

A Methodology for Thermal Characterization of High-Power Liquid-Cooled Servers

Pardeep Shahi

NVIDIA Corporation

2788 San Tomas Expressway, Santa Clara, California.

pshahi@nvidia.com

Ali Heydari

NVIDIA Corporation

Satyam Saini

Mechanical and Aerospace Engineering Department,

The University of Texas at Arlington,

Pratik Bansode

Mechanical and Aerospace Engineering Department,

The University of Texas at Arlington,

Uschas Chowdhury

NVIDIA Corporation

Harold Miyamura

NVIDIA Corporation

Ashwin Siddharth

Mechanical and Aerospace Engineering Department,

The University of Texas at Arlington,

Amirreza Niazmand

Mechanical and Aerospace Engineering Department,
The University of Texas at Arlington,

Himanshu Modi

Mechanical and Aerospace Engineering Department,
The University of Texas at Arlington

Dereje Agonafer

Mechanical and Aerospace Engineering Department,
The University of Texas at Arlington,

Mohammad Tradat

NVIDIA Corporation

Jeremy Rodriguez

NVIDIA Corporation

Keywords: Electronics Cooling, Liquid Cooling, Data Center, High Performance Computing

Abstract

Effective cooling is crucial for high power liquid cooled servers to ensure optimal performance and reliability of components. Thermal characterization is necessary to ensure that the cooling system functions as intended, is energy-efficient, and minimizes downtime. In this study, a proposed methodology for thermal characterization of a high-power liquid-cooled server is presented. The server layout includes multiple thermal test vehicle (TTV) setups equipped with direct-to-chip cold plates, with two or more connected in series to form a TTV cooling loop. These cooling loops are connected in parallel to the supply and return plenums of the server manifold, which includes a chassis-level flow distribution manifold. To obtain accurate measurements, two identical server prototypes are instrumented with sensors for coolant flow rate and temperature measurements for every TTV cooling loop. Four ultrasonic flow sensors are installed in the flow verification server to measure the coolant flow rate to each TTV cooling loop. In the thermal verification server, thermistors are installed at the outlet of each TTV cooling loop to log temperature measurements. The amount of heat captured by the coolant in each TTV cooling loop is subsequently estimated based on the flow rates determined from the flow verification server. This methodology enables precise characterization of the thermal

performance of high-power liquid-cooled servers, ensuring optimal functionality, energy efficiency, and minimized downtime.

1. Introduction

Efficient data center cooling is one of the most important issues that need addressing, especially due to continuously increasing power densities at rack-level. The rising processing demands due to cryptocurrency mining and machine learning applications have ushered in a new era of high-performance computing servers that have further exacerbated the challenge of efficient ITE (Information Technology Equipment) thermal management. The increasing performance demands are being met by packing more and smaller transistors on the chips and chips 3-D stacking to achieve increased packaging integration densities (Chainer, et al., 2017). This series of multiscale subsystems involved in heat transfer starting from the transistors and interconnects at the chip level to server and rack level heat rejection mediums also add to the complexity of the existing heat transfer challenges. The statistics show that nearly 4.8 billion people were accessing the internet that represents approximately 61% of the total world population (Internet Live Stats, 2022). This involves millions of hours of email exchange and online browsing which solely relies on streamlined information flow through the data centers. The power densities have exceeded 10 kW and are predicted to cross 30 kW and beyond in future (Hasan, et al., 2013). This further emphasizes the need for efficient data center thermal management and data center uptime.

Historically, data centers have relied on traditional compressor-based cooling using the forced air convection to remove heat fluxes from the ITE. These data centers can expend up to 30% of the total energy consumed towards equipment cooling only (Alliance to Save Energy, 2007). With improvements in airflow paths and data center infrastructure, the cooling efficiencies are being removed but this adds to both data center CapEx and OpEx figures. ASHRAE (American Society of Heating, Refrigeration and Air-Conditioning Engineers) expanded its cooling envelope guidelines to accommodate ambient air-cooling techniques to reduce the cooling energy costs (ASHRAE, 2011). This limits the use of compressors and chillers to cool down the air for a considerable time of the year, especially in favorable climates. But this, in turn, may lead to higher ambient fan power consumption because of elevated processor junction temperatures (David, et al., 2012) (Parida, et al. 2016). This is because the fan power consumption is directly proportional to both the ambient temperatures and processor junction temperatures. Rear door heat exchangers are also deployed in data centers for low as well as for high density rack applications where the heat from the rear of the racks is rejected into the facility chilled water (Shalom, et al., 2022) (Schmidt and Iyengar, 2010). (Fakhim et al., 2011) showed the inefficiencies in air-cooled data centers such as hot air recirculation and air leakages lead to local hot-spots and can have a detrimental impact on the ITE thermal management. Also, with an increasing impetus towards green data centers and green ITE, data centers have started the transition towards efficient cooling technologies with additional benefits like waste heat recovery (Ardito, et al., 2014; Ebrahimi, et al., 2014).

As discussed, several factors such as low thermal conductivity, low specific heat capacity, and larger values of thermal resistance contribute towards the low efficiency of air cooling for high heat flux cooling. During the past decade, Direct Liquid Cooling (DLC) has again become one of the most popular methods. In this type of cooling, the coolant comes in direct contact with the

electronic components and since the coolant being dielectric, it provides an electrical insulation (Bar-Cohen, et al., 2006). The advantage of this type of method is that there is no sealed enclosure and piping is required to direct and maintain the fluid flow (Tuma, 2008). This type of cooling method involves phase change phenomena which yields uniform temperature profile because of associated latent heat transfer. Although direct contact between the coolant and heat source surface reduces thermal resistance, the thermo-physical characteristic of dielectric coolant is much lower than water, thus this method can be hardly considered as a heat transfer technique compared to conventional indirect liquid cooling solutions (Kheirabadi and Groulx, 2016). In this type of cooling approach, higher maintenance is required for preventing fluid losses and degassing the system of infiltrating air and humidity (El-Genk, 2012)(Tuma, 2011). Water-based dielectric coolants like ethylene/propylene glycol offer noticeably higher heat transfer capabilities (ASHRAE 2014). In indirect liquid cooling, the coolant is specifically targeted to the major heat-generating components like CPU/GPU and the memory units. The coolant is circulated through special mini/microchannel heat sinks or cold plates which are typically made from copper. One of the major benefits of this type of cooling is the ease of retrofitting the cold plates on existing air-cooled servers. These assemblies can be field replaceable assemblies can be used instead of conventional air-cooled heat sinks with a few changes to the data center infrastructure such as installing coolant distribution units, manifolds to provide coolant to the racks, etc. The main factor that effects this system is typically the cold plate performance. The cold plate performance is in turn related to parameters like the material, coolant inlet temperature, coolant flow rates, and channel geometry (Kandlikar, et al., 2009). The effect of these parameters on ITE cooling has been extensively studied in the current literature, Addagatla, et al. showed the impact of various inlet supply water temperatures on a hybrid air-liquid cooled server and characterized the thermal performance of a 2OU (OpenU) web server (Addagatla, et al., 2015). In this study, (Druzhinin, et al., 2012) designed a cold plate with low heat-resistance for an effective cooling with only 20-30°C between coolant and electronic components of the server. An earlier study (Shahi, et al., 2022), proposed a new cold plate design where the cold plate consists of four different fin sections and the flow through each section is passively controlled using bimetallic strip which response based on the outlet temperature from the fin section. The dependance of power efficiency on the inlet temperature of the coolant at individual server level was studied by (Moskovsky, et al., 2016) and showed that the power performance ratio drops the efficiency by 10% from 19°C to 65°C due to increase in the leakage current in chipset component and reduction of processor frequency. A recent study showed that the thermal performance for 2OU liquid cooled server experimentally and showed that savings in pumping power (Shahi, et al., 2022) can be achieved by dynamically varying the flow rate (Shahi, et al., 2021) and to vary the flow a flow control device (Shahi, et al., 2022) was designed and developed to vary across each server. A CFD study (Modi, et al., 2022) performed transient simulations, showing a control strategy to maintain a given outlet temperature for different power consumptions of a component and changing the flow rate using valves at the outlet, this led to upto 56% of savings in term of pumping power. A comparative study (Sahini, et al., 2017; Shahi, et al., 2022) evaluated the critical inlet temperature for air and liquid-cooled servers, the thermal resistance was also calculated by changing the coolant flow rate and chip power, (Ramakrishnan, et al., 2017) and the heat transfer coefficient was obtained from the resistance values using well-established relations. A control strategy was developed to stabilize the fluid supply temperature for a 450 kW CDU using PG25% coolant at very low heat loads using

combination of flow control valves and PID to reduce fluctuations of the three-way valve used in CDU and thus avoiding its failure (Heydari, et al., 2022). A study (Zeighami, et al., 2014) showed that the overall cost of ownership of data center can be substantially reduced by the adoption of hybrid cooled server. A study on the power consumption by a retrofitted water-cooled supercomputer (Ovaska, et al., 2016) showed that for 1 GigaFlop/s of computing performance there was an 11-15% drop in power consumption with water-cooled CPUs as compared with air-cooled CPUs.

While there are multiple accounts in the literature on improving the efficiencies of components of the liquid cooling system, i.e., cold plate optimization, distributed and centralized pumping configuration etc., a detailed account of a methodology to characterize a high-power liquid-cooled server is missing. This study aimed to characterize a multiple cold plate high-powered liquid-cooled loop for its thermal design. The contribution of this study lies in measuring the average percentage of heat removed at three different heat flux values and three different flow rates. To achieve this, mock electronic packages (TTVs) were developed using ceramic heaters and attached to each of the cold plates. The temperature of each TTV was measured using embedded thermocouples. Thermistors were also placed in each of the secondary loops within the server to calculate the heat captured by the coolant (25% Propylene Glycol) assuming that the inlet temperature remained constant until the coolant reached the cold plates. The thermal shadowing effect in each cooling loop was monitored to determine the efficiency of the cooling loop design. Additionally, the flow rates entering each of the cold plates were monitored using a separate identical cooling loop fitted with flow sensors. The results of this study indicate that 95% of the total heat was dissipated to the coolant with no external airflow under ambient conditions. A maximum change in thermal resistance was observed when the flow rate was increased from 4.5 lpm to 8 lpm, but the thermal resistance did not vary appreciably with increasing heat flux. Therefore, this study provides valuable insights into the thermal design of a multiple cold plate high-powered liquid-cooled loop.

2 Experimental Setup and Methodology

The setup for these experiments required several components to be assembled and, in this section, a detailed explanation is provided for the components assembled and tested. The components are as follows:

- (1) Quick Disconnects (QD): Server is connected with the rack manifold with the help of quick disconnects.
- (2) Each server consists of six cooling loops (four GPU and two NV switch cooling loops). In each loop, the cold plates are installed in series within the server.
- (3) The cooling fluid in the cooling loops is distributed by a server manifold.
- (4) Coolant Distribution Unit (CDU) is used to pump the coolant to the server in the rack.

Each cooling loop consists of at least two TTVs connected in series, from the front to the rear, and these TTVs are then connected to the server manifold. Multiple cooling loops are connected in parallel to the server-level flow distribution manifold. The server-level flow distribution manifold includes primary supply and return pipes for distributing coolant from the rack

manifold. The assembled temperature and flow verification servers were placed in the upper 1/3rd part of a standard 19-inch rack for experimentation. The flow verification server is present at the top shelf, followed by the temperature verification server under it, as shown in Figure 1. The rationale behind placing the two servers towards the top and adjacent to each other is that the flow rate difference in the last two ports of a rack manifold is not substantial compared to any other pair of ports.

The thermal verification server has thermistors installed at the inlet and outlet of the cold plate, as shown in Figure 2. These thermistors, labeled as TH, are installed on the NV and GPU cold plates. In the flow verification server, as shown in Figure 3, four flow sensors (FS) are installed. All sensor and thermistor readings were recorded through Keysight KT-DAQ970A-Data Acquisition System.

The assembled temperature and flow verification servers were then placed in the upper 1/3rd part of a standard 19-inch rack for experimentation. The flow verification server is located on the top shelf, followed by the temperature verification server underneath it. Each GPU cooling loop consists of a pair of TTVs connected in series, front to rear, with the loop's supply and return ends connected to the server manifold. Multiple such loops are connected in parallel to the server-level manifold, which includes the primary supply and return pipes facilitating coolant distribution from the rack manifold. The TTVs present at the inlet of the cooling loop are denoted by odd-numbered GPUs in Figure 3, and the TTVs at the rear or outlet of the cooling loops are denoted by even-numbered GPUs.

A total of eighteen experiments were performed at two different inlet temperatures and three different flow rates and total power consumption levels by the server. The power consumption levels were varied to represent power consumption levels of the current and future trends in GPU and CPU power consumptions (Chainer, et al., 2017). Coolant flow rates were varied between 4.5-8 lpm by changing the pumping power directly from the CDU. The coolant inlet temperatures of 35°C and 45°C were varied by changing the percentage valve opening within the CDU. This valve changes the heat exchange rate between the primary and the secondary cooling loops by varying the chilled water flow rate from the primary side. The power consumption of the TTVs was varied for power three power density values of 20 W/cm², 32 W/cm² and, 40 W/cm².

Uncertainty Analysis:

Experimental measurements and methods inherently include errors. Therefore, it is necessary to conduct an uncertainty analysis to validate the accuracy and reliability of the results. The study at hand involves experimental uncertainties pertaining to power, temperature, flow rate, and pressure drop. Table 1 displays the parameters of uncertainty that were ascertained through the calibration methodologies.

3 Results and Discussion

3.1 Analysis of Temperature Variations of Thermal Test Vehicles

Figure 4 shows the variation of TTV temperatures for a power density of 20W/cm² per GPU. As seen in the plot, a consistent trend in maximum temperatures is observed for the front and rear

GPUs with increasing flow rates at both coolant inlet temperatures. A maximum temperature of 66°C was obtained at GPU4 in the temperature verification server. The consistency in the temperatures obtained implies the efficient distribution of the coolant to each of the TTVs in all the cooling loops in the temperature verification server.

Figure 5 and Figure 6 show the variation of the TTV temperatures with varying rack flow rates and coolant inlet temperatures for 32 W/cm² and 40 W/cm² of power density per TTV, respectively. A maximum temperature of 78°C and 83°C was observed for these power levels. It was noted that the 4th TTV consistently shows higher temperatures as compared to the other TTVs for all power levels and flow conditions. This can be validated by the fact that the loop where GPU4 is located receives a slightly lower coolant flow rate. A reason behind this can be a variability in coolant flow distribution due to pressure drop variation in server manifold.

The flow rates entering each of the GPUs in the cooling loops were obtained using ultrasonic flow sensors retrofitted in the cooling loops. An assessment of the flow rates to each of the parallel loops in the server shows that the flow distribution from the server manifold is symmetrical. This means that the coolant flow rate in the extreme and the middle of the parallel loops stays the same as seen in Table 2.

3.2 Thermal Resistance Analysis of the Cold Plates

An energy balance was carried out for the power supplied to the TTV and the heat carried away by the coolant to calculate the thermal resistance of the cold plates. The thermal resistance calculation assisted in quantifying the convective behavior of the cold plates used in the experiments. The value of R_{th} also depicts the efficiency with which the cold plate assembly is performed on the server and the effect of sensor retrofitting in the cooling loops. The cold plate R_{th} is calculated using the first law of thermodynamics as:

$$R_{th} = \frac{T_{out} - T_{in}}{Q} \quad (1)$$

To calculate the temperature, rise in the coolant across the cold plate, thermistors were placed at the outlet of each TTV. The inlet temperatures were assumed to be constant from the inlet of the server manifold to the inlet of the front cold plates. The variation of the thermal resistance at the two inlet temperatures for different flow conditions is shown in Figure 7, Figure 8, and Figure 9.

The TTVs located at the extreme ends, GPU1 and GPU7, show a variation of less than 5% between their R_{th} values. A variation of 8.3% and 7.4% is observed among R_{th} values in the rear and the front cold plates, respectively. This variation was concluded to be due to a slight variation in coolant flow rates entering these cooling loops. It can also be due to a slight difference in contact pressure or the amount of TIM (Thermal Interface Material) between the cold plate and the TTV that can alter conductive heat flow. The issue of contact pressure variation was neglected as all the cold plates were assembled using a standard torque value with a digital torque wrench. Measurement errors were also minimal since the base of the custom-made TTVs was well insulated using insulating pads.

A maximum drop of 24-26% was observed in the cold plate thermal resistance as the rack flow rate increased from 4.5 lpm to 8 lpm for TTV 7 for the three power levels tested. This drop of thermal resistance reduces as the inlet temperature of the coolant is increased to 45°C from 35°C by 8-10%. However, a maximum reduction of 1.8% was observed as the TTV power increased from 20W/cm² and 40W/cm² for TTV3. Such variation has also been observed in other studies where a 2.7% variation was seen as the power was increased by more than double from its base value (Internet Live Stats, 2021). Thus, it can be concluded that the cooling loop assembly and the TTVs were designed efficiently for dissipating high power loads in flow and thermal limits tested.

Table 3 and Table 4 show the error in the sensible as calculated using equation 2 for the entire server and to each TTV. Q_{total} represents the sensible heat for each TTV calculated using the thermal data from thermistors placed at the outlet of each of the TTVs. Q_{overall} represents the sensible heat calculated from sensors placed at the inlet and out of each server. The last column in both the tables represents the difference in these values of sensible heats to represent the heat losses occurring from the piping and bare copper of the cold plates. It can be seen that a very small portion of the total power gets dissipated across the server which implies that most of the heat is carried away by the coolant. This error in heat dissipation showed a similar trend for other flow rates and was within 5% of the Q_{total} value.

$$Q = \dot{m}C_p\Delta T \quad (2)$$

Figure 11, Figure 12 and Figure 13 shows the variation in the temperature rise or thermal shadowing effect between the two cold plates observed for various thermal and flow condition cases tested. It was analyzed that as the power density on the TTVs is doubled, the temperature difference between the cold plates increases by 140% at 8 lpm flow rate and 35°C inlet temperature. The same value doubles for the same flow rate at 45°C inlet temperature of the coolant.

The local variation of the thermal shadowing value lies within 1-3°C, which implies that each of the cold plate loops in the server receives similar and sufficient flow rates. Also, the variation between the temperature difference values reduces as the flow rate increase from 4.5 to 8 lpm.

4 Conclusions

Direct-to-chip liquid cooling has quickly become one of the most popular in the last decade to dissipate high heat fluxes, especially for GPU-based server platforms. Efficient implementation of Direct-to-chip liquid cooling in existing air-cooled infrastructure will require careful system and rack-level thermal and flow characterizations. A significant number of components such as manifolds, disconnects, cold plates, CDUs make it difficult to characterize these systems.

A methodology for the detailed thermal assessment of high-power density liquid-cooled servers was presented and experimentally verified. Two servers with identical cooling loops and cold plates retrofitted with thermal and flow sensors were used to characterize the thermal performance of the cooling loop. Thermal verification was done by analyzing the total sensible heat dissipated by the coolant in the cooling loop of the temperature verification server. To calculate the sensible heat, the flow rates were monitored using the flow verification server and the heat dissipated was calculated using the first law of thermodynamics. The results showed

significantly low values of thermal resistance as compared to various studies published in the literature for the high-power densities used for testing. The measured temperatures of the TTVs measured using embedded thermocouples in the TTVs show that safe temperatures can be maintained even at high power densities for sufficient coolant flow rate values.

References

- Addagatla, A., Fernandes, J., Mani, D., Agonafer, D., and Mulay, V., Effect of warm water cooling for an isolated hybrid liquid cooled server, *Proc. Of 31st Thermal Measurement, modeling & Management Symposium (SEMI-THERM) Conference*, San Francisco, CA, pp. 203-207, 2015. DOI: 10.1109/SEMI-THERM.2015.7100161
- Alissa, H. A., Nemati, K., Sammakia, B. G., Schneebeil, K., Schmidt, R. R., & Seymour, M. J.. Chip to facility ramifications of containment solution on it airflow and uptime. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol 6, nos. 1, pp. 67-78, 2016. DOI: 10.1109/TCPMT.2015.2508453
- Ardito, L., and Green, I. T. Available data and guidelines for reducing energy consumption in IT systems, *Sustainable Computing: Informatics and Systems*, vol. 4, nos. 24-32, pp. 24-32, 2014. DOI: 10.1016/j.suscom.2013.09.001
- ASHRAE Technical Committee. Thermal guidelines for data processing environments-expanded data center classes and usage guidance. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, accessed June 10, 2022, from <http://www.ashrae.org/>, 2011.
- Bar-Cohen, A., Arik, M., & Ohadi, M. (2006). Direct liquid cooling of high flux micro and nano electronic components. *Proceedings of the IEEE*, vol. 94, nos. 8, pp. 1549-1570, 2006. DOI: 10.1109/JPROC.2006.879791
- Beatty, L., Liquid cooling guidelines. *ASHRAE Journal*, vol. 56, nos. 10, 2014.
- Brown, R. et al., Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-4, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA, Tech. Rep. LBNL-363E, Aug. 2007.
- Chainer, T. J., Schultz, M. D., Parida, P. R., and Gaynes, M. A. Improving data center energy efficiency with advanced thermal management. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 7, nos. 8, pp. 1228-1239, 2017. DOI: 10.1109/TCPMT.2017.2661700
- David, M. P., Iyengar, M., Parida, P., Simons, R., Schultz, M., Gaynes, M., Scmidth, R., and Chainer, T. Experimental characterization of an energy efficient chiller-less data center test facility with warm water cooled servers. *Proc. of 28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, San Francisco, CA, pp. 232-237, 2012. DOI: 10.1109/STHERM.2012.6188853
- Druzhinin, E., Shmelev, A., Moskovsky, A., Migal, Y., Mironov, V., and Semin, A. High temperature coolant demonstrated for a computational cluster. *Proc. Of International*

- Conference on High Performance Computing & Simulation (HPCS)*, Innsbruck, Austria, pp. 814-817, 2016. DOI: 10.1109/HPCSim.2016.7568418
- Ebrahimi, K., Jones, G. F., and Fleischer, A. S. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews*, vol **31**, pp. 622-638, 2014. DOI: 10.1016/j.rser.2013.12.007
- El-Genk, M. S.. Immersion cooling nucleate boiling of high power computer chips. *Energy conversion and management*, vol **53**, nos. 1, pp. 205-218, 2012. DOI: 10.1016/j.enconman.2011.08.008
- Fakhim, B., Behnia, M., Armfield, S. W., and Srinarayana, N. Cooling solutions in an operational data centre: A case study. *Applied thermal engineering*, vol **31**, nos. 14-15, pp. 2279-2291, 2011. DOI: 10.1016/j.applthermaleng.2011.03.025
- Heydari, A., Shahi, P., Radmard, V., Eslami, B., Chowdhury, U., Hinge, C., ... & Rodriguez, J., A Control Strategy for Minimizing Temperature Fluctuations in High Power Liquid to Liquid CDUs Operated at Very Low Heat Loads. *Proc. Of International Electronic Packaging Technical Conference and Exhibition*, vol. **86557**, pp. V001T01A011, 2022. DOI: 10.1115/IPACK2022-97434
- Internet Live Statistics, accessed Feb. 27, 2022, from <http://www.internetlivestats.com/>, 2022
- Kandlikar, S. G., and Hayner, C. N, Liquid cooled cold plates for industrial high-power electronic devices—thermal design and manufacturing considerations. *Heat transfer engineering*, vol **30**, nos. 12, pp. 918-930, 2009. DOI: 10.1080/01457630902837343
- Kheirabadi, A. C., & Groulx, D.. Cooling of server electronics: A design review of existing technology. *Applied Thermal Engineering*, vol. **105**, pp. 622-638, 2016. DOI: 10.1016/j.applthermaleng.2016.03.056
- Modi, H., Shahi, P., Sivakumar, A., Saini, S., Bansode, P., Shalom, V., ... & Agonafer, D., Transient CFD Analysis of Dynamic Liquid-Cooling Implementation at Rack Level. *Proc. of International Electronic Packaging Technical Conference and Exhibition*, vol. **86557**, pp. V001T01A012, 2022. DOI: 10.1115/IPACK2022-97443
- Moskovsky, A. A., Druzhinin, E. A., Shmelev, A. B., Mironov, V. V., and Semin, A. Server Level Liquid Cooling: Do Higher System Temperatures Improve Energy Efficiency?. *Supercomputing frontiers and innovations*, vol **3**, nos. 1, pp. 67-74, 2016. DOI: 10.14529/jsfi160104
- Parida, P. R., Vega, A., Buyuktosunoglu, A., Bose, P., and Chainer, T, Embedded two phase liquid cooling for increasing computational efficiency. *Proc. of 15th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Las Vegas, NV, pp. 326-336, 2016. DOI: 10.1109/ITHERM.2016.7517567
- Ramakrishnan, B., Hadad, Y., Alkharabsheh, S., Chiarot, P. R., and Sammakia, B. Thermal analysis of cold plate for direct liquid cooling of high performance servers. *Journal of Electronic Packaging*, vol **141**, nos. 4, 2019. DOI: 10.1115/1.4044130

- Ramakrishnan, B., Alkharabsheh, S., Hadad, Y., Sammakia, B., Chiarot, P. R., Seymour, M., and Tipton, R. Experimental characterization of a cold plate used in warm water cooling of data centers. *Proc. of 33rd Thermal Measurement, Modeling & Management Symposium (SEMI-THERM)*, San Jose, CA, pp. 191-196, 2017. DOI: 10.1109/SEMI-THERM.2017.7896929
- Sahini, M., Chowdhury, U., Siddarth, A., Pradip, T., Agonafer, D., Zeigham, R., Metcalf, J. and Branton, S., Comparative study of high ambient inlet temperature effects on the performance of air vs. liquid cooled IT equipment, *Proc. of 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Orlando, FL, pp. 544-550, 2017. DOI: 10.1109/ITHERM.2017.7992534
- Schmidt, R., & Iyengar, M., Server rack rear door heat exchanger and the new ASHRAE recommended environmental guidelines. *Proc. Of International Electronic Packaging Technical Conference and Exhibition*, vol. **43604**, pp. 851-862, 2009. DOI: 10.1115/InterPACK2009-89212
- Schmidt, R., Steinke, M., & Shelnutt, A., Moderating the Impact of Integrating Water-Cooled Servers Into Data Centers. *ASHRAE Journal*, vol **61**, 2019.
- Shahi, P., Deshmukh, A. P., Hurnekar, H. Y., Saini, S., Bansode, P., Kasukurthy, R., and Agonafer, D. (2022). Design, Development, and Characterization of a Flow Control Device for Dynamic Cooling of Liquid-Cooled Servers. *Journal of Electronic Packaging*, vol **144**, nos 4, 2022. DOI: 10.1115/1.4052324
- Shahi, P., Deshmukh, A., Hurnekar, H., Saini, S., Bansode, P., and Agonafer, D. Numerical Investigation on Effect of Target Coolant Delivery in Liquid-Cooled Microchannel Heat Sinks. *Journal of Enhanced Heat Transfer*, ISSN 1065-5131 2022. DOI: 10.1615/JEnhHeatTransf.2022044339
- Shahi, P., Hurnekar, H., Deshmukh, A., Saini, S., Bansode, P., Kasukurthy, R., and Agonafer, D. Assessment of Pump Power Savings at Rack level for Dynamic Direct-to-Chip Liquid Cooling Using a Novel Flow Control Device. *Journal of Enhanced Heat Transfer*, ISSN 1065-5131, 2022. DOI: 10.1615/JEnhHeatTransf.2022044476
- Shahi, P., Mathew, A., Saini, S., Bansode, P., Kasukurthy, R., and Agonafer, D., Assessment of Reliability Enhancement in High-Power CPUs and GPUs Using Dynamic Direct-to-Chip Liquid Cooling. *Journal of Enhanced Heat Transfer*, vol **29**, nos. 8, 2022. DOI: 10.1615/JEnhHeatTransf.2022043858
- Shahi, P., Saini, S., Bansode, P., and Agonafer, D, A comparative study of energy savings in a liquid-cooled server by dynamic control of coolant flow rate at server level. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol **11**, nos. 4 pp. 616-624, 2021. DOI: 10.1109/TCPMT.2021.3067045
- Shalom Simon, V., Modi, H., Sivaraju, K. B., Bansode, P., Saini, S., Shahi, P., ... & Agonafer, D., Feasibility Study of Rear Door Heat Exchanger for a High Capacity Data Center. *Proc. Of International Electronic Packaging Technical Conference and Exhibition*, vol. **86557**, pp. V001T01A018, 2022. DOI: 10.1115/IPACK2022-97494

- Tuma, P. E.. Fluoroketone C₂F₅C(O)CF(CF₃)₂ as a heat transfer fluid for passive and pumped 2-phase applications. *Proc. Of Twenty-fourth Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, pp. 173-179, 2008. DOI: 10.1109/STHERM.2008.4509386
- Tuma, P. E.. Design considerations relating to non-thermal aspects of passive 2-phase immersion cooling. *Proc. Of 27th Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, pp. 1-9, 2011. DOI: 10.1109/STHERM.2011.5767224
- Ovaska, S. J., Dragseth, R. E., and Hanssen, S. A, Impact of retrofitted CPU water cooling on supercomputer performance and power consumption. *Proc. Of SoutheastCon*, pp. 1-2, 2016. DOI: 10.1109/SECON.2016.7506669
- Zeighami, R., Saunders, W. A., Coles, H., and Branton, S., Thermal performance modeling of hybrid liquid-air cooled servers. *Proc. Of Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, pp. 583-587, 2014. DOI: 10.1109/ITHERM.2014.6892333

List of Figures

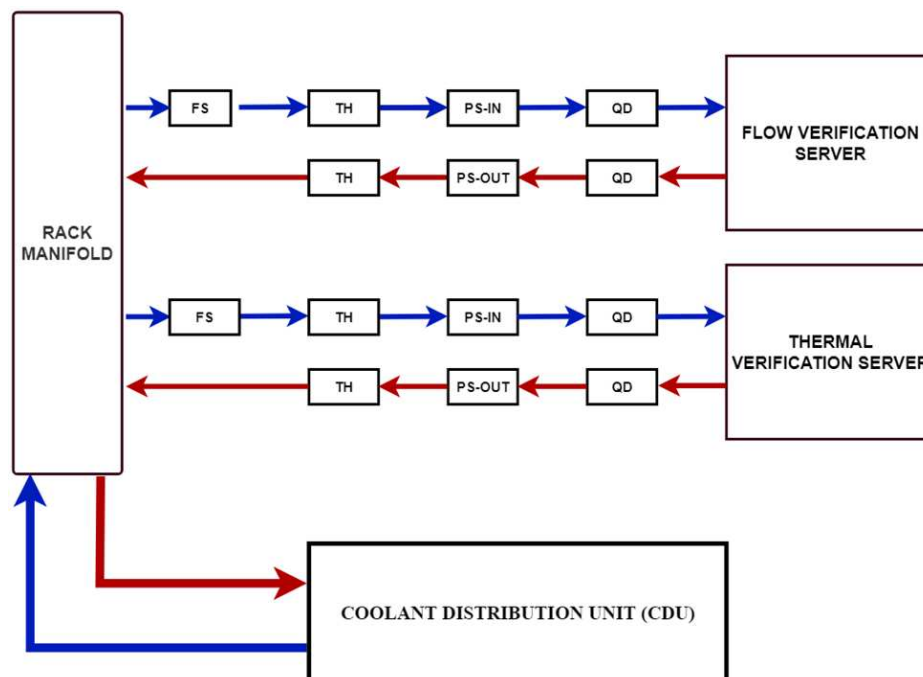


Fig. 1. Schematic of the thermal and flow verification serves placed in the rack for experimental testing

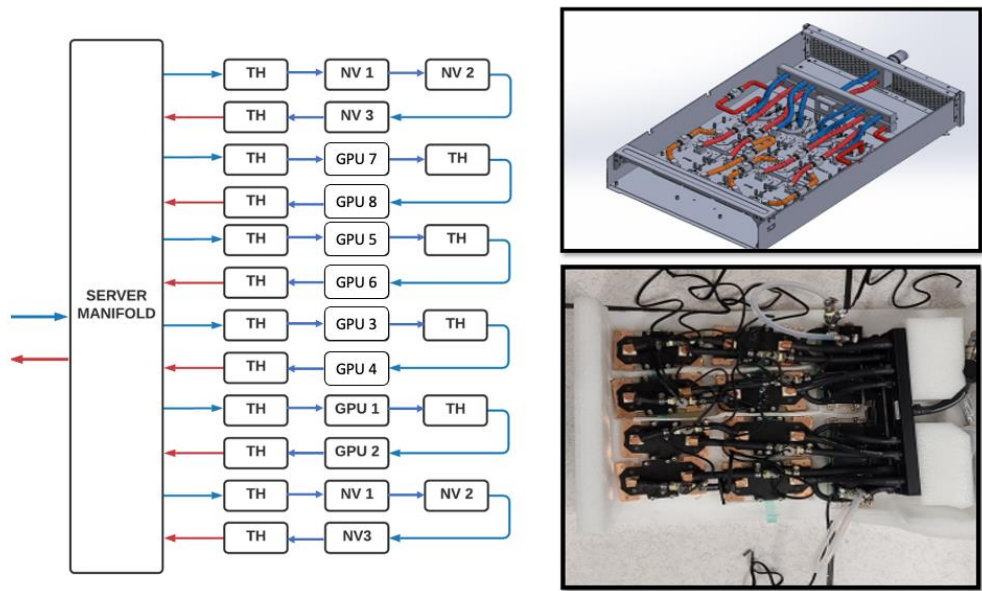


Fig. 2. Schematic of thermal verification server showing the placement of thermal sensors in the cooling loop

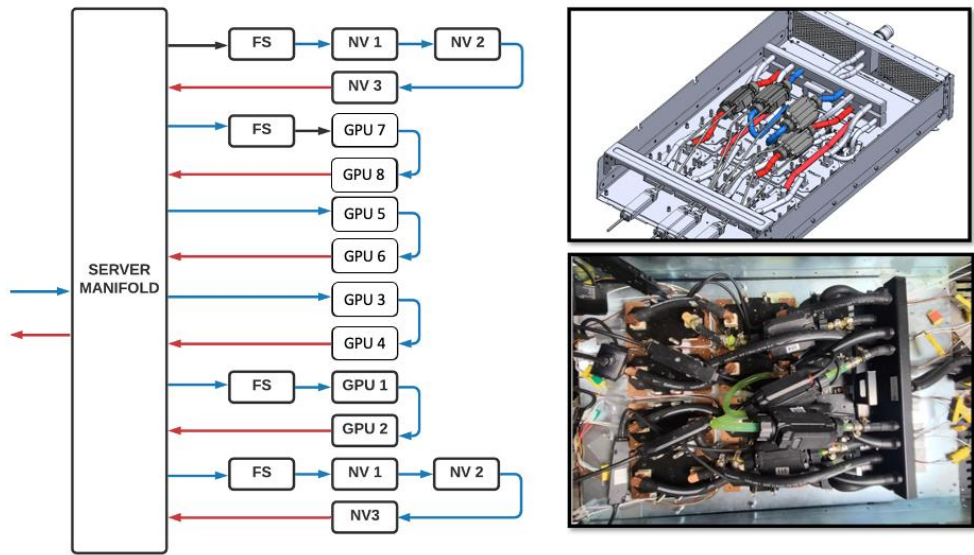


Fig. 3. Schematic of flow verification server showing the placement of flow sensors in the cooling loop

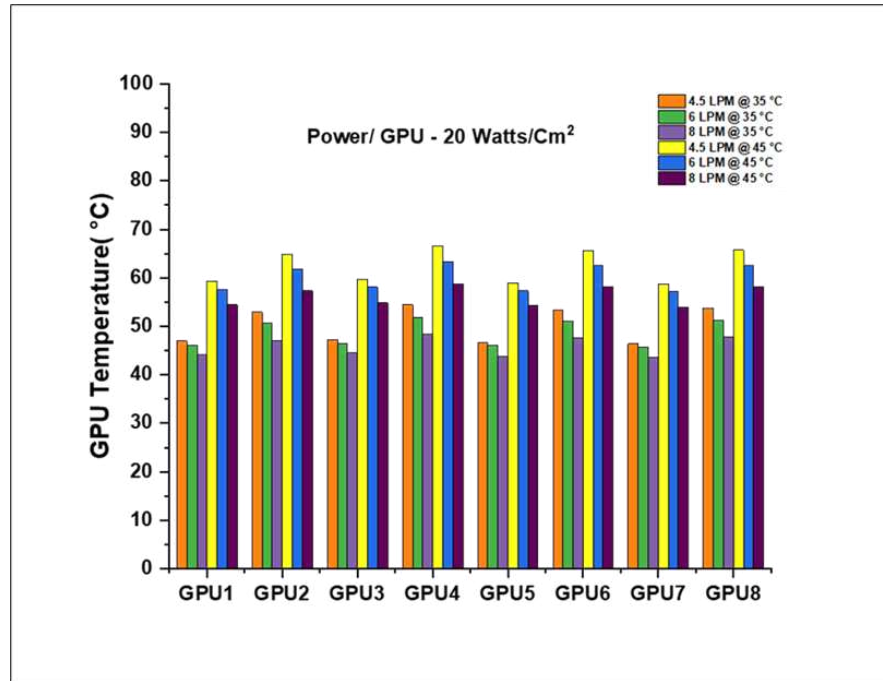


Fig. 4. Variation of TTV temperatures for a power level of 20W/cm² per TTV at different inlet temperatures and flow rates

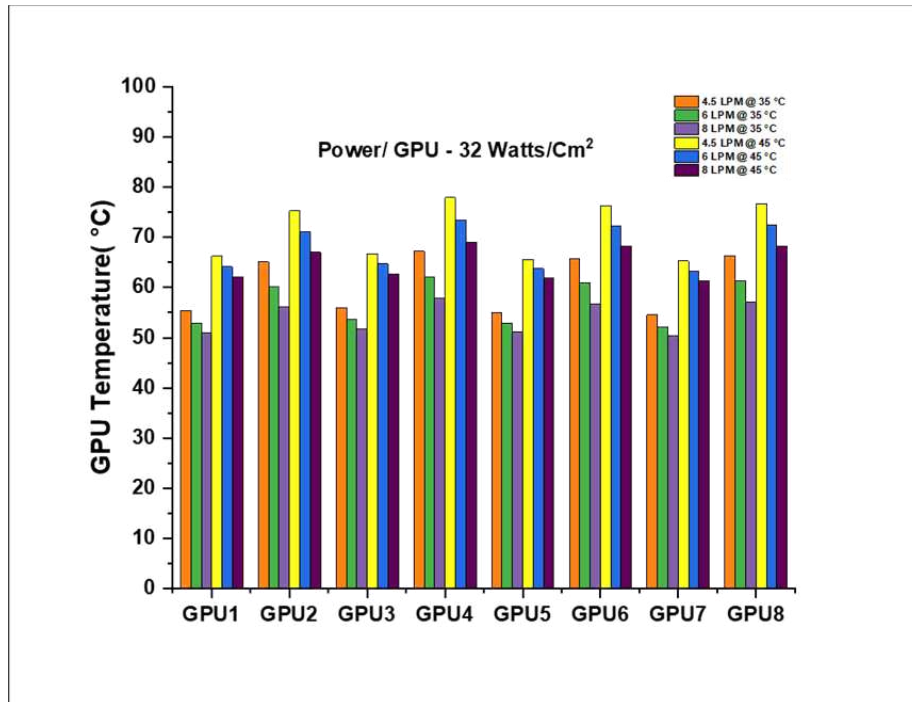


Fig. 5. Variation of TTV temperatures for a power level of 32W/cm² per TTV at different inlet temperatures and flow rates

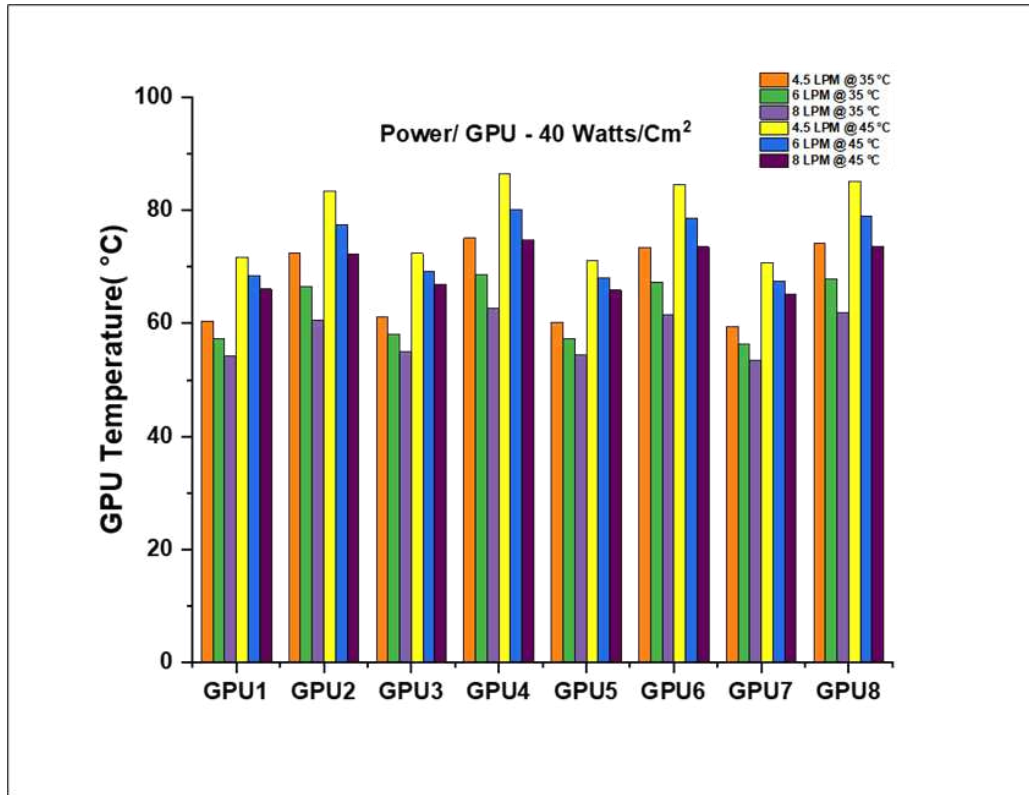


Fig. 6. Variation of TTV temperatures for a power level of 40W/cm² per TTV at different inlet temperatures and flow rates

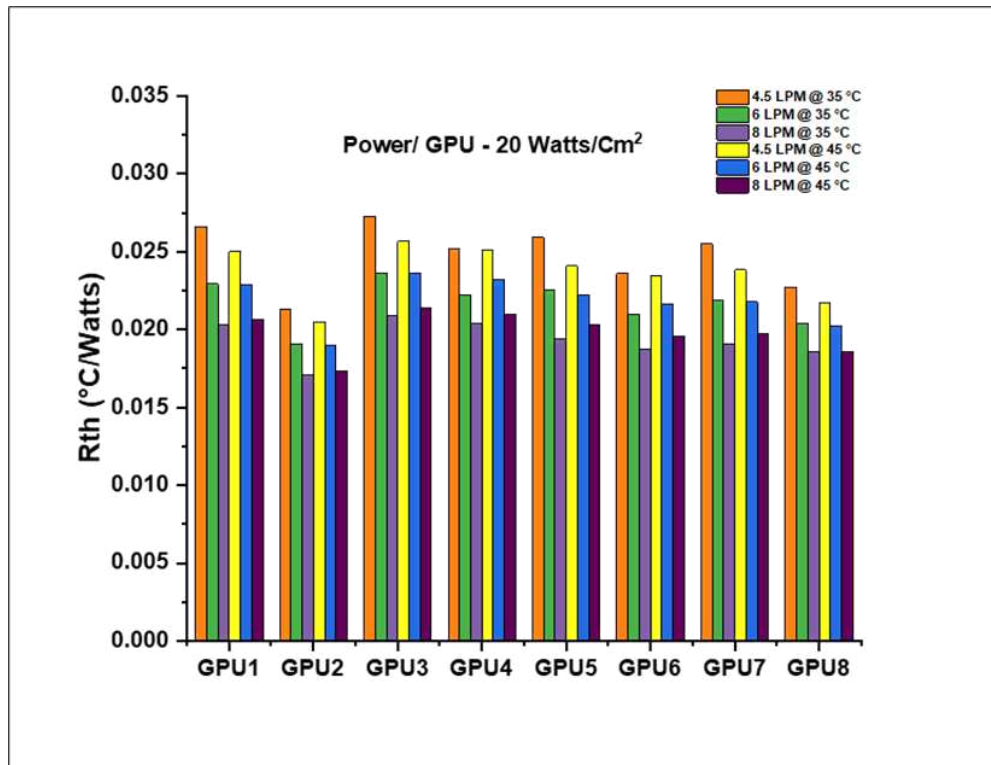


Fig. 7. Variation of cold plate thermal resistance for a power level of 20W/cm² per TTV at

different inlet temperatures and flow rates

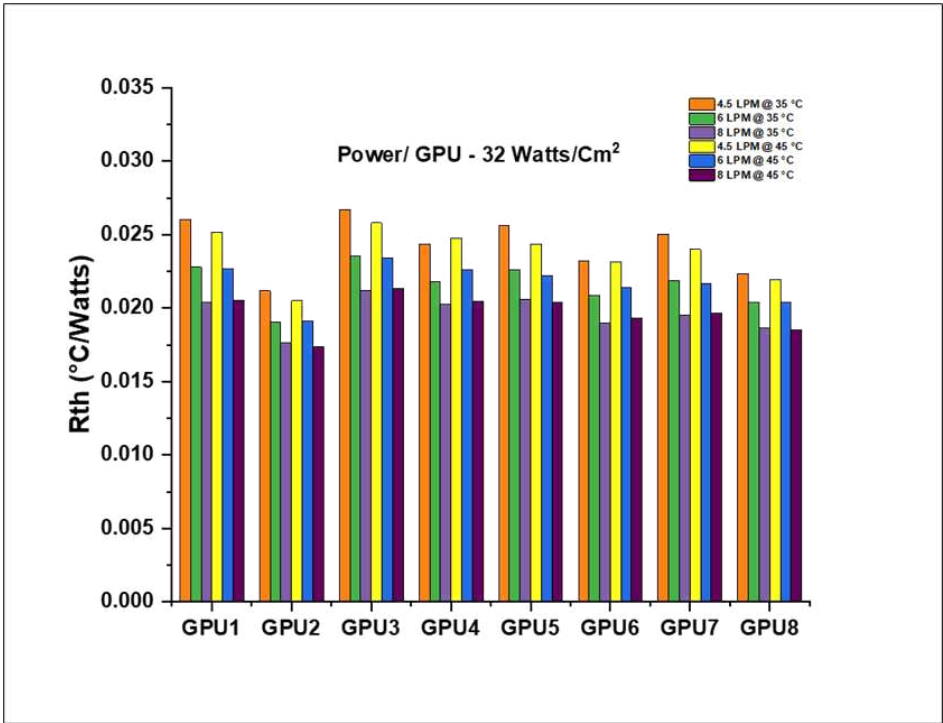


Fig. 8. Variation of cold plate thermal resistance for a power level of 32W/cm² per TTV at different inlet temperatures and flow rates

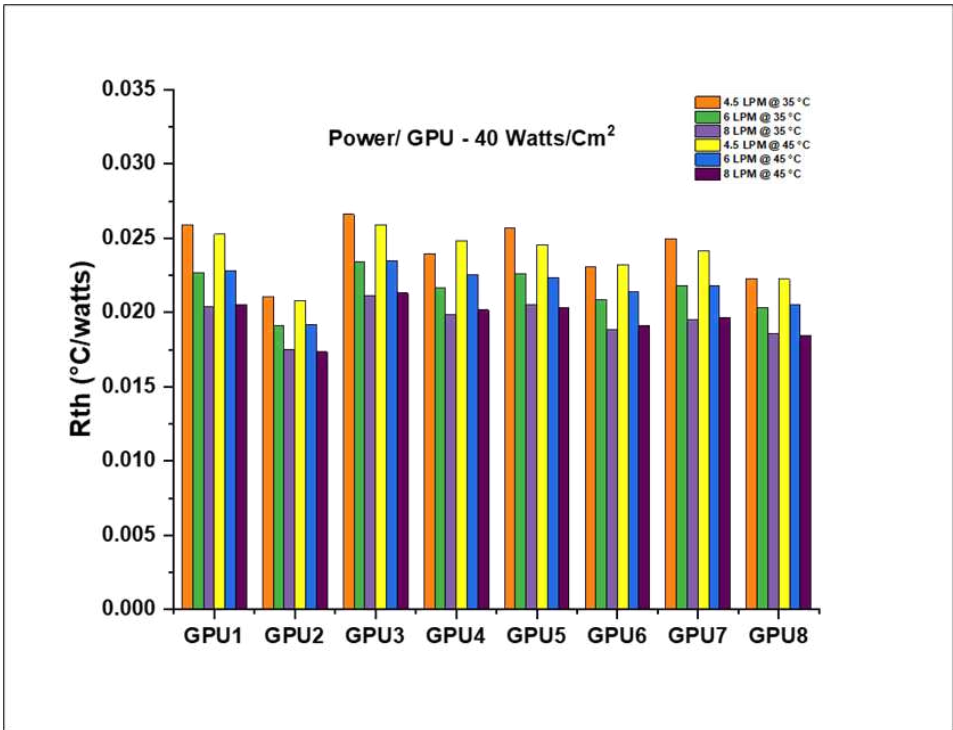


Fig. 9. Variation of cold plate thermal resistance for a power level of 40W/cm² per TTV at different inlet temperatures and flow rates

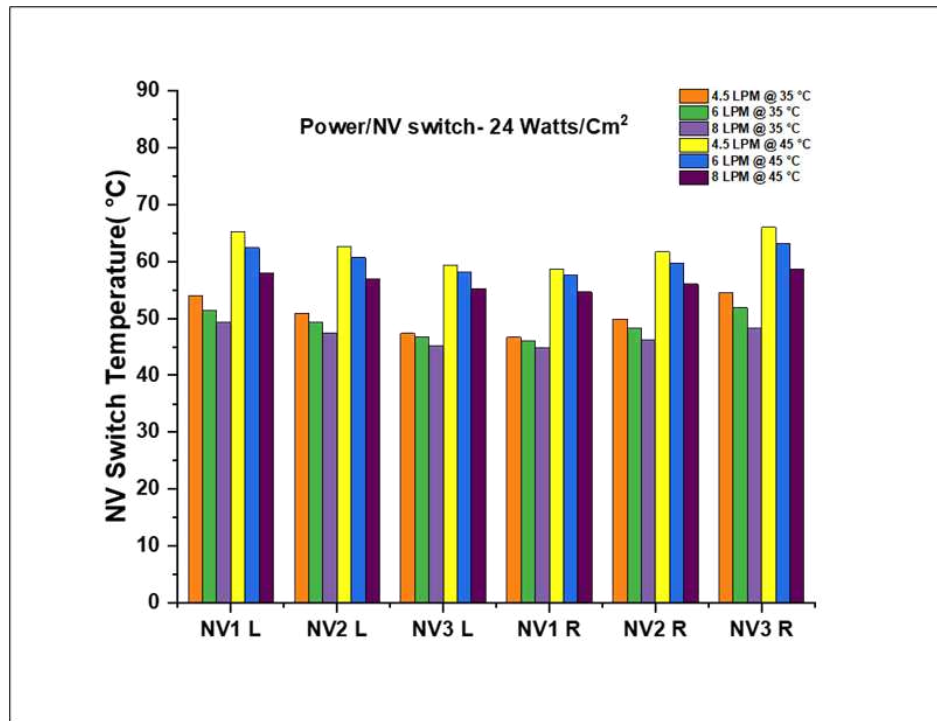


Fig. 10. Variation of TTV temperatures for a power level of 24W/cm² per TTV at different inlet temperatures and flow rates

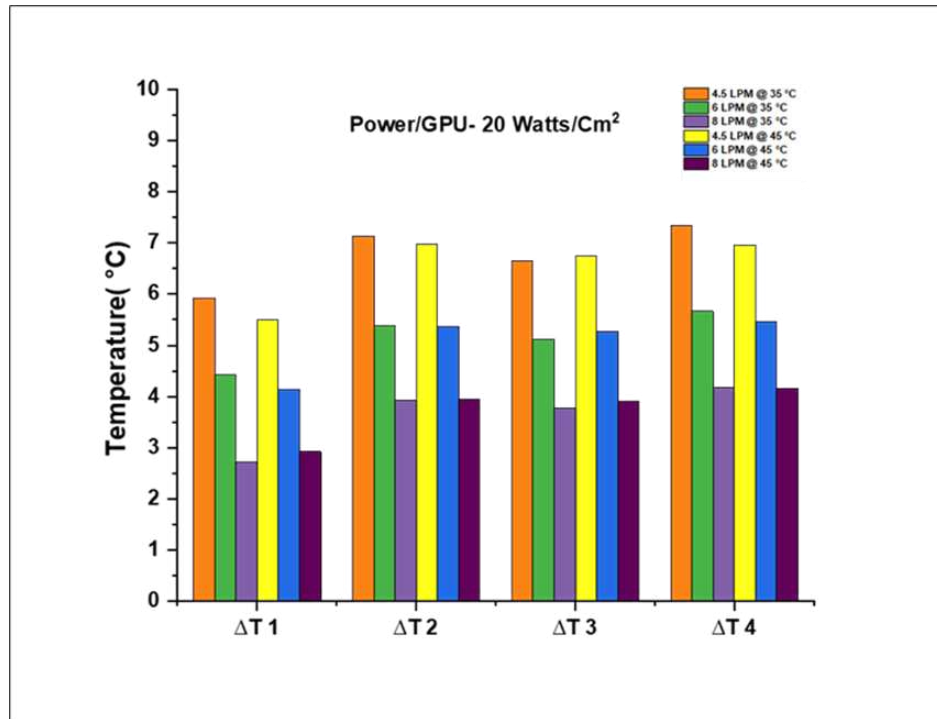


Fig. 11. Variation of the thermal shadowing between the cold plates in series for a power density of 20W/cm² per TTV at different inlet temperatures and flow rates

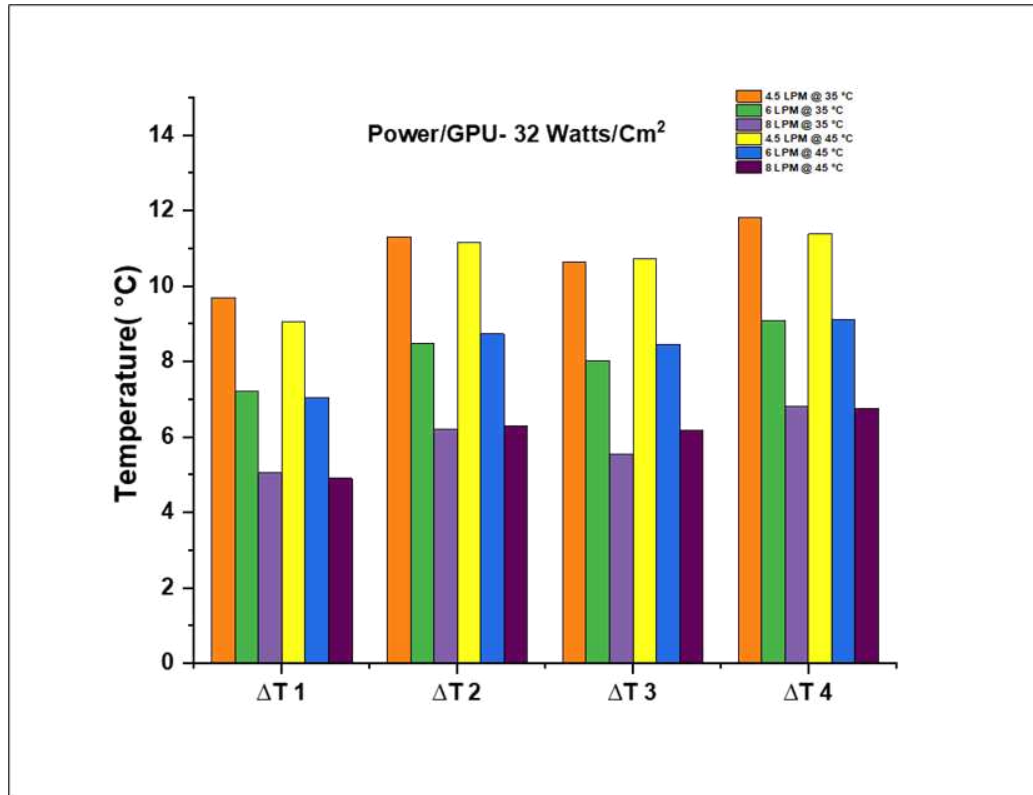


Fig. 12. Variation of the thermal shadowing between the cold plates in series for a power density of 32W/cm² per TTV at different inlet temperatures and flow rates

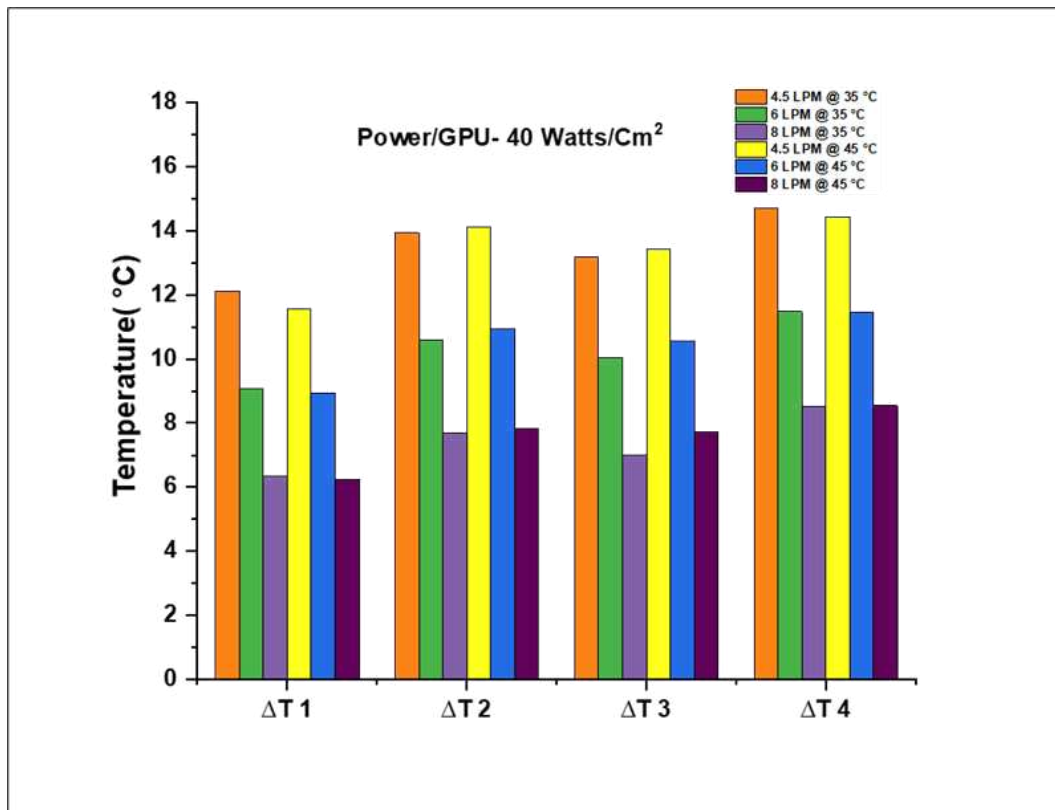


Fig. 13. Variation of the thermal shadowing between the cold plates in series for a power density of 40W/cm² per TTV at different inlet temperatures and flow rates

List of Tables

TABLE 1: Uncertainty Analysis

Equipment	Uncertainty Values
Flow sensor	0.3%
Pressure sensor	0.25%
K type thermocouple	± 0.7°C
10k Ω thermistors	±2°C
Power meter	±0.3% of the reading

TABLE 2: Flow Rates in Each Loop in the server verified using ultrasonic flow sensor in flow verification sensor

NVL Loop (lpm)	GPU1 Loop (lpm)	GPU2 Loop (lpm)	GPU3 Loop (lpm)	GPU4 Loop (lpm)	NVR Loop (lpm)
0.64	1.16	1.11	1.11	1.16	0.77

TABLE 3: Server Power Calculated from energy balance and electrical power given to each TTV for 35°C inlet temperature for 6 LPM Flow Rate

Power/ GPU (W/cm ²)	Q GPU1 (kW)	Q GPU2 (kW)	Q GPU3 (kW)	Q GPU4 (kW)	Q GPU5 (kW)	Q GPU6 (kW)	Q GPU7 (kW)	Q GPU8 (kW)	Q NVL (kW)	Q NVR (kW)	Q Total (kW)	Q Overall (kW)	Error (%)
20	0.51	0.51	0.46	0.52	0.45	0.46	0.51	0.51	0.43	0.49	4.86	4.73	2.87
32	0.82	0.78	0.75	0.81	0.72	0.73	0.82	0.78	0.46	0.52	7.19	7.01	2.46
40	1.01	0.97	0.94	0.99	0.90	0.90	1.04	0.95	0.47	0.53	8.70	8.60	1.15

TABLE 4: Server Power Calculated from Energy Balance and Electrical Power given to each TTV for 45°C Inlet Temperature and 6 LPM flow rate

Power/ GPU (W/cm²)	Q GPU1 (kW)	Q GPU2 (kW)	Q GPU3 (kW)	Q GPU4 (kW)	Q GPU5 (kW)	Q GPU6 (kW)	Q GPU7 (kW)	Q GPU8 (kW)	Q NVL (kW)	Q NVR (kW)	Q Total (kW)	Q Overall (kW)	Error (%)
20	0.47	0.50	0.41	0.54	0.41	0.48	0.48	0.48	0.40	0.45	4.63	4.68	1.03
32	0.77	0.78	0.70	0.82	0.68	0.74	0.78	0.76	0.40	0.45	6.88	7.01	1.85
40	0.96	0.96	0.88	1.01	0.86	0.92	0.98	0.94	0.41	0.47	8.39	8.49	1.25