

PhD Dissertation Defense Announcement
Mechanical and Aerospace Engineering Department
University of Texas at Arlington

**COLD PLATE BASED DYNAMIC LIQUID COOLING AT DATA
CENTER AND CHIP LEVEL**

By

PARDEEP SHAHI

Thesis Advisor: Dr. Dereje Agonafer

10:00 Monday, 06/27/2022

[Teams Link](#)

Abstract

The rising demand for high-performance central and graphical processing units has resulted in the need for more efficient thermal management techniques like direct-to-chip liquid cooling. Direct Liquid Cooling using cold plates is one of the most efficient and investigated cooling technology since the 1980s. Major data and cloud providers like IBM, Microsoft, and Google are actively deploying liquid-cooled data center infrastructure due to rising computational demands but its performance and efficiency can be further be enhanced using dynamic cooling technologies.

At the chip level since the early '60s, based on Moore's law, transistor density has been doubling every generation resulting in increased power density. Eventually, in the early '90s, we moved from constant voltage to constant electric field and corresponding constant power for a given area during technology changes. Dennard's model of voltage scaling and corresponding constant power ceased, ending improved performance gains in the early 2000s that again required techniques to mitigate increased power and corresponding temperature. The performance gain is being achieved by using multi-core processors, leading to non-uniform power distribution and localized high temperatures making cooling very challenging.

At rack level the coolant distribution to multiple racks housing such computing systems is an important factor for efficient cooling. Most liquid-cooled data centers provision their IT equipment with a constant coolant flow rate through multi-channel heat sinks mounted on the processing units. A redundant cooling flow rate is provided based on the maximum anticipated heat loads from the processing units to be cooled. This, in turn, consumes significant pumping power even when the computing systems are working at their lowest possible IT loads.

This study will address the issues as mentioned above for future direct to chip liquid cooled data centers. First the benchtop experiments were performed on open compute (OU) server. In this study 2OU server is used to show the effect of variable flow rate (0.2, 0.4, 0.6 LPM) on the core temperature, DIMMs (Dual In-line Memory Modules) temperature, Platform Controller Hub (PCH) temperature, cooling power at Ideal, 50% and 100% IT load at inlet temperatures of 25, 30, 35, 40 and 45 ° C, which falls within the ASHRAE liquid-cooled envelope, W4. For rack level dynamic cooling the novel active Flow control device (FCD) is designed to control the coolant flow rates at the server level. The dynamic cooling will result in pumping power savings by controlling the flow rates based on server utilization. The proposed FCD design contains a V-cut ball valve connected to a micro servo motor. The valve position is varied to change the flow rate through the valve by servo motor actuation based on pre-decided rotational angles. FCD working was validated by varying flow rates and pressure drop across the device by varying the valve position using both CFD and experiments at bench top. Further the experiments were performed at rack level using this FCD and control strategy was tested for maximum pumping power saving of 87%. For chip level dynamic cooling to reduce thermal gradient on the chip, the passive bimetallic based cold plate is designed. The proposed dynamic cold plate design is a 3-part assembly that is divided into three main parts. The bottom part of the plate contains 4 different sections of parallel copper microchannels through which the coolant flows, extracting the heat. The middle part of the plate has both inlet and outlet passage with bimetallic strips to control flow rate in each section based on coolant temperature in that section, and the top section of the plate is the plastic sealed cover plate. Dynamic cold plate working was validated by using CFD and it was observed that maximum of 62% temperature gradient can be reduced on the chip.

Further, Liquid to liquid heat exchangers used in liquid-cooled data centers is also referred to as coolant distribution units (CDUs). Most of these CDUs selected by the data center operator is based on the heat load of the data center and the available head with that CDU. A 450-kW liquid-cooled CDU is used, and propylene glycol 25% is used as a coolant. Typical CDUs are designed to operate at 20 to 30% of the rated heat load to achieve a stable secondary coolant supply temperature. The present study will investigate the operations of CDU at very low heat loads, like 1% to 10% of the CDU's rated capacity. At these low loads, large fluctuations in secondary side supply temperature were observed. This large fluctuation can lead to the failure of the 3-way valve used in CDUs at the primary side. In this paper, a control strategy is developed to stabilize the secondary supply temperature within ± 0.5 °C at very low loads using the combination of a flow control valve on the primary side and PID control settings within the CDU.

The present study provides an in-depth analysis of hydraulic transients when rack-level flow control valves are used with and without flow control. The operating conditions of the CDU are varied for different parameters such as a constant flow rate, a constant differential pressure, and a constant pump speed. Furthermore, hydraulic transient is examined when the cooling loop modules are decommissioned from the rack one by one. The effect of this step-by-step decommissioning is assessed on the CDU operation and other racks. The pressure drop based control strategy has been developed to maintain the same flow rate in the remaining servers in the rack when some cooling loop modules are decommissioned