Liquid Cooling Heat Capture Ratio in the Attaway and Manzano Supercomputers

2021-03-19

Dr. Steve Harrington
Trevor Irwin

5900 Sea Lion Place
Suite 150
Carlsbad, CA
92010
(760) 476-3429

Executive Summary

The heat capture ratio was measured at the server and data center scale for two top 100 supercomputers using negative pressure direct-to-chip liquid cooling. At the server level, 68% - 82% heat capture ratios were measured with coolant close to ambient air temperature. Reducing fan speed and using cooler water increased the heat capture ratio. At the datacenter level, 66% and 78% heat capture ratios were measured. This study showed that the heat capture ratio is a function of the difference between the coolant and the air temperature, the CPU power, the cooling fan flow rate, and the motherboard and server design.

Contents

Introduction

The *heat capture ratio* is the ratio between heat absorbed by the liquid cooling system and power into the system. It is used to estimate how much air cooling and liquid cooling a system will require. It is calculated as the product of the temperature difference in the coolant, the mass flow rate, and the heat capacity of the coolant, divided by the electrical power into the server

The heat capture ratio of direct-to-chip systems depends not only on the design of the cold plate, motherboard, and server, but also on the operating conditions such as coolant temperatures and server workload. In general, lower server workload, higher air temperatures and lower coolant temperatures will be associated with higher heat capture ratios. A lower thermal resistance cold plate will increase the heat capture ratio. Slowing down the fans will also increase the heat capture ratio.

Any heat not absorbed by the liquid cooling system will need to be removed via air cooling. However, since the processor cooling usually drives the air-cooling requirement, the air flow can be reduced and the air temperature can be increased and much less air cooling is required. A typical direct-to-chip liquid-cooled server will liquid cool CPUs and GPUs, and may also cool DIMMs, VRMs and NIC hardware. The negative pressure approach makes it easier to cool the additional hardware, as the fittings, connectors, and tubing are inexpensive and simple to install.

The following report summarizes the performance of two liquid cooled supercomputers, Attaway and Manzano, that were recently installed at Sandia National Laboratories (SNL). The supercomputers were built by Penguin Computing and were outfitted with Chilldyne's Cool-Flo™ direct-to-chip liquid cooling technology. These warm-water liquid cooling systems – and a single server borrowed from each – will be the focus of this analysis of the heat capture ratio.

Sandia National Laboratory System Overview

In October 2018 Sandia National Labs (SNL), a United States Department of Energy (DOE) national laboratory completed construction on a new data center. SNL has actively been searching for new and innovative cooling ideas to reduce the energy use of the new computer systems that are being installed.

The Attaway supercomputer cluster includes 1,488 compute nodes consuming a total of 645 kW of system power. Three CDUs and six smart failover valves provide N+1 redundancy to cool the 2.6 petaflop system outfitted with Chilldyne's hybrid liquid-air heat sinks.

Table 1: Sandia National Lab's Attaway HPC System Key Attributes

"Attaway"	Manufacturer: Penguin Computing
Cores:	52,920
Memory:	282,240 GB
Processor:	Xeon Gold 6140 18-core @ 2.3 Ghz, 140 watts TDP
Interconnect:	Intel Omni-Path
Performance	
Linpack Performance (Rmax)	2,724.52 TFlop/s

Figure 1 Attaway cluster with Chilldyne CDU on the left end

The Manzano system uses the same liquid cooling system at higher server power.

Table 2: Sandia National Lab's Manzano HPC System Key Attributes

"Manzano"	Manufacturer: Penguin Computing
Cores:	71,424
Memory:	285,696 GB
Processor:	Xeon Platinum 8268 24C 2.9GHz, 205 watts TDP
Interconnect:	Intel Omni-Path
Performance	
Linpack Performance (Rmax)	4,281 TFlop/s

Cool-Flo™ Liquid Cooling Technology

The Attaway and Manzano systems use Chilldyne's Cool-Flo™ direct-to-chip negative pressure liquid cooling technology. Chilldyne's negative pressure system operates under a vacuum which allows for leak-free operation. The supply lines are held at low vacuum (about -4 inHg or -13 kPa) while the return lines are held at higher vacuum (about -18 inHg or -70 kPa.) The pressure difference induces flow through the system. Because the entire system is held under vacuum, it is impossible for coolant to leak out of the lines and into a server. Instead, a damaged seal or a loose fitting leaks air into the lines, and the air is purged by the CDUs.

The Cool-Flo cooling solution uses hybrid liquid-air heat sinks that minimize the thermal resistance to the coolant while allowing for air-cooled backup. Liquid cooling is distributed by rack manifolds with server flow control to prevent air leaks in one server from affecting any other server. The systems have each been designed with smart failover valves to enable N+1 CDUs to provide redundant cooling.

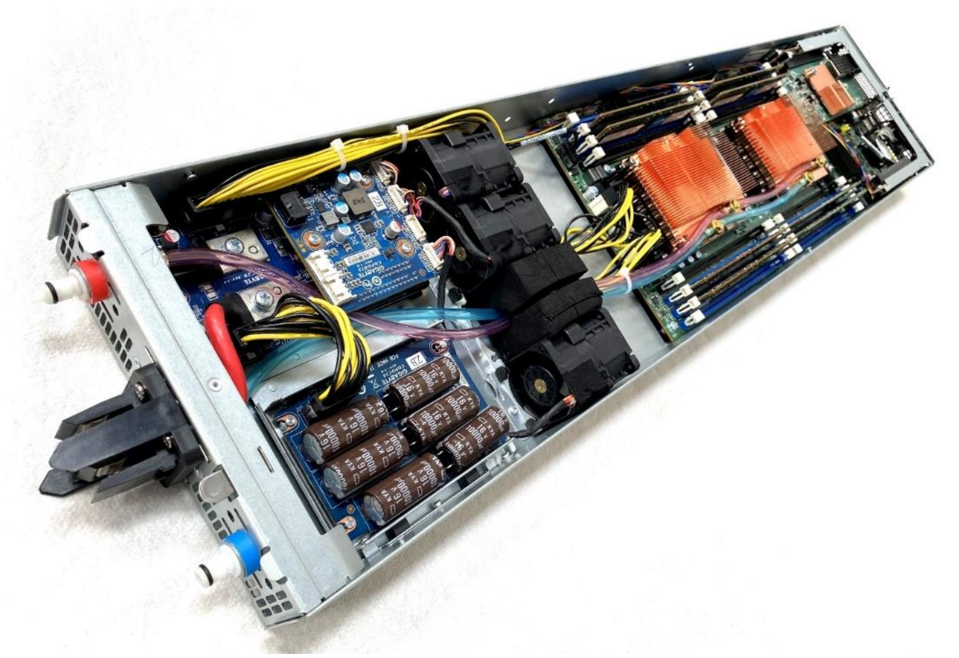

Figure 2 Relion server with two Xeon Gold processors with Cool-Flo™ hybrid liquid-air heat sinks

Understanding Heat Capture Ratio

In the context of liquid cooling, the *heat capture ratio* is a useful metric for understanding how efficiently a datacenter makes use of its liquid cooling system. The heat capture ratio is defined as the amount of heat captured by the liquid cooling system as a fraction of the electric power input to the servers or the data center. This ratio depends on the server workload and on coolant and air temperatures. It also depends on the server fan speed. If a server using hybrid liquid-air heat sinks is lightly loaded, the air is hot, and the coolant is cold, then the heat capture ratio can be over 100%. This means the air coming out of the server is colder than the air coming in.

While heat capture ratio tells us something of the efficiency of a system, it does not provide the entire story. Like the fuel economy of an automobile, the heat capture ratio is a function of many different factors. The design of the system is important, but so is the way it's operated. Just as your car gets great

MPG going downhill, the server heat capture ratio will vary depending on how it is used. At this point there is no standard for server load, air temperature and coolant temperature used to determine the heat capture ratio. For the purposes of this paper, we are primarily using fully loaded servers and air and water temperatures that are the same.

Measuring Heat Capture Ratio

Heat capture ratio will naturally influence a datacenter's power usage effectiveness, or PUE. Datacenter PUE is usually defined as follows:

Equation 1 Power usage effectiveness

$$PUE = \frac{Total\ Facility\ Energy}{IT\ Equipment\ Energy}$$

Like PUE, heat capture ratio is expressed as a ratio of energy, but heat capture ratio strictly describes the rejection of thermal energy from the servers, while PUE encompasses other consumers of energy, such as lighting.

Heat capture ratio not only varies as a function of the design and operation of the system; it can also vary based on how the ratio is measured. Heat capture ratio can be measured at the single server level – as in the case study above – by measuring the power into the water and dividing by the DC power consumption, as in Equation 2:

Equation 2 Heat capture ratio (DC power)

$$Heat\ Capture\ Ratio_{Server} = \frac{heat\ rate\ into\ coolant}{DC\ power\ into\ server} = \frac{\dot{m} * c_p * \Delta T}{I * V}$$

Where $\dot{m}$ is the mass flow rate of the coolant, c_p is the coolant specific heat, ΔT is the coolant temperature change, and I and V are the DC current and voltage input to the server respectively.

This power measurement encompasses *all* of the DC components within the server, not just those that are directly liquid cooled. However, a cold plate is not necessarily required for the liquid cooling system to take power away from a component. By means of conduction – either through the motherboard, chassis, or air – heat can be transferred from a component that is not actively liquid cooled into the cold plate of a component that is. From there, it's a quick journey into the coolant. Hence, there is potential for greater than 100% heat capture ratio if heat is absorbed not only from other components but also from the ambient air.

The formulation of heat capture ratio in Equation 2 is useful for single server testing and datacenter predictions. It neglects the inefficiency of the power supplies and other non-liquid cooled assets. Single server testing is used to optimize the server performance, but what matters is the entire system. For the entire system we can measure the total flow rate and temperature change across the entire liquid coolant loop and calculate the heat transfer. Chilldyne CDUs report these measurements automatically.

Equation 3 Heat capture ratio (AC power)

$$Heat\ Capture\ Ratio_{Datacenter} = \frac{\sum CDU\ heat\ rates}{AC\ power\ consumption\ of\ compute\ racks * PSU\ efficiency}$$

This formulation of the heat capture ratio includes consumption from other assets in the compute racks, such as power supplies, switches and servers which are not liquid cooled. If the power supplies are efficient and the non-compute assets are lightly loaded, then the single server performance gives a good estimate of overall heat capture ratio.

Single Server Testing

Chilldyne conducted a benchmark test on a single server equivalent to those used in the Attaway installation. The test established baselines for performance under air cooling and under liquid cooling with the stock fan speed profile. Then, the fan speed was manually reduced. At the lowest fan speed, the coolant supply temperature was reduced.

Attaway Installation

To perform the test, Chilldyne outfitted a Penguin Computing Tundra Relion XO1132e server with Cool-Flo hybrid liquid-air heat sinks (see Table 3). The server was connected to a coolant distribution unit and outfitted with sensors to measure total DC power consumption, flow, and coolant temperature in and out, as well as component temperatures. A stress test program was run for each test, and the heat capture ratio was measured in each of five different combinations of fan speed and liquid coolant temperatures.

Table 3 Attaway server configuration

Type	Make	Model	Notes
Chassis	Penguin	Tundra Relion XO1132e	
Motherboard	Intel	S2600BP	Buchannan Pass
CPU 0	Intel	Xeon Gold 6140	18 core, 36 thread, 140 W (TDP)
CPU 1	Intel	Xeon Gold 6140	18 core, 36 thread, 140 W (TDP)
RAM	Micron	MTA9ASF1G72PZ-2G9E1UG	12x 8 GB, 96 GB Total
NIC	Intel	Omni-Path HFI Silicon 100 Series	100HFA016LS
OS	CentOS	CentOS Linux 7.6.1810	
Benchmark	Intel	Intel Distribution for LINPACK* Benchmark (shared-memory version)	Package ID: `l_mklb_p_2018.3.011`

The results demonstrate that there is significant variation in heat capture ratio with fan speed and with coolant supply temperature. Using the motherboard's default fan control profile for air cooling and supply coolant temperatures equivalent to the ambient air temperature (25 °C), the Tundra Relion server captured 68% of heat generated at 8800 RPM. With fan speeds reduced to 6000 RPM (32% slower), the heat capture ratio increased to 72%. With fan speeds reduced to 3000 RPM, the heat capture ratio increased again to almost 77%. By dropping the supply coolant temperature to 10 °C below ambient, an additional 3 percentage points of heat capture (for a total of 80%) could be achieved. Results are tabulated in Table 4.

Table 4 Single server heat capture test results for Attaway

Measurement	**Test 1**	***Test 2 (as installed)***	**Test 3**	**Test 4**	**Test 5**
Fan Speed (RPM)	**20500**	***8800***	**6000**	**3000**	**3000**
Ambient (°C)	25	*25*	25	25	25
Coolant (°C)	25	*25*	25	25	15
Core (°C)	78.5	*46.5*	46	46.5	41.5
Heat Sink (°C)	70	*31*	31.3	33.6	28.5
DC Power (W)	415	*365*	360	358	362
Flow Rate (GPH)	n/a	*7.5*	7.5	7.5	7.6
In/Out Delta (°C)	n/a	*7.5*	7.85	8.33	8.66
Heat Capture (W)	n/a	*247.5*	259.0	274.9	289.6
Heat Capture (%)	**n/a**	***67.8%***	**72.0%**	**76.8%**	**80.0%**

As the fan speed is reduced, the air flow through the server is reduced, so less heat flows into the air. As the coolant supply temperature is reduced, the temperature difference between the coolant and heat exchangers increases, causing more heat to flow into the coolant and therefore an increase in heat capture. If the CPU is cooled below the air temperature, the liquid cooling extracts heat from the motherboard and the VRMs, DIMMs etc. thereby increasing the heat capture ratio.

The heat capture ratio as a function of fan speed is plotted in Figure 3. Even at the lowest fan speed, the air-cooled components were within their recommended operating temperatures. The result of reducing coolant supply temperatures demonstrates that lower heat sink temperatures improve the heat capture ratio. Therefore, reduced heat sink thermal resistance – which reduces heat sink temperatures – would also improve the heat capture ratio.

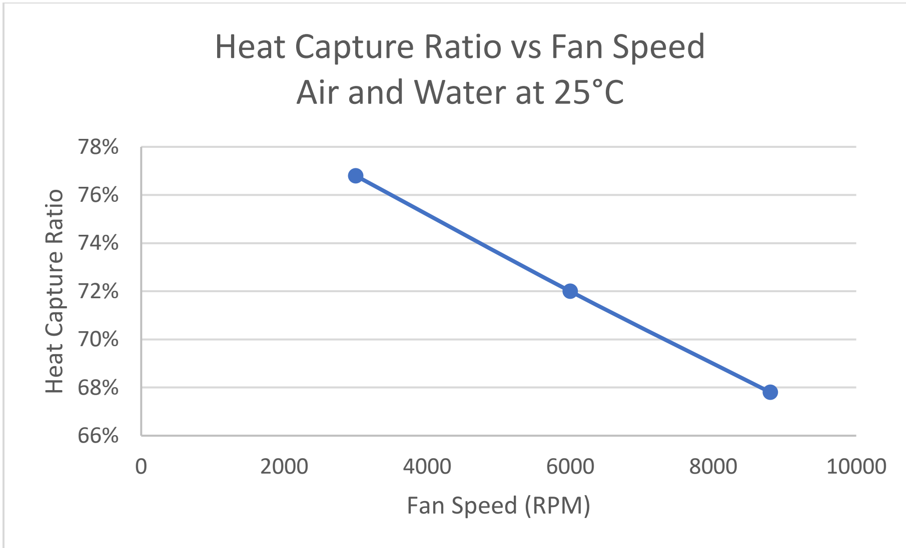

Figure 3 Single server heat capture as a function of fan speed

Manzano Installation

The same cooling system with new higher flow manifolds was used for the Manzano system. Based on the agreement between single-server testing and cluster performance, we predicted that a cluster with 205-watt processors will capture up to 82% of the server heat into the liquid cooling system. As CPU power increases, so will the heat capture ratio. The Linpack Performance was the same with liquid cooling as when cooled by air only. Table 5 lists the results of the single server testing in Manzano.

Table 5: Heat capture test results for Manzano

Measurement	Baseline – Air Only	*Liquid Cooled*
Fan Speed (RPM)	21800 RPM	*5400 RPM*
Ambient (°C)	22.5 °C	*22.5 °C*
Coolant (°C)	n/a	*22.8*
Core (°C)	84	*56*
Heat Sink (°C)	n/a	*39.5*
DC Power (W)	521	*496*
Flow Rate (GPH)	n/a	*7.6*
In/Out Delta (°C)	n/a	*12.2*
Heat Capture (W)	n/a	*407*
Heat Capture (%)	**n/a**	***82.2%***

Heat Capture Ratio and Warm-Water Cooling Systems

If reducing the coolant temperature increases heat capture ratio, why not use chilled water? Chilled water for cooling is inefficient and expensive. Both the Attaway and Manzano systems take advantage of warm-water liquid cooling. Ultimately, all of the heat rejected by a datacenter has to go outside. By running a datacenter with warm coolant, it is easier to reject that heat to the outdoors, especially when the outdoor temperature is high. The datacenters at Sandia National Laboratories use thermosyphons to provide the cooling water and the installation is designed around the temperature that they provide.

Heat Capture Ratios at Attaway and Manzano

The coolant flow rate in the Attaway supercomputer at Sandia National Laboratories is between 195 and 205 LPM per CDU when three CDUs are running, and between 275 and 285 LPM per CDU when two CDUs are running. Ambient air and cooling water supply temperatures were measured at 22 °C (72 °F). The feedback-controlled facility water return temperature from the CDUs was set to 30 °C (85 °F). (1)

Collectively, the CDUs measured 360 kW of heat removal under a typical Linpack test, across 1488 nodes each with two 140 W CPUs. The DC bus bars drew 542 kW of AC power. Using Equation 3:

$$Heat\ Capture\ Ratio_{Attaway} = \frac{129.0\ kW + 125.8\ kW + 105.3\ kW}{542\ kW} = \frac{360.1\ kW}{542\ kW} = 66\%$$

The Manzano CDUs measured 532 kW of heat removed, while using 684 kW of AC power. The 1488 compute nodes were each outfitted with two 205 W CPUs.

$$Heat\ Capture\ Ratio_{Manzano} = \frac{176.0\ kW + 180.6\ kW + 175.2\ kW}{684\ kW} = \frac{531.8}{684\ kW} = 78\%$$

Comparing Heat Capture Ratios

The heat capture ratio calculated using total datacenter heat rates and power was 2-4% lower than the single server testing predictions. The heat capture ratio at the data center level was measured while running typical HPC loads, while the single server testing was performed with Linpack, so there may be different loads on the servers. Also, the data center scale results include the power draw from the air-cooled power supplies, network switches and the management racks. A prediction of the data center scale heat load can be made more accurately if the power supply efficiency and maintenance rack power dissipation is known.

Conclusions

Heat capture ratios measured at the server level can be used to determine the expected liquid and air-cooling loads on the data center. The heat capture ratio, like automotive fuel economy, depends on the way the servers are configured and used. The liquid cooling system will generally be more efficient than the air-cooling system, so the best data center efficiency can be achieved by:

- Using cold plates or hybrid liquid-air heat sinks with low thermal resistance. This will result in low CPU/GPU temperatures so that the processor won't heat up the motherboard.
- Using the coldest coolant that can be achieved without using mechanical chillers, while remaining above the dewpoint. Colder temperatures increase the heat capture ratio *and* reduce the CPU leakage current.

- Slowing down the fans by adjusting the BIOS, while maintaining the temperature of non-liquid cooled components below specified limits. Slower fans reduce the heat flow into the air *and* use less power.

As the power draw of newly designed processors continues to increase, as motherboard makers adjust their BIOS for liquid cooling, and as power supplies become more efficient, we expect the heat capture ratio to increase. Getting the last few percentage points of heat into the liquid may not be cost effective, so the system design must be balanced between cost, efficiency and reliability. The best thermal resistance is achieved with direct-to-chip water cooling, other approaches may not be able to cool future high-power chip designs.

References

1. **Martinez, Dave M. Smith and David J.** *Cooling Performance Testing of Attaway's Negative Pressure CDU.* Albuqurque : Sandia National Labs, 2020. SAND2020-6888R.