ELECTRIC POWER
RESEARCH INSTITUTE

Evaluation of Liquid Cooling Technologies for Servers

Energy Efficiency Opportunities in Small Data Centers

2020 TECHNICAL REPORT

Evaluation of Liquid Cooling Technologies for Servers

Energy Efficiency Opportunities in Small Data Centers

3002019755

Final Report, September 2020

EPRI Project Manager
A. Amarnath

ELECTRIC POWER RESEARCH INSTITUTE
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ PO Box 10412, Palo Alto, California 94303-0813 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

THIS DOCUMENT WAS PREPARED BY THE ORGANIZATION(S) NAMED BELOW AS AN ACCOUNT OF WORK SPONSORED OR COSPONSORED BY THE ELECTRIC POWER RESEARCH INSTITUTE, INC. (EPRI). NEITHER EPRI, ANY MEMBER OF EPRI, ANY COSPONSOR, THE ORGANIZATION(S) BELOW, NOR ANY PERSON ACTING ON BEHALF OF ANY OF THEM:

(A) MAKES ANY WARRANTY OR REPRESENTATION WHATSOEVER, EXPRESS OR IMPLIED, (I) WITH RESPECT TO THE USE OF ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT, INCLUDING MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, OR (II) THAT SUCH USE DOES NOT INFRINGE ON OR INTERFERE WITH PRIVATELY OWNED RIGHTS, INCLUDING ANY PARTY'S INTELLECTUAL PROPERTY, OR (III) THAT THIS DOCUMENT IS SUITABLE TO ANY PARTICULAR USER'S CIRCUMSTANCE; OR

(B) ASSUMES RESPONSIBILITY FOR ANY DAMAGES OR OTHER LIABILITY WHATSOEVER (INCLUDING ANY CONSEQUENTIAL DAMAGES, EVEN IF EPRI OR ANY EPRI REPRESENTATIVE HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES) RESULTING FROM YOUR SELECTION OR USE OF THIS DOCUMENT OR ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT.

REFERENCE HEREIN TO ANY SPECIFIC COMMERCIAL PRODUCT, PROCESS, OR SERVICE BY ITS TRADE NAME, TRADEMARK, MANUFACTURER, OR OTHERWISE, DOES NOT NECESSARILY CONSTITUTE OR IMPLY ITS ENDORSEMENT, RECOMMENDATION, OR FAVORING BY EPRI.

THE ELECTRIC POWER RESEARCH INSTITUTE (EPRI) PREPARED THIS REPORT.

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

Electric Power Research Institute, EPRI, and TOGETHER…SHAPING THE FUTURE OF ELECTRICITY are registered service marks of the Electric Power Research Institute, Inc.

Copyright © 2020 Electric Power Research Institute, Inc. All rights reserved.

ACKNOWLEDGMENTS

The Electric Power Research Institute (EPRI) prepared this report.

Principal Investigators
M. Sweeney
M. Khattar[1]

This report describes research sponsored by EPRI.

Southern California Edison's Emerging Products group funded this project. It was developed as part of Southern California Edison's Emerging Technologies Program. This technology evaluation was conducted by Micah Sweeney and Mukesh Khattar at the Electric Power Research Institute (EPRI) with guidance from Jay Madden and Paul Delaney of Southern California Edison.

The EPRI team would like to acknowledge their colleague and dear friend, Dr. Mukesh Khattar, Principal Investigator, who guided the project team with technical rigor. Dr. Khattar passed away unexpectedly in February 2020. We dedicate this project and many others in his memory and career legacy that included innovative work providing energy, environmental and societal benefits.

---

[1] Deceased February 13, 2020

This publication is a corporate document that should be cited in the literature in the following manner:

Evaluation of Liquid Cooling Technologies for Servers: Energy Efficiency Opportunities in Small Data Centers. EPRI, Palo Alto, CA: 2020. 3002019755.

ABSTRACT

Data centers are among the most energy-intensive categories of buildings and provide an opportunity for utilities that seek to incentivize energy efficiency. In large and small data centers, liquid cooling technology presents an opportunity for energy savings over conventional air-cooled designs. However, this approach has met with limited success, primarily in high-density and/or high-performance applications. Recently, several technology innovations have become available that have the potential to expand the use of liquid cooling to more mainstream applications, especially in small data centers.

This report summarizes several liquid cooling technologies that are on the market—server immersion, direct-to-chip (cold plates), and encapsulated servers—and the adoption barriers that each face. One direct-to-chip technology was identified that addresses the barrier of leakage by using negative pressure (suction) to circulate coolant through the server. This technology was evaluated in a test environment on standard server equipment and compared side-by-side with the baseline air-cooled approach. The efficiency of the liquid cooling system is assessed, as well as its impact on server power due to lower chip temperature, and an overall impact on data center energy use is calculated. The results from this testing provide insight into the efficiency of this liquid cooling technology when applied in small data centers and include recommendations for their consideration in utility efficiency programs.

Keywords

Energy efficiency
Small data centers
Liquid cooling
Direct-to-chip cooling
Power utilization effectiveness (PUE)

Deliverable Number: 3002019755

Product Type: Technical Report

Product Title: Evaluation of Liquid Cooling Technologies for Servers: Energy Efficiency Opportunities in Small Data Centers

---

PRIMARY AUDIENCE: Researchers in emerging end-use technologies, managers of utility energy efficiency programs

SECONDARY AUDIENCE: Operators of small data centers and facilities staff.

KEY RESEARCH QUESTION

The objective of this project is to evaluate potential energy savings offered by emerging liquid cooling technologies in small data centers. This report summarizes several liquid cooling technologies that are on the market, selects a technology for testing, provides the test procedure for evaluating the selected liquid cooling technology in a test environment, and presents results with detailed analysis and recommendations.

RESEARCH OVERVIEW

Liquid cooling has presented an opportunity for energy savings over conventional air-cooled designs, yet this approach has met with limited success, primarily in high-density and/or high-performance applications. Recently, several technology innovations have become available that have the potential to expand the use of liquid cooling to more mainstream applications. One liquid cooling technology was identified that addresses the primary barrier to adoption: potential liquid leaking in ICT equipment. This technology was evaluated in a laboratory setting to assess its performance in a side-by-side comparison with a conventional air-cooled design.

KEY FINDINGS

- Several liquid cooling technologies are available on the market today, including server immersion cooling, direct-to-chip (cold plates), and encapsulated servers. Though each of these offer substantial energy savings compared to traditional air-cooled technology, each of these technologies has significant barriers to adoption in mainstream applications. These are primarily related to limited compatibility with available servers and risk (real and perceived) of potential leakage.
- One emerging technology was identified that supplies liquid under negative pressure directly to cold plates attached to hardware components within server equipment. These cold plates can be retrofitted to conventional air-cooled systems, including processors, memory, etc. The negative-pressure design dramatically reduces the risk of leaks by pulling coolant under a vacuum at negative operating pressures. The technology solution is available in custom sizes, from 300 kW to as small as 15 kW cooling, making it compatible with small and embedded (mainstream) data centers.
- Results from testing with a cold plate on the CPU alone indicate that this technology removed about 45%, 30%, and 15% of the heat generated at three inlet water temperatures (67°F, 86°F, and 102°F, respectively). The amount of heat removed by liquid depends on inlet liquid temperature and flow rate, surrounding air temperature and which heat generating IT equipment components are connected with cold plates.
- Liquid cooling also reduced server power as the CPUs operated at lower junction temperatures. The overall power of the server was reduced by about 8%, 6%, and 3% with low, medium, and high facility inlet water temperatures (67°F, 86°F and 102°F), respectively.

EXECUTIVE SUMMARY

WHY THIS MATTERS AND HOW TO APPLY RESULTS

The results of this study indicate that liquid cooling technology has the potential to save energy in typical data centers and may offer significant savings in some applications. As demonstrated in this evaluation, test data can be applied to existing air-cooled data centers to estimate the energy savings of this technology as a retrofit. This allows the estimation of energy savings as a custom measure in a utility's incentive program for energy efficiency.

LEARNING AND ENGAGEMENT OPPORTUNITIES

- Readers of this report may be interested in joining EPRI's Data Center Collaborative supplemental project, which will evaluate the energy use of small data centers in greater depth and seeks to engage this market for greater effectiveness in utility incentive programs for energy efficiency. Refer to EPRI Supplemental Project Notice 3002007545, *Data Center Collaborative*, February 2016.
- Readers who seek additional technologies for saving energy in data centers should consult the following:
    - EPRI Technical Report 3002010547, *Energy Efficiency and Heat Recovery in Embedded Server Rooms and Small Data Centers Using Variable Refrigerant Flow Cooling Systems*, December 2017.
    - EPRI Technical Report 3002005665, *Pumped Refrigerant Economizers: Emerging Technology for Efficient Data Center Cooling*, December 2015.
    - EPRI Technical Report 3002003809, *Efficient Data Centers: Line-Interactive UPS and DC Power*, December 2014.
    - EPRI Technical Report 3002001360, *Emerging Technologies for Efficient Data Centers: Uninterruptible Power Supply Eco Mode, Liquid Cooling, and Evaporative Cooling*, December 2013.

EPRI CONTACTS: Ammi Amarnath, Sr Technical Executive, aamarnath@epri.com; Micah Sweeney, Project Engineer/Scientist III, msweeney@epri.com; Ron Domitrovic, Program Manager, rdomitrovic@epri.com

PROGRAM: Program 170: Customer Technologies

ACRONYMS AND ABBREVIATIONS

ASHRAE	American Society for Heating, Refrigeration and Air-Conditioning Engineers
BTU	British thermal unit
CDU	Cooling distribution unit
CPU	Central processing unit
CRAC	Computer room air conditioner
CRAH	Computer room air handler
DX	Direct expansion
HPC	High-performance computing
ICT	Information and Communications Technology
IT	Information technology
LBNL	Lawrence Berkeley National Laboratory
pPUE	Partial PUE (e.g., for cooling only)
PUE	Power usage effectiveness
VFD	Variable frequency drive

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

Cooling systems play a critical role in maintaining operations in data centers. After the servers themselves, cooling systems are typically the second largest electrical load in data centers. For this reason, cooling systems are often the target of energy efficiency improvements. Most data centers rely on air as the medium for removal of heat from servers. Yet due to the relatively low specific heat of air (0.24 Btu/lb-°F at 68°F, about 1/4th that of water) and low density of air (~13.5 Cu.Ft./lb vs. ~0.002 Cu.Ft./lb for water), a significant volume of air must be cooled and moved past the electronic circuitry to sufficiently remove heat from the powered components.

In a typical data center, the cooling system accounts for 30-40% of the total energy consumption of the data center (for an average Power Utilization Effectiveness, PUE = 1.9). The vast majority of data centers are air-cooled, although the energy intensity of the cooling system varies depending on the method of heat rejection. Large facilities may use a chiller plant to provide cooling via a chilled-water system, where smaller data centers often rely on direct-expansion (DX) air conditioning equipment. Best-in-class data centers rely on outdoor conditions to provide free cooling with an economizer, which can provide sufficient cooling year-round in some climates.

Liquid cooling has been investigated previously as a more efficient transport medium for cooling computer hardware. In the past, this technology has been limited to high-performance computing (HPC), where the density of the information technology (IT) equipment generates a high degree of local heat, which is hard to remove with air cooling alone. With the growing energy consumption of data centers globally, a number of products have been introduced that apply liquid cooling to more mainstream computing applications.

A number of businesses offer products that take advantage of the efficiency of liquid for cooling servers. The following is a summary of the products currently available, although availability may be limited in the U.S. for some of the technologies discussed. This is not an exhaustive list of available products in the market, but only a representative list of technologies that are currently available.

2

ASSESSMENT OF AVAILABLE LIQUID COOLING TECHNOLOGY

There are several liquid cooling solutions for data centers available on the market today. This study considers only the technologies that involve the use of a liquid to directly remove heat from servers. The use of chilled-water systems to cool air within the data center—for example with computer room air handling (CRAH) units or other air-to-water heat exchangers—is outside the scope of this study. This chapter provides a summary of liquid cooling products available for data centers today.

Server Immersion Cooling

One liquid cooling technology that has received media attention in recent years is server immersion cooling, which was evaluated by Pacific Gas & Electric under Emerging Technology project ET13PGE1101 [1]. This technology fully immerses the entire rack of information and communications technology (ICT) equipment in a non-conductive coolant, which is often mineral oil. This technology can be used with any server hardware. To prepare the hardware for immersion, internal fans are removed, the thermal grease used on heat sinks is replaced with an insoluble alternative, and spinning hard disk drives are sealed. The server rack is usually oriented such that heat can rise through each server via convective currents, eliminating the need for a coolant circulator through the oil immersion tank.

In addition to immersion tanks for the servers, this system requires a pump and control module that contains a liquid-to-liquid heat exchanger for removing heat from the mineral oil. Each module is rated to remove up to 100 kW of IT load and a maximum of four racks. When evaporative cooling towers or dry radiators are used for final heat rejection, immersion cooling is claimed to use up to 95% less energy than conventional DX air conditioning.

Despite the added liquid cooling infrastructure, immersion cooling has the potential to reduce installation costs compared to air-cooled designs. Without any air cooling, the data center does not require any airflow management measures, such as a raised floor, aisle containment, blanking panels, etc. This eliminates some of the infrastructure needs of a new data center build. This can make immersion cooling attractive for a new data center or capacity expansion in a building with existing cooling tower or chilled water supply.

In addition to cooling system energy savings, liquid cooling can offer a reduction in server power. First, removal of the server fans—which are attributed to server power because they are contained within the server and PSU—eliminates 5-10% of the server power. Second, server power draw is further reduced from lower operating temperature of the central processing unit (CPU), due to the dependence of processor leakage current on operating temperature. Vendor white papers cite a 35°F reduction of measured CPU temperature when immersed in coolant. In total, immersion cooling reduces server power by 10 to 20%. Combined with savings in cooling, this technology yields as much as 50% reduction in total data center energy consumption.

The primary barrier to use of this technology is its impact on manufacturers' warrantees on hardware. In most cases, immersion in mineral oil and the requisite hardware modifications void the warrantee on the IT hardware provided by the manufacturers. In some cases, this is a fatal obstruction that dissuades a data center operator from considering this technology. This is because many companies with small and medium-sized data centers enter into a service agreement with the vendor or hardware manufacturer to provide replacement equipment and support in case of failure. As of August 2014, three vendors have agreed to support equipment immersed in one manufacturer's immersion system. In addition, the vendor has partnered with a third-party warrantee and support provider to offer supplemental coverage.

Direct-to-Chip Liquid Cooling

Several vendors offer products that use direct-to-chip liquid cooling, whereby a liquid-filled heat sink is attached directly to heat-producing components such as processors (CPU and GPU) and memory. Similar to immersion cooling, this requires removal of the processor heat sinks and thermal grease. However, most of the server fans are left in place to provide cooling to the internal components besides the processors. Hence, direct-to-chip liquid cooling only captures a portion of the total server heat load and does not completely eliminate the need for air cooling.

As such, direct-to-chip liquid cooling does not provide savings as large as immersion cooling. The vendors claim energy savings greater than 50% of typical cooling energy. Testing at Lawrence Berkeley National Laboratory (LBNL) confirms that total energy savings are about 20% of the data center total [2]. LBNL evaluated operation of the system at various water temperatures, calculating the percentage of the total heat load capture by the liquid cooling system and estimating the overall savings from models of the energy use of various arrangements of a chilled-water cooling system (using a retrofitted cooling tower or dry cooler and/or the chiller to cool the servers). Results from this testing revealed that the greatest energy savings occurs at maximum IT load.

A Canadian manufacturer also offers direct contact liquid cooling technology for OEM (equipment manufacturers), enterprise, and enthusiast applications (typically for gaming). For enterprise servers, two basic configurations are offered: liquid-to-liquid and liquid-to-air. The liquid-to-liquid approach uses cold plates to remove heat from CPU, GPU and memory. Its liquid-to-liquid heat exchanger can remove up to 650 kW (at 30°C / 86°F facility water), supporting 5 to 20 racks while consuming less than 4.2 kW. It also offers smaller sizes: an 80 kW (4U) and 40 kW (1U) units. (1U refers to one industry standard Rack Unit, which is 1.75" tall and 19" wide [44.45 mm by 482.6 mm], and 4U refers to a system of equal width and 4x the height.) For data centers without facility water, 10 kW and 20 kW liquid-to-air heat exchangers, which remove heat from the IT equipment and dump it in the data center, are available to increase rack density without investing in a hydronic system.

The primary barrier to adoption of this technology is fear of liquid leakage in the IT equipment, and higher initial costs associated with providing liquid distribution and containment systems that prevent leakage.

Negative Pressure Direct-to-Chip

One vendor offers a direct-to-chip liquid cooling product that is unique in its negative-pressure system. The basic heat transfer configuration is very similar to other direct-to-chip cooling products, with coolant routed through cold-plate heat exchangers that are directly attached to the heat generating components. The system uses treated water as the coolant to the servers, with lubricants added to improve the operation of pumps and valves. The liquid-filled heat exchangers replace the conventional air-cooled heat sinks on the CPUs, and must be designed to match the dimensions and thermal design power (TDP) of the CPU heat sink being replaced. Heat sinks can be specified that minimize the amount of air cooling that occurs simultaneously. Alternatively, finned heat exchangers can be used to allow some amount of air cooling to occur, providing some level of cooling redundancy.

The primary differentiating feature of this system is its use of a negative-pressure (suction) liquid loop to the servers. With the coolant under negative pressure, the risk of coolant leaks is significantly reduced. The company likes to demonstrate this advantage by cutting the coolant lines to the server during normal operation. Due to the negative pressure design, there are no leaks from severed coolant lines. Moreover, coolant connectors can be removed without leaks while filled with liquid. This feature nearly eliminates that concern for leaks, one of the primary barriers to adoption of liquid cooling technology. This also allows the use of less expensive components for liquid distribution (tubing, fittings, etc.) which can be plastic instead of high pressure steel.

Many early adopters of liquid cooling technology have expressed frustration at the leak-prevention measures that must be taken to service other liquid-cooled servers. In addition, the initial cost of liquid distribution and quick connects for leak prevention can be quite high. The negative-pressure system boasts that it is low-risk, low-cost, compatible with a wide range of conventional air-cooled servers, and allows hassle-free server maintenance.

Negative-pressure operation is enabled by the unique design of the coolant distribution unit (CDU), which pulls coolant through the servers by creating a vacuum on the coolant reservoir. Pumping action is achieved by using a two-chamber design; when one reservoir fills with coolant, valves are changed so that the server coolant lines are connected to a second chamber that is under vacuum and emptied of coolant.

Final heat rejection is similar to the other liquid cooling systems and can be accomplished with a chilled-water system, cooling tower water, or dry cooler heat exchanger. In either case, the server-side liquid loop is isolated from facility water. Heat is exchanged with facility water at the CDU, which is offered in a 300 kW (as well as in other custom sizes) design packaged into the dimensions of a standard server rack.

Since the coolant is in direct contact with the heat-generating CPUs, the temperature of outlet coolant can exceed the temperature of the hot aisle—perhaps as high as 100°F (38°C). As these temperatures, heat can be recaptured from the liquid for reuse elsewhere in the building. For example, this temperature can be used to preheat water for water heating or space conditioning, or to provide relatively low-temperature heating (e.g. melting ice / snow from outdoor walkways in cold climates).

Encapsulated Servers

One UK-based manufacturer offers products that fully submerge server hardware in a different way. This technology is offered in custom blade servers that are sealed and filled with a liquid coolant. Currently the manufacturer works with several IT manufacturers to offer custom, liquid-cooled server hardware. Two basic configurations are available (as of 2017): a standalone portable tower, and standard blade server chassis available in 24U and 52U sizes that can support up to 27 blade servers in a rack. Yet this cooling solution is not compatible with standard server hardware. The primary barrier to adoption of this technology is the limited server hardware options available.

The coolant used is an engineered fluid that is noted for its phase-changing properties. Yet the variety of coolant used in these products has a high boiling point and remains a liquid in this system. The system offers very high density of IT hardware, with up to 48 blades per rack. The platform is rated to 20 kW of IT load, which can be cooled with as little as 107 W. In total the vendor has claimed 97% savings over baseline cooling (which consumes 4 kW to cool the 20 kW rack), plus 20% savings in IT power (presumably from the smaller server platform and reduced CPU temperature), for 50% overall savings. A case study was published from a university field site cites 80% cooling system energy savings.

Another manufacturer offers a product very similar to these encapsulated servers. Currently four models are offered: single-node blade servers; a 2U, 4-node cloud server; a 2U, dual-node for GPU applications; and a rugged chassis for harsh environments. The vendor has claimed HVAC savings of 70% or more, with overall power consumption and footprint reduced by 40-60% in early literature. Few details are offered of the system, which are evidently custom solutions per customer needs. A vendor white paper concludes that cooling system energy savings of up to 98%, for total savings of 28%. The system is shown to lower CPU operating temperature by 20° to 30°C (36° to 54°F). The primary barrier of this technology is the limited hardware options available, with only custom designs offered and no off-the-shelf products.

Custom Solutions

Another company offers a custom blade server platform that uses liquid-filled cold plates attached directly to custom blade servers. This custom hardware platform was designed for efficiency. Networking and server input/output is handled by a backplane that is shared between 16 blade servers. High-voltage (380 V) dc power is supplied to the servers, taking advantage of reduced conversion stages and lower distribution current for maximum efficiency.

A side-by-side demonstration found that this direct-touch liquid cooling technology consumed the least energy to cool the server hardware when compared to various close-coupled cooling technologies [3]. These findings support the vendor's claim of up to 95% savings in cooling energy, and up to 50% overall energy savings. However, this technology is still evolving, as few prototypes have been produced and only custom solutions are available from the manufacturer.

In 2016 another company released the first liquid-cooled rack based on the Open Compute Platform (an open-source design for a high-efficiency server developed by an industry consortium). This HPC design enables very high density, supporting up to 108 servers and 100 kW or thermal load per rack. While notable for integrating liquid cooling into an open-source hardware, this product is primarily applicable to high-performance and high-density applications.

3
TECHNICAL APPROACH

Recognizing the potential for energy savings offered by liquid cooling, a liquid cooling technology was evaluated in a laboratory environment to assess its cooling performance and energy efficiency. The negative-pressure, direct-to-chip technology was selected for evaluation, as discussed below. This chapter presents the testing procedure, including the equipment tested, instrumentation used, and test method.

Equipment Under Test

The negative-pressure, direct-to-chip technology was selected for evaluation in this study since it was identified as a suitable liquid cooling technology for small and embedded data centers. Two characteristics were identified that make this technology more suitable for smaller data centers. First, the ability to operate with conventional, air-cooled server models allows a small data center to acquire off-the-shelf equipment, which typically use lower cost, commodity servers. Second, this technology is offered in a smaller 15-kW model, designed to showcase the technology before a major investment. In addition, this technology offers a key advantage over other liquid cooling approaches through the inherent low-leak design of the negative-pressure coolant loop.

The specifications of the equipment under test are listed below:

- Cooling capacity: 15 kW of server load
- Cooling water: up to 3 gpm (12 lpm) at 41-104°F (5-40°C), with 5 psi (35 kPa) pressure difference
- Facility water input: 59-86°F (15-30°C)
- Vacuum: up to 2 cfm at 20 in Hg (30 lpm at -70 kPa) can be delivered
- CDU dimensions: 8.75” (h), 21” (w), 32” (d) (22.2 x 53.3 x 81.3 cm) mounted near servers. Weight: 52 lbs (23.6 kg)
- Vacuum pump dimensions: 8.75” (h), 21” (w), 26” (d) (22.2 x 53.3 x 66.0 cm) in rack or under floor. Weight: 45 lbs (20.4 kg)
- Operating environment: 41-104°F (5-40°C), 0-95% (non-condensing), 0-6000 ft (0-2000 m) elevation
- Onboard monitoring of water temperature, flow rate and heat transfer
- Onboard diagnostics and leak detection
- Web-based interface allows remote operation and monitoring

Figure 3-1
Coolant Distribution Unit (CDU) and Vacuum Pump of Negative-Pressure, Direct-to-Chip System

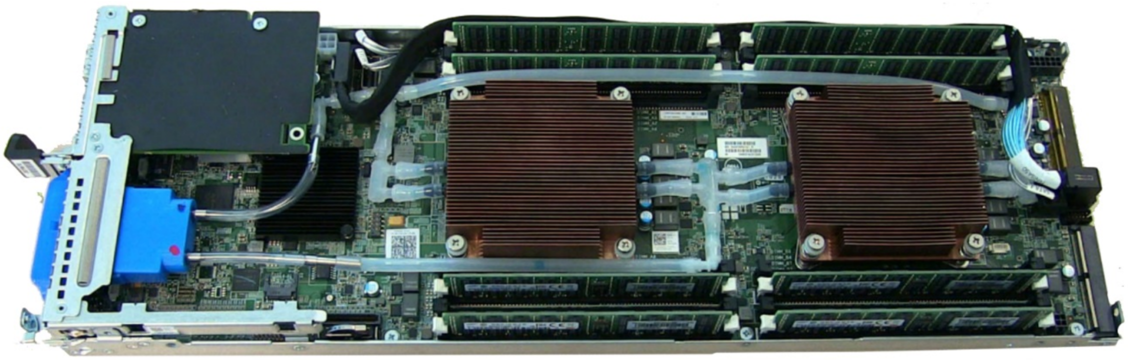

Figure 3-2
Direct-to-Chip Heat Sinks

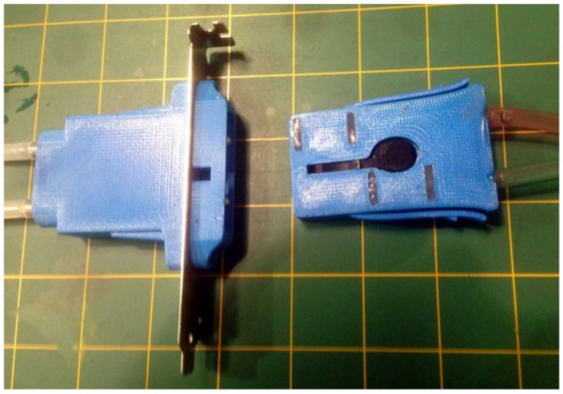

Figure 3-3
Drip-Less Coolant Connector to Server

Test Load

To provide server load on the liquid cooling system in this testing, two HP DL360 G5 servers were used for laboratory testing. Although these models were of older generation (produced in 2008) they provide a reasonable representative of baseline, air-cooled servers. Each 1U server is equipped with two Intel Xeon X5460 quad-core CPUs, that are each rated for a thermal design power (TDP) of 120 W.

- CPU: two Intel Xeon X5460 (quad-core, 3.16 GHz, 120 W each)
- Memory (RAM): 4 GB (four 1 GB modules) PC2-5300
- Storage (disk drives) – two 73 GB (10k rpm) SAS hard-disk drives
- Power supply: one 700 W baseline unit (not 80-PLUS certified)

Servers were loaded with Prime95, which uses a "Torture Test" of solving Fast Fourier Transforms (FFTs) to emulate full-load on CPU and memory. This placed a relatively constant and even load on the two servers under near maximum power. In addition, the temperature of the CPUs and speed of the fans was monitored using CoreTemp and Speedfan applications, respectively. The total server power was measured, but due to the vintage of the hardware, the server monitoring software was unable to record the power of the CPUs alone.

Instrumentation Plan

Electrical and thermal parameters of the liquid-cooled and air-cooled servers were monitored, as shown in Table 3-1.

Table 3-1
Data Collection Points

Data Point	Unit
Air-Cooled Server	
Server power	Watts
Inlet air temperature	°C
Exhaust air temperature	°C
Liquid-Cooled Server	
Server power	Watts
Inlet air temperature	°C
Exhaust air temperature	°C
Coolant flow rate	L/min
Liquid inlet temperature	°C
Liquid outlet temperature	°C

Figure 3-4 shows a diagram of the data collection and equipment under test.

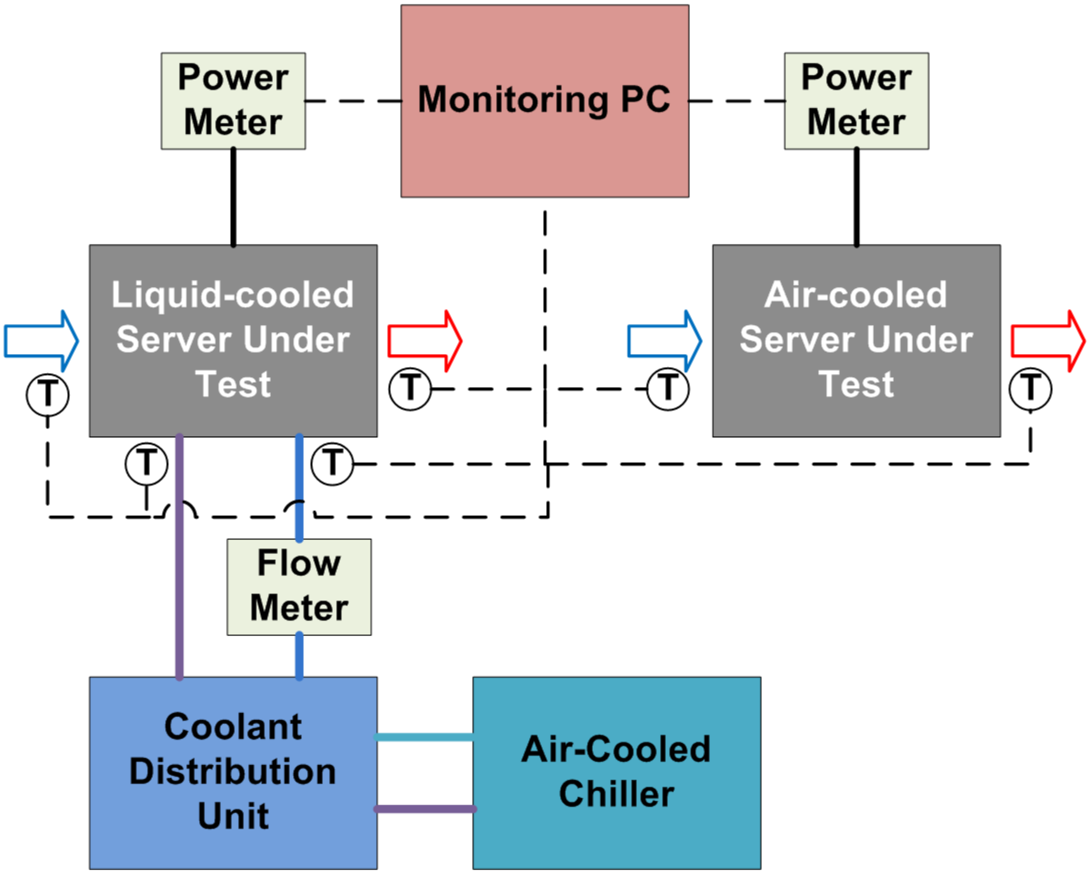

Figure 3-4
Data Collection Diagram

Data Monitoring Equipment

The equipment used for metering and data collection at the field site is listed below:

- Power meter: Elkor WattsOn (revenue grade)
- Current transformers (CTs): Continental Controls Accu-CT (15 A, revenue grade)
- Air temperature: Type-T thermocouple
- Liquid temperature: calibrated thermistors
- Liquid flow rate: calibrated orifice and delta pressure transducer

Table 3-2 lists the accuracy of the sensors used in this study.

Table 3-2
Accuracy of Sensors Used

Instrument	Accuracy
Elkor WattsOn	<0.2% @ 25°C
Accu-CT	±0.75%
Type-T Thermocouple	±1.0°C@ -200° to 400°C

Figure 3-5 shows the test setup with each piece of equipment identified.

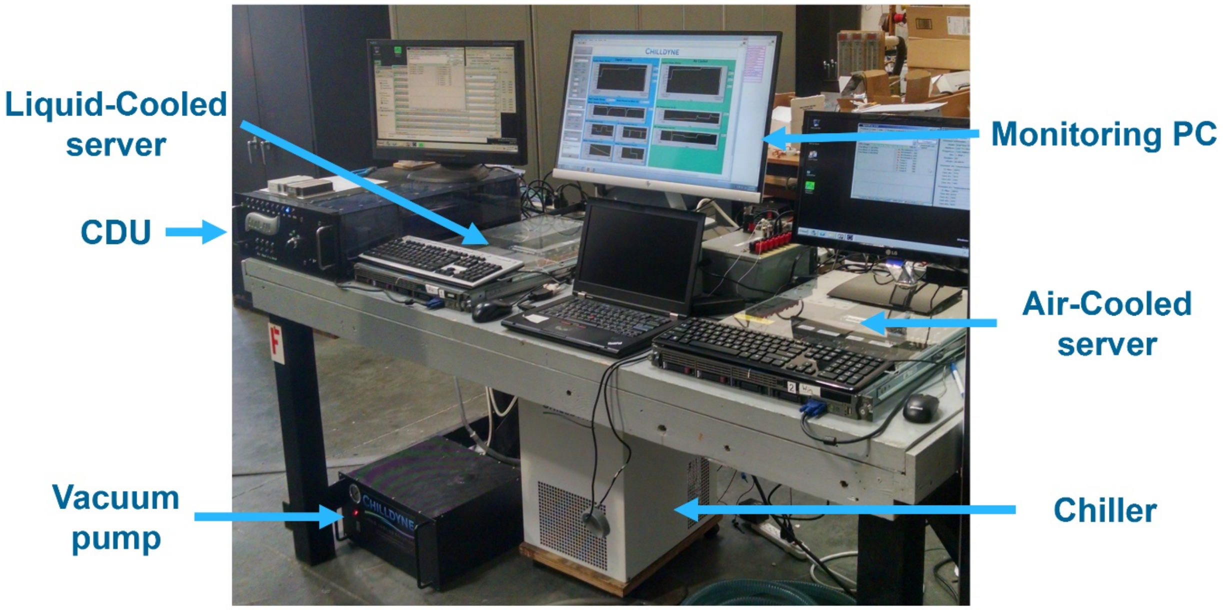

Figure 3-5
Test Setup

Figure 3-6 shows the location of the sensors measuring the temperature of inlet and exhaust air on each server.

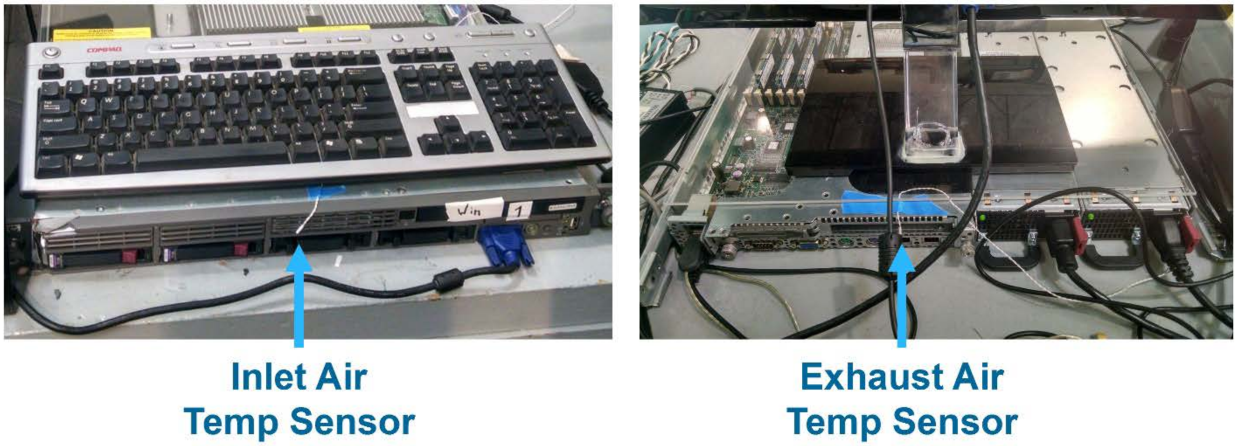

Figure 3-6
Air Temperature Sensor Locations

Test Method

Testing was conducted in a laboratory environment to evaluate the ability of the liquid cooling system to remove heat from a server. Side-by-side testing was conducted using one liquid-cooled and one air-cooled server of identical models, using Prime95 to place nearly maximum load on the server CPU and memory. As noted above, electrical and thermal parameters of both servers were monitored during the testing.

Performance of the liquid cooling system was evaluated under a range of coolant flows and facility water temperatures. A small chiller was used to control the temperature of facility water, and coolant flow to the servers was controlled by a small throttle valve in the coolant loop. Table 3-3 shows the operating conditions that the system was operated at under this testing. Facility water temperatures were used to represent a range of operation, from an efficiently operating chiller (Low), the maximum typical operation of a cooling tower (Medium), to a potential heat recovery operation (High). Similarly, coolant flow to the server was adjusted from its lowest stable limit to its highest flow rate, with a mid-point selected for comparison. Such flow rates roughly correspond to temperature rise across the server of nominal 10°, 20°, and 30°F (5.6°, 11.1°, 16.7°C).

Table 3-3
Matrix of Test Conditions

Control Parameter	Low	Medium	High
Nominal Facility Water Temperature	64.4°F (18.0°C)	85°F (29.4°C)	104°F (40.0°C)
Nominal Coolant Flow Rate	0.8 gph (0.05 lpm)	3.2 gph (0.2 lpm)	6.3 gph (0.38 lpm)

The test procedure was as follows: first the servers, liquid cooling and data monitoring systems were powered. The chiller was set to deliver a specific facility water temperature and allowed roughly 30 minutes to reach equilibrium. During this period, the load was added to the servers, and coolant flow to the servers was increased to maximum flow rate. Once thermal equilibrium was reached (at least 5 minutes' operation at a fixed coolant flow rate), data was collected for about 5 minutes of continuous operation. After data collection was completed, coolant flow was decreased to the next measurement point. After all three points were collected, chilled water temperature was increased and the chiller was operated for at least 30 minutes to reach thermal equilibrium.

4

RESULTS AND DISCUSSION

Table 4-1 shows the results from testing liquid cooling as described in the prior section, showing parameters that were measured—liquid flow rate, temperature, power—and those that were calculated—the amount of heat capture by coolant and reduction in server power. Server power reduction is calculated for each of the test points in comparison with the air-cooled server, which was measured to use 413W on average during testing (with the same load). Note that liquid flow rate was manually adjusted to represent minimum (about 0.5 gph / 0.03 lpm), maximum (about 6.0 gph / 0.4 lpm), and a mid-point (about 3.0 gph / 0.2 lpm). Coolant inlet temperature was determined by the chilled water delivered to the CDU at 68°, 86°, and 104°F (18°, 29.4°, and 40°C).

Table 4-1
Results from Liquid Cooling Testing

Liquid Flow Rate (gph)	Liquid Heat Transfer (Watts)	Heat Capture (%)	Server Power (Watts)	Server Power Savings (%)	Coolant Inlet Temp (°F)	Coolant Exit Temp (°F)	Coolant Delta-T (°F)
5.7	175.1	46.8%	373.9	9.6%	67.4	80.0	12.7
3.0	153.1	40.4%	379.2	8.3%	67.7	88.7	20.9
0.4	29.6	7.2%	409.3	1.0%	73.3	101.9	28.6
6.0	130.0	33.9%	384.0	7.1%	86.2	95.1	8.9
3.4	123.4	31.7%	389.0	5.9%	86.0	100.9	14.9
0.5	21.9	5.4%	406.3	1.7%	85.9	105.2	19.3
6.0	74.4	18.7%	397.5	3.8%	103.6	108.7	5.1
3.0	60.2	15.0%	401.0	3.0%	103.1	111.2	8.1
1.4	30.9	7.6%	405.5	1.9%	102.0	111.2	9.2

Figure 4-1 shows the amount of heat removed by the liquid cooling system as a function of flow rate. Each point on the plot represents data collection under steady-state conditions after the system had reached thermal equilibrium, and the overlaid color represents the temperature of coolant into the server. It was discovered that maintaining consistent coolant flow at low levels was a significant challenge, due to the periodic dip in coolant flow as the CDU transitioned from its primary fluid reservoir to its secondary reservoir.

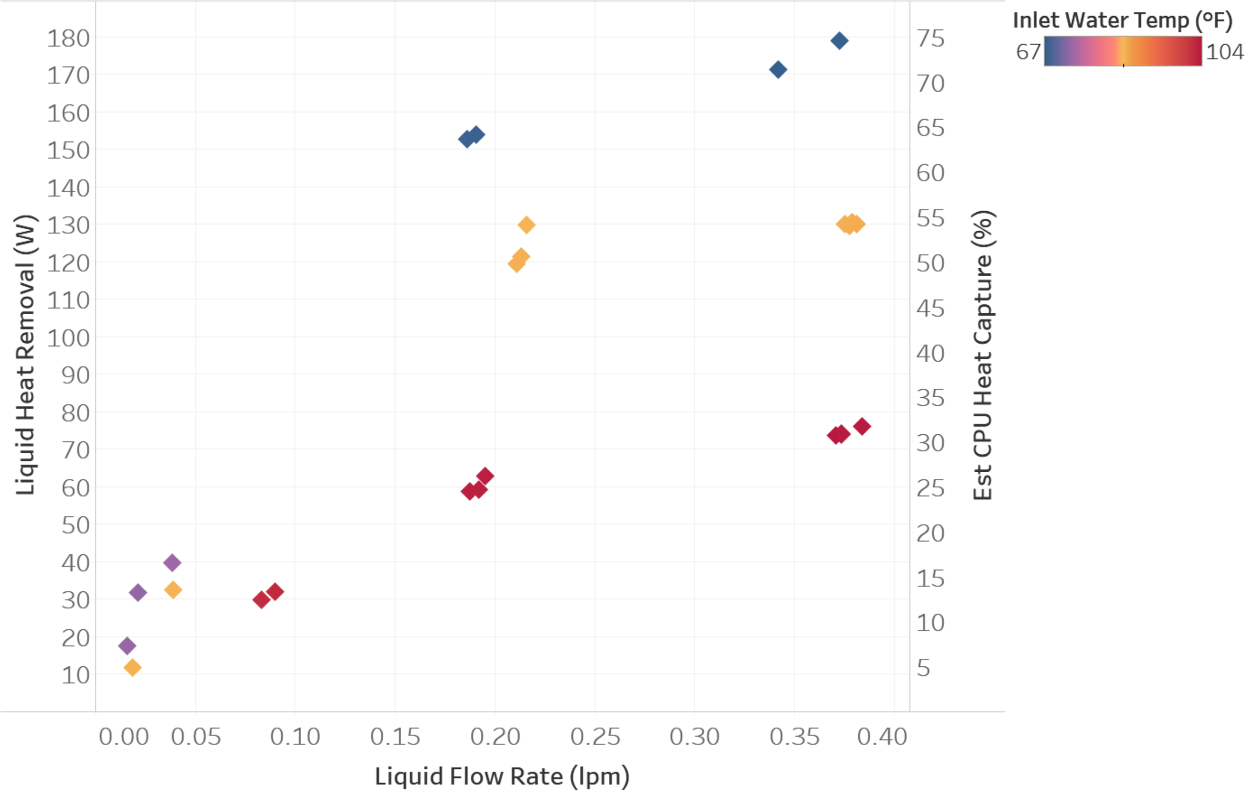

Figure 4-1
Heat Removal V Liquid Flow Rate

Figure 4-1 shows a positive relationship between flow rate and heat removal. Moreover, the data suggest a dependence on inlet water temperature. For flow rates below 1.6 gph (0.1 lpm), heat removal is shown to fall off dramatically. For this reason, it is not recommended to operate the system below 1.6 gph, and performance at these low flow levels will be ignored in further analysis.

Figure 4-2 shows the measured power consumption of the liquid-cooled server as a function of the temperature of coolant into the server for coolant flow above 1.6 gph. Data points indicate operation at steady state, colored by the average temperature of the coolant leaving the server. The average power consumption of the air-cooled server (413.4 W) is plotted as a constant reference line to show the energy savings from the server alone.

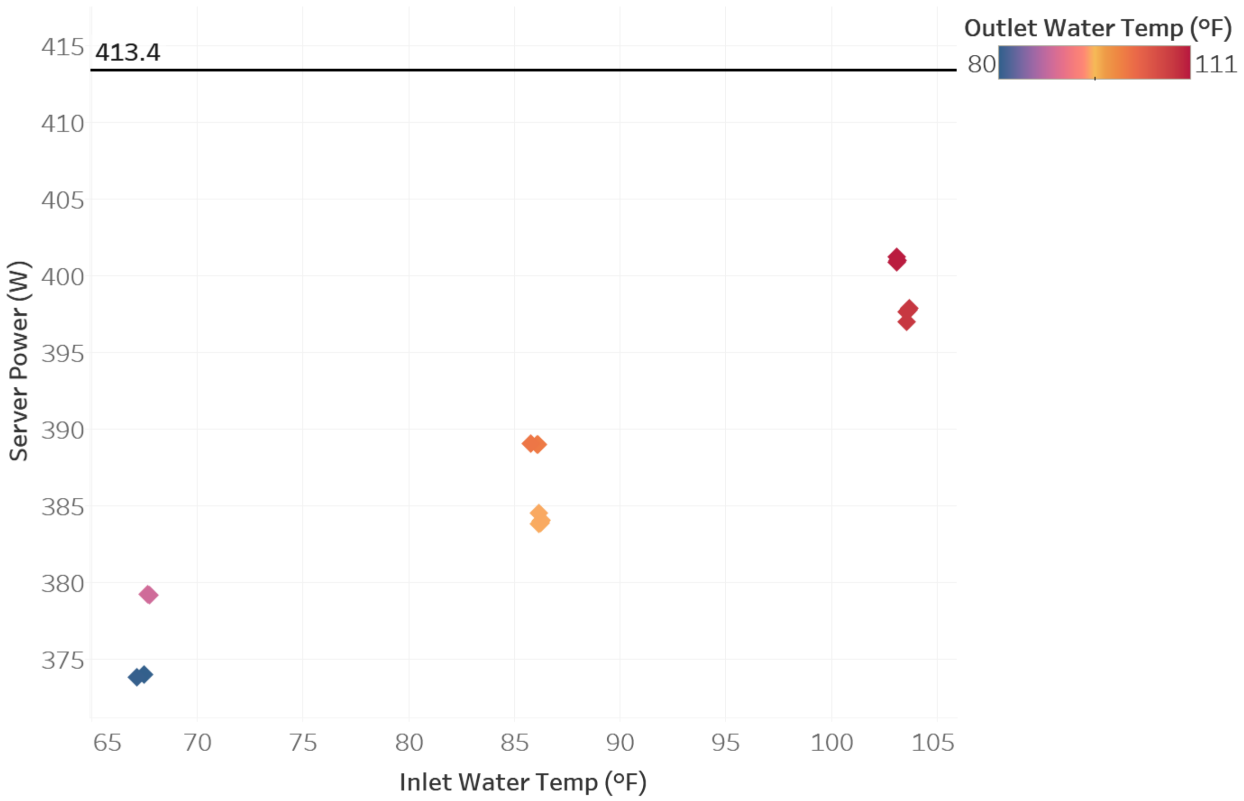

Figure 4-2
Server Power V Inlet Water Temperature

Figure 4-2 shows a linear dependence between inlet coolant temperature and server power. However, a secondary effect from outlet coolant temperature can be seen. To analyze this in greater detail, Figure 4-3 plots the average power of the liquid-cooled server against the temperature of coolant leaving the server. Again, this data does not include operation at low flow rates, and the average power of the air-cooled server is plotted for reference.

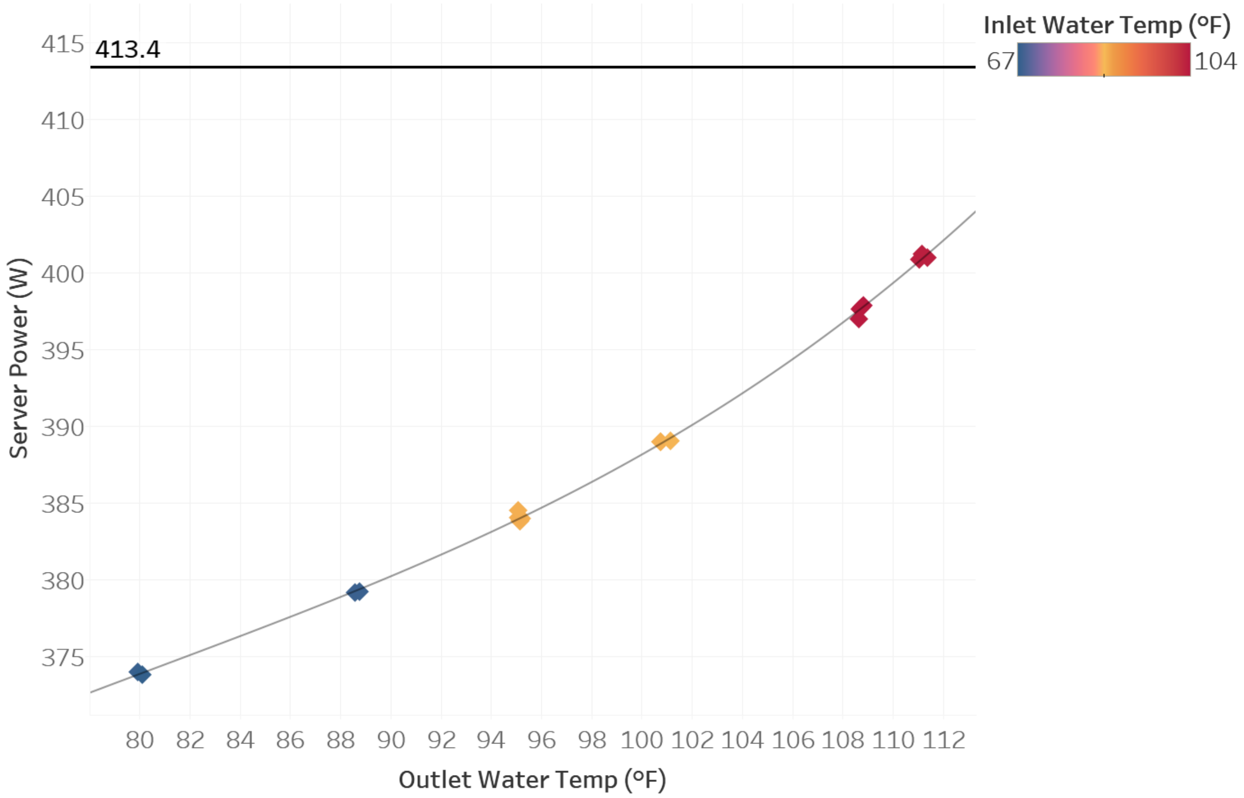

Figure 4-3
Server Power V. Outlet Water Temperature

Figure 4-3 shows a more direct relationship between outlet coolant temperature and server power. A best fitting trend line is plotted that describes the relationship as a cubic function ($ax^3 + bx^2 + cx + d$) with a p-value <0.0001 (indicating very strong statistical significance). This relationship illustrates the impact that CPU temperature has on CPU power, with outlet temperature providing a better understanding of average CPU temperature than inlet temperature.

These data demonstrate the energy savings that liquid cooling can provide from simply operating the CPUs at lower temperature than air-cooled operation. Moreover, when the CPUs are directly cooled by liquid, server fan speed does not increase under the heavy load used in this testing, as opposed to what was observed with the air-cooled server. The lowest operating power of the liquid-cooled system was measured to be 373.8 W, a savings of 39.6 W (9.6%) over the baseline air-cooled server with identical model and workload.

For comparison, Figure 4-4 shows the rise in coolant temperature across the server against liquid flow rate for all test points, including low-flow testing. The temperature of inlet water is overlaid for each point. Note that temperature rise increases for low inlet temperature. This suggests that operating at lower inlet temperatures increases the total amount of heat removed by the coolant, which was shown in Figure 4-1, except for low-flow operation.

5

CONCLUSION

In summary, liquid cooling technologies promise substantial energy savings over conventional data center cooling systems. The cooling only partial-PUE of heat rejection using liquid cooling technologies is typically less than 1.05. Moreover, improvements in cooling efficacy allow components to operate at lower temperature, reducing losses. In this study liquid cooling technology was found to reduce overall data center energy consumption by 14%.

Liquid cooling also offers the ability to expand IT capacity in existing data centers of all sizes. By reducing ICT power use, liquid cooling can release power capacity for additional ICT equipment. This attribute of liquid cooling can reduce power upgrade requirements as well as associated costs, especially in locations that are constrained in power supply. In many cases it is possible and practical to accommodate liquid cooling in existing air-cooled data centers. Liquid-cooled racks can be placed alongside conventional air-cooled equipment if the technology used has a low risk of potential leaks, such as the negative-pressure system evaluated in this study.

Recommendations

The results of this study indicate that liquid cooling technology has the potential to save energy in typical data centers, and may offer significant savings in some applications. As demonstrated in this evaluation, test data can be applied to existing air-cooled data centers to estimate the energy savings of this technology as a retrofit. This allows the estimation of energy savings as a custom measure in a utility's incentive program for energy efficiency. The Green Grid recently released a spreadsheet-based tool for estimating the total cost of ownership of liquid-cooled data centers [5], which may provide a standard tool that utility program administrators can use to validate compliance.

While this testing evaluated the savings potential of this technology in a controlled environment, additional testing is needed with a scaled-up installation to evaluate the performance and efficiency in a production-scale environment. Such testing will corroborate the heat capture performance of this technology in a data center environment, both the impact on existing air-cooled equipment and server power use under more realistic workloads.

In addition, further testing could explore the opportunity for performance gains through use of liquid cooling. Vendors of this technology cite the ability to "overclock" server hardware when using liquid cooling, much like gaming PCs. As found in this study, direct-touch liquid cooling allows the CPU to operate at lower temperature than air cooling. This suggests that greater performance could be attained from a liquid-cooled CPU if its operational parameters were adjusted—primarily CPU frequency. Additional testing of this operation could validate the concept of enhanced efficiency through a potential gain in performance.

One of the primary barriers to adoption of liquid cooling technology is data center operators' familiarity and comfort with the technology. Many perceive liquid cooling as risky, since it has the potential to cause significant damage due to leaks. Moreover, some consider the up-front cost and effort to adopt a liquid cooling system to outweigh the benefits. Since both of these concerns—while valid—relate to user perceptions rather than technical challenges, they suggest that an educational campaign may directly address some of these concerns and promote adoption of this technology.

6
REFERENCES

[1] Pacific Gas and Electric Company (PG&E), "Submersion Cooling Evaluation," ET Project Number: ET13PGE1101, San Francisco, CA, 2014.

[2] H. Coles and S. Greenberg, "Direct Liquid Cooling for Electronic Equipment," Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, 2014.

[3] H. Coles, "Demonstration of Rack-Mounted Computer Equipment Cooling Solutions," Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, 2014.

[4] Integral Group, "Energy Efficiency Baselines for Data Centers: Statewide Customized New Construction and Customized Retrofit Incentive Programs," Commissioned by the California Investor-Owned Utilities, Oakland, CA, 2013.

[5] The Green Grid, "Liquid Cooling Total Cost of Ownership Calculations Tool," 2018. [Online]. Available: https://www.thegreengrid.org/en/resources/library-and-tools/468-Liquid-Cooling-Total-Cost-of-Ownership-Calculation-Tool.

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI members represent 90% of the electricity generated and delivered in the United States with international participation extending to nearly 40 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; Dallas, Texas; Lenox, Mass.; and Washington, D.C.

Together...Shaping the Future of Electricity

Program:

Customer Technologies

© 2020 Electric Power Research Institute (EPRI), Inc. All rights reserved. Electric Power Research Institute, EPRI, and TOGETHER...SHAPING THE FUTURE OF ELECTRICITY are registered service marks of the Electric Power Research Institute, Inc.

3002019755

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com