperature is above a programmable maximum. The dry cooler has the advantage that it is a closed-loop system, which does not depend on evaporation for cooling. This is advantageous in very warm and humid climates such as those found in Mississippi. It also reduces water consumption since it does not require the use of municipal water to perform evaporative cooling. The pumping station includes redundant pumps with controller circuitry for automatic failover in the case of a pump failure..

IV. Results and Analysis

The information presented in this section was collected during multiple runs of the Linpack benchmark during acceptance testing. All air and fluid temperature, fluid flow rate, fluid pressure, and fluid heat load data points were collected at one minute intervals on each rack CDU unit and averaged between all five of the units. All processor and coprocessor temperature data points were collected via the ACE management software at either one minute or ten minute intervals. It was possible to obtain temperatures for each core of each processor, but the coprocessors provided only a single temperature reading. In all of the following test cases, facility fluid flow rate was held relatively constant, averaging 586 ml/s (9.3 gpm) per rack, except as noted. Linpack was run in hybrid mode such that both processors and coprocessors were performing computations. The system was tested in a datacenter that operates in a production environment. There are other large computer systems in the datacenter, and as a production facility, it is not sufficiently instrumented to collect many data points that might be desirable, such as system noise levels and PUE.

During evaluation of various liquid cooled technologies during its most recent procurement cycle, Mississippi State University personnel were initially skeptical as to whether it was possible to install a warm water cooled system using only free cooling in their warm subtropical environment. The first step in evaluating the possibility was to gather historical temperature data for the Starkville, Mississippi area. A dataset was obtained through the MSU Geosciences Department that provided hourly weather recordings for a seven year period from 2006 through 2013 (Fig. 1). After analyzing the data and evaluating average and maximum highs and lows for each day of the year over that seven year span, it was determined that except in a few peak (typically less than 30) hours per year , local temperature conditions would allow for free cooling with a well-designed warm water cooling solution. In order to maintain system functionality even during these rare peak temperature spikes the decision was made to add a small heat exchanger with the datacenter chilled water loop that would automatically activate if the input temperature of the system’s water loop reaches the maximum allowable threshold.

It is not surprising that CPU temperatures affect system performance, so cooling a processor with hot water may at first appear counterintuitive. However, the most important factors in the efficacy of any cooling system are the difference in temperature between the heat source and the heat sink and the efficiency of the heat transfer medium. It has previously been established that water is an excellent medium for heat transfer, so it is only necessary to establish

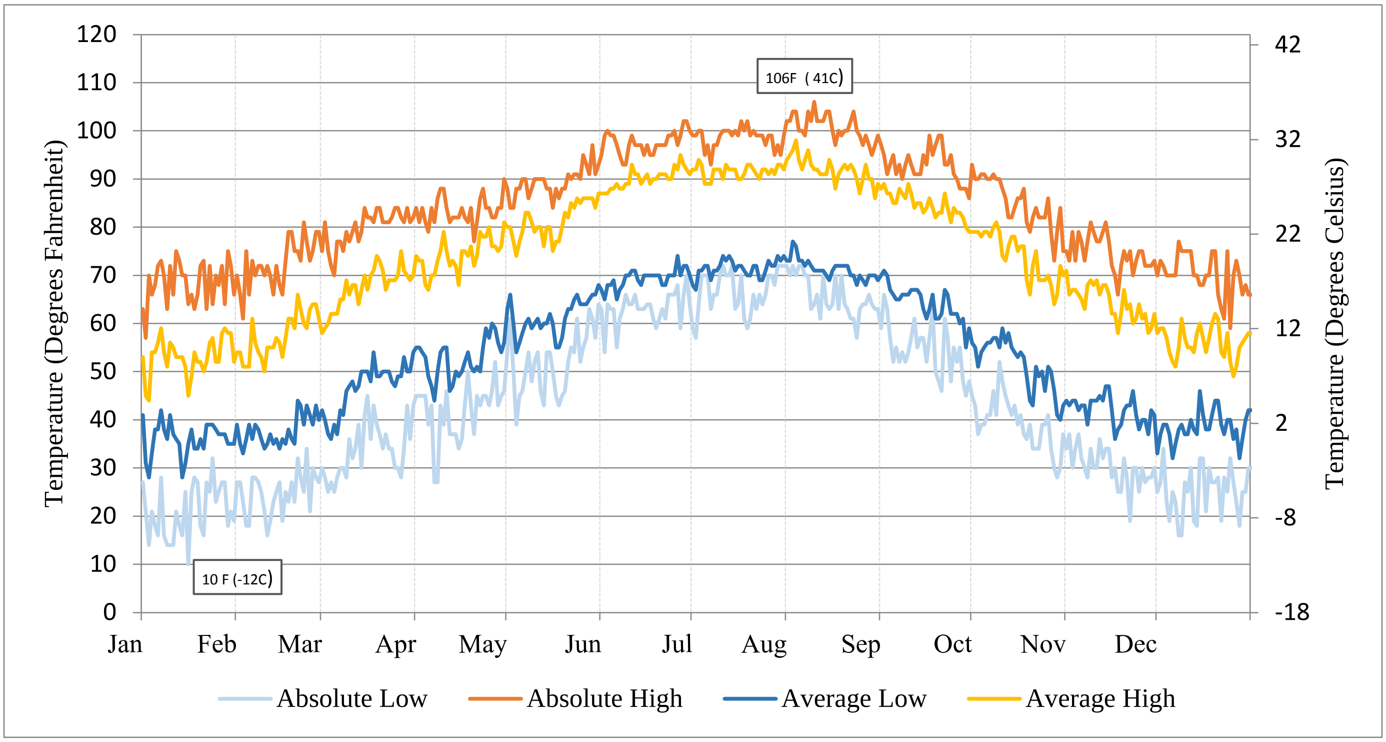

FIGURE I. CUMULATIVE DAILY TEMPERATURE EXTREMES IN STARKVILLE, MISSISSIPPI (2006-2013)

that the differential temperatures (ΔT) are sufficient to cool the heat generating components. Intel recommends an operational temperature range (T_{case}) of 5°C to 82°C [1] for the E5-2860 v2 processors in the MSU system. The Xeon Phi coprocessors will begin throttling ($T_{throttle}$) at 104°C [2]. Fluid temperatures of up to the recommended maximum of 40°C are still less than half the maximum operational temperature of the heat sources, and so are more than sufficient to provide the required cooling capacity.

There are only two significant ways to affect processor temperatures from a customer standpoint: change the temperature of the facility supply fluid or change the flow rate of that fluid. From a facility point of view, balancing supply fluid temperature to computational performance is an important consideration. By varying input fluid temperatures, it was possible to generate multiple Linpack tests with various average processor core temperatures and then analyze system performance from both computational and cooling efficiency perspectives.

In these test cases, with a constant facility fluid flow rate of approximately 586 ml/s (9.3 gpm), a best Linpack performance of 80.45% computational efficiency was achieved with a supply fluid temperature of 13.9°C (Fig. 2). In this case, computational efficiency is calculated as actual performance on the Linpack benchmark (R_{max}) divided by the theoretical peak performance (R_{peak}) of the system. No additional benefit was seen by lowering the input fluid temperatures, which were tested as low as 8.3°C. Computational efficiency decreased just over 4% with increased input fluid temperatures up to 27°C, with computational efficiency remaining above 76% in all test cases. With input fluid temperatures above 37°C, computational efficiency remained around 70%. Computational performance remained relatively consistent over a wide range of supply temperatures with no benefit provided by cooling the supply fluid below approximately 15°C.

The system's computational efficiency ranged from 80.45% to 70.37% across the entire test range of input fluid temperatures (8.3°C to 37.4°C). To put these results into some perspective, there were fourteen systems on the November 2013 Top 500 Supercomputer Sites List [3] that indicated the use of Intel Xeon Phi coprocessors in their system configuration. Only six of those systems had a computational efficiency greater than 70.37%, which was the worst performance recorded in these tests. Only one system had an efficiency greater than 80.45%, which was the best performance recorded in these tests. That system was the Intel Endeavour cluster which produced a computational efficiency of 81.29%. That system uses the Xeon Phi 7110 coprocessor, which provides greater performance, but has a higher TDP of 300W instead of the 225W TDP of the Xeon Phi 5110Ps in the MSU system. So, while there is some change in system performance based on input fluid temperatures, the worst case performance tests with water cooled Xeon Phis still compares very favorably with traditional air-cooled Xeon Phi implementations.

It is critical in any system design that components operate within their required thermal envelope, especially components such as the processors, coprocessors and memory. To better understand how supply fluid temperatures may affect system computational performance, it is useful to study processor and coprocessor core temperatures as a function of computational performance. Even though all components remained well within their

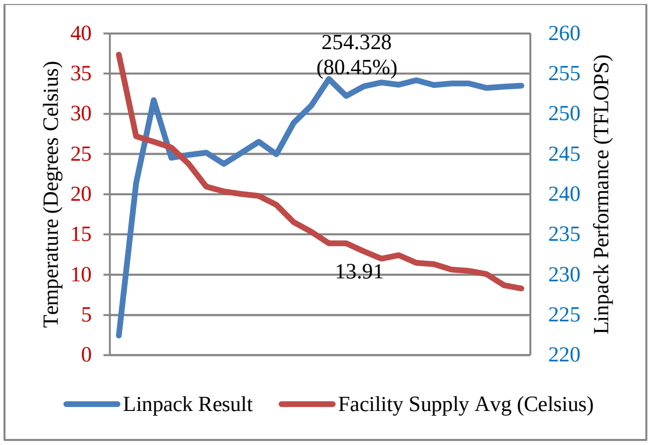

FIGURE II. FACILITY SUPPLY TEMPERATURE VS. LINPACK PERFORMANCE

recommended thermal profile during all testing sessions, some variation in system performance was recorded. The best computational performance was achieved when average processor core temperatures were below 21°C and coprocessor cores were below 54°C. Driving these temperatures lower did not improve system performance. Warmer temperatures did have a slight impact on performance, but even after increasing processor core temperatures to 47°C and coprocessors to 65°C, system performance remained within 95% of the best case performance (Fig. 3).

Intel states that some computationally intensive applications, such as Linpack, may cause the Xeon Phi 5110P to temporarily draw up to 245W of power. It also states that if the power surge is above 236W for more than 300ms, the coprocessor will reduce its operating frequency by approximately 100 MHz. If the power surge is greater than 245W for more than 50ms, the coprocessor controller will assert a PROCHOT_N signal to the coprocessor, which reduces the clock frequency to its minimum values. [2] During testing, it was noted that in many cases some number of the system's Xeon Phis had at least briefly reduced their clock rate from a reported frequency of 1053MHz to a reported frequency of 947 MHz. At no point did any of them throttle to the apparent minimum clock rate of 842 MHz. There is a correlation between the coprocessor core temperature and the likelihood that it will throttle its frequency. Since higher temperatures typically increase chip leakage currents at the same clock speeds, warmer core temperatures may cause higher power consumption, which in this case leads to a reduced operating frequency. Although research into this subject is ongoing, based simply on the excellent computational efficiency achieved on the Linpack benchmark by the MSU system, it appears that its liquid cool coprocessors may be operating at a lower temperature than those in other large air-cooled systems, and therefore throttling less often and producing better average computational efficiencies.

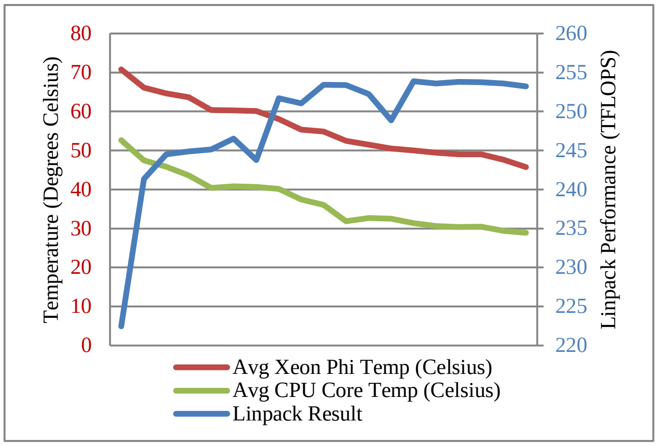

FIGURE III. CPU AND XEON PHI TEMPERATURE VS. LINPACK PERFORMANCE

In order to better understand the difference in performance between the air cooled and liquid cooled systems, a series of ten consecutive single node Linpack runs were made on two liquid cooled compute nodes and two air cooled login nodes. The processor and coprocessor type as well as the amount of memory was the same in all nodes. The input facility fluid temperature for the liquid cooled nodes was 25°C. The input air temperature for the air cooled nodes was 16°C. During these tests, the liquid cooled nodes performed on average 10% faster than the air cooled nodes (2.01 TFLOPS vs. 1.82 TFLOPS). During the tests, the liquid cooled coprocessors had an average temperature over the duration of the tests of 62.5°C and a maximum recorded temperature of 79°C. The air cooled coprocessors had an average temperature of 72.75°C and a maximum recorded temperature of 85°C. At no time during these test cases did the liquid cooled coprocessors throttle their frequency. There was some frequency throttling in the air cooled coprocessors. Although this was a small and overly simple test case, it does seem to confirm that even when operating with fairly warm input fluid temperatures, the liquid cooled Xeon Phi coprocessors are operating at a cooler temperature, and yielding better performance than the air cooled versions. While throttling effects were observed at high input fluid temperatures, it is very likely that in computationally intensive workloads, the observed effect on an air cooled system would be greater.

From an operational perspective, perhaps the most important metric is how the facility supply fluid temperature affects the temperature of the processors and coprocessors. As expected, lower facility fluid temperatures produced lower processor temperatures. Lower fluid temperatures also produced a higher difference in temperature (ΔT) between the fluid and the processors. In tests with an average facility fluid temperature of 37°C, the system maintained a secondary loop average temperature of 42°C. Even with these very warm temperatures, the processor cores averaged less than 53°C, and the coprocessors averaged less than 71°C. Facility supply temperatures were

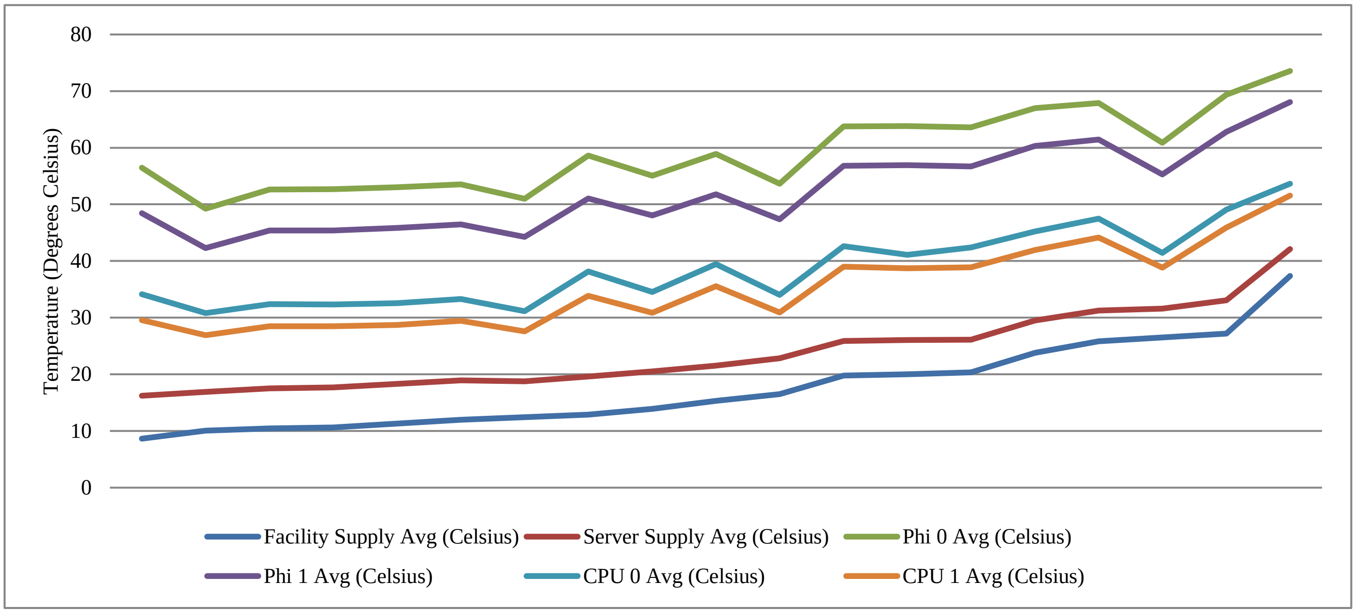

FIGURE IV. CPU AND XEON PHI CORE TEMPERATURE WITH SUPPLY FLUID TEMPERATURE

tested between approximately 9°C and 37°C. This yielded processor core average temperatures between approximately 32°C and 53°C and an average ΔT of 23°C down to 15°C. The coprocessors ranged from 52°C to 71°C over the same facility supply fluid ranges with a ΔT of 44°C down to 33°C (Fig. 4).

Each node in the MSU system actually has two secondary server fluid loops. The first supplies water to the processors and memory. It enters the node, passes through the first memory bank, then onto the first processor then over the second processor, and through the second memory bank before leaving the node. The second loop supplies water to the Xeon Phis, passing over one, then the other in series. Because of this design, it should be expected that the second processor (or coprocessor) in the fluid loop will operate at a warmer temperature than the first one since its input fluid has already been warmed by the component. Testing has proven that while this is true, the effect is somewhat less than might be expected. Across the range of input fluid temperatures tested, the average ΔT between CPU 0 and CPU 1 ranged from 2.1° to 4.5°C. The two Xeon Phis in each node had an average ΔT of 5.5° - 8.1C°. In both cases the differential between the first and second component was reduced as the supply fluid temperature increased. This indicates that cooling the components inside a node sequentially with a single secondary fluid loop actually works quite well, and the minor temperature differences seem to be a fair tradeoff for the reduced expense and complexity that would be required if they were plumbed in some other fashion.

Given the hybrid cooling aspect of the CS300-LC, there may be various factors in the system environment that could potentially affect the efficiency of the liquid cooling system. A constant input air temperature of around 16°C was provided during testing through traditional Freon-based CRAC units. Under load, the output air temperature of the system as measured inside the rear of the racks was also fairly constant around 32°C for all input fluid temperatures. However, because the CS300-LC has fewer, slower fans than are required by air-cooled systems or systems with rear door heat exchangers, the total volume of warm air created was significantly decreased.

The power consumed by the system (in kW) during testing was recorded at the iSCB in each sub-rack. These values were then totaled and averaged on one minute intervals. The rack CDUs each report a heat load (in kW), which may be calculated as a function of the difference of input and output fluid temperatures and the flow rate of the fluid. By comparing the heat load eliminated by the fluid as reported by the rack CDUs to the total power consumed by the system, it is possible to calculate the percentage of heat eliminated by the fluid. This may be expressed as the system's liquid cooling efficiency. Over the range of test cases, cooling efficiency was remarkably stable, averaging 76% of power consumed being eliminated by the liquid. With very cool supply fluid temperatures (below 10°C), cooling efficiencies as high as 84% were achieved. Cooling efficiency decreased slightly to around 70% with warmer input fluid temperatures of nearly 30°C, and as the input fluid temperature approached the maximum recommended 40°C, cooling efficiency held at a fairly consistent 66% (Fig. 5).

As mentioned previously, the flow rate of the facility supply fluid was held relatively constant for most tests, averaging 586 ml/s (9.3 gpm) per rack. However, much lower flow rates are possible. In order to study the effects of lower facility flow rates on system cooling performance, per rack flows were reduced first to an average of around

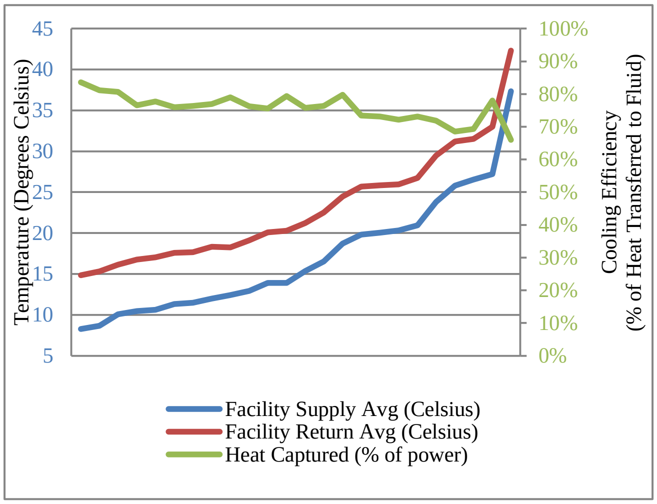

FIGURE V. FACILITY FLUID TEMPERATURES VS. COOLING EFFICIENCY

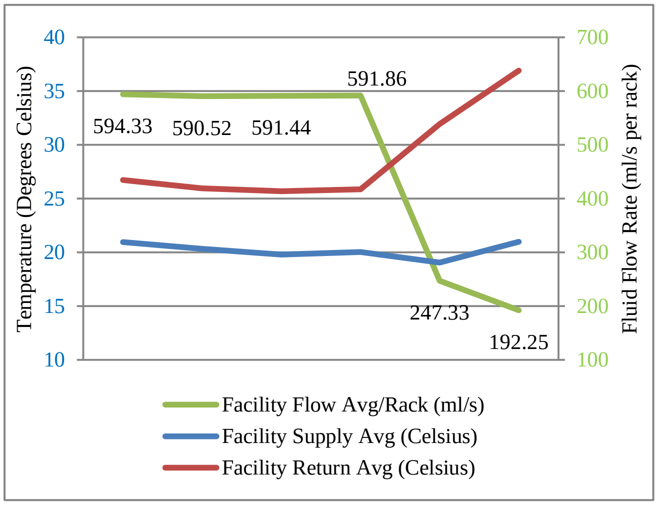

FIGURE VI. EFFECTS OF FLOW RATE REDUCTION

247 ml/s (3.9 gpm), then to around 192 ml/s (3.0 gpm). During these tests, facility supply fluid temperatures remained in the 19-21°C range and the results were compared to tests with the same facility supply temperature range with unrestricted flow. With unrestricted flow and a given supply fluid temperature range, the system averaged 73% cooling efficiency. After reducing the flow roughly 60%, the cooling efficiency was only reduced to 68%. Reducing the flow almost 70% still yielded 64% cooling efficiency. An interesting and potentially useful side effect of flow reduction is that the ΔT between supply and return fluid flows increased from less than 6°C to almost 16°C. In this test case, the fluid on the supply side was 21°C with a return side temperature of almost 37°C, making it an excellent candidate for use in waste heat recovery processes which could, in turn, provide building heat (Fig. 6).

V. FUTURE DIRECTIONS

In future work, we will attempt to derive datacenter PUE and compare the performances of LC and AC systems. This work will include a comparison of internal system temperatures and their relationship to system computational performance.

At the system level, Cray will continue to improve the integration of the CDU's PLC into the chassis management module and system management infrastructure. It will provide better monitoring, guidance and control of processor throttling states. Currently, flow rates are constant on the IT loop, but Cray plans to explore the pros and cons of implementing variable flow rates in controlling throttling states and non-uniform loads.

Currently, about 20% to 30% of the heat generated still must be removed by traditional air cooled methods. Going forward Cray envisions more and more components like the network interface controller, memory, and Northbridge may be integrated into the socket, reducing the miscellaneous heat load from the board. By attaching chilled water rear doors to the racks, the system can be made room neutral. Such a system would still require chilled water and heat transfer from the IT load to the chilled door would still happen through air. Another choice may be to have a heat pipe embedded in the cold plate and attached to the rest of the low intensity heat producing components (e.g., VRMs, Northbridge, etc.). On the other end, the heat pipe would transfer heat to the IT liquid loop. Another choice could be to fully insulate an enclosed chassis with an internal air circulation system with an attached liquid to air heat exchanger on the IT loop.

VI. CONCLUSIONS

We see clear correlation between facility inlet water temperatures and processor/accelerator efficiencies due to throttling effects in the coprocessors; however this throttling appears to be more directly related to the coprocessor power consumption than directly to temperature. In fact, indications are that the warm water cooled coprocessors actually throttle less and perform better than air cooled versions during computationally intensive work loads. Facility fluid inlet temperatures near the maximum recommended temperature cause as much as a 10% reduction in Linpack efficiency over the best test cases at cooler temperatures. This effect would likely not be experienced in a system without coprocessors.

Due to the high thermal capacity of water, case temperature differences between the first and the second processor / accelerator are much smaller than in the air cooled version. Even with very high facility input fluid temperatures and the system at 100% computational load, all processors and coprocessors remained well within the manufacturer's recommended temperature range.

As 70% to 80% of the heat generated by the system is captured through the dry cooler without taxing the datacenter CRAC/CRAH cooling system, datacenter IT capacity can be more than doubled without expanding traditional cooling systems, thereby saving datacenter capital expenditures. In addition, the operating power requirement of a dry cooler and water circulation pumps is a small fraction of equivalent power required by a traditional CRAC unit with an electrical compressor. This results in substantial operational expenditure savings, and therefore contributes to datacenter PUE. There is also further energy savings in chassis fan power compared to an air cooled system. An indirect benefit of LC implementation is substantial reduction in system noise levels (> 10DB) compared to the air cooled version.

Although MSU personnel were initially skeptical about the ability to free cool a direct warm water system in their subtropical climate, preliminary results with artificially elevated fluid input temperatures and 100% computational load has proven that the system is stable, reliable, and remains well within all operating parameters. This proves that free cooling should no longer be considered an option only for those in cool climates.

REFERENCES

[1] Intel Corporation, "Intel Xeon Processor E5-1600/E5-2600/E5-4600 v2 Product Families Datasheet - Volume One of Two, Reference Number 329187-003," Intel Corporation, 2014.

[2] Intel Corporation, "Intel Xeon Phi Coprocessor Datasheet, Reference Number 328209-002EN," Intel Corporation, 2013.

[3] The TOP500 Project, "The TOP500 Supercomputer Sites," November 2013. [Online]. Available: *http://www.top500.org/lists/2013/11*.