# MINIMIZING POWER CONSUMPTION AT MODULE, SERVER AND RACK-LEVELS WITHIN A DATA CENTER THROUGH DESIGN AND ENERGY-EFFICIENT OPERATION OF DYNAMIC COOLING SOLUTIONS

by

JOHN EDWARD FERNANDES

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2015

# ACKNOWLEDGEMENTS

I would like to thank Dr. Dereje Agonafer for his continued mentoring and support over the course of my doctoral degree. He is primarily responsible for encouraging me to work on industry-funded projects and understanding the importance of networking by attending conferences and workshops.

I would also like to thank Dr. Haji-Sheikh, Dr. Nomura and Dr. Woods for serving on my dissertation committee. I owe a great deal of gratitude to Dr. Veerendra Mulay of Facebook Inc. for not only serving as an external committee member, but also being my mentor with regards to industry-related topics. As a former member of the EMNSPC, he was a constant source of encouragement and was instrumental in my search for employment.

I cannot thank Ms. Sally Thompson enough for her assistance with administrative matters or otherwise during my time with the EMNSPC. Our conversations always served as a stress reliever and her concern for my well-being was genuine. Special thanks to Rick Eiland for not only co-leading our funded projects, but also lending an ear and providing input on much of my research. I am also thankful to Marianna Vallejo, Saeed Ghalambor and other EMNSPC team members for their support.

Most importantly, I would like to thank my family for supporting my decision to pursue a doctoral degree. My father (Edward Fernandes) and brother (Gregory Fernandes) have always been there to provide both emotional and financial support over the past five years. My mother (Fatima Fernandes), who is no longer with us, is responsible for who I am today.

March 23, 2015

# ABSTRACT

## MINIMIZING POWER CONSUMPTION AT MODULE, SERVER AND RACK-LEVELS WITHIN A DATA CENTER THROUGH DESIGN AND ENERGY-EFFICIENT OPERATION OF DYNAMIC COOLING SOLUTIONS

John Edward Fernandes, PhD

The University of Texas at Arlington, 2015

Supervising Professor: Dereje Agonafer

Data center energy consumption continues to increase with the proliferation in online services such as social networking, banking, entertainment, cloud computing, etc. Recent estimates show that associated power consumption accounts for around 2% and 1.8% of electricity production in the United States and worldwide respectively. In addition to growth in the industry, power densities continue to rise in IT equipment increasing the criticality of thermal management in data center operation. Cooling systems account for around a third of overall power consumption requiring more energy-efficient solutions across all levels within the facility.

Liquid cooling of high power modules using cold plates has been around since the early 1980s. However, till today, designs remain fairly static and fail to adapt to variations in power dissipation at the device. This can be alleviated through the introduction of a “dynamic” cold plate design. Through implementation of sensing and control, the solution can distribute available resources based on local cooling requirements. A high-power multi-chip module (MCM) platform is chosen as reference for the design of such a solution. In-depth computational fluid dynamics (CFD) analysis is conducted to select appropriate heat transfer surfaces and predict thermal performance of the cold plate, when assembled with the MCM. A cost-effective MCM thermal test vehicle is assembled to enable experimental testing of both dynamic and static cold plates. Components of a liquid cooling test bench and control system are selected and permit future evaluation of thermal performance and energy-efficiency of both solutions.

The current trend in air cooling of data centers involves higher ambient temperatures to maximize use of free cooling. However, power consumption at the server-level may increase due to elevated fan activity and CPU leakage current. Minimizing power consumption of web servers (1.5U profile) is achieved by studying the effect of ambient temperature on performance and investigating means to improve chassis fan control through accurate selection of CPU target temperature. Multiple servers are instrumented and deployed in a test bed data center and are subjected to different air supply temperatures and fan speeds to achieve the same. Limits of energy-efficient operation and available savings will be discussed.

Fan efficiency is known to increase with size. Departing from conventional server designs, wherein fans are installed within the chassis, and consolidating air moving devices at the rear of a rack or ‘stack’ of servers permits increase in size and cooling efficiency. Preliminary studies have shown that replacing server-enclosed 60mm units with a rear-mounted wall of larger fans (80mm or 120mm) enables savings in fan power of the order of 50%. A methodology for row-wise control of such rack-level fans, with the purpose of simulating an actual product, is previewed and savings are reported. In addition, performance under real-life scenarios such as non-uniform loads and fan failure is investigated. Each rack-level setup has distinct advantages. However, selecting between configurations would necessitate a compromise between efficiency, redundancy and cost.

TABLE OF CONTENTS

## LIST OF ILLUSTRATIONS

## LIST OF TABLES

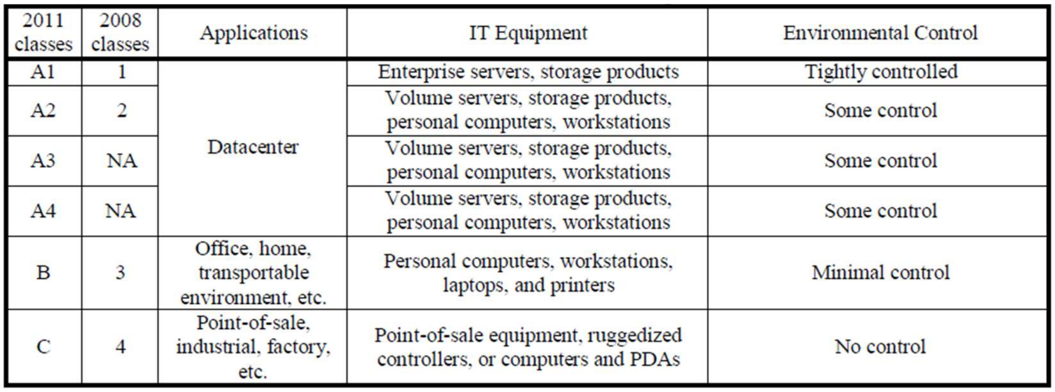
# CHAPTER 1

# INTRODUCTION

A data center is a dedicated space that centralizes an organization's IT equipment and operations, and is responsible for processing, storing and exchanging digital information. The primary contents of a data center can be broken down as follows:

- IT equipment – actual equipment responsible for managing data. This includes compute servers that process the information, storage servers that store the information and networking equipment that serve to enable communication across servers within the facility.
- Support Infrastructure – equipment responsible for maintaining operation of IT counterparts. Two primary subcomponents are power delivery and conversion equipment that power and maintain uninterrupted operation of IT equipment, and cooling systems that control the operating environment as per requirement.

Table 1-1 Environmental Class Definitions [1]



| 2011 classes | 2008 classes | Applications | IT Equipment | Environmental Control |
|---|---|---|---|---|
| A1 | 1 | Datacenter | Enterprise servers, storage products | Tightly controlled |
| A2 | 2 | | Volume servers, storage products, personal computers, workstations | Some control |
| A3 | NA | | Volume servers, storage products, personal computers, workstations | Some control |
| A4 | NA | | Volume servers, storage products, personal computers, workstations | Some control |
| B | 3 | Office, home, transportable environment, etc. | Personal computers, workstations, laptops, and printers | Minimal control |
| C | 4 | Point-of-sale, industrial, factory, etc. | Point-of-sale equipment, ruggedized controllers, or computers and PDAs | No control |

Due to the critical function and nature of such facilities there are stringent guidelines governing their maintenance and operation. ASHRAE TC 9.9 [1] specifies environmental classes that govern product operating conditions based on the mission critical nature of the data center and type of equipment contained within as shown in Table 1-1. By expanding the number of classes (A1 to A4, as opposed to 1 and 2 in 2008), the end user can operate the facility based on whether the primary attribute is reliability/uptime or energy-efficiency.

## 1.1 Data Center Power Trends and Cooling

The technical committee (TC) 9.9 in ASHRAE is primarily responsible for establishing guidelines for operation, and tracking and publishing trends of data center facilities as reference for operators worldwide. In recent years, data center equipment power trends [2] have started to level in comparison to the sharp rise in power densities (heat load per product footprint, W/m$^2$) observed between 1997 and 2005. However, as seen in Figure 1-1, heat loads for high density IT equipment continue to increase and this trend is expected to endure till 2014.

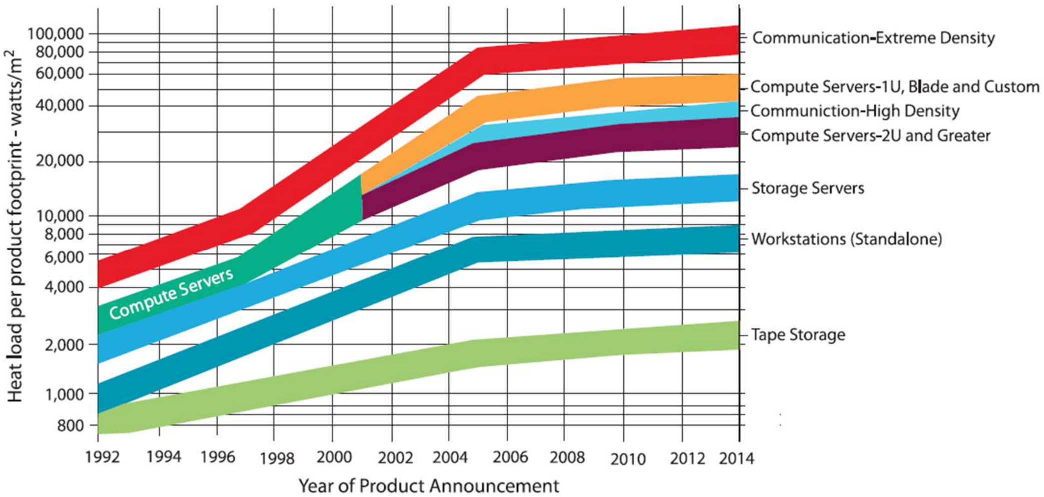


Figure 1-1 Data center equipment power trends [2]

Microelectronics forms the basis of most IT hardware, and in general, silicon devices reach their functional limitations in the 85 to 105°C range and experience permanent damage when operated at temperatures 15 to 25°C higher than that [3]. Therefore, depending on the power density of equipment installed, an operator may choose to use either air or liquid cooling. The suggested environmental operating ranges by the class of data center for both air and liquid cooling are shown in Figure 1-2 and Table 1-2 respectively [1].

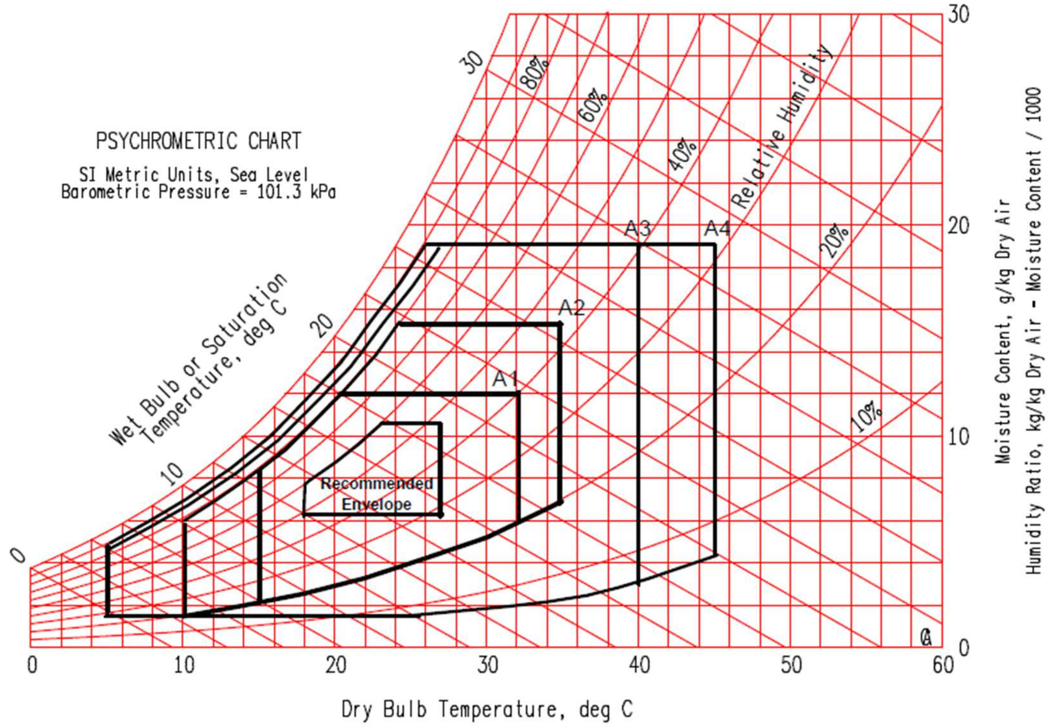


Figure 1-2 Environmental envelopes based on the class of data center [1]

Table 1-2 Guidelines for liquid cooled IT equipment [1]

| **Liquid Cooling Classes** | **Typical Infrastructure Design** | | **Facility Supply Water Temperature** |
|---|---|---|---|
| | **Main Cooling Equipment** | **Supplementary Cooling Equipment** | |
| W1 | Chiller / Cooling Tower | Water-side Economizer (cooling tower or dry cooler) | 2 – 17°C |
| W2 | | | 2 – 27°C |
| W3 | Cooling Tower | Chiller | 2 – 32°C |
| W4 | Water-side Economizer (cooling tower or dry cooler) | N/A | 2 – 45°C |
| W5 | Building Heating System | Cooling Tower | > 45°C |

## 1.2 Data Center Growth and Energy Consumption

The growth and dependence of global commerce, social interaction, news sources and other industries on information technology systems over the last decade has contributed to the rise of large data centers. These facilities are responsible for a significant portion of national and global energy consumption. Such are the implications that, it has been reported, the national energy usage by data centers more than doubled between 2000 and 2005 and it was projected that consumption would continue to rise over the course of the following five years [4]. By 2010, it was reported that data centers accounted for around 2% (between 1.7% and 2.2%) of the total national electricity consumption [5]. In the United States, this figure continues to grow with electricity usage increasing by 8.7% between and 2011 and 2012, while projected growth is expected to be around 9.8% over the next year. Electricity usage worldwide was reported to be around 1.8% or a corresponding power consumption of 322 TWh [6].

With this industry continuing to grow and strain the national electricity grid, there is a need to target energy efficiencies within the data center. Power Usage Effectiveness (PUE) is a common metric used to gauge energy-efficiency of operation of the facility.

$$PUE = \frac{Total\ Facility\ Power}{IT\ Equipment\ Power} \qquad (1\text{-}2)$$

A recent survey [7] reported that the average PUE is around 2.9 with only 20% of surveyed facilities recording a value of less than 2.0. Thus, majority of data centers in North America operate extremely inefficiently with IT equipment accounting for less than half the total power consumption (support infrastructure responsible for majority of the remaining). In addition, predictions indicate that power and cooling related expenditure is increasingly becoming a larger portion of a data center's Total Cost of Ownership (TCO) [8]. It is therefore imperative to reduce data center power consumption and operating cost by increasing the efficiencies of power distribution and cooling systems.

## 1.3 Targeting Efficiencies in Cooling Infrastructure

With a sizeable portion (around 30%) of typical data center power consumption attributed to cooling [9], which is categorized as a parasitic load, it has become vital that energy savings and efficiencies be pursued in these components at various levels within the data center facility. Typical cooling infrastructure in a traditional data center facility [10] and breakdown of corresponding energy consumption [11] are shown in Figure 1-3.

Many applications exist that help improve air cooling efficiency of legacy data centers such as aisle containment, blanking panels, chimneys, floor grommets, etc. Green field data centers possess an added benefit with easier implementation of economization which takes advantage of outdoor conditions to reduce chiller operating hours [12, 13]. A comprehensive review of techniques that can be utilized for energy-efficient operation of data centers can be found in [14].

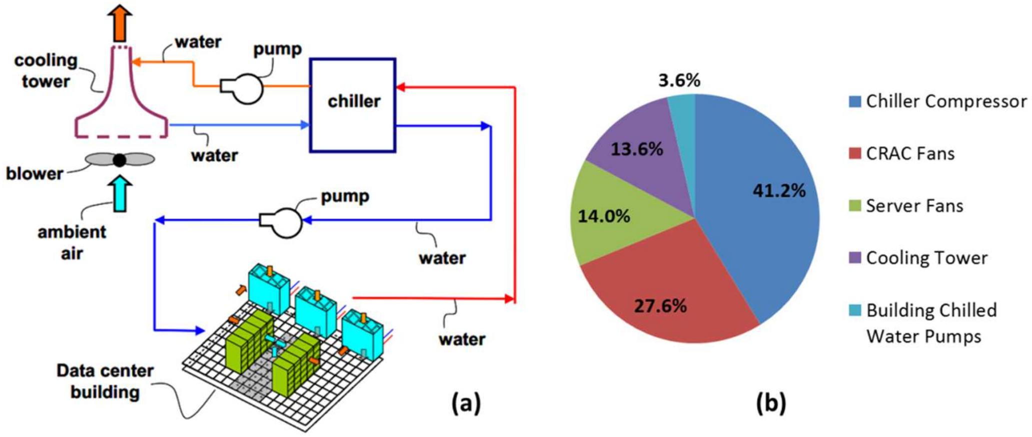


Figure 1-3 (a) Traditional data center cooling infrastructure [10] and (b) associated energy breakdown [11]

Research outlined in the current report primarily involves module, server, and rack-level cooling and solutions that target efficient operation by minimizing overall power consumption. At these levels the effect of operating temperature on power consumption is sizeable. The continuing advancement of microelectronics is driving down gate lengths by increasing the number of transistors on a given footprint of Silicon. Thus, as shown in Figure 1-4, high performance CMOS logic devices are becoming increasingly sensitive to operating temperature due to the effect of leakage current (static power) [15, 16]. Static power, as opposed to dynamic power responsible for switching of transistors, is a non-functional component of logic operation and therefore must be minimized.

In summary, the following trends in the data center and microelectronics industries serve as motivation for the current studies:

- Promotion of free cooling and/or economization systems may expose IT equipment to a wider range of operating temperatures that traditional facilities

- Continuing advancement in microelectronics technology has given rise to non-uniform power distribution at the device (module)
- Need to make cooling systems at every level within the facility more energy efficient

Thus, ***dynamic*** solutions that adjust the amount of cooling required for maintaining energy-efficient device operating temperatures are crucial at module, server, and rack-levels. Of equal importance is the need to understand the effect of these changes at higher levels within the data center facility.

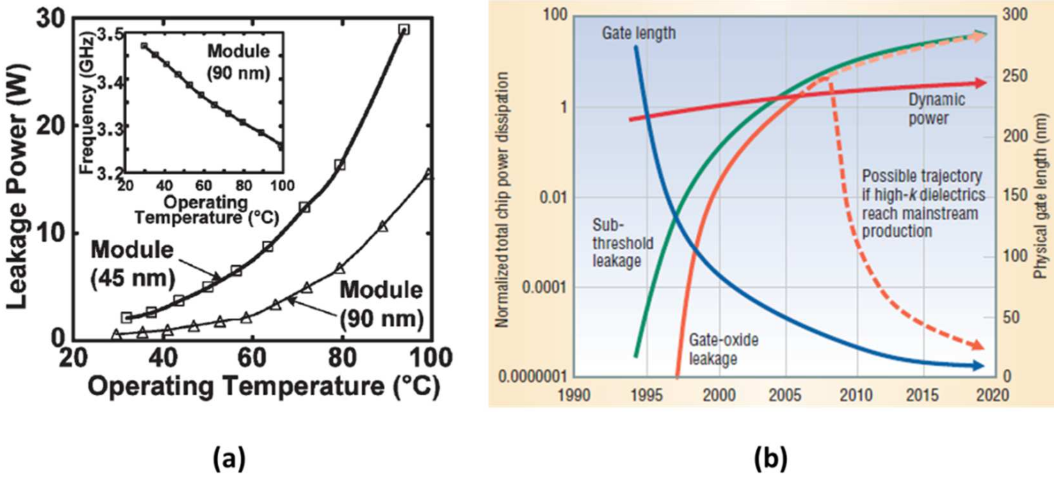


Figure 1-4 Variation of static (leakage) power versus (a) temperature [15] and (b) gate length [16]

## 1.4 Scope of the Work

### *1.4.1 Objectives*

The objectives of this dissertation are as follows:

- Module-level: Dynamic liquid cooling for high power devices

    - Propose the concept of a dynamic cold plate
    - Conduct extensive CFD analyses to evaluate the solution designed for a reference multi-chip module
    - Outline components and procedures for experimental testing of the dynamic cold plate
- Server-level: Energy-efficient operation of web servers
    - Test servers at different loads across a wide range of rack inlet temperatures (RITs)
        - Report effect of increased RIT on server power, cooling, efficiency, acoustics and reliability
    - Understand tradeoff between cooling power and device leakage current
        - Suggest improvements to the existing fan control scheme
- Rack-level: Effectiveness of rack-level fans
    - Outline a methodology for rack-level fan control
    - Report performance and savings available under a variety of test/operating conditions

### *1.4.2 Layout*

Work carried out to accomplish the aforementioned objectives is organized into six subsequent chapters. A comprehensive review of literature pertaining to history of and advancements in single-phase liquid cooling of high power modules, current techniques in dynamic thermal management and recent studies and observations made in raising coolant supply temperatures in data centers is presented in chapter 2. Specifics of the procedure followed for design and evaluation of a dynamic cold plate for a high power multi-chip module are presented in chapter 3. CFD analyses to establish appropriate flow distribution across sections of the cold plate as well as evaluate thermal performance when assembled with the MCM are previewed. Requirements for experimental testing of the proposed solution are explained in detail.

Chapter 4 provides a description of the web servers employed in both server and rack-level studies. Details of the setup and procedures followed for stressing these rackmount units in a test-bed data center are provided. Effects of operating web servers across the ASHRAE A4 envelope are outlined in chapter 5. The upper limit of RIT based on individual factors that, together, govern settings at the data center level is discussed. Continuing work in the preceding chapter, improvements in chassis fan control to extend the limit of RIT is outlined in chapter 6. Individual effects of device static power and fan power on total server power consumption are also discussed.

The extent of reduction in cooling power through deployment of larger, more-efficient fans at the rear of a stack of web servers is reported in chapter 7. Different factors that may determine which of the two rack-level configurations is deployed are also discussed.

# CHAPTER 2

# LITERATURE REVIEW

This section contains a review of single-phase liquid cooling using cold plates, dynamic thermal management and effects of raising supply temperatures in a data center.

## 2.1 Single-phase Liquid Cooling of High Power Modules

Liquid cooling has grown in acceptance for cooling novel, high-powered microelectronic devices [17]. Cold plates employing water as the coolant are one of the many prominent liquid cooling solutions available. One of the earliest applications of such a solution could be found in the IBM 3081 mainframe computer, announced in 1982, which incorporated water-cooled thermal conduction modules (TCMs) [18, 19]. The cold plate served to remove the heat dissipated by around a hundred chips mounted on the glass ceramic substrate, or a total module power of up to 2000W, within the body of the TCM [20, 21]. However, with the departure from bipolar devices and the introduction of CMOS technology in the early 1990's, air cooling was a more cost-effective method due to the significant reduction in heat dissipation. Despite the change in technology, module powers began to increase and, around a decade later, reached the levels of its bipolar counterpart and liquid cooling was once again being considered for thermal management of microelectronic devices [22].

Using water over air cooling for high power devices has multiple advantages like greater heat carrying capacity, targeted delivery and lower transport power. In addition, servers are more energy-efficient when operating at higher utilization (or higher power dissipation) [23] making liquids more appropriate for heat transfer while maintaining desired operating temperatures. A comprehensive review of cold plates employed for thermal management of high density servers can be found in [24]. Recent publications involve the use of cold plates for higher-temperature water cooling to enable substantial energy savings [10]. Application of CFD analysis to target improvements in existing cold plate designs is not uncommon. Fernandes et al. [25] reported a methodology for multi-design variable optimization of a water-cooled cold plate while employing user defined functions to fix the pumping power. Width and height of the serpentine channels in an IBM ES/9000 cold plate were varied with the objective of minimizing external thermal resistance. The numerical model was validated by comparison with published experimental data [26] and the predicted thermal resistance (baseline design) was found to be in excellent agreement. Fernandes et al. [27] previewed a methodology for rapid evaluation of a custom cold plate for cooling a high power MCM through alternating employment of CFD analysis and flow network modeling. This procedure permits effective parametric or optimization studies with minimal computational time and resources. Novel designs have also been previewed to extend performance of static cold plates. Remsburg [28] reported a solution that availed of impingement heat transfer along with non-linear fin patterns for thermal management of a 1080W IGBT/diode assembly. When compared to conventional designs, from formed tube to machined pin fins, the proposed solution reported visibly lower maximum temperatures for a series of fixed flow rates. This was attributed to a combination of improved heat transfer due to impingement and reduced pressure drop through optimized fin design.

It is to be noted that, in the aforementioned publications, all cold plates have a static design by nature and do not possess the capability to respond to non-uniform heat dissipation by modulating the cooling resource accordingly. In this report, the proposed unique “dynamic cold plate” concept aims to improve on conventional static designs through implementation of sensing and control, by redistributing the liquid across its body to counter varying power dissipation across the device being cooled.

## 2.2 Dynamic Thermal Management Techniques

Effective and energy-efficient thermal management may be implemented in different forms including, but not limited to, microarchitecture restructuring, workload distribution and prediction, and variable cooling. Karajgikar et al. [29] presented multi-objective optimization of the Pentium IV microarchitecture by repositioning functional units to minimize temperature with minimal performance penalty. From a thermal standpoint, it was advantageous to scatter high power units and locate them at the periphery. Coskun et al. [30] presented proactive temperature balancing, through dynamic job allocation, implemented on an UltraSPARC T1 processor. This predictive technique achieved significant reduction in temperature gradients and hot spots in comparison to reactive systems. In a following study [31], a proactive management technique was implemented to 3D chip architecture with variable flow liquid cooling. Results showed that integration of variable cooling decreased frequency of hot spots by an additional 20% over 75% reported through implementation of the predictive method. Coskun et al. [32] evaluated 3D stacked architectures with variable flow cooling to meet temperature limitation and job scheduling for balancing gradients. These implementations promoted energy-efficient operation by visibly reducing both cooling and overall energy consumption. In addition, multiple patents [33, 34, 35, 36] describing control systems and schemes for dynamic cooling have been published.

In the data center industry, dynamic cooling of CPUs using single-phase liquids is enabled by integrating low power, electromagnetically coupled pumps into cold plates such that the flow rate can be modulated with respect to the device's operating temperature. Many manufacturers offer such solutions in either closed-loop or rack-integrated formats [37, 38]. A more widely investigated aspect of (passive) dynamic cooling using liquids is two-phase solutions. Tuma [39] discussed the benefits of immersion (passive cooling) of servers in a two-phase dielectric fluid from an economic and environmental standpoint. This method, however, requires a redesign of a data center's layout and construction, and may not be ideal for most data center owners. Ohadi et al. [40] presented a study that compared air, liquid and two-phase cooling of CPUs in a data center environment. While both liquid and two-phase solutions have superior performance as compared to the air counterpart, the author conceded that a lack of industrial standards may slow the adoption of such technologies. The perception of cost is also an issue when considering liquid cooling.

In comparison, dynamic thermal management using air cooling is more predominant in the data center industry. A common feature at the server-level is speed control of chassis fans as a function of device temperature or power consumption [3]. Multiple existing techniques in dynamic thermal management by power and cooling control beyond the server-level will not be discussed or previewed as they do not factor in to the current study.

Regardless of whether air or liquid is used for heat transport, dynamic thermal management will assist in raising coolant supply temperature thereby reducing energy consumption at higher levels in the data center facility.

## 2.3 Raising the Supply Temperature

A current practice in the industry involves raising coolant supply temperatures to take advantage of outdoor conditions and maximize use of free cooling. As a result, ASHRAE once again expanded envelopes (see Figure 1-2 and Table 1-2) and introduced additional environmental classes (see Table 1-1) to encourage application of air and water-side economization.

Iyengar et al. [10] experimented on a 100% water cooled server rack, supported by an outdoor dry cooler, with server inlet water temperatures reaching as high as 45°C. The expected benefits were determined to be an equivalent 90% reduction in cooling energy as compared to conventional refrigeration based systems. Coles et al. [41] reported maximum liquid coolant supply temperatures based on two types of primary cooling equipment at 16 locations across the country. When adopting cooling towers, a maximum temperature of 32°C was found to be adequate for servers regardless of liquid or air cooling at the chip. Employing a dry cooler raised the maximum value to 43°C; however, only liquid cooling of CPUs could be employed with an adequate safety margin.

Through a simplified thermodynamic evaluation, Patterson [42] demonstrated the effect of data center temperature on energy efficiency. By increasing the ambient temperature by 10°C from a baseline of 20°C, efficiency of the chiller could be improved but overall power consumption went up due to leakage and fan effects at the server level. Application of an economizer as opposed to a chiller was recommended for reducing energy costs for warmer room setpoints in a data center. El-Sayed et al. [43] reported that higher operating temperatures have a smaller effect on system reliability as generally assumed. The effect of temperature (below 50°C) on disk failure rates showed a linear growth rate as compared to conventional exponential models and was dependent on neither age nor utilization. Increase in power consumption by higher ambient temperatures was mainly attributed to fan power consumption and more sophisticated algorithms were recommended.

Eiland et al. [44] investigated use of light mineral oil for direct liquid cooling of a web server oriented horizontally in a large acrylic tank. Promotion of high coolant supply temperatures at low flow rates was observed to be most efficient, predominantly due to reduction of oil viscosity with increasing temperature. Addagatla et al. [45] previewed complete liquid cooling of a server with active cold plates attached to CPUs and an internal, fan-assisted radiator for recirculated air-cooling of auxiliary components. CFD analysis was employed to design a duct tailored for recirculating internal air flow and performance of a prototype was tested experimentally. Coolant supply temperatures as high as 45°C were investigated and, at maximum load, cooling power was reported to be around 1.54% to 2.4% of IT power. Fernandes et al. [46] investigated a similar cooling configuration deployed at the rack-level with six servers. With a distributed pumping setup, at higher loads, cooling power was reported to be around 3% of IT power. Thus, at a larger scale, performance similar to a single server setup was observed. In addition, failing a single pump and fan in a server delivered less than 10% increase in component operating temperatures, with up to 50% reduction in cooling power. The configuration was deemed overprovisioned, with further efficiencies possible through improved design.

Few of the main concerns regarding higher operating temperatures are effects on server power consumption and equipment reliability. Strutt et al. [47] provided a review of established publications discussing the same. Intel [48] conducted a proof of concept test by comparing mechanical cooling and airside economization systems used for thermal management of identical 500ft$^2$ facilities with 900 production blades (200W/ft$^2$) over a period of 10 months. In the facility using economization, supply temperatures varied between 64°F and 92°F, humidity experienced rapid changes between 4% and 90%, and a layer of dust formed on IT equipment and room interior. It was reported that under server utilizations of around 90%, failure rates in the economization layout was around 0.6% higher than the DX counterpart. This was considered minimal. In addition, reduced operation of the DX system in the economized test bed reduced cooling power from 111.78kW (100% DX) to 28.6kW; or a 74% decrease. ASHRAE [1] has published data on estimated failure rates of volume servers under continuous operation for different air supply temperatures (see Figure 2-1). These values are reported assuming an X factor of 1 for continuous operation (24 X 7 X 365) at 20°C. In addition, discussions of factors that influence failure rates for specific components within IT equipment are available [49, 50, 51, 52, 53, 54].

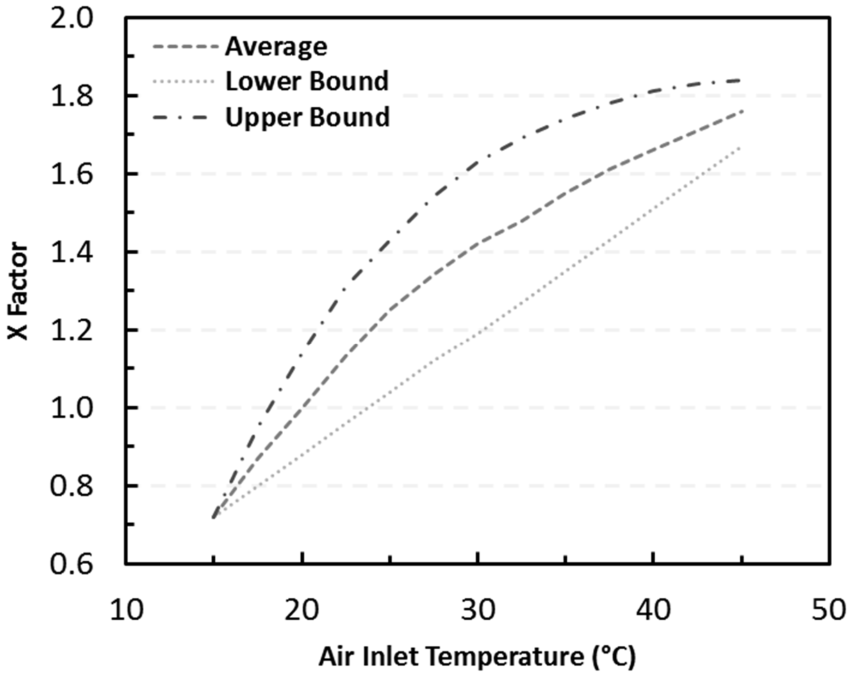


Figure 2-1 Relative failure rates for volume servers (X factors) at different inlet temperatures for continuous operation [1]

Thus, it can be concluded that substantial savings exist through increase in supply temperature and improved design and implementation of dynamic cooling solutions at lower levels within the facility.

# CHAPTER 3

# DYNAMIC COLD PLATE FOR A HIGH POWER MULTI-CHIP MODULE

The need for energy-efficiency in data centers has coincided with continuing trends of increasing microprocessor power densities and non-uniform temperature distributions, which pose a significant challenge to the cooling requirements of high power devices. Post-Pentium III era of microprocessors introduced non-uniform power distribution at the die with varying power densities assigned to different functional units. This gave rise to localized regions of high temperature known as 'hot spots' [55]. Thus, a substantially large temperature difference can be observed across the surface of a device which is detrimental to its performance and reliability. As a result, conventional static cooling solutions have to be designed to cool these high temperature regions which increase the thermal budget and in-turn cost of cooling these devices. In addition, rackmount servers are the most energy efficient when they operate close to maximum utilization [23]. Therefore, the primary requirements of next-generation solutions are high power cooling and selective distribution of resources for promotion of uniform device temperatures. This requires a detailed understanding and integration of microprocessor architecture, electronic packaging, and control systems to produce a robust solution.

## 3.1 The Concept

To meet aforementioned requirements, conventional static cold plate designs require integration of control devices and schemes to enable targeted delivery of cooling resources based on requirement. The layout and operation of the dynamic cold plate can be easily visualized as seen in Figure 3-1.

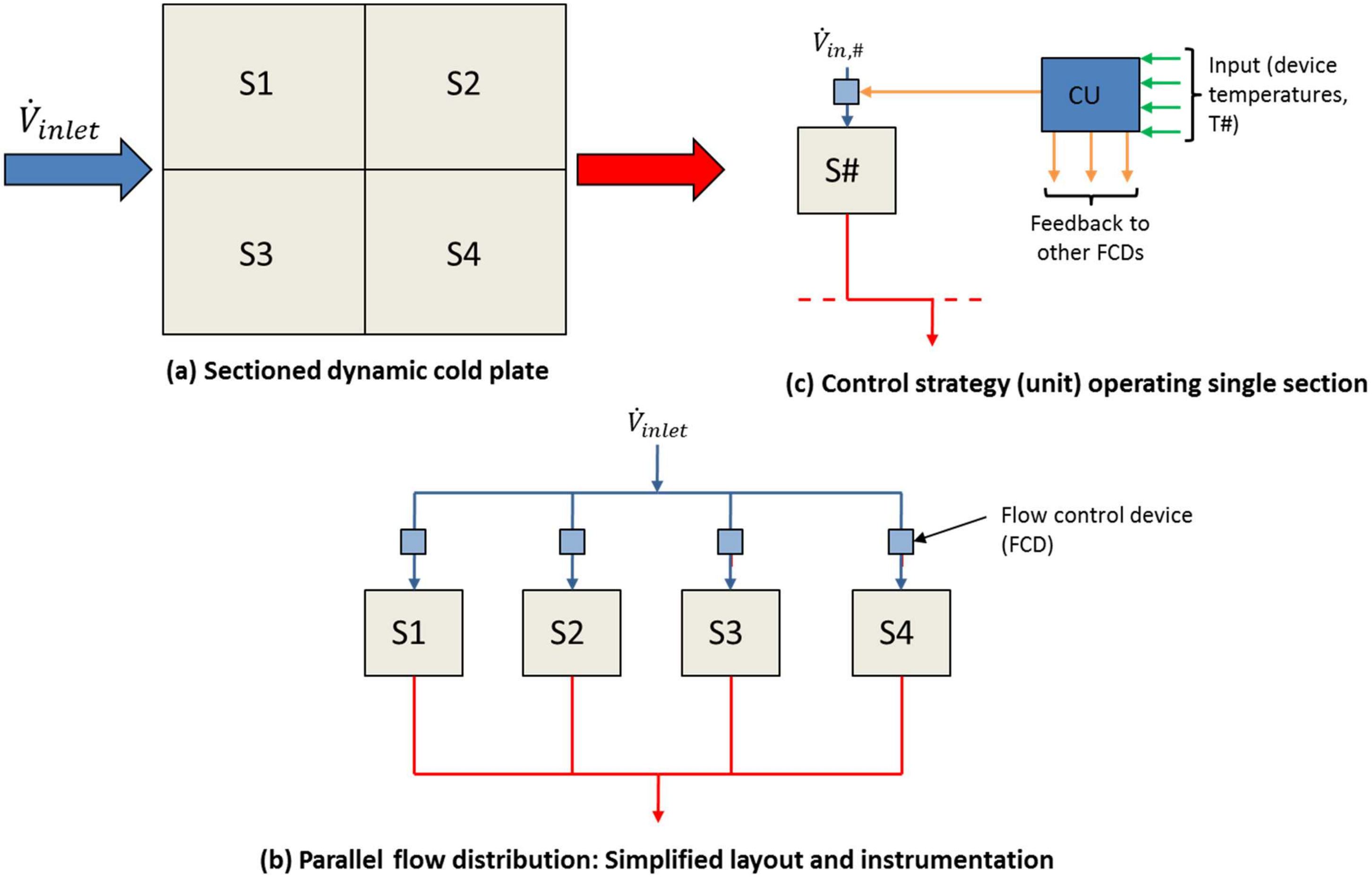
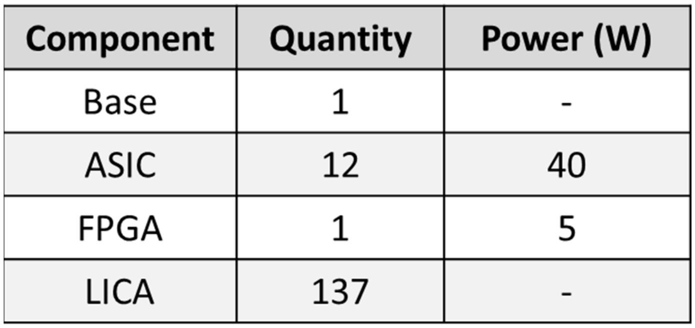

$\dot{V}_{inlet}$
S1
S2
S3
S4
(a) Sectioned dynamic cold plate
$\dot{V}_{in,\#}$
S#
CU
Input (device temperatures, T#)
Feedback to other FCDs
(c) Control strategy (unit) operating single section
$\dot{V}_{inlet}$
Flow control device (FCD)
S1
S2
S3
S4
(b) Parallel flow distribution: Simplified layout and instrumentation

Figure 3-1 Conceptual layout for a cold plate with segregated flow control

Depending on the complexity of the device and its power map, the footprint of the cold plate is divided into individual channel sections, as shown in Figure 3-1(a). Each section has a different inlet conduit (see Figure 3-1(b)) that is fed by the main point of ingress to the cold plate, making each independent from the others. In order to counter different power dissipations beneath the active region of each section, the introduction of sensing and control is necessary. By utilizing readings from temperature diodes integrated into the dies, or embedded thermocouples within the cold plate body, an indication of power dissipation variation may be established between different sections. These representative readings are then fed into a control unit (shown in Figure 3-1(c)) wherein a preset algorithm determines the magnitude/proportion of flow that needs to be distributed to each section based on cooling requirement. This signal is then fed into a FCD that regulates flow as per requirement. Thus flow to each section is controlled in real-time depending on representative temperature readings to promote lower temperature differences between individual sections. It is imperative that the proposed design be scalable to ensure application for high power devices with different footprints.

Table 3-1 Details of MCM components



| Component | Quantity | Power (W) |
|---|---|---|
| Base | 1 | - |
| ASIC | 12 | 40 |
| FPGA | 1 | 5 |
| LICA | 137 | - |

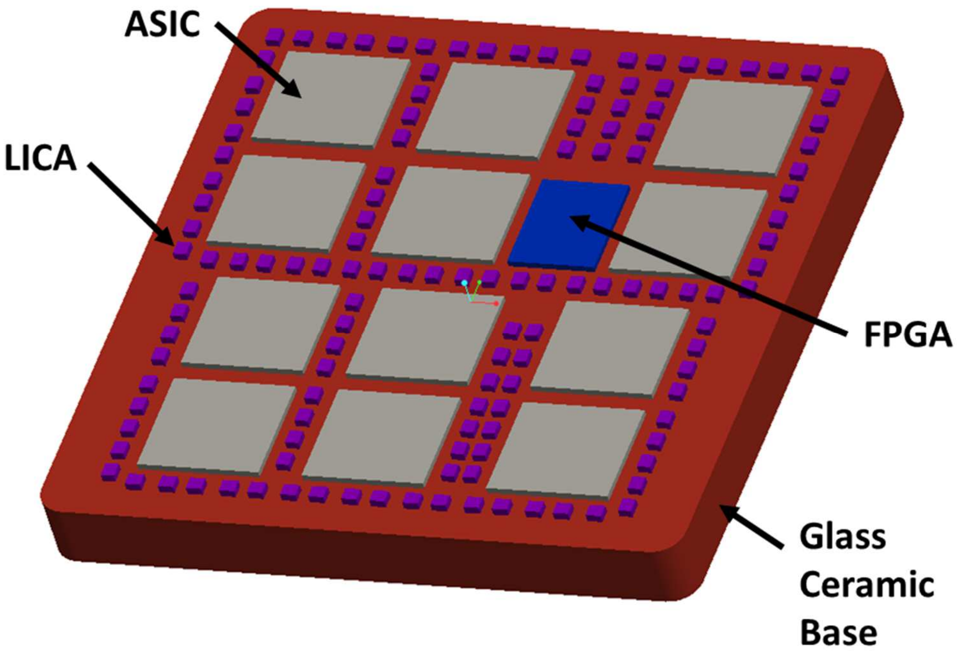


Figure 3-2 Multi-chip module and its components

## 3.2 Reference MCM Platform

In order to apply the preceding concept, a reference platform is required for design of such a dynamic solution. Figure 3-2 shows a high-power mutli-chip module (MCM), provided by Endicott Interconnect Technologies Inc. (now i3 Electronics, Inc. [56]), that serves this purpose. The module is populated with an array of surface mounted components (ASICs, LICAs and a FPGA) and setup to have a maximum power dissipation of 485 watts over a 78mm by 78mm footprint. A list of component power specifications can be found in Table 3-1. It is apparent that the only heat generating components of interest are the ASICs (application-specific integrated circuit) and FPGA (field-programmable gate array). The former is 14.71mm x 13.31mm x 0.8mm in size and the latter is 10.50mm x 12.70mm x 0.8mm in dimension. The LICAs, while much greater in number, do not dissipate (significant) heat and are therefore disregarded for the purpose of this study. A copper heat spreader, designed to account for disparity in component heights and spreading of heat, is not considered in this study to take advantage of multiple spatially-separated heat sources on the module.

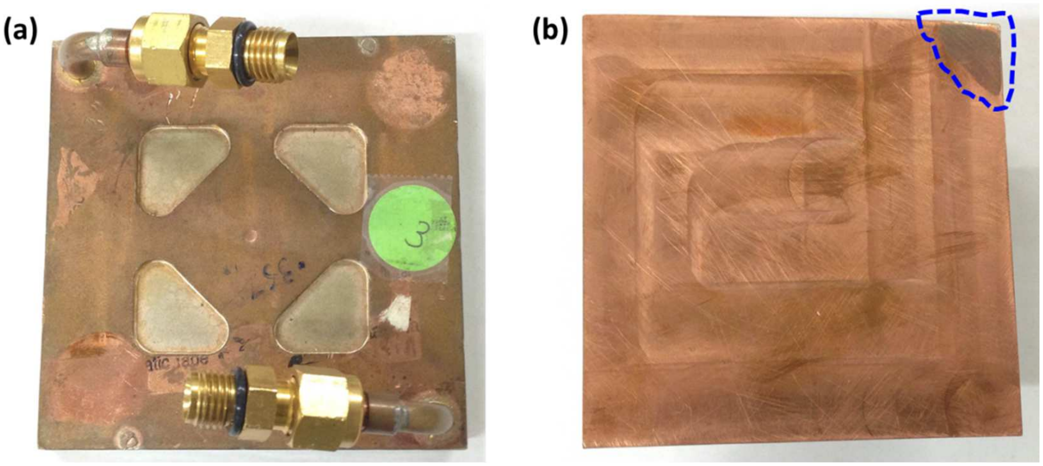


Figure 3-3 (a) Top and (b) bottom views of original cold plate

Table 3-2 Specifications of the original cold plate

| Parameter | Value |
|---|---|
| Size | 92mm × 92mm × 12mm |
| Thermal Resistance | 0.0152°C/W |
| Flow Rate | 2lpm (3.33E-5$m^3$/s) |
| Pressure Drop | 2.5psi (17236.89Pa) |
| Pumping Power | 0.575W |

### *3.2.1 Original cold plate*

The MCM, being a product in service, was assembled with and thermally managed by a static cold plate. Top and bottom views of this solution are shown in Figure

3-3. The cold plate has a copper body measuring 92mm x 92mm x 12mm (excluding ports) with multiple layers assembled by brazing to prevent leaks. Single points of ingress and egress of coolant are setup in this design, similar to the proposed concept. SAE/JIC fittings at the inlet and outlet facilitate connection at the system-level (see Figure 3-3(a)). Internal circuiting of coolant within the body of the cold plate is unknown. Experimental testing should provide insight whether the design is based on serpentine channeling, parallel networks or otherwise. The base of the cold plate was milled to 0.002” planarity to prevent detrimental performance of thermal interface material (TIM) during testing. This operation was implemented to counter surface non-uniformity that may have developed over time as seen in Figure 3-4. A small portion of the base, highlighted by the dotted blue line in Figure 3-3(b), was left unaffected to prevent significant removal of copper and possible leaks. However, this section would not come in contact with components and is left unaltered.

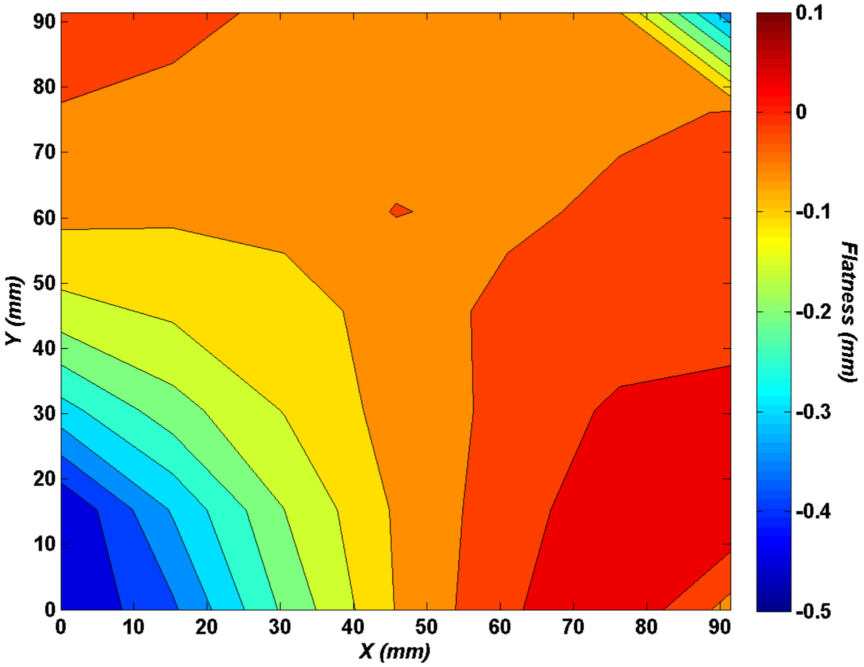


Figure 3-4 Bottom surface non-uniformity of original cold plate

## 3.3 Requirements of Manufacturing

Irrespective of air or liquid, heat sinks with effective heat transfer surfaces require less pumping power to provide required cooling. Cold plates with fins manufactured using micro deformation technology (MDT) have been reported to deliver lower thermal resistance per pressure drop as compared to conventional straight fins [57]. MDT is a non-subtractive method of machining employed by Wolverine Tube, Inc. [58] as seen in Figure 3-5. Straight fins fabricated using this method form heat transfer surfaces of the dynamic cold plate. Constraints associated with MDT are listed in Table 3-2.

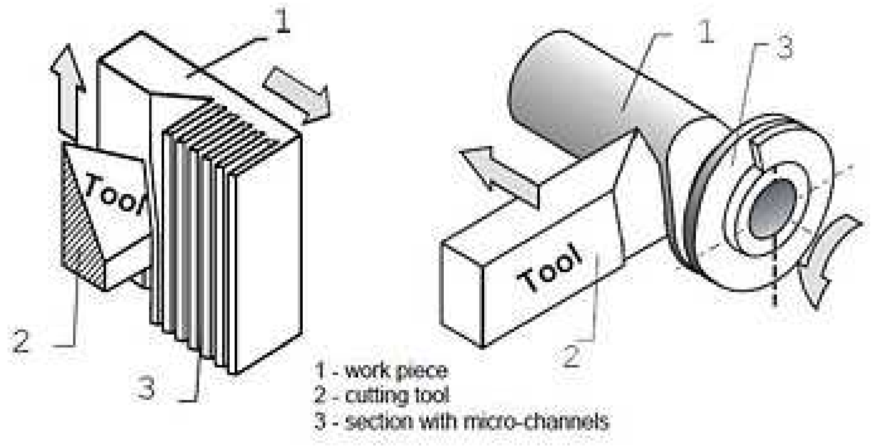
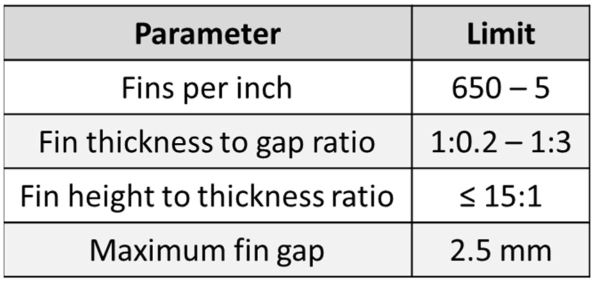


Figure 3-5 Simplified depiction of manufacturing using MDT [58]

Table 3-3 MDT manufacturing constraints [58]



| Parameter | Limit |
|---|---|
| Fins per inch | 650 – 5 |
| Fin thickness to gap ratio | 1:0.2 – 1:3 |
| Fin height to thickness ratio | ≤ 15:1 |
| Maximum fin gap | 2.5 mm |

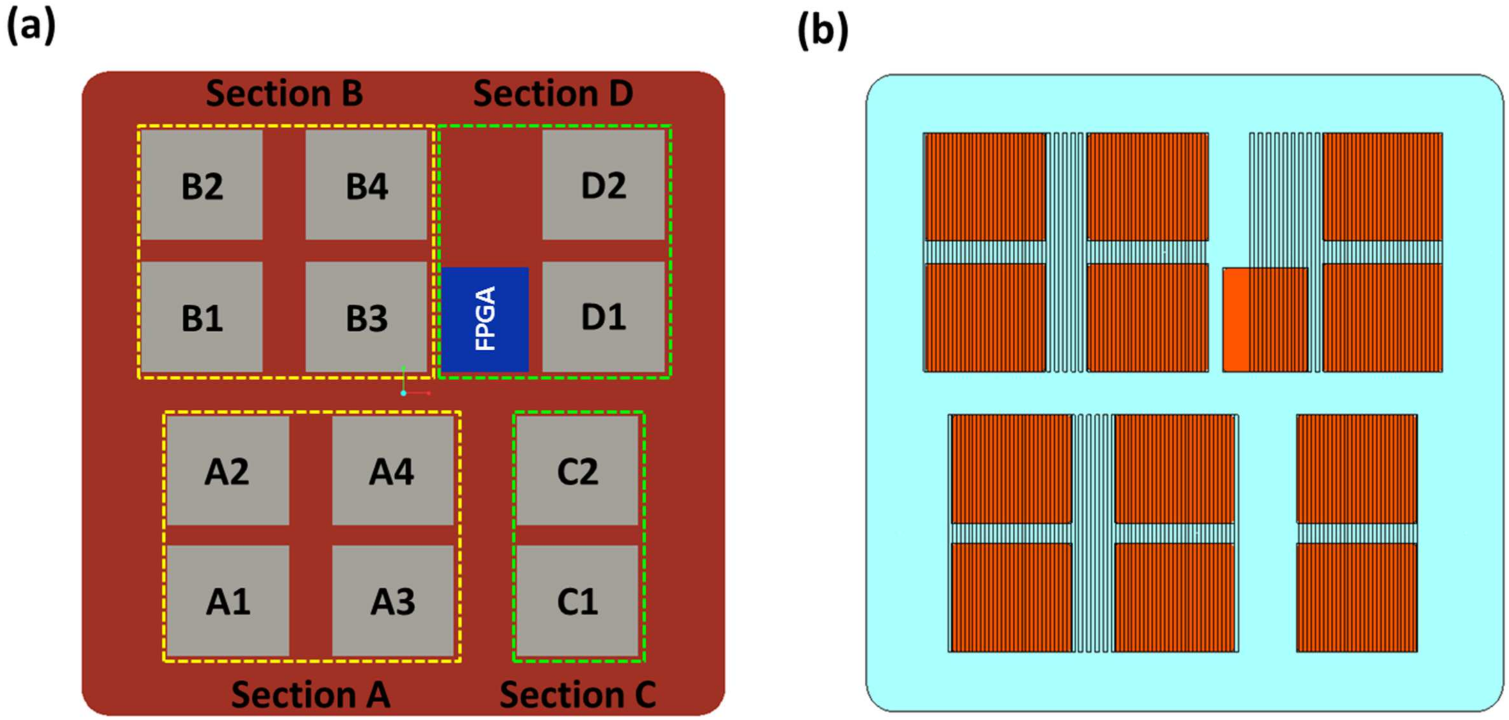


Figure 3-6 Sectioning of (a) components based on proximity and (b) corresponding heat transfer surfaces of proposed cold plate

### 3.3.1 Setting zones

As described in section 3.1, sectioning of the solution based on local heat dissipation forms that basis of the dynamic solution. Heat generating components of the MCM (13 in total) were grouped into four distinct regions based on proximity as seen in Figure 3-6(a). These clusters, henceforth referred to as sections, were named A to D and individual components were numbered starting from the bottom left corner and moving in an ‘N’ pattern. Sections A and B have a similar footprint and account for maximum power dissipation of 160W. Section D is dissimilar to C as it accommodates the FPGA. Flow of coolant to each section must be independent and, therefore, have isolated islands of fins. This is shown in Figure 3-6(b). To improve cost effectiveness of manufacturing the base plate, fins were fabricated from a leading edge before a fly cutting operation created the aforementioned islands. As a result, in some sections, fins extended beyond the span of components. In section D, FPGA area was partially accounted for. This was to provide space for walls to isolate flow from section A. In addition, this component dissipates a relatively lower power of 5W. The envisioned cold plate design has a footprint of 90mm x 90mm.

## 3.4 CFD Analysis for Cold Plate Design

Before samples could be fabricated for experimental testing, detailed CFD analysis was imperative to determine if flow distribution and thermal performance of selected heat transfer surfaces were per requirement. An acrylic cover was assembled above the base plate described in section 3.3.1 and had portions machined away to permit fluid flow through each individual section. Figure 3-7 shows details of circuiting employed to deliver coolant to fins in section B and facilitate exit of transferred heat. The inlet (labeled 1), threaded to facilitate a barbed fitting, introduces coolant at a fixed flow rate to the section. The fluid then travels in the transverse direction through a rib (labeled 2) before impinging at the center of each channel between adjacent fins. Trapezoidal design of this part helps promote uniform flow distribution across all channels (labeled 3). Common conduits at opposite ends of each channel direct flow to the outlet (labeled 4). Varying dimensions of these components were constrained (width of conduits) due to the close proximity of sections as well as minimum thickness of isolating walls.

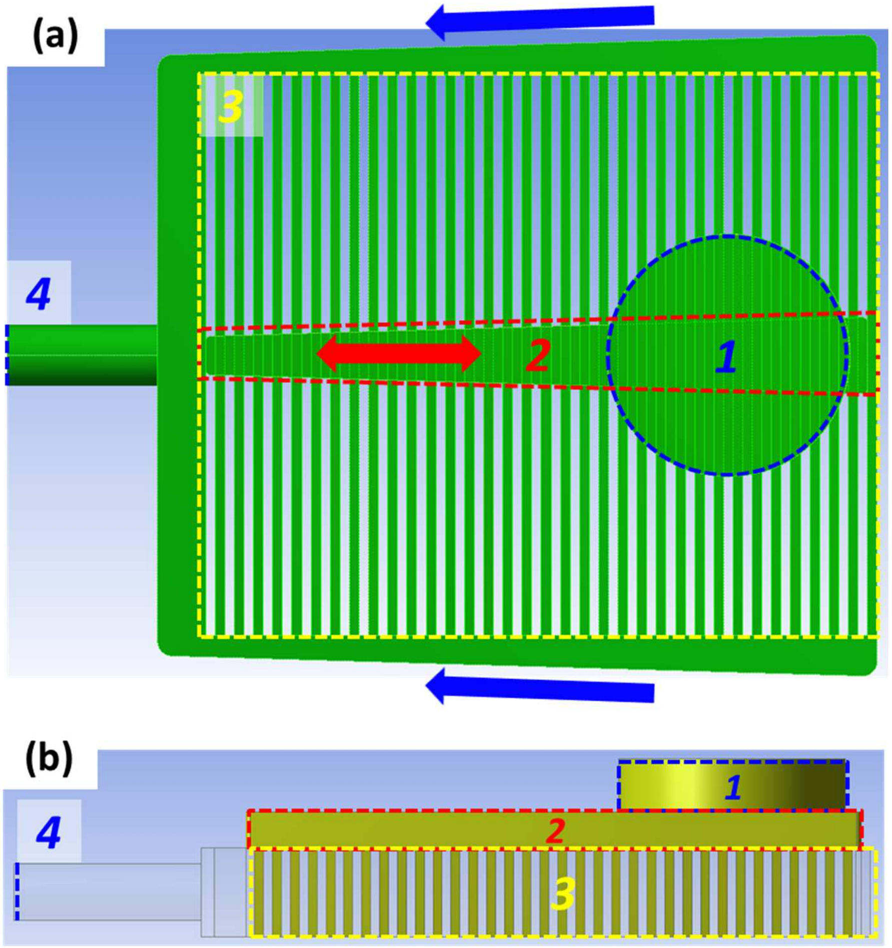


Figure 3-7 Circuiting of fluid flow for uniform distribution (section B) shown in (a) top and (b) front views

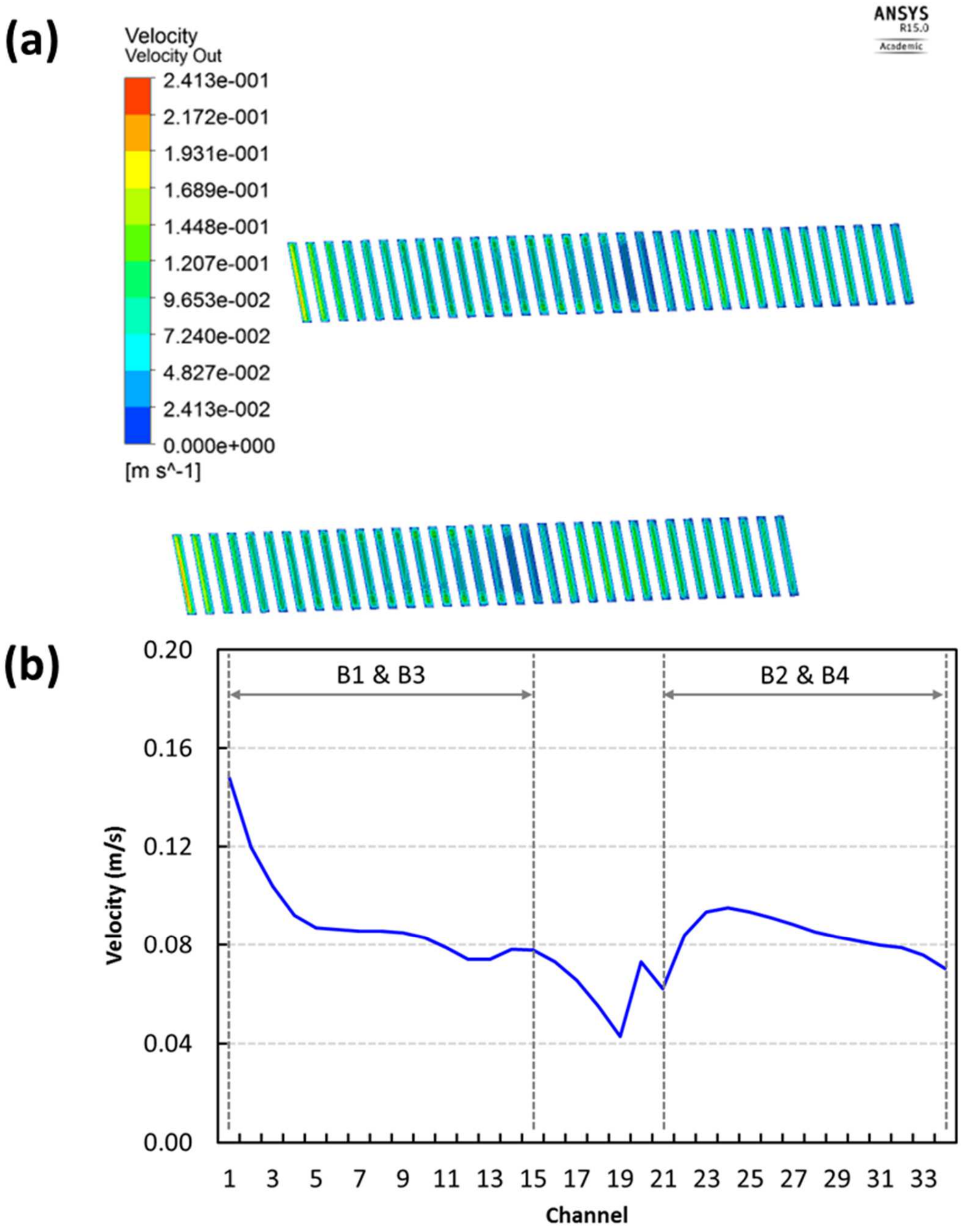


Figure 3-8 (a) Velocity contours at exits and (b) distribution of average velocity at the exit of each channel

### *3.4.1 Analyzing flow distribution within a section*

Since each section was designed as described above, results from analysis of region B could be applied to the remaining three. ANSYS FLUENT [59] and Workbench [60] were employed to analyze flow distribution across individual channels. Lu et al. [61] showed that increasing the number of channels (in parallel) reduces flow maldistribution. Based on constraints of MDT fabrication, 0.5mm thick plate fins (suggested minimum) with 1mm pitch were employed. The objective of this analysis was to counter the asymmetric locations of the inlet and outlet by varying the transverse location of the inlet, and dimensions of the trapezoidal rib (width of long and short parallel sides) and conduits (width at wider end) to promote uniform flow distribution across all channels (34 in section B). Cut-cell Cartesian method of meshing was employed to create predominantly flow-aligned elements. The solution became grid independent with around 1.6 million elements and a maximum skewness of 0.946. This corresponded to 4 cells across each channel. An inlet flow rate of 0.5lpm was specified, and the laminar viscous model and standard pressure discretization scheme were employed. The design depicted in Figure 3-7, one of many analyzed variations, delivered suitable flow distribution as seen in Figure 3-8. Average velocity in fins adjacent to regions of heat dissipation was fairly uniform (see Figure 3-8(b)) with a coefficient of variation of 0.18. Circuiting of sections A, C and D were setup similar to region B.

### *3.4.2 Thermal performance of cold plate with assembled MCM*

The proposed cold plate, complete with the acrylic cover, is shown in Figure 3-9(a). Thermal performance of the solution, assembled with the MCM, over a range of flow rates was studied using 6SigmaET [62]. Geometry of the assembly used in this analysis is depicted in Figure 3-9(b).

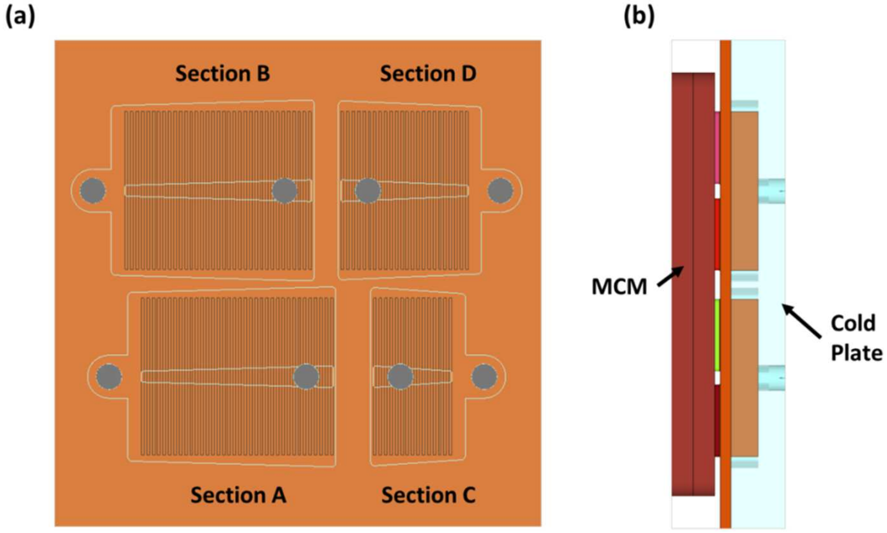


Figure 3-9 (a) Top and (b) left views of MCM assembly

Thermal grease was accounted for between the top surface of each component and the cold plate base. Each TIM layer had a thickness of 50um and a bulk thermal conductivity of 3.45W/m-K, assuming an interfacial pressure of 75psi. Cartesian mesh with a cell count of around 16 million was found to be adequate to ensure that results were grid independent. Grid control objects (non-conformal meshing) were employed to focus regions of high grid density in the acrylic cover (domain of fluid flow). Results are outlined in Figure 3-10. Sections have similar impedance curves owing to the design, with regions A and B provided relatively higher flow to account for greater heat dissipation. Range of flow rates considered were based on the rule-of-thumb that states, for efficient indirect liquid cooling, around 2kW to 3kW of heat should be dissipated by 1gpm of coolant flow. Using the available impedance curves and predicting the pressure drops for a total flow rate of 2lpm (summation of all sections) gives at total pumping power of 0.05W. This prediction is an order of magnitude lower than the quantity reported in Table 3-2. Thus, expectations are that savings in flow power with respect to the original solution may be available through deployment of the proposed cold plate.

Figure 3-10(b) reports variation of heater temperatures (components in each section with maximum values) with coolant flow rates. The observed range of operating temperatures (58°C to 66°C) for sections A, B and D were sufficiently close. This is an indication of effective design. Relatively lower heater temperatures in section C were a result of a thinner section with flow from the inlet impinging near the transverse center of both devices. In addition, reported temperatures were 50% lower than derating temperatures of the heaters (150°C) [63]. A fabricated sample of the finalized cold plate design is shown in Figure A.1.

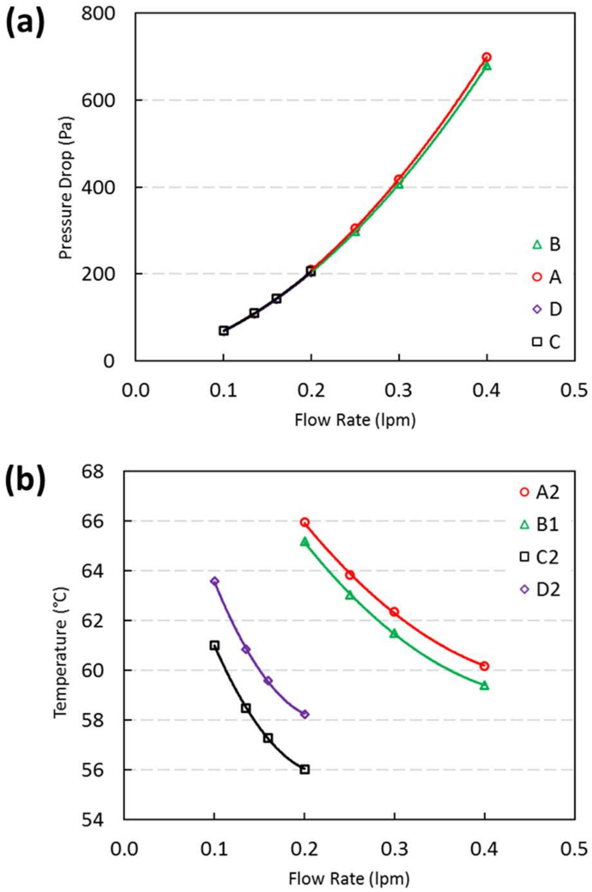


Figure 3-10 (a) Pressure drop in different sections and (b) maximum heater temperatures in each section versus coolant flow rates

## 3.5 Preparing for Experimental Testing

With both original and proposed cold plates available for testing, and electromagnetically commutated pumps used as external FCDs, details of the control system needed to be finalized. In addition, since a functioning MCM was not available and thermal test vehicles (TTV) are monetarily unfeasible, a mock MCM TTV was designed and fabricated for experimental testing. Details of a liquid cooling test bench are also finalized and previewed. Finally, series of tests to be conducted on both cold plates and resultant parameters of interest are outlined.

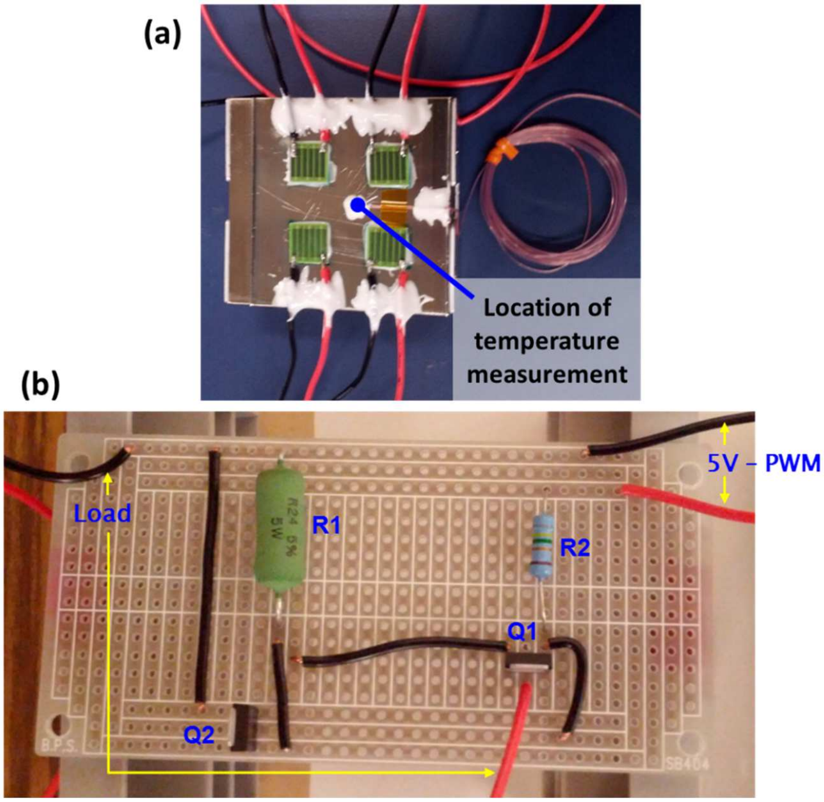


Figure 3-11 (a) Bottom surface of heat sink showing epoxied heaters, leads and thermocouple; (b) heater control circuit and its components

### *3.5.1 Initial study to evaluate the control system*

In the previewed test setup, the control system is not only responsible for modulating the FCDs based on temperature measurements but also moderating heat dissipation in the ASICs (FPGA power does not change). Thus, testing can be sufficiently automated saving both resources and time while simplifying data reduction. A simple study was conducted to confirm concepts and components outlined for the control system. To imitate the eventual system, four thick film heaters were epoxied to the base of an aluminum plate fin heat sink as seen in Figure 3-11(a). A thermocouple was epoxied at the center of the heated region to measure base temperature. Leads from each resistor were setup such that all four heaters can be powered by a single source. A 4-wire 60mm DC fan was attached on top of the heat sink to provide cooling. Figure 3-11(b) shows a current control circuit used to modulate heater power dissipation through a PWM signal. The load (thick film resistors in parallel) was connected as indicated and the PWM signal was generated by an Arduino microcontroller board [64]. BJT (bipolar junction transistor) Q1 limited the collector (or load) current based on the PWM signal applied at the base. If current beyond a certain limit was permitted to flow, the voltage drop across resistor R1 would trip the base of Q2, which limits the base current of Q1. This, in turn, lowered current through the load. Figure A.2 shows all components used in this study. A LabVIEW [65] code read the heat sink base temperature through a data logger and programed the Arduino to send PWM signals from digital I/O pins. Inputs to the code were specifications of deadband control of thermocouple readings (high and low temperatures; rate of increase and decrease of fan duty cycles) and heater trace file (variation of PWM with time). A screenshot of the front panel can be found in Figure A.3. Figure 3-12 shows specifics of a short test run to evaluate this setup.

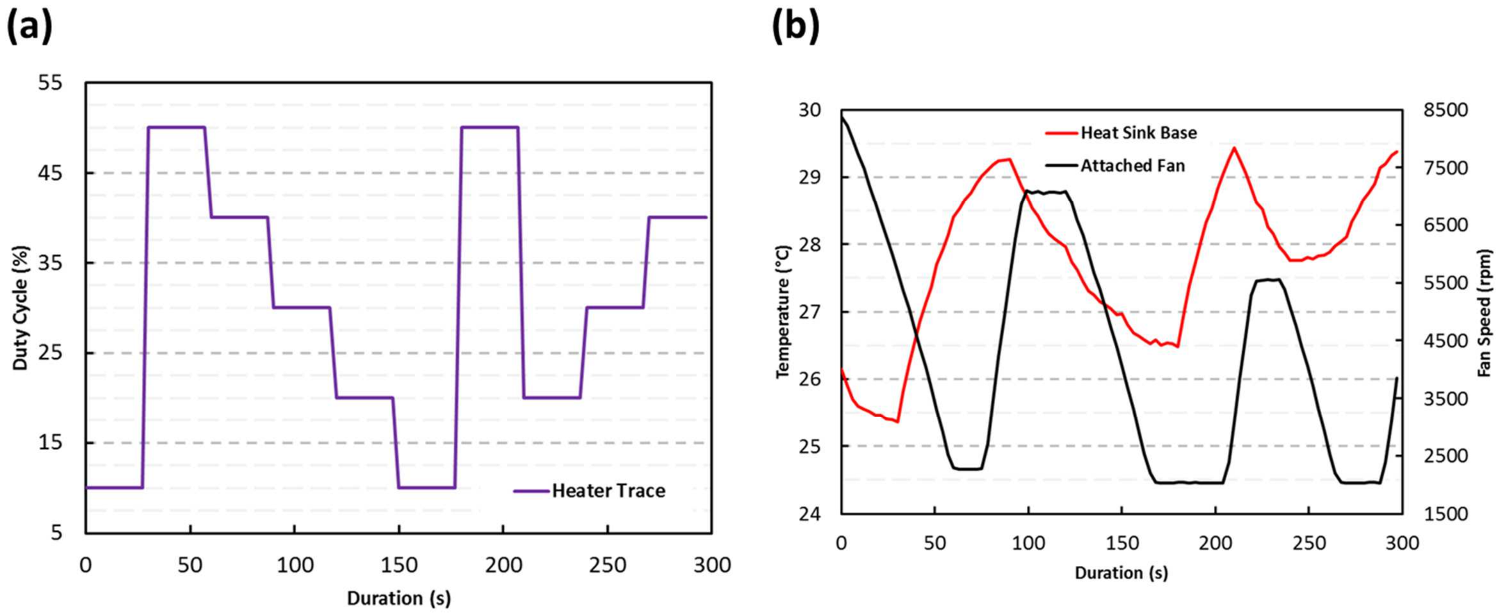


Figure 3-12 (a) Heater duty cycle and (b) corresponding heat sink base temperature and fan speeds

Heater power dissipation was controlled by varying the duty cycle as seen in Figure 3-12(a). The deadband was setup to maintain temperature between 28°C and 29°C by increasing or decreasing fan duty cycle by 10% per second or 5% per second respectively. Effects of these settings are seen in Figure 3-12(b). It was observed that the measured temperature oscillates as the range of fan speed between dead zones continues to shrink. At the end of the run, variation of heat sink temperature was practically limited to the span of the deadband. Thus, a simple test setup was proven to be capable of modulating both power and cooling resources as intended.

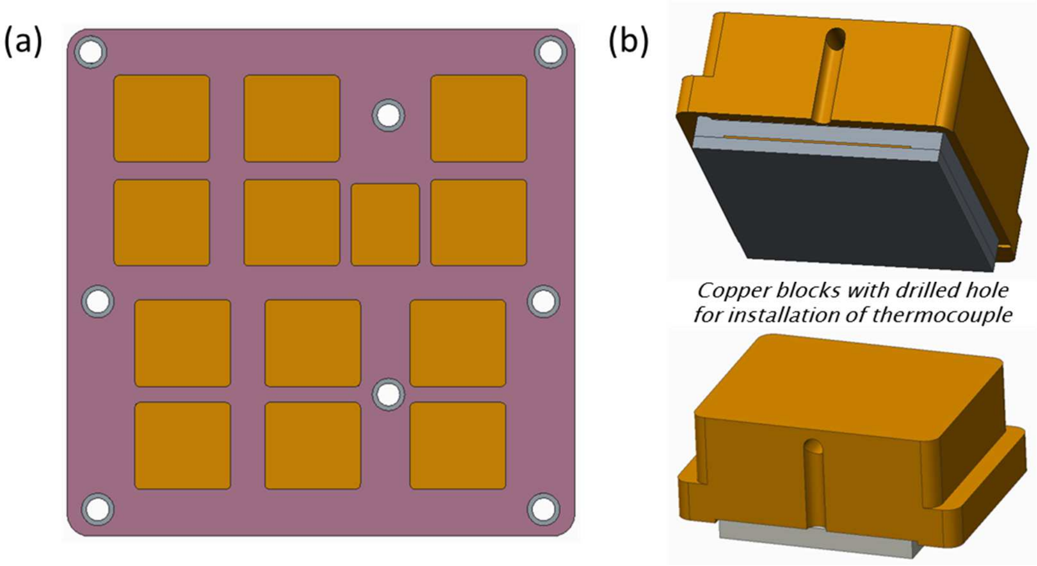

(a)
(b)
Copper blocks with drilled hole
for installation of thermocouple

Figure 3-13 (a) Assembled MCM TTV and (b) copper blocks with resistive heaters simulating ASICs and FPGA

### *3.5.2 The MCM TTV*

A thermal test vehicle, dimensionally identical to the MCM as well as capable of independent power dissipation within heat generating components is required for experimental testing. Figure 3-13 shows details of the mock MCM designed to serve this purpose. Thick film heaters (12.7mm x 12.7mm) [63] soldered to the base of copper blocks (accurate footprints) simulate the ASICs and FPGA. A hole was drilled midway into the side of each block such that the attached thermocouple measures temperature 1mm below the center of the top surface. These measurements are intended to represent case temperatures of ASICs and the FPGA. The base of the MCM was machined from acrylic and has two parts. The top half ensures that installed copper blocks only extend 0.8mm above the base (similar to component heights). The bottom half prevents blocks from moving when either cold plate is attached with applied pressure. In addition, holes in the bottom half of the base facilitate exit of heater leads and thermocouples from each individual block.

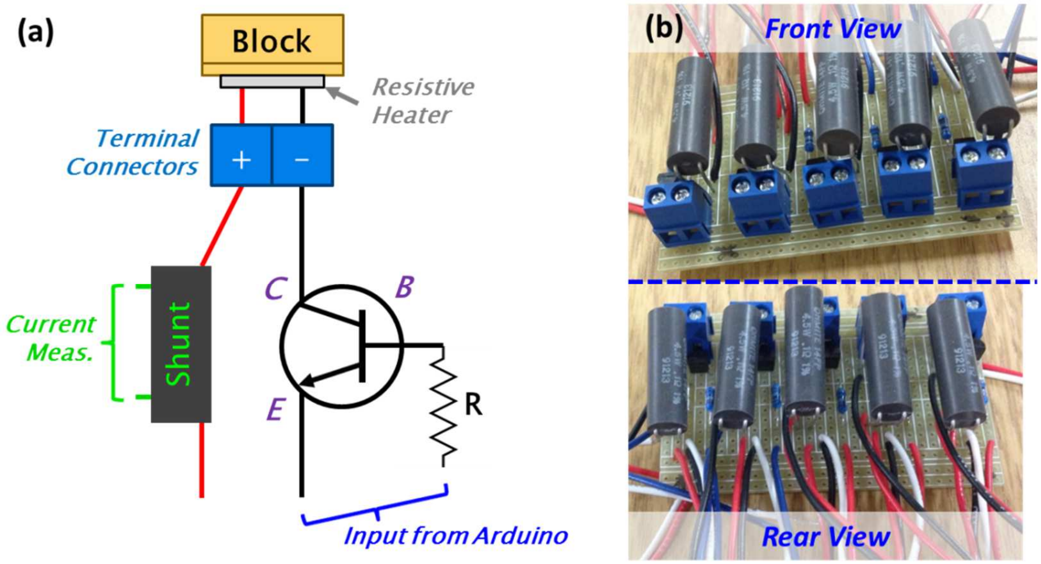


Figure 3-14 (a) Simplified depiction of a single heater control circuit and, (b) an assembled board with components for controlling five heaters (sections C and D)

Controlling power dissipation of each heater is achieved similar to details in section 3.5.1. Figure 3-14(b) shows a prototype board with assembled components designed to control five heaters (sections C and D). Simplified depiction of a single circuit is seen in Figure 3-14(a). A shunt resistor measures current drawn by the heater based on voltage drop readings made by the data acquisition (DAQ) unit. In addition, an NPN BJT modulates current flowing through the circuit based on a PWM signal from the microcontroller board. Additional leads at the terminal connectors measure the voltage across the resistive heater for power measurement. A board similar to the one pictured in Figure 3-14(b) with eight individual circuits was assembled for blocks in sections A and B.

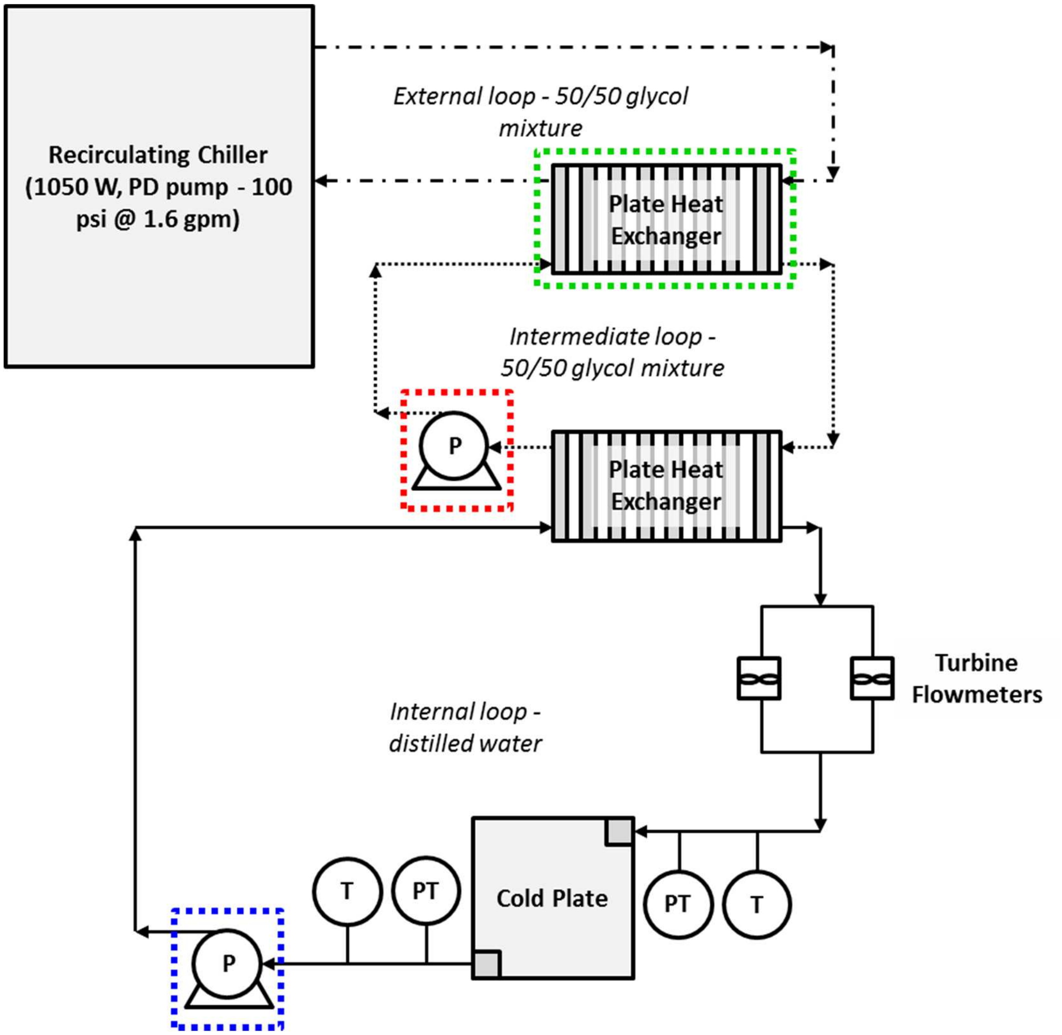

Recirculating Chiller (1050 W, PD pump - 100 psi @ 1.6 gpm)
External loop - 50/50 glycol mixture
Plate Heat Exchanger
Intermediate loop - 50/50 glycol mixture
P
Plate Heat Exchanger
Turbine Flowmeters
Internal loop - distilled water
T
PT
Cold Plate
PT
T
P

Figure 3-15 Coolant circuit for testing both original and dynamic cold plates

### *3.5.3 Liquid cooling test bench*

A simplified depiction of the test bench setup to evaluate both cold plates is depicted in Figure 3-15. A Kinetics RS33AO11 recirculating chiller drives flow through the external loop and cools the plate heat exchanger (HEX) highlighted by the dotted green line. The chiller is equipped with a positive displacement pump capable of pumping up to 1.6gpm of coolant at 100psi and a temperature range of -15°C to 75°C. As these units are known to drift (variation in temperature), the HEX provides substantial thermal capacitance to prevent transmission to subsequent loops. A DC 4-wire pump [66], highlighted by the red dotted line, drives flow through the two HEXs in the intermediate loop. Similarly, a separate pump (highlighted by the blue dotted line) controls flow of distilled water through all components in the internal loop. Turbine flowmeters arranged in parallel measure the flow rate of water cooled by the plate HEX. Temperature and pressure differences across the cold plate are measured using K-type thermocouple probes and pressure transducers. The pump in the internal loop is primarily responsible for maintaining a fixed flow rate during testing. The pump in the intermediate loop controls the inlet temperature of water to the cold plate by modulating flow rate between the two heat exchangers. Temperature and flow rate readings are input to the LabVIEW code that in-turn controls both pumps. Inlet temperatures to the cold plate as low as 15°C are targeted during testing. Water and glycol mixture (50/50) is employed in both intermediate and external loops to enable near zero temperatures to account for heat loss.

### *3.5.4 Test matrix*

The MCM TTV is subjected to different loads when testing both original (OCP) and dynamic cold plates (DCP). Note that FPGA power dissipation does not change

(5W). Five different cases of uniform loading are considered. These are when all ASICs dissipate 5W, 10W, 20W, 30W and 40W respectively. In contrast, twelve different cases of non-uniform loading are outlined. Herein, a single ASIC is set to dissipate 40W alone while all remaining units idle at 5W each. Thus, a total of seventeen different MCM TTV loads are cycled during a single cooling condition (fixed flow rate and inlet temperature to cold plate).

Number of cases in the cooling setup varies based on which cold plate is being tested. For the original solution, four different inlet temperatures are considered (15°C, 25°C, 35°C and 45°C). In addition, four different flow rates through the internal loop are studied (1lpm to 4lpm in increments of 1lpm). Thus, a total of sixteen unique cooling conditions are investigated while testing the original cold plate. Primary results of interest are device temperatures as a function of pumping power for different water inlet temperatures.

Results of OCP testing are used to determine a maximum device temperature. Individual pumps (FCDs) servicing each section modulate flow to achieve said temperature using deadband control. Therefore, the only variable in the cooling setup is water temperature at the inlet to the DCP (15°C to 45°C in increments of 10°C). Variation of device temperatures with individual section flow rates for different water inlet temperatures are obtained from these tests. Measured flow rates can be converted into pumping powers by determining the impedance curve for each section (separate tests).

### 3.6 Future Work

With the MCM TTV and control circuits assembled, and components of the liquid cooling bench purchased, performance evaluation of both cold plates is imminent. The cooling loops are to be setup as specified and a LabVIEW code automating the entire procedure must be generated. Savings in either pumping power or inlet temperature through application of the dynamic cold plate (when compared to the original solution) for thermal management of a high-power multi-chip module will be reported.

# CHAPTER 4

# EXPERIMENTAL SETUP AND PROCEDURES FOR EVALUATING WEB SERVERS

## 4.1 Servers Under Study

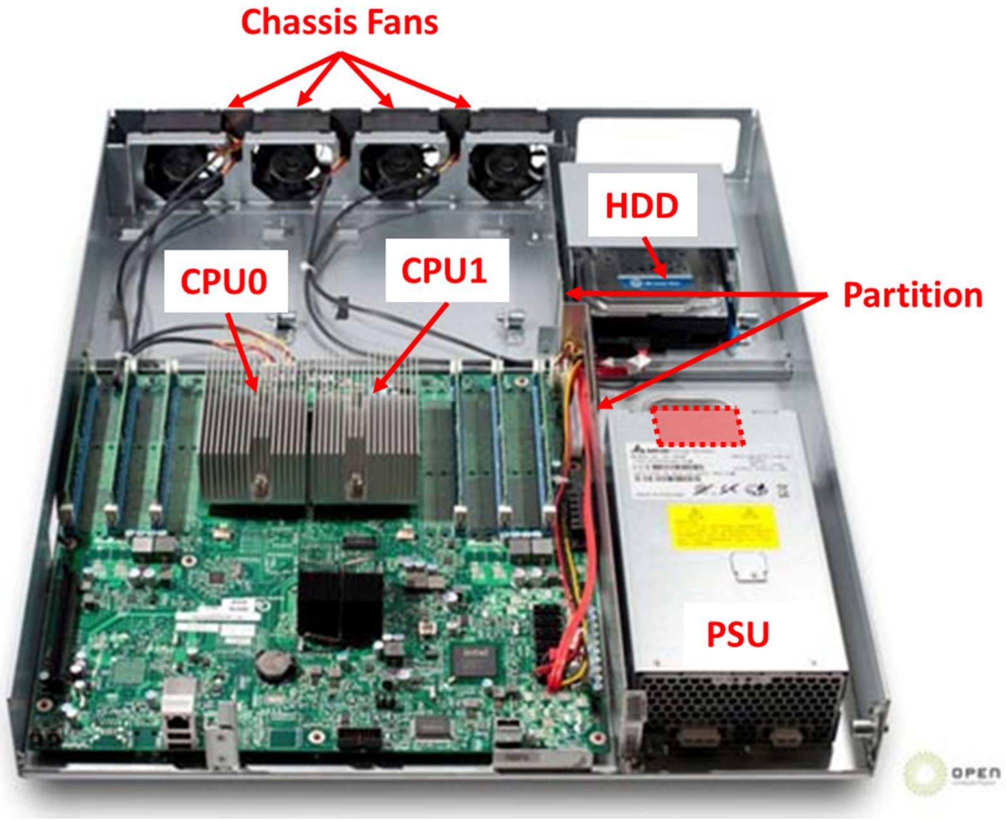


Figure 4-1 Intel-based Open Compute server with locations of components of interest (air duct removed for visual purposes) [67]

Figure 4-1 shows an Intel-based Open Compute server [67] used in this study. This 1.5U (1U = 1.75”) rackmount unit has two CPUs, each with a rated thermal design power (TDP) of 95W [46], and an installed memory capacity of 12GB. Four 60mm DC fans are installed within the chassis to provide cooling to the motherboard and its critical components. Both processors represent the principal heat load within the system and their temperatures drive a native algorithm, explained in detail in the following subsection, which controls speed of the chassis fans using a pulse width modulation (PWM) signal. As seen in Fig. 1, a sheet metal partition isolates the flow through the motherboard from flow through the power supply unit (PSU) and hard drive seen at the right. The '80 Plus Platinum' rated PSU (efficiency of 94.6%) has an integrated 60mm fan that regulates air flow through this section based on inlet temperature [69]. This fan idles at 30% duty cycle at temperatures below 25°C and ramps linearly up to 45°C, beyond which it runs at full capacity. Thus, air flow through each section is modulated independently. Three such servers with minor part variations are considered to simulate a multiple-vendor or supplier scenario commonly seen in data centers. Table 4-1 outlines differences in installed components across these systems.

Table 4-1 Component/Vendor configuration differences between the three servers under study

| Server | Component | Notes / Specifications |
|---|---|---|
| A | DIMM | 2 GB; single-sided; 1333 MHz |
| | Hard Disk Drive | 250 GB; 7200 rpm; 5.5 W |
| | Chassis Fan (each) | 5.64 W; 43.3 cfm; 0.763 inch $H_2O$; 9000 rpm |
| | Power Supply Unit | 450 W; 80Plus Platinum |
| B | DIMM | 2 GB; dual-sided; 1333 MHz |
| | Hard Disk Drive | 250 GB; 7200 rpm; 5.5 W (*same as A*) |
| | Chassis Fan (each) | 5.64 W; 37.1 cfm; 0.622 inch $H_2O$; 8000 rpm |
| | Power Supply Unit | 700 W; 80Plus Platinum |
| C | DIMM | 2 GB; dual-sided; 1333 MHz (*same as B*) |
| | Hard Disk Drive | 500 GB; 7200 rpm; 10.6 W |
| | Chassis Fan (each) | 5.64 W; 37.1 cfm; 0.622 inch $H_2O$; 8000 rpm (*same as B*) |
| | Power Supply Unit | 700 W; 80Plus Platinum |

*4.1.1 Chassis fan control scheme*

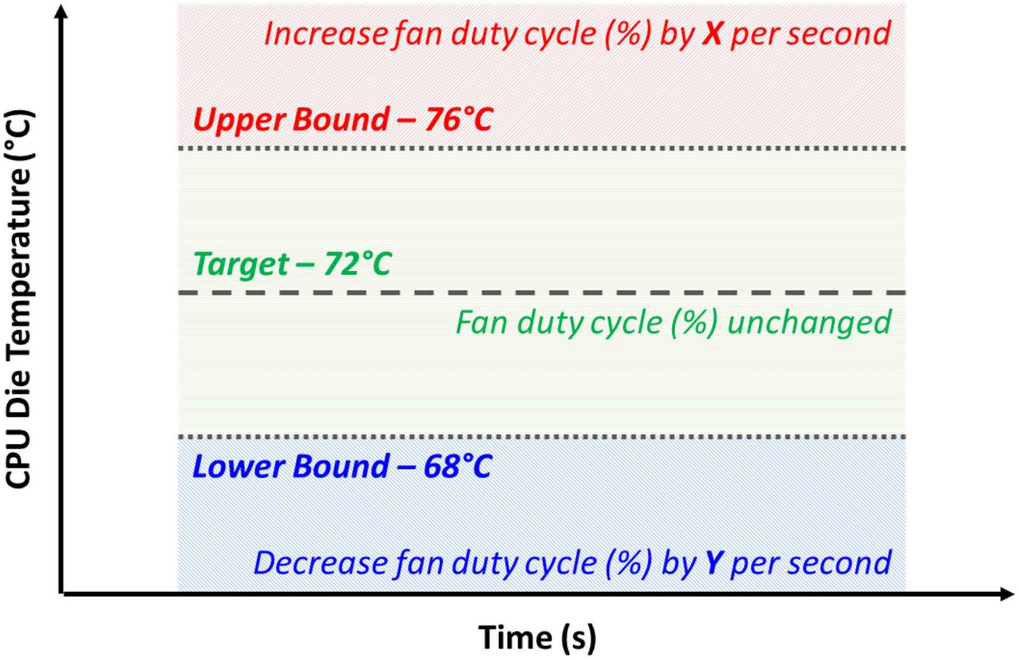

Increase fan duty cycle (%) by X per second
Upper Bound – 76°C
Target – 72°C
Fan duty cycle (%) unchanged
Lower Bound – 68°C
Decrease fan duty cycle (%) by Y per second
CPU Die Temperature (°C)
Time (s)

Figure 4-2 Chassis fan speed control setup to maintain CPU temperature within a predetermined range of the target value

Chassis fan speeds are modulated based on CPU temperature using the principle of deadband control. This technique is commonly employed in voltage regulation. The deadband represents a domain within which a signal experiences no change from its present value. Application of this method to the servers under study is illustrated in Figure 4-2. Purpose of this setup is to minimize cooling power unless CPU operating temperature (control parameter) increases beyond a predefined range. This range consists of a buffer or tolerance coupled to a target (desired) temperature within which the deadband exists. Fan speeds change only when the control parameter is either below or above the lower and upper bounds respectively. Operating outside the deadband for extended periods will cause either the fans to slow down to idling speeds (below lower bound) or ramp up at a steady rate to maximum capacity (above upper bound). In short, objective of this control scheme is to promote operation within this predetermined range of CPU die temperature with a view to minimizing cooling power. Default parameters of the deadband scheme are specified in Figure 4-2.

## 4.2 Test Setup

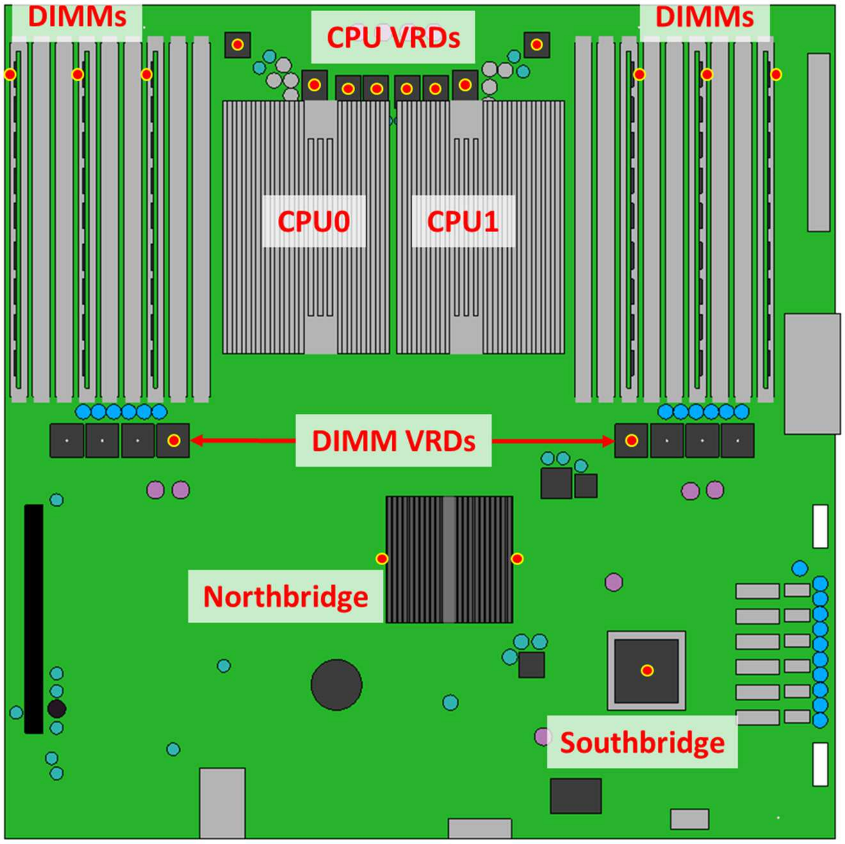
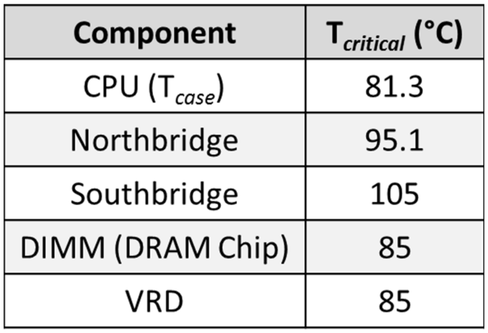


Figure 4-3 Location of installed thermocouples (red dots with yellow outline) for surface temperature measurements of critical motherboard components

In order to study operating conditions and health of critical server components across the thermal envelope, thermocouples are installed to facilitate surface temperature monitoring during testing. Figure 4-3 depicts locations of said measurement and corresponding devices of interest. Table 4-2 outlines critical temperatures below which these components need to be operated. VRDs (voltage regulators for both CPUs and DIMMs) are sensitive to temperature and throttle performance if operated beyond 85°C. Case temperature of the DRAM chip [70, 71] furthest downstream on all DIMMs is measured to account for the effect of thermal shadowing. When dual sided DIMMs are employed (servers B and C), thermocouples are installed on sides facing the CPU heat sinks. The Northbridge chipset, which has a TDP of 27.1W [72], is equipped with an aluminum plate fin heat sink. As the case cannot be accessed directly, surface measurements on either side of the heat sink base in the longitudinal direction are facilitated. Similar to VRDs and DIMMs, a thermocouple at the exposed center measures case temperature of the Southbridge [73]. A total of nineteen surface measurements are made per server.

Table 4-2 Thermal limitations of critical motherboard components



| Component | $T_{critical}$ (°C) |
|---|---|
| CPU ($T_{case}$) | 81.3 |
| Northbridge | 95.1 |
| Southbridge | 105 |
| DIMM (DRAM Chip) | 85 |
| VRD | 85 |

Servers under study, named A to C from top to bottom, are installed in the central column of a triplet rack [74] as seen in Figure 4-4. Each rack mount unit has an adjacent simulated server (1.5U profile; equivalent to a load bank) that operates at a constant thermal and electrical load of around 440W. A total capacity of around 21kW is installed within the rack to ensure adequate heating in the room to permit testing at high air supply temperatures. Blanking panels are installed at the rear of empty slots to prevent recirculation of exhaust air. 277VAC supply to each web server is measured independently using a power meter. A total of 57 thermocouple readings from the three units are measured by a data acquisition (DAQ) unit. Dry bulb temperature and relative humidity of air entering each server is recorded by a USB ambient conditions logger. A workstation, set up adjacent to the rack, communicates with all metering equipment and web servers to ensure a common timestamp across all measurements for effective data reduction.

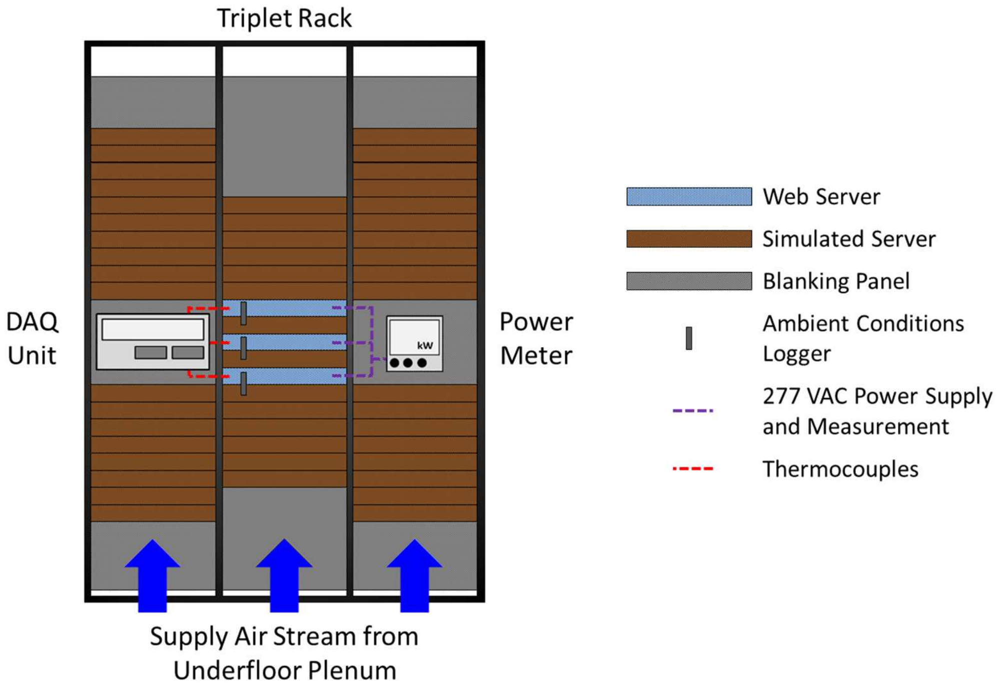

Triplet Rack
DAQ Unit
kW
Power Meter
Web Server
Simulated Server
Blanking Panel
Ambient Conditions Logger
277 VAC Power Supply and Measurement
Thermocouples
Supply Air Stream from Underfloor Plenum

Figure 4-4 Simplified depiction of the test setup with power metering and data acquisition equipment

The rack is set up in a 625 square foot test-bed data center facility equipped with a 12 inch raised floor and a 44 ton computer room air conditioning (CRAC) unit for cooling. The CRAC pressurizes an under-floor plenum with conditioned air which is ejected at the inlet to the rack through perforated tiles. Hot aisle containment curtains mitigate mixing of return air with the supply stream and direct flow exiting the rear of the rack to a dropped ceiling. Negative pressure generated by the CRAC blowers draws in return air to complete the flow cycle. Controller settings in the CRAC are adjusted to achieve supply temperature set points between 15°C and 45°C in increments of 5°C. Additional settings to ensure that operating relative humidity is maintained between 10% and 90% (per specifications of server manufacturers) for extended periods are confirmed by ambient logger measurements over the course of testing.

## 4.3 Test Procedures

In order to study server performance with the preexisting chassis fan control as well as interaction between static and fan powers, at different rack inlet temperatures, two sets of tests need to be conducted. The differentiating factor between these two cases concerns whether or not chassis fans are controlled by the server. Following subsections will explain these dissimilarities in detail. In the first set of tests, chassis fan speeds are modulated 'internally' based on deadband control. Henceforth, results or comments pertaining to these tests will be alluded to as when the fans are internally controlled (or for the sake of depicting data or reference, with the abbreviation 'INT').

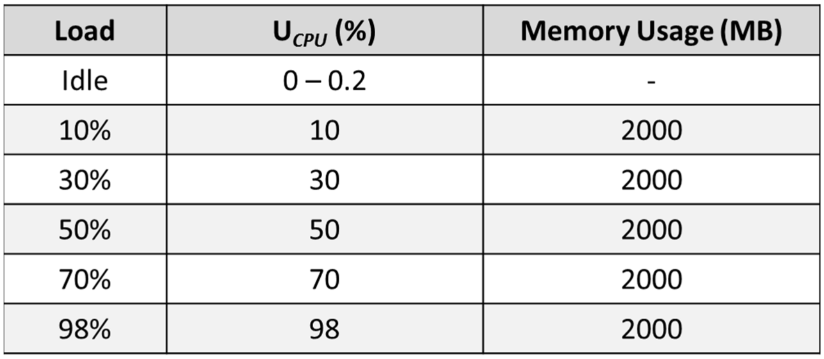
### *4.3.1 Stressing servers with internal fan control*

These rack mount units are configured with CPU and memory resources to function as web servers. Applications usually deployed on these systems are found to utilize relatively more CPU as compared to memory. Table 4-3 outlines different simulated loads setup to run on these servers to mirror operation in a data center. Synthetic load generator lookbusy [75] is employed to create loads as previously outlined. A bash script, executed on each server at the start of a test, automates a procedure wherein each load is applied (in order) for a duration of 30 minutes with 30 minutes of idling between two loads (with the exception of the initial Idle load that ran for 60 minutes). This sequence of stressing the server from idling to maximum loads is repeated three times before a test is concluded. This is to ensure that the results obtained are repeatable. Native Linux commands *mpstat* [76] and *free* [77] are employed to measure CPU utilization ($U_{CPU}$) and memory usage over the course of each test and ensure loads are per requirement. An internal diagnostic tool provided by the manufacturer provides readings from each CPU's digital temperature sensors (DTS) and tachometer input (rotations per minute) from each fan in the motherboard section. Along with server power and component surface temperature measurements, these represent data of primary interest in this study. Steady state operation was achieved within the first 20 minutes of each load and individual measurements made within the last 10 minutes are averaged before reporting. Results presented in this study are reported in terms of average values across three runs (from each server) for each test condition (specific rack inlet temperature and server load).

### *4.3.2. Modification for external fan control*

Interaction between static and chassis fan powers as well as their influence on total server power consumption can be effectively studied when chassis fans are modulated between their maximum and minimum speeds in short increments. This is accomplished by modifying the aforementioned experimental setup to accommodate an arbitrary waveform generator (see Figure 4-5) which, when connected to chassis fans in all three servers, produces a PWM signal to control these units externally. Thus, throughout a given test, fans in the motherboard section operate at a fixed speed before changing duty cycles between tests. For a given rack inlet temperature, chassis fan duty cycles are varied between idling (0.1%) and 99.9% in increments of 4%. Henceforth, results or comments pertaining to these tests will be alluded to as when the fans are externally controlled (or for the sake of depicting data or reference, with the abbreviation 'EXT').

Table 4-3 Different loads under which servers were stressed



| Load | $U_{CPU}$ (%) | Memory Usage (MB) |
|---|---|---|
| Idle | 0 – 0.2 | - |
| 10% | 10 | 2000 |
| 30% | 30 | 2000 |
| 50% | 50 | 2000 |
| 70% | 70 | 2000 |
| 98% | 98 | 2000 |

Introducing idling periods between adjacent server loads in a test are necessary to force chassis fans to idle before a subsequent computational load when fans are controlled internally. However, when chassis fans operate at a fixed duty cycle, a true steady state in operating temperatures and power consumption is achieved and stabilizing fans through intermittent idling periods is unnecessary. Therefore, the bash script previously described is modified to generate server loads in series. In addition, steady states are attained faster in this case (less than 10 minutes) and the duration of each stress level is shortened to 20 minutes. Data averaged over the last 10 minutes of each test is reported as steady state values. These modifications shorten the duration of testing at a given fan duty cycle to seven hours. A LabVIEW [65] code, in conjunction with the scripts, automates testing such that three different duty cycles can be evaluated in tandem over a day.

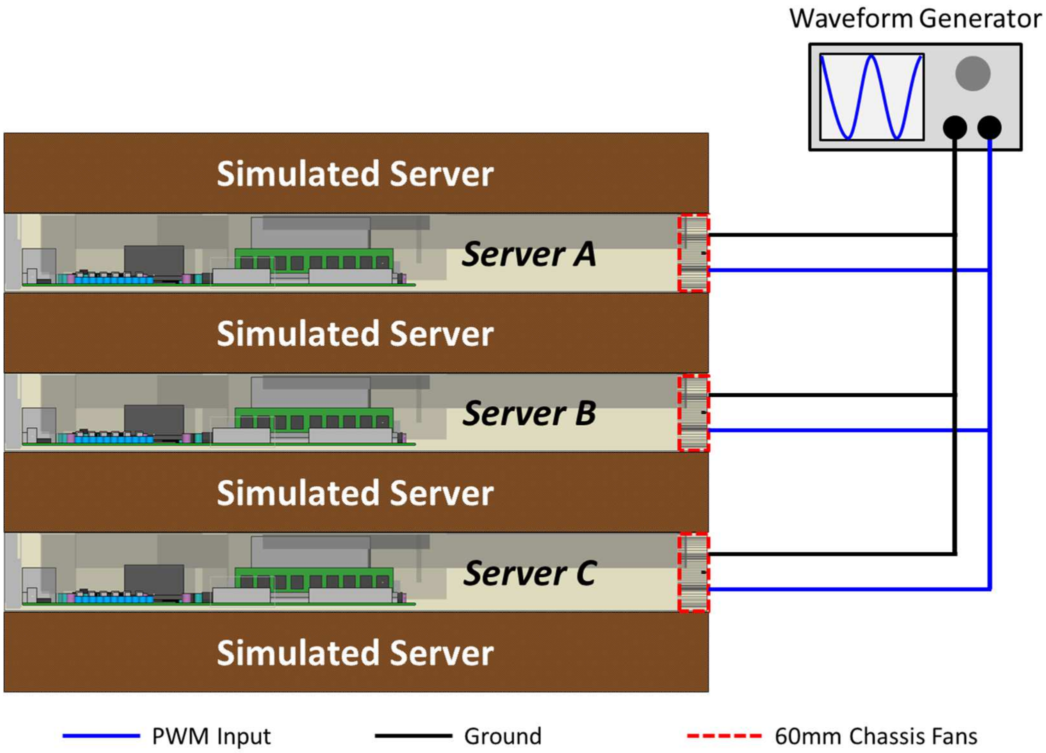


Figure 4-5 Server chassis fan speeds synchronously controlled using an arbitrary waveform generator during testing

## 4.4 Supplementary Tests

Testing three web servers across the A4 envelope provides the following operating parameters under different computational loads – power consumption, component temperatures and fan speeds. Derived parameters, essential to facility operation, such as cooling power, air flow rate and noise pressure levels for each server can be calculated based on corresponding fan speed and rack inlet temperature. Following tests briefly explain how these correlations are calculated.

### *4.4.1 Power (MB fans and PSU fan)*

As previously outlined, fans downstream of the motherboard and in the PSU operate independently of each other and, as such, measurements for power consumption or flow rate as a function of duty cycle (or speed) are conducted separately. In either case, the fans are disconnected from their respective headers and powered and controlled externally using a benchtop DC power supply and waveform generator. Standard specifications for input pulse width modulation (PWM) signals are followed [78]. A total of forty readings recorded over two minutes are averaged to report power consumption at a given speed. Fan duty cycle is varied between idling (0% for chassis fans and 30% for PSU fan) and full capacity (100%) in short increments to ensure the resultant cubic polynomial correlation provides a coefficient of determination greater than 0.99. Manufacturer's specifications for idling and maximum fan speeds are used to generate a linear relation between duty cycles (%) and speed (rpm). Thus, for chassis fans, speeds recorded during server testing are used to calculate corresponding power consumption. Similarly, for the PSU fan, rack inlet temperature is used as previously explained. Both figures are combined to report cooling power consumption.

### *4.4.2 Flow rate*

Similar to the previous test, with fans powered and controlled externally, server motherboard section and PSU are respectively installed in an air flow bench to measure air flow rate (cubic feet per minute) as a function of duty cycle. For a detailed explanation of the test procedure followed when using this equipment, please refer to [79, 80]. Thus, for each server, two quadratic polynomial correlations are generated. Once again, fan speeds and rack inlet temperatures from server testing are used to calculate corresponding volumetric flow rates. Both figures are combined to report flow rate through the entire server.

### *4.4.3 Acoustics*

Unlike the aforementioned measurements, recording parameters for two sets of fans separately before combining them is not effective in the case of acoustics. A parametric study wherein one variable is changed while the other is held constant is better suited. Since PSU fan speed varies with rack inlet temperatures between 25°C and 45°C, a separate series of tests were conducted for each 5°C increment. In each case, speed of chassis fans is varied between idle and maximum duty cycles in increments of 25%. A sound metering application [81] is employed to measure each data point. The microphone is positioned 1m from the center of the leading edge of the server. Total of three runs are conducted to ensure repeatability. Thus, a polynomial correlation between sound pressure levels (dB) and fan duty cycle (%) is generated for each server and rack inlet temperature. Fan speeds from server testing for a given load and RIT are used to calculate corresponding noise levels.

### 4.5 Notice regarding Test Conditions

Blowers in the CRAC unit provided more flow than the installed servers (both web and simulated units) required. This may have caused results presented in proceeding two chapters to be skewed due to effects of overcooling (forcing more air to flow through the servers). Kindly take note of this before utilizing results for other studies. However, aspects of data analysis and interpretation still apply.

# CHAPTER 5

# EFFECTS OF ELEVATED RACK INLET TEMPERATURES

In traditional data center cooling system architectures, chillers account for around 41% of total cooling power consumption [11]. The refrigeration plant operates inefficiently and has typical COP (coefficient of performance) values of around 3 to 9 [82]. As previously mentioned, a current practice in the industry involves raising coolant supply temperatures to reduce chiller hours and/or take advantage of outdoor conditions and maximize use of free cooling. Data center designs that eliminate the need for refrigeration systems by relying on evaporative cooling and air-side economization have been gaining more widespread adoption over the last decade [83]. These facilities have been reported to operate extremely efficiently with PUE values around 1.08 [84].

While the studies outlined in Section 2.3 have proven that raising data hall temperature is of advantage, limited information is available regarding the effects at server-level under such conditions. Specifically, although reports on analytical modeling of data centers [85, 86] are widespread, accurate representation of server operation or performance is relatively absent. Information regarding trends in power consumption and reliability is available; however, a study that accounts for all parameters pertaining to server performance has yet to be conducted. Of utmost importance is the need to understand the interdependent operation of servers and the facility. In addition, effects under varying server utilizations are never accounted for or reported. This investigation outlined the effect of increased RIT at a variety of server loads on server power consumption, cooling or air flow rate, efficiency, acoustics and component reliability. Reported data can be employed to determine the optimal range of RIT for minimizing operating costs for a given data center utilization. An important notice regarding test conditions and results can be found in Section 4.5.

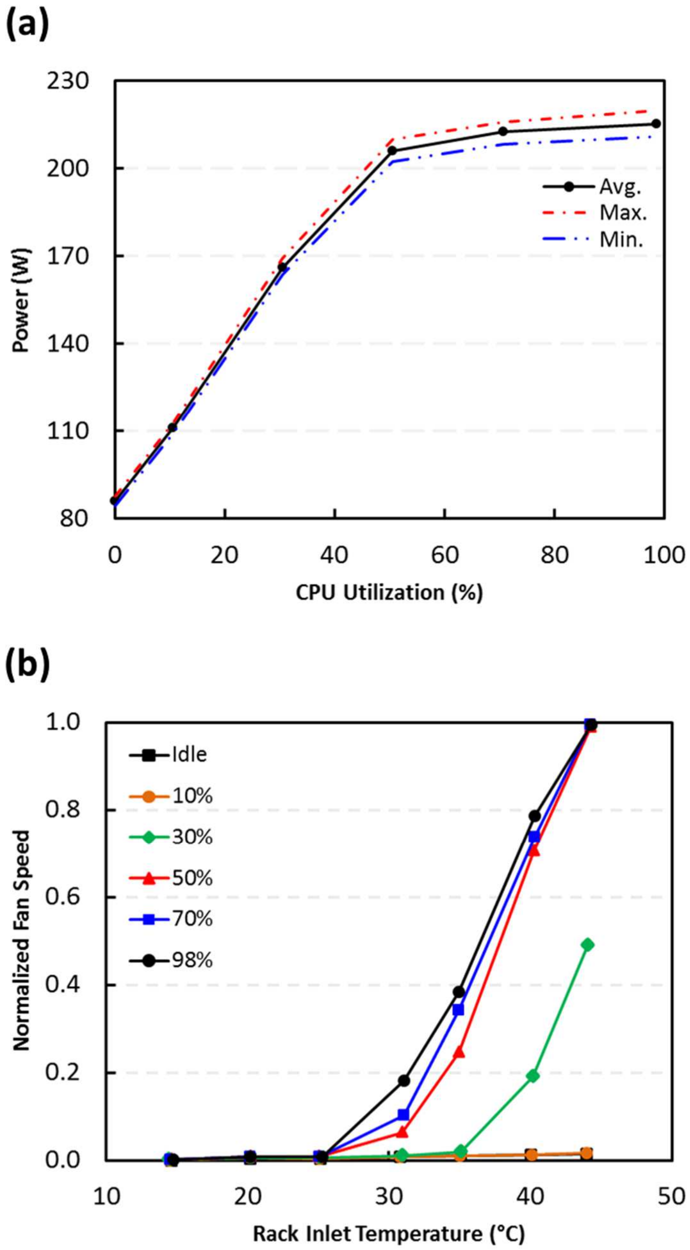


Figure 5-1 Variation of (a) total server power consumption with CPU utilization for rack inlet temperature of 15°C and, (b) chassis fan speed with RIT for all tested loads

A total of six different loads were executed on each server for a given air supply temperature. Figure 5-1(a) illustrates variation of server power consumption with applied computational load at 15°C ambient conditions. Minor variation in heat dissipation was visible at loads greater than 50%. Power consumption at 30% load could be interpreted as midway between idling and 98% values. In addition, data centers have been reported to typically operate at CPU utilizations of around 30% [23]. Thus, to avoid clutter in plotting results, values for idle, 30% and 98% loads were considered only. However, as seen in Figure 5-1(b), from the perspective of fan performance and its derived parameters, the discussion (text) also considered 50% utilization as noticeable difference in operation at RIT of 30°C and 35°C was observed with respect to 98% load. Apart from power consumption, no difference was observed between idle and 10% cases and effects at 70% could be approximated based on results for 50% (30°C) and 98% loads (35°C and higher RITs).

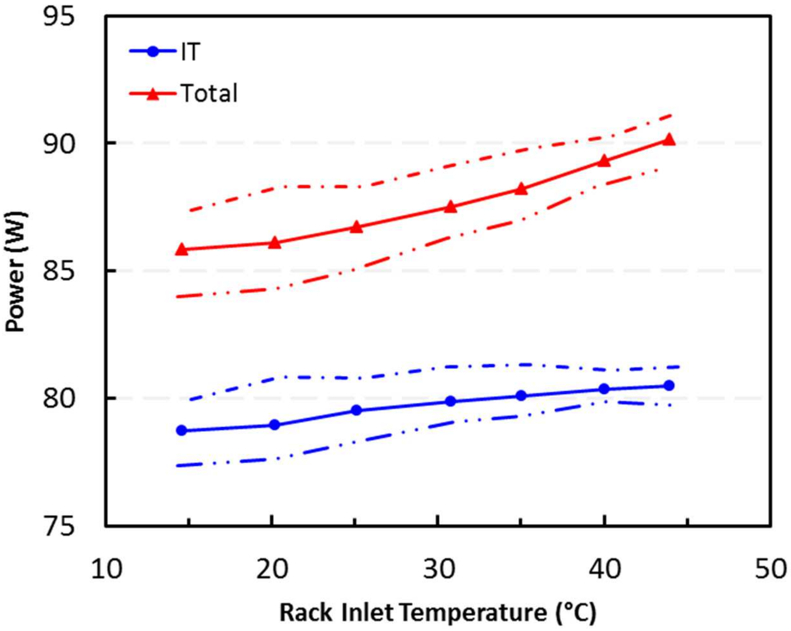


Figure 5-2 Variation of IT and total server powers with RIT for idle load

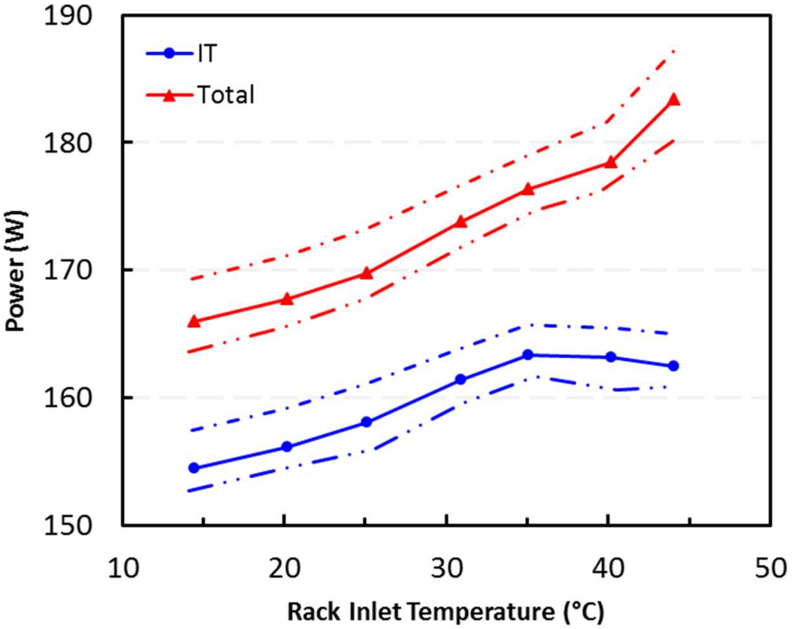


Figure 5-3 Variation of IT and total server power with RIT for 30% utilization

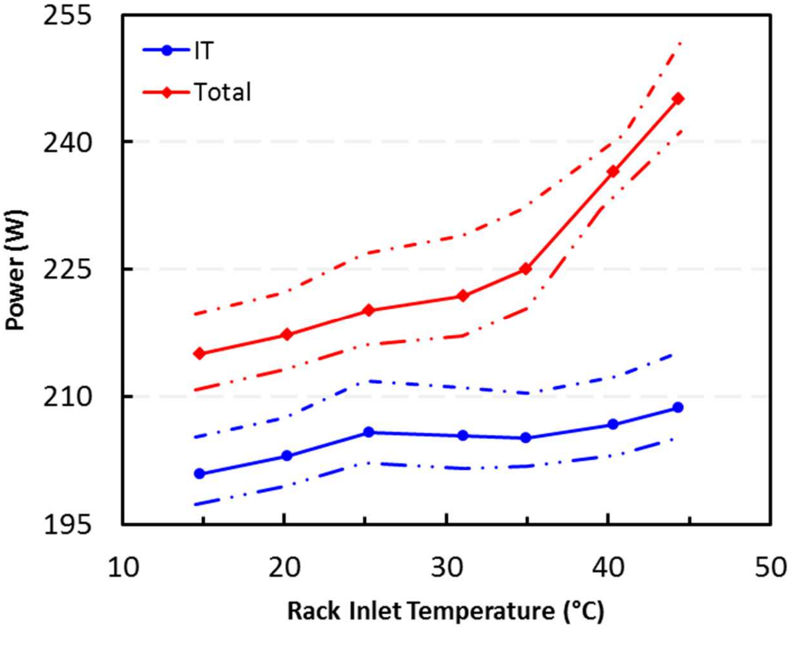


Figure 5-4 Variation of IT and total server power with RIT for 98% load

## 5.1 Power

Measurements reported by the power meter represent total consumption by each individual server. Based on PSU efficiency (assumed to be 94.6% at all loads), chassis fan speeds and rack inlet temperature, overheads such as cooling and PSU losses can be accounted for to report a pure IT power consumption at all test cases. Figures 5-2 to 5-4 shows variation of both values for different inlet temperatures. Note that solid lines with markers represent average values, single dotted and dashed lines denote the maximum, and double dotted and dashed lines represent minimum values for the three servers. When idling, IT power consumption was seen to steadily rise with RIT due to the effects of increasing CPU temperature. This is evident as seen in Figure 5-5. Overhead (difference between total and IT power) increased from 9% up to 12% beyond ambient conditions of 25°C due to corresponding increase in PSU fan speeds. The native chassis fan control scheme engaged beyond RITs of 35°C and 30°C for 30% and greater than 50% loads respectively as seen in Figure 5-5. Correspondingly, IT power was observed to plateau due to reduced effect of leakage (static power). However, overhead grew rapidly beyond these temperatures due to increased chassis fan speeds to maintain the CPU case between 62°C and 68°C.

Since total server power is a significant contributor to a data center's operational expenditure (OpEx) it is critical to discuss increase with RIT while leveraging reduced chiller operation (traditional facilities) or water consumption (economizer based designs). These savings should not be significantly diminished or eclipsed by operating costs of IT equipment. At idling and 10% loads, total server power consumption was reported to increase by 2% to 7% (with respect to values at 15°C) at RITs greater than 30°C. For utilizations of 30% and more, total server power increased gradually from 2% and up to

14% beyond air supply temperatures of 25°C. This information could be used to set the data hall temperature range with a view to maximizing savings in facility power consumption. Specifically, considering an increase in server power consumption of 5% to be an upper bound, then irrespective of IT load, temperature of air entering racks should be at or below 30°C. Interestingly, total power at 30% utilization is responsible for the aforementioned upper limit as the increase at 35°C supply temperature is around 6.3%. Facility energy modeling may be greatly enhanced with the detailed server level information presented here.

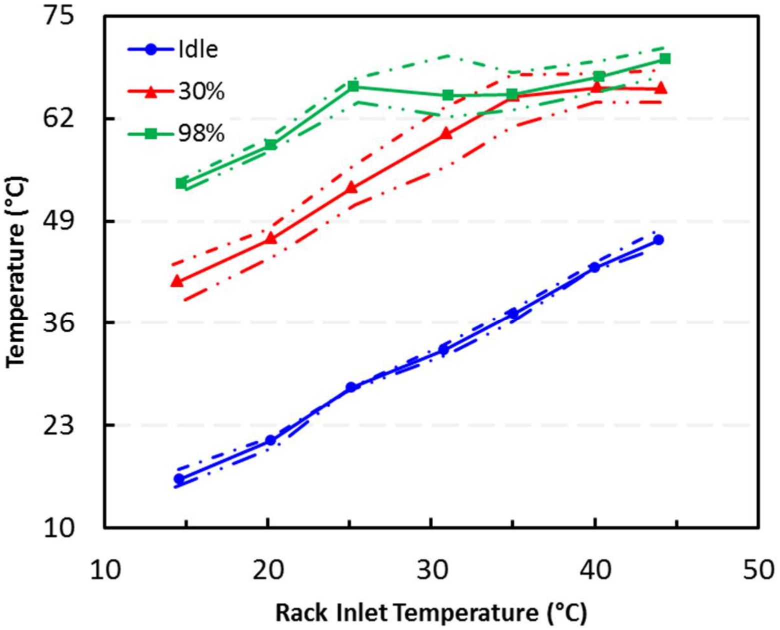


Figure 5-5 Variation of CPU case temperature with supply temperature

## 5.2 Cooling

Unlike power consumption of IT equipment, air flow rate through individual servers has a direct influence on operation of facility-level cooling systems. Figure 5-6 shows variation of the aforementioned parameter with RIT. For idling and 10% loads, air flow rate through each server steadily rose beyond 25°C. This rise was only attributed to increased PSU fan operation at elevated ambient conditions since the chassis fans remained at idle speeds over the RIT range. For 30% utilization, noticeable rise in flow rate was observed beyond 35°C due to elevated chassis fan speeds. When compared to lower loads (Idle and 10%), the penalty was around 41% at 40°C and 99% at 44°C respectively. At higher server utilizations (50% to 98%), a similar breakaway trend was observed beyond RITs of 25°C. However, corresponding penalties (with respect to idling or 10% utilizations) were much larger than those seen for 30% loads with required flow rate more than doubling beyond supply temperatures of 40°C and 35°C for 50% and 98% utilizations respectively.

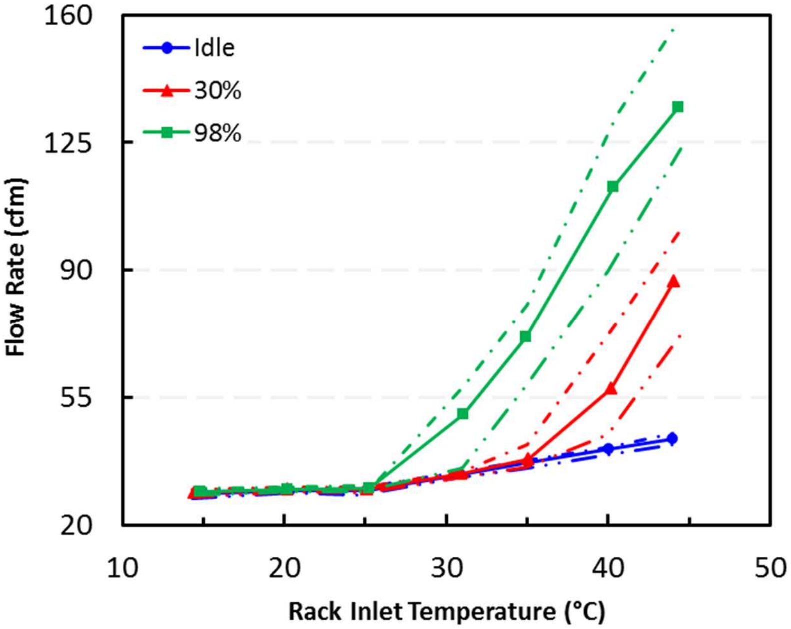


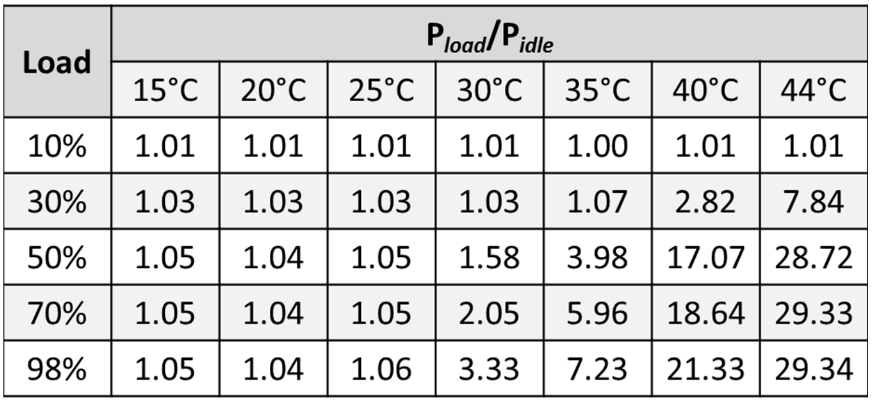
Figure 5-6 Variation of air flow rate per server for different RIT and loads

This was significant because power consumption of blowers in CRACs (traditional facilities) or fan walls (economizer based designs) vary with delivered flow rate based on the following fan law:

$$P_2 = P_1 \times \left({}^{Q_2}/_{Q_1}\right)^3 \tag{5-1}$$

P represents power consumption of the blowers corresponding to a delivered flow rate of Q. Table 5-1 outlines the ratio of perceived facility blower power at higher server loads with respect to idling conditions across the ASHRAE A4 envelope.

Table 5-1 Ratio of facility fan power to corresponding value at idle load for different rack inlet temperatures and server loads



| Load | $P_{load}/P_{idle}$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 15°C | 20°C | 25°C | 30°C | 35°C | 40°C | 44°C |
| 10% | 1.01 | 1.01 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 |
| 30% | 1.03 | 1.03 | 1.03 | 1.03 | 1.07 | 2.82 | 7.84 |
| 50% | 1.05 | 1.04 | 1.05 | 1.58 | 3.98 | 17.07 | 28.72 |
| 70% | 1.05 | 1.04 | 1.05 | 2.05 | 5.96 | 18.64 | 29.33 |
| 98% | 1.05 | 1.04 | 1.06 | 3.33 | 7.23 | 21.33 | 29.34 |

As observed, it would become prohibitive to operate the data center beyond an ambient condition of 30°C for an average utilization of 50% or greater. Thus, similar to the consideration for server power, increased flow rate should not usurp targeted savings through raised RIT. If a restriction of 100% rise in blower power over corresponding values for idling or 10% loads is considered, then for utilizations at or below 50%, data hall set point could vary up to 30°C. However, for higher loads, RIT must be maintained at or below 25°C. DCIM (data center infrastructure management) solutions can be setup to account for these restrictions with intent to minimize OpEx. For example, RIT may be allowed to drift to higher levels during known periods of low compute utilization or remain more strictly controlled at peak operating periods. Figure 5-7 shows variation of fan power consumption per server with air supply temperature.

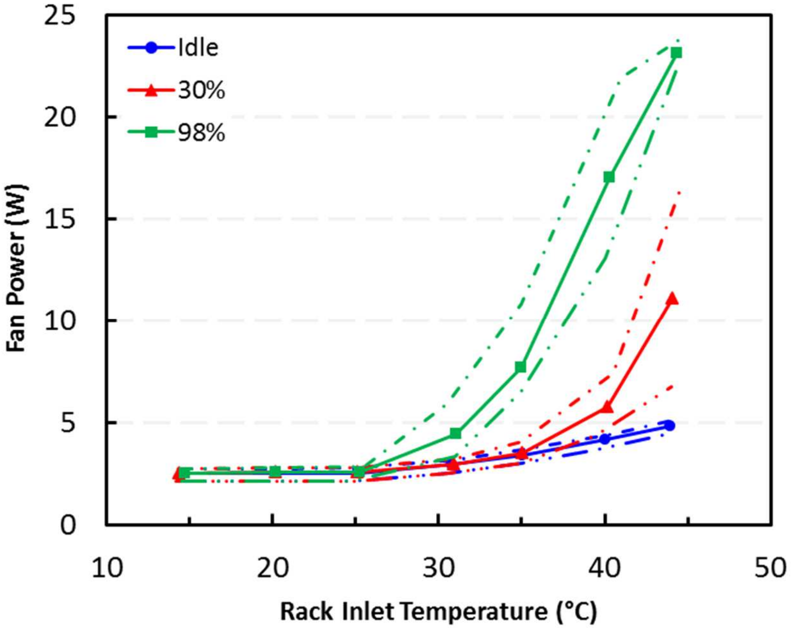


Figure 5-7 Variation of fan power consumption per server fir different supply temperatures

## 5.3 Efficiency

While efficiency metrics do not provide as clear a picture of the effects of raising RIT as the previous two subsections, they may provide insight that might influence decisions eventually made. One such parameter, evaluated at the server-level, is the partial PUE (power usage effectiveness) as defined below.

$$pPUE = \frac{Total\ server\ power}{Server\ IT\ power} \tag{5-2}$$

It is apparent that reporting a pPUE as close to unity is advantageous. Conversely, reducing overhead (PSU losses and cooling power) to zero is beneficial. Figure 8 shows the variation of pPUE with rack inlet temperature. Comparatively, idling seemed more inefficient when compared to higher utilizations at ambient conditions of 30°C or below. This was because of lower IT powers recorded. However, since this idling trend is representative of the server's innate configuration, it can be used to draw conclusions regarding the limits of energy-efficient operation. In this case, irrespective of the load, rack inlet temperatures must be limited to 35°C.

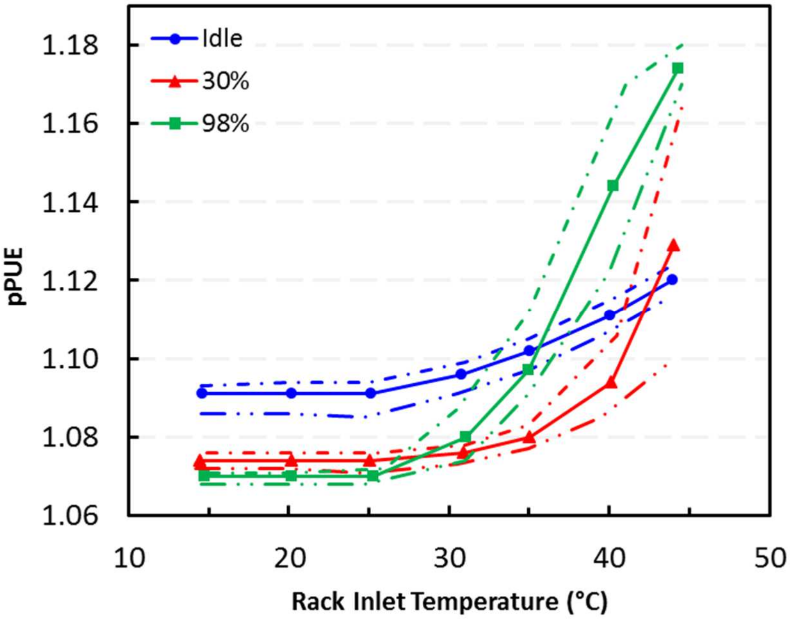


Figure 5-8 Partial PUE calculated at the server-level and its variation with air supply temperature

## 5.4 Acoustics

Regulating agencies such as OSHA (Occupational Safety and Health Administration) and NIOSH (National Institute for Occupational Safety and Health) set standards for healthy working environments across different industries. Limits for permissible exposure duration of 8 hours for sound pressure levels are 90dBA and 85dBA according to OSHA [87] and NIOSH [88] respectively. This study is focused on noise levels per server as opposed to effects of concurrent operation of IT equipment. The authors encourage readers to employ reported values with standard correlations to predict additive effects of multiple servers in operation. Figure 5-9 shows recorded sound pressure as a function of air inlet temperature and computational loads. Owing to relatively low increase in chassis fan speeds for 30% server utilization, a similar trend to idling and 10% loads was observed as noise from the PSU fan dominated measurements. However, at higher utilizations, sharp rise in chassis fan speeds generated noticeable increase in sound levels from aforementioned trends for RITs beyond 30°C and 25°C for 50% and 98% loads respectively. The corresponding penalties (with respect to idling or 10% utilizations) amounted to a rise of around 18% to 35%. Irrespective of stress levels, for a single server, reported sound pressure levels met regulation standards.

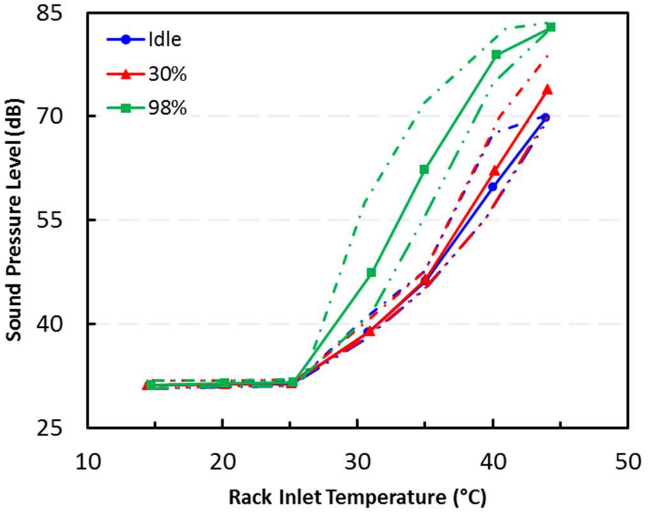


Figure 5-9 Effect of rack inlet temperature on single server noise levels

## 5.5 Reliability

Operating temperatures of crucial server components recorded during testing provide an indication of whether critical limitations (see Table 4-2) were met across the entire thermal envelope. As previously discussed, the chassis fan control scheme was engaged when CPU case temperatures rise beyond 68°C. The highest recorded value (at 98% load with 44°C air entering the server) was still 10°C below the critical limit specified by the device manufacturer. Also, CPU failures are uncommon as infant mortality, a significant factor, is typically accounted for during 'burn-in' tests before deploying servers in a data center. Therefore, this section will focus on operating temperatures for other critical components and comment on reliability when applicable. Note that the maximum of all thermocouple measurements of a given type is reported in the following sections.

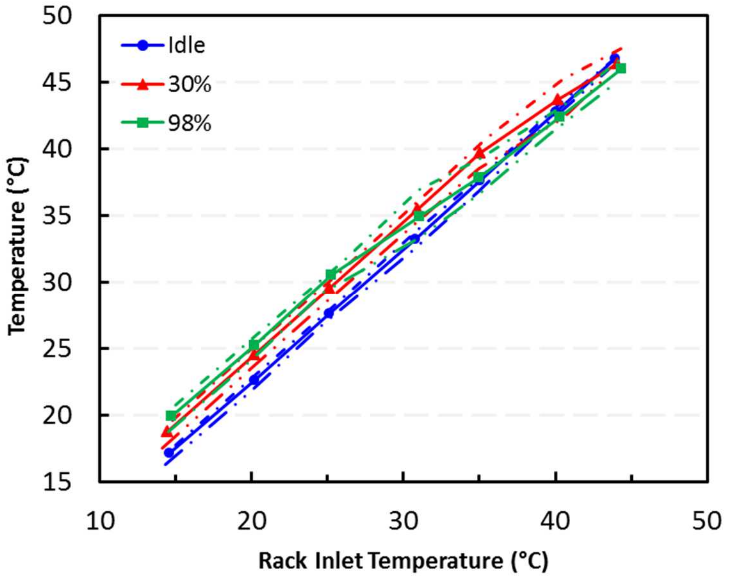


Figure 5-10 Variation of maximum surface temperature for DRAM chips with RIT

### *5.5.1 Memory*

Figure 5-10 reports a linear variation in DRAM case temperature with RIT. Minor differences were visible between different utilizations, which was to be expected, as the server loads were CPU intensive with only 20% of DIMM usage tested. Highest recorded values were more than 30°C lower than the specified critical case temperature and throttling of performance was not imminent. In addition, in terms of reliability or failure, a recent study [49] has shown that DRAM errors are strongly correlated with CPU utilization and DIMM usage (not utilization) as opposed to operating temperature. Similar error rates were reported for a temperature difference of 20°C under a given utilization. In conclusion, the chassis cooling system maintained the DIMMs at very reliable operating temperatures even at elevated RIT.

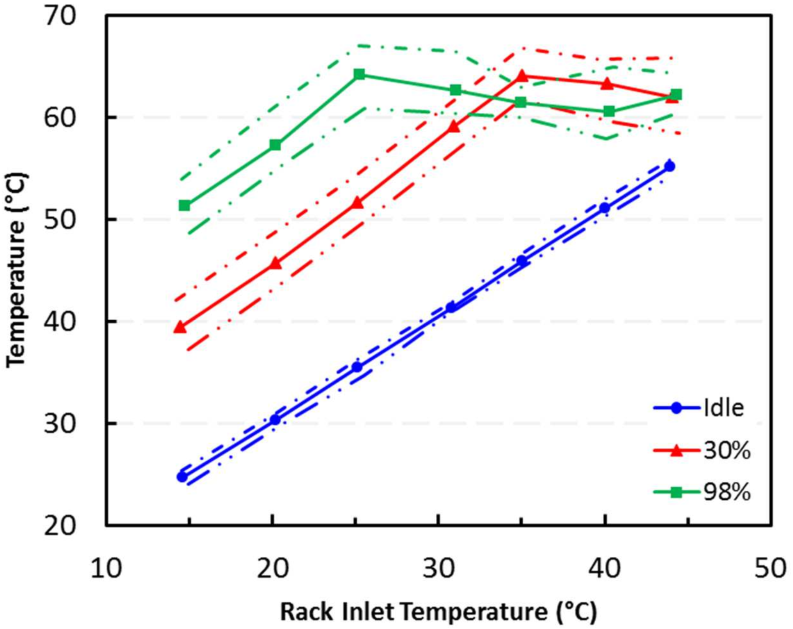


Figure 5-11 Variation of maximum surface temperature for VRDs with air supply temperature

*5.5.2 VRDs*

Regardless of the test case, CPU VRDs were found to operate at higher temperatures as compared to DIMM counterparts owing to the effect of thermal shadowing from CPUs. Since the motherboard section of the server is equipped with a duct that directs majority of air flow through CPU heat sinks, VRDs (see Figure 5-11) showed similar temperature trends as seen in Figure 5-5. Regardless of RIT or server load, these devices operated more than 15°C below their specified limit.

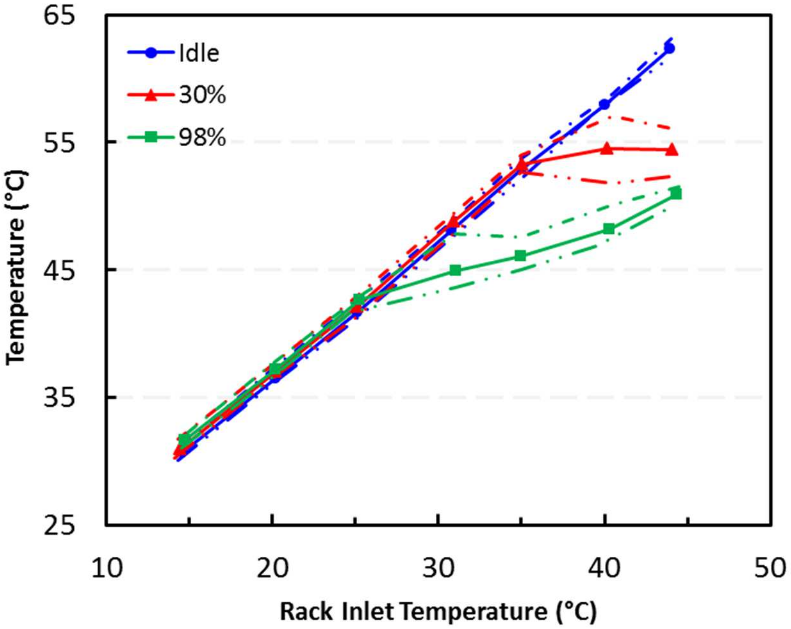


Figure 5-12 Variation if maximum surface temperature for Northbridge with RIT

*5.5.3 Northbridge*

This chipset is equipped with an aluminum heat sink to dissipate a reported TDP of 27.1W. Thermocouples equipped on both sides of the heat sink base require consideration of conduction resistances only to discuss possible case temperatures. Figure 5-12 reports variation in these values with RIT for different utilizations. Owing to the presence of extended surfaces, trends for higher utilizations were seen to plateau under the influence of increased air flow rate. Maximum values were seen under idling and 10% loads due to chassis fan control not yet being engaged. Irrespective of utilization, assuming a conservative difference of 10°C between the heat sink base and chipset case, this device operated 20°C lower than the specified limit.

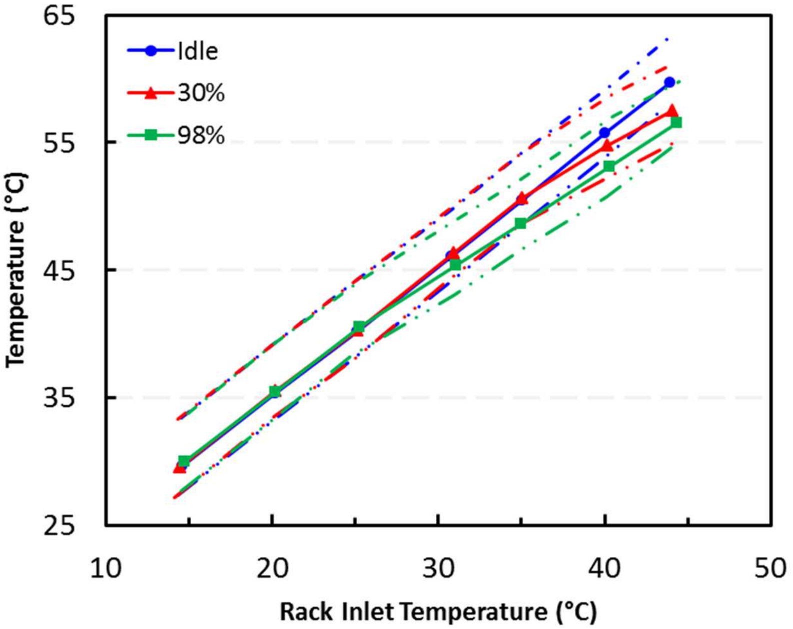


Figure 5-13 Variation of maximum surface temperature for Southbridge with air supply temperature

*5.5.4 Southbridge*

Owing to lower TDP (4.5W) and relatively large case area (around 25mm by 25mm), this chipset is not equipped with a heat sink to aid heat transfer. A single thermocouple installed at the center of the case reported the operating temperature as seen in Fig. 5-13. Once again, similar to DIMM temperatures and absence of extended surfaces, a linear trend in reported values was seen with RIT. Minimal effect due to increased air flow was seen at higher utilizations. A large thermal overhead, of the order of around 40°C, was seen for case temperature irrespective of server load.

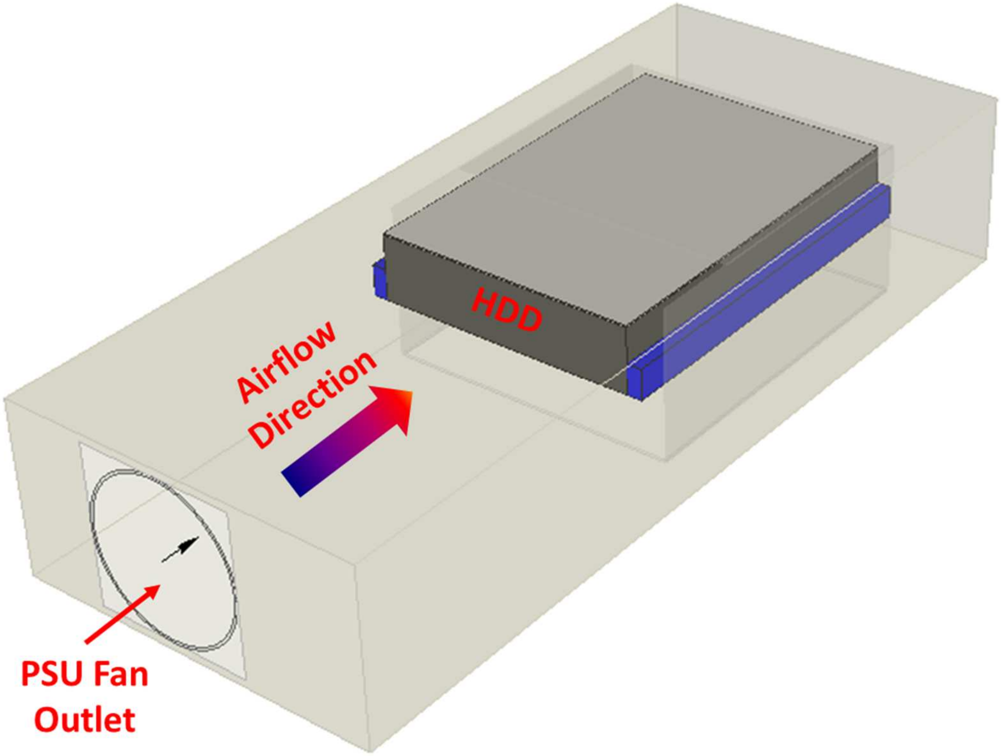


Figure 5-14 Simple CFD model used to predict HDD case temperature

### 5.5.5 Hard disk drive

A recent study [50] reported that hard disk drives accounted for around 70% of hardware component errors in a large data center based on two years of failure data. Also reported was that correlation of errors was stronger with temperature than utilization. HDD case temperature was predicted based on results of a simple CFD model (see Figure 5-14). Exhaust air stream from the PSU was modeled along with HDD power dissipation for maximum utilization (read and write operation) to generate a trend of case-to-ambient temperature difference with respect to PSU flow rate. Thus, based on RIT and PSU heat loss, case temperature of the HDD could be predicted. Despite choosing two different HDDs, manufacturers reported the same AFR (annualized failure rate) of 0.73%

for case temperatures of 40°C or below. An acceleration factor (AF) of two for every 15°C rise beyond the specified limit was applied to predict the resultant AFR as seen in Fig. 5-15. The large spread between maximum and minimum values owed to significant difference in HDD powers (10.6W and 5.5W) as well as different flow rate trends with duty cycle for each server PSU. A sharp rise in AFR was seen beyond rack inlet temperatures of 30°C, with failure rate at 44°C supply nearly doubling when compared to 15°C. The corresponding penalty for 30°C was 22.5%. Interestingly, HDD reliability would not depend on utilization.

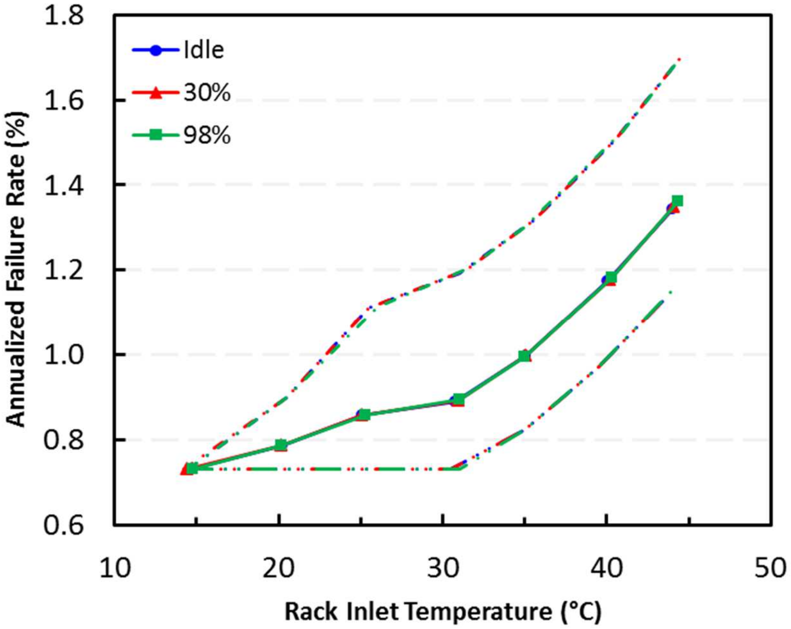


Figure 5-15 Effect of RIT on hard disk drive AFR

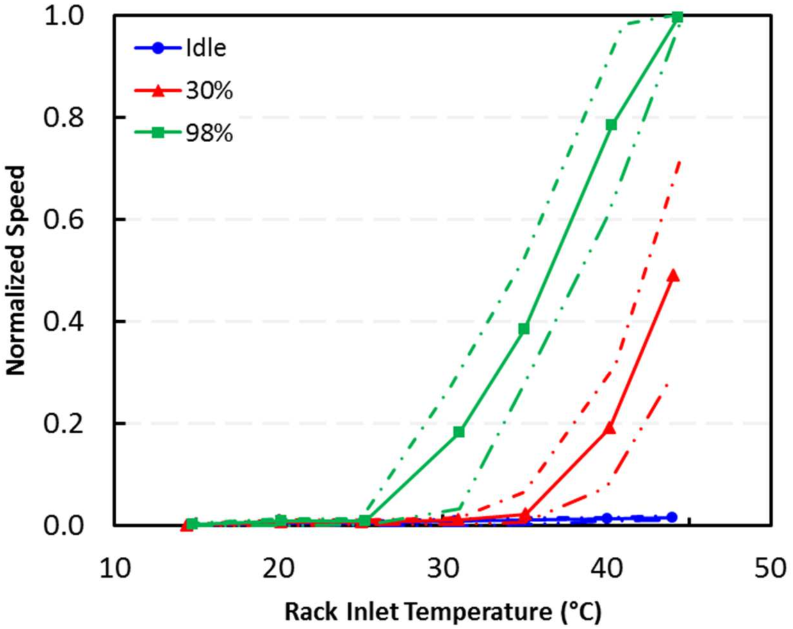


Figure 5-16 Motherboard or chassis fan speeds as a function of RIT and server load

*5.5.6 Fans*

Owing to the presence of moving parts, fans are a possible source of failure or downtime in rackmount servers. The most common failure mechanism in fans is bearing lubricant deterioration [89, 90, 91]. Bearing manufacturers [90] provide correlations for life expectance of fan bearings using general purpose grease as seen below.

$$\log L_{50} = 6.54 - 2.6\frac{n}{N_{max}} - \left(0.025 - 0.012\frac{n}{N_{max}}\right)T \quad (5\text{-}2)$$

$L_{50}$ is the time at which 50% of bearings will fail, n and $N_{max}$ are current and maximum speeds respectively in rotations per minute and T is operating temperature (°C). Figure 5-16 shows variation in chassis fan speed, normalized with respect to the rated (maximum) value, for different server loads and RITs. At low utilizations (idling and 10%), no increase in fan operation was observed and bearing life would be dependent on the approach temperature of air based on motherboard power dissipation. For 30% server load, fan speed would be the dominant factor for air supply temperatures beyond 35°C. Similarly, this breakaway point was 30°C and 25°C for 50% and 98% utilizations respectively. Thus, fan life would be dependent on both RIT and server load.

## 5.6 Other Considerations

Another factor to consider when choosing the set point of the data hall within the ASHRAE A4 thermal envelope is human comfort. OSHA [92] defines the range for ‘thermal comfort’ with ambient temperature controlled between 20°C and 24.5°C and relative humidity between 20% and 60%. While this is a severely constrained range within which to operate a data center efficiently, comfort of personnel working within the facility could become a concern beyond 30°C and affect operational expenditure. However, detailed discussion of this topic is beyond the scope of this study, but an important factor to acknowledge. Regardless, selecting the optimal range for RIT would require an evaluation of total cost of ownership (TCO) while considering all aforementioned parameters. There exists a tradeoff between deploying greater number of servers (increased capital expenditure, CapEx) operating at lower utilizations with higher RITs (decreased OpEx) or less servers (decreased CapEx) operating at higher utilizations with lower RITs (increased OpEx). Results presented in this study can support estimates made towards operational expenditure.

# CHAPTER 6

## TRADEOFF BETWEEN COOLING POWER AND LEAKAGE CURRENT

As outlined in Section 2.3, a current practice in the industry to reduce cooling power consumption involves raising coolant supply temperatures to reduce chiller hours and/or take advantage of outdoor conditions and maximize use of free cooling. In air cooled servers a common side-effect of raising the ambient temperature is the increase in power consumption due to elevated fan speeds [43]. In addition, higher coolant supply temperatures to servers may also lead to elevated CPU operating temperatures and, in turn, increased device power. This is attributed to leakage current or static power within a device. Leakage current, per its name, pertains to current that leaks through transistors during off states and represents loss of useful energy. For a detailed understanding of this term and its components please refer to [16]. In contrast, dynamic power relates to useful work involved during switching operations in a device. Static and dynamic powers together constitute CPU power consumption. Studies [93, 94] have shown that at elevated operating temperatures proportion of static to total power can be greater than 50%. Krishnan et al. [95] reported that leakage current increases exponentially with temperature and subthreshold leakage power reduces by 50% with a temperature drop of 30°C. The issue of static power is further exacerbated by the advancement of microelectronics and reduction in gate lengths. Haensch et al. [96] reported that, approaching the 65nm technology and lower, leakage power density grows more dominant and eventually equals its dynamic counterpart.

From a thermal standpoint, static power can be lowered by reducing operating temperature of the device (increase provided cooling). Lin et al. [15] showed that a tradeoff exists between device and cooling powers that minimizes system power consumption. Increasing cooling beyond this minimum point lowers energy efficiency. In this study, web servers outfitted as per Section 4 (based on relatively current 32nm lithography) investigated this phenomena with a view to minimizing total server power consumption. To account for aforementioned trends in energy-efficient cooling (economization in data centers), these rackmount systems were tested across the ASHRAE A4 envelope with RIT varied between 15°C and 45°C in increments of 10°C. An important notice regarding test conditions and results can be found in Section 4.5.

Effect of CPU operating temperature on leakage current was accounted for by reporting IT power consumption per server. This term was calculated by subtracting overheads such as PSU inefficiency and fan power, and chassis fan power from total server power consumption. For any given server, variation in IT ($P_{IT}$) and chassis fan ($P_{f, MB}$) powers with CPU die temperature ($T_{die}$) formed the basis of understanding the contribution of each term to total server power consumption ($P_{Total}$).

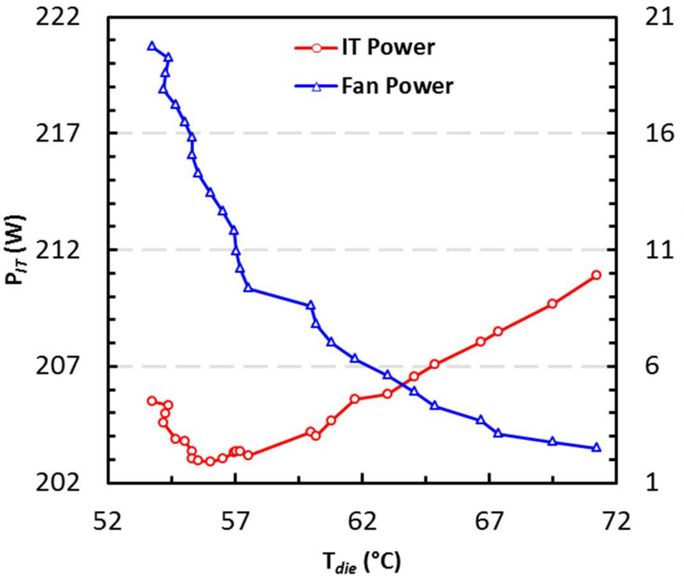


Figure 6-1 IT and chassis fan power versus die temperatures for server B at 98% load with inlet temperture of air around 25°C

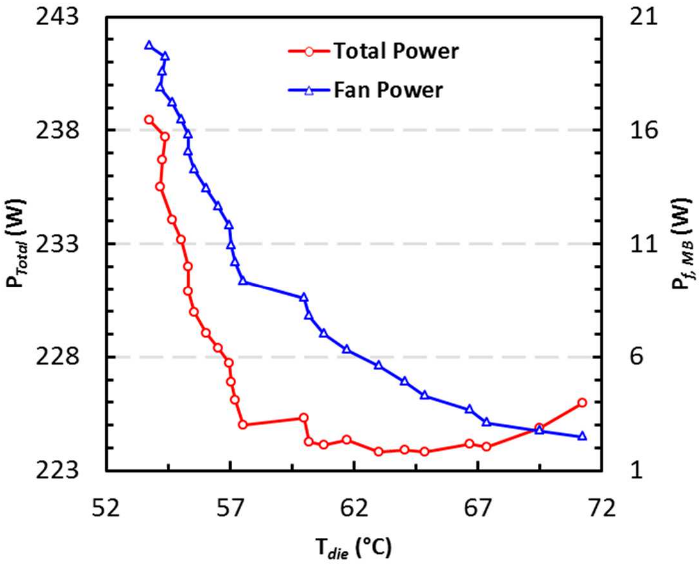


Figure 6-2 Total server and chassis fan powers versus CPU die temperatures for server B at 98% load with RIT of around 25°C

## 6.1 Understanding and Comparing Effects of Static and Fan Powers

As previously explained, for a given RIT, each server was subjected to a series of loads at a fixed chassis fan speed. By externally varying fan duty cycle from maximum to minimum in short increments, different CPU operating temperatures were obtained and their effect on IT and chassis fan powers was studied. Figure 6-1 shows variation of aforementioned terms for server B at 98% load with RIT of 25°C. When considering maximum variation (difference between extreme values) in both IT and chassis fan powers, it was observed that the latter was greater than the former by a factor of around 2.2; calculated as follows.

$$\text{Factor} = \frac{Max\left(P_{f,MB}\right) - Min\left(P_{f,MB}\right)}{Max\left(P_{IT}\right) - Min\left(P_{IT}\right)} \tag{5-2}$$

This value was found to vary between 1.9 and 9.1 for different servers, loads and RITs. Thus, for servers under consideration, chassis fan power exerted greater influence than static or IT power consumption. Figure 6-2, similar to 6-1 except for the inclusion of total server power, reflects the observation made in the previous statement. At the lower end of CPU die temperatures a sharp drop in $P_{Total}$ was observed corresponding to similar behavior in $P_{f,\ MB}$ and minor variation in $P_{IT}$. However, as $T_{die}$ increased with reducing chassis fan power, variation in $P_{f,\ MB}$ and $P_{IT}$ were observed to be similar (while opposing) with minor deviation in total server power consumption. This behavior of evening $P_{Total}$ at higher CPU die temperatures was observed only at server loads greater than 50%. At lower server utilizations (30% and below), minimum $P_{Total}$ was observed at the highest $T_{die}$. This was attributed to lower CPU and static power consumptions. A distinct minimum in total server power was seldom observed and discussions henceforth focus more on improving chassis fan operation to extract savings other than $P_{Total}$.

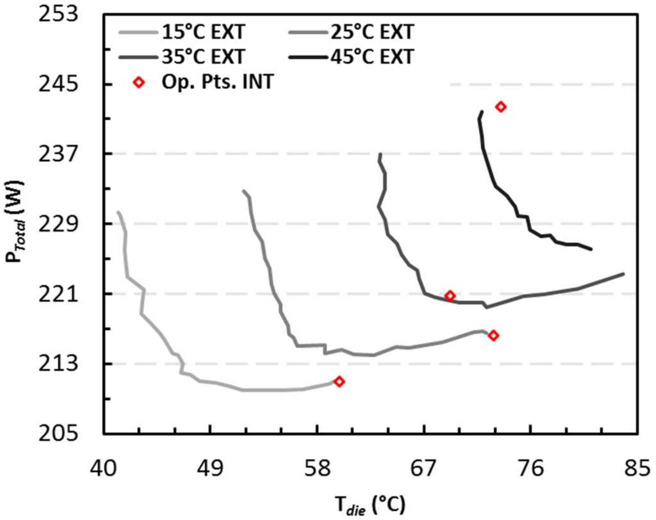


Figure 6-3 Variation of total server power with CPU die temperature for EXT tests at different temperatures and corresponding operating points from INT trials for server A at 98% computational load

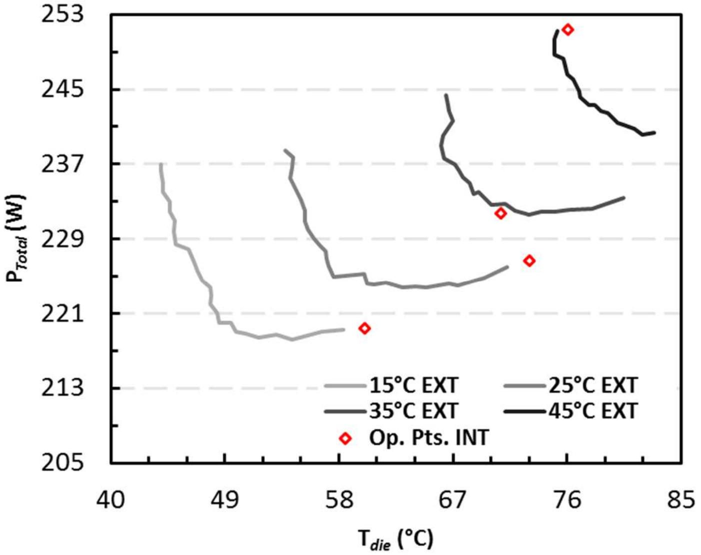


Figure 6-4 Variation of total server power with CPU die temperature for EXT tests at different temperatures and corresponding operating points from INT trials for server B at 98% computational load

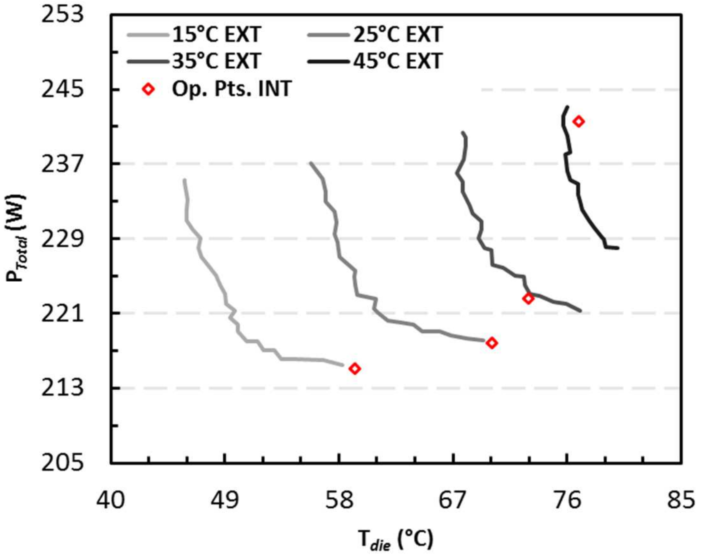


Figure 6-5 Variation of total server power with CPU die temperature for EXT tests at different temperatures and corresponding operating points from INT trials for server C at 98% computational load

## 6.2 Comparison of Results from INT and EXT Testing

When chassis fans were internally controlled, different operating points (averaged values) were obtained for a combination of server loads and RIT. These results were taken from Section 5. A comparison of operating parameters between tests wherein fans were controlled both externally and internally for 98% server utilization is shown in Figures 6-3 to 6-5. Operating points were observed to seldom intersect with distributions from external testing and could be attributed to factors such as minor variation in RITs, different chassis fans used in both tests and effects of averaging steady state data. At air supply temperatures of 15°C and 25°C, chassis fans idled and servers operated at the trailing edge of the (EXT) distribution. At 98% load, minor savings in $P_{Total}$ were available for servers A and B by reducing CPU operating temperatures. It was observed that irrespective of server or computational load, these savings were of the order of 0.5% to 1.5% (with respect to the operating points) and represent minimal improvement as well as values within uncertainty of experimental measurements. In addition, corresponding reduction in $T_{die}$ was accompanied by an increase in chassis fan speed and air flow rate, effects of which are discussed in detail later. For high server loads (50% and above), the native fan control scheme was observed to visibly engage at RITs of 35°C and 45°C. Similar to lower RITs, at 35°C marginal reduction in total server power may be available with change in die temperature. However, in this case, an increase in $T_{die}$ is sought with corresponding reduction in chassis fan speeds. In contrast to lower RITs, at 45°C chassis fans were found to operate at maximum capacity to ensure that reliable CPU temperatures were maintained. Substantial savings in $P_{Total}$ were available, at utilizations of 50% or more and irrespective of server, of around 4.4%

to 7.2% with a corresponding increase in die temperature of the order of 3°C to 7.6°C. Chassis fan speeds, in turn, reported reductions of around 41% to 58%. It was apparent that at reasonable RITs (35°C and below), savings exist in forms other than total server power such as reduced fan speed and air flow rate through the server. However, additional considerations such as reliability of components needed to be accommodated as discussed below.

### 6.3 Modifying Chassis Fan Control Parameters and its Effects

As seen in Figures 6-3 to 6-5, specifically for higher loads (50% or more) and RITs of 35°C and above, increasing $T_{die}$ within reasonable limits may provide savings in power and/or air flow rate requirements. The latter is important is it has substantial influence on facility cooling power consumption. Increasing CPU operating temperature required modification of the chassis fan control parameters. In addition, the influence on chassis fan operation (specifically, reduction in speed and air flow rate) and subsequent effect on component temperatures needed to be accounted for. As such, with reference to Table 4-2, limits for operating temperatures of components needed to be established when considering modification of control parameters. Maximum case temperatures 10°C lower than specified critical values were considered for all VRDs (75°C), DRAM chips (75°C) and the Southbridge chipset (95°C) to ensure continuous and reliable operation. Since thermocouples were mounted at opposite edges of the heat sink base installed on the Northbridge chipset, operating temperatures were targeted to be 20°C lower (75.1°C) than the specified limit. The additional 10°C margin of safety accommodated for conduction resistances through the thermal grease and heat sink base. The device manufacturer (ODM) specified that CPU case temperatures must not exceed 81.3°C.

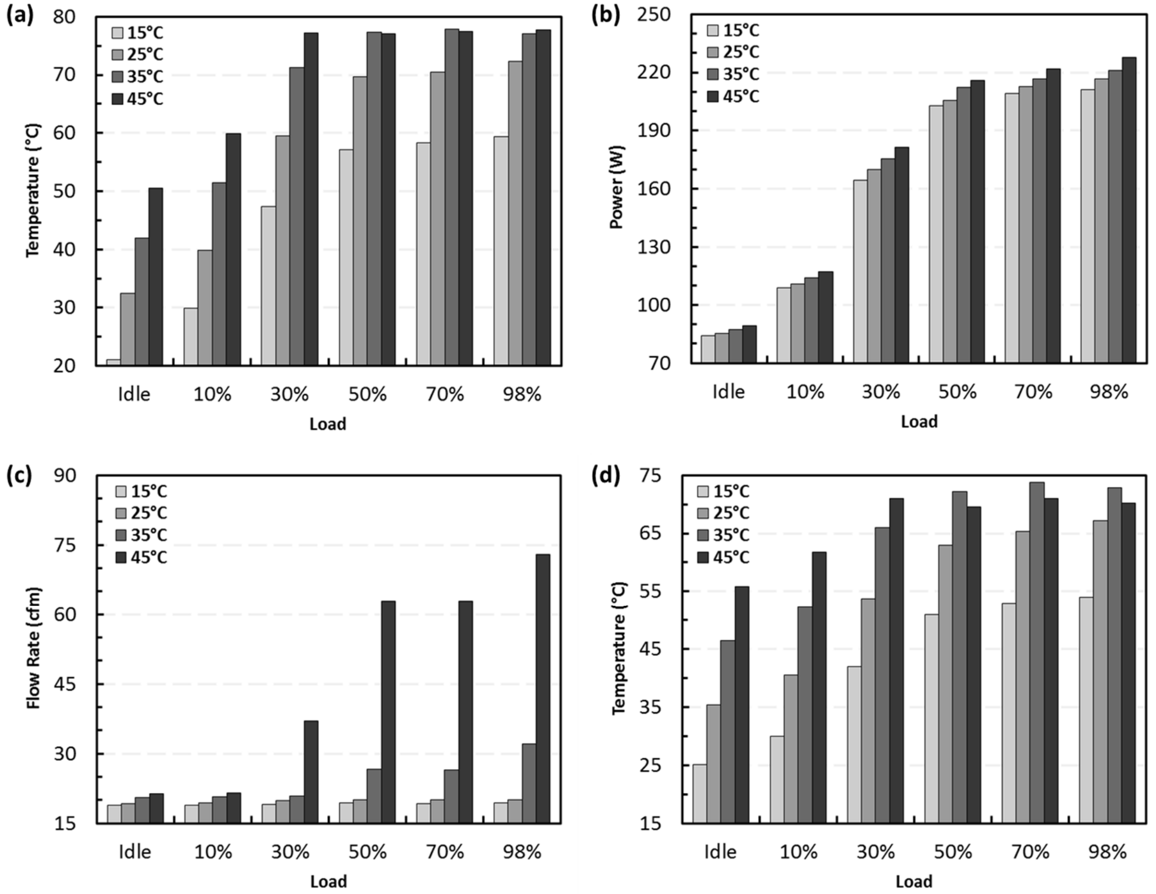


Figure 6-6 Effect of updated control parameters on (a) CPU die temperature, (b) total server power, (c) flow rate, and (d) maximum VRD temperature for different RITs and loads for server A

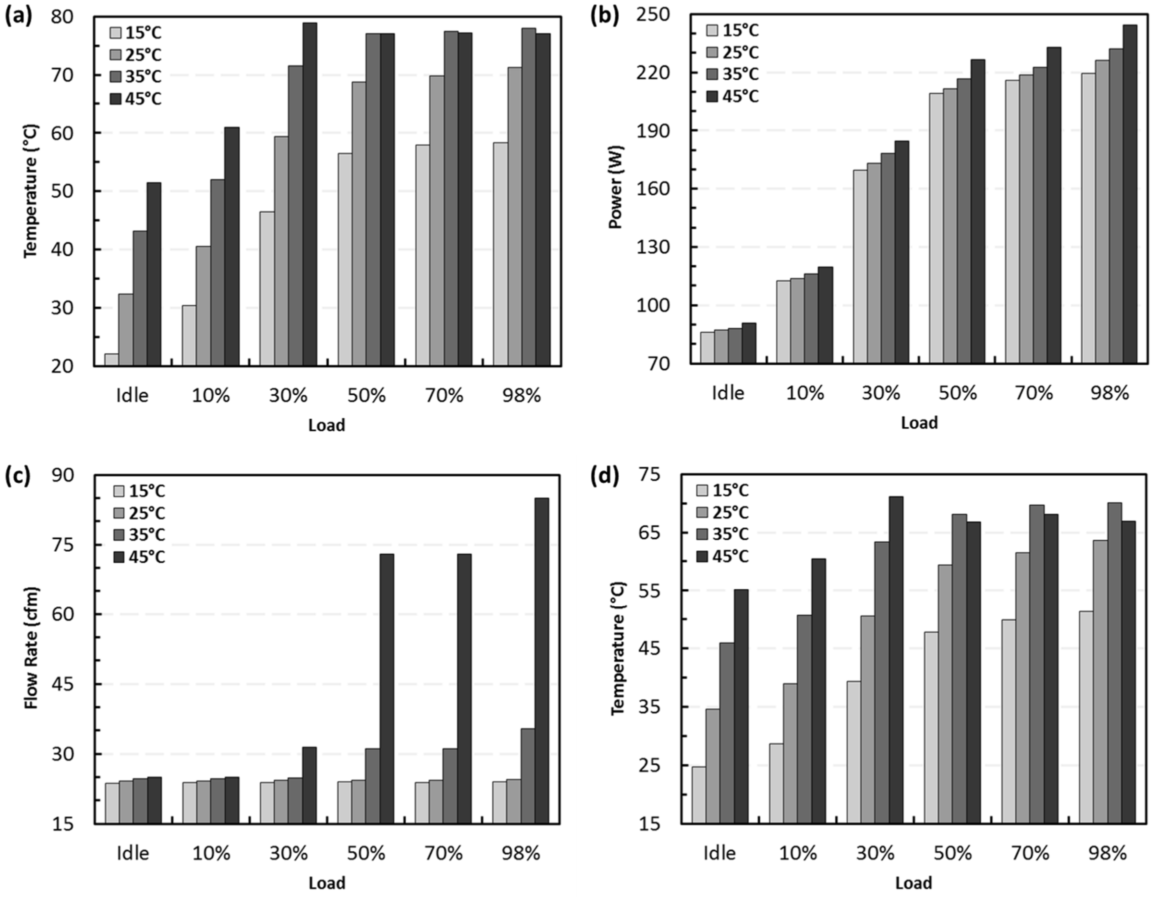


Figure 6-7 Effect of updated control parameters on (a) CPU die temperature, (b) total server power, (c) flow rate, and (d) maximum VRD temperature for different RITs and loads for server B

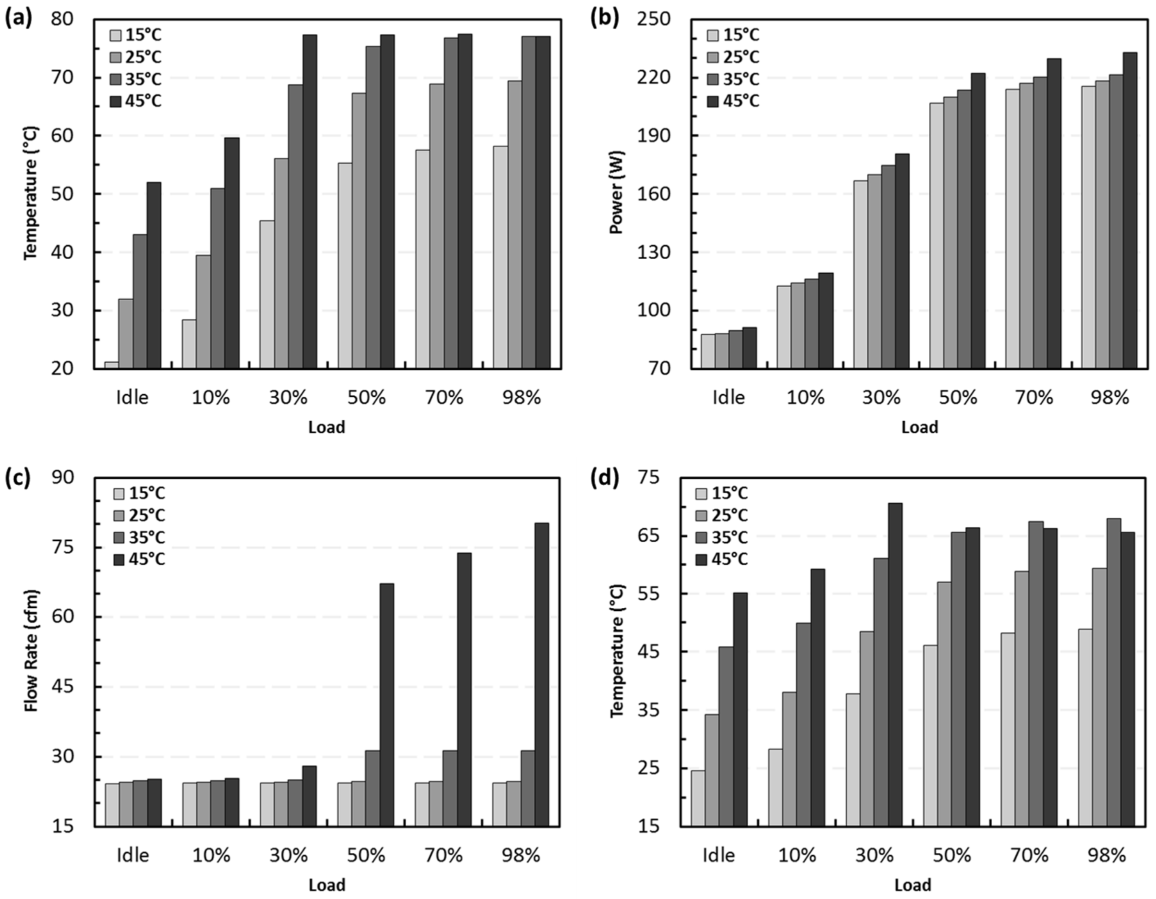


Figure 6-8 Effect of updated control parameters on (a) CPU die temperature, (b) total server power, (c) flow rate, and (d) maximum VRD temperature for different RITs and loads for server C

However, since DTS readings were reported at the die and the temperature difference between sensor and case needed to be accounted for, we assumed a maximum CPU die temperature of 80°C. Considering the chassis fan control scheme being employed, this specification would correspond to the upper limit of the deadband. With the aforementioned limitations in mind and data available (average values) from testing (EXT), it was observed that CPU die and VRD temperatures are the primary factors that control chassis fan operation. Other components in the motherboard operated well within chosen safety limitations. A target CPU temperature of 77°C was found to be adequate while ensuring all cooling requirements were met (see Figures 6-6 to 6-8). This represented an increase of 5°C for all chassis fan control parameters specified in Figure 4-2 and corresponding savings are outlined in Table 6-1.

Table 6-1 Comparison of server performance for higher CPU die target temperature with respect to default settings

| Server | Load (%) | RIT (°C) | $T_{die}$ (°C) | $P_{Total}$ (W) | $Q_S$ (cfm) | SPL (dB) | $N_{f,\,MB}$ (rpm) |
|---|---|---|---|---|---|---|---|
| A | 30 | 45 | 7.4% | 0.3% | -30.1% | 0.1% | -39.4% |
| | 50 | 35 | 8.9% | 0.4% | -30.0% | -6.7% | -36.9% |
| | | 45 | 6.4% | -6.8% | -42.6% | -15.2% | -48.9% |
| | 70 | 35 | 11.5% | 0.0% | -37.0% | -11.2% | -44.6% |
| | | 45 | 6.8% | -7.0% | -43.7% | -14.8% | -49.9% |
| | 98 | 35 | 11.4% | 0.1% | -37.6% | -18.2% | -44.6% |
| | | 45 | 5.7% | -6.0% | -37.1% | -13.5% | -42.6% |
| B | 30 | 45 | 9.7% | -0.7% | -40.3% | -5.7% | -48.1% |
| | 50 | 35 | 6.6% | -0.2% | -23.6% | -16.9% | -26.7% |
| | | 45 | 3.5% | -4.2% | -25.7% | -8.4% | -29.0% |
| | 70 | 35 | 9.9% | -0.9% | -36.4% | -27.0% | -40.3% |
| | | 45 | 3.0% | -4.4% | -25.7% | -8.4% | -28.8% |
| | 98 | 35 | 10.3% | 0.3% | -33.9% | -25.3% | -37.5% |
| | | 45 | 1.3% | -2.8% | -16.2% | -4.6% | -18.2% |
| C | 30 | 45 | 8.8% | -1.6% | -45.2% | -8.7% | -52.8% |
| | 50 | 35 | 3.5% | -0.6% | -20.0% | -12.3% | -22.5% |
| | | 45 | 2.2% | -5.5% | -30.4% | -8.2% | -33.7% |
| | 70 | 35 | 6.2% | -0.6% | -28.8% | -19.3% | -31.9% |
| | | 45 | 1.1% | -4.4% | -25.1% | -6.1% | -27.8% |
| | 98 | 35 | 5.5% | -0.6% | -27.8% | -18.5% | -30.9% |
| | | 45 | 0.2% | -3.5% | -19.7% | -4.2% | -21.9% |

Note that percentage savings were calculated with respect to operating points when fans were controlled by the server (native scheme). Loads and air supply temperatures not specified in this table reported either marginal savings in power or air flow rate, or operating conditions similar to the original scheme. Similar to observations made from Fig. 7, reasonable savings in total server power were only available at a RIT of 45°C and high computational loads (50% or more). Magnitude of reduction in $P_{Total}$ varied from server to server and could be attributed to the type of chassis fans employed. Therefore, savings reported for servers B and C were similar. However, the primary improvement attained by increasing the chassis fan control parameters was the reduction in fan speeds ($N_{f,\ MB}$). Irrespective of load or RIT, chassis fan speeds were lowered by 18% to 50%. Corresponding reductions in sound pressure levels were reported to be as high as 27%. For servers B and C, greater decline in server noise was observed at an air supply temperature of 35°C. Savings in air flow rate through each server was calculated to be around 16% to 45%. Since the data hall within which IT equipment resides is pressurized by blowers in cooling systems setup at the facility, reduction in server-level flow rate requirements affect facility blower power requirements based on Equation 5-1.

Table 5 outlines the percentage of server-level air flow rate and perceived facility fan power consumption due to the increased chassis fan control parameters when compared to the original scheme or default settings. Due to the nature of the aforementioned fan law, a reduction in flow rate of around 16% corresponds to around 40% reduction in facility blower power. At the reported loads and air supply temperatures, substantial savings in $P_{blower}$ were available of up to 83%.

Table 6-2 Savings in air flow rate through the server and facility fan power due to higher target temperature with respect to default setting

| Load (%) | RIT (°C) | Server A | | Server B | | Server C | |
|---|---|---|---|---|---|---|---|
| | | $Q_S$ (cfm) | $P_{blower}$ (W) | $Q_S$ (cfm) | $P_{blower}$ (W) | $Q_S$ (cfm) | $P_{blower}$ (W) |
| 30 | 45 | 30.1% | 65.8% | 40.3% | 78.8% | 45.2% | 83.6% |
| 50 | 35 | 30.0% | 65.6% | 23.6% | 55.5% | 20.0% | 48.9% |
| | 45 | 42.6% | 81.1% | 25.7% | 59.0% | 30.4% | 66.2% |
| 70 | 35 | 37.0% | 75.0% | 36.4% | 74.2% | 28.8% | 63.9% |
| | 45 | 43.7% | 82.1% | 25.7% | 58.9% | 25.1% | 57.9% |
| 98 | 35 | 37.6% | 75.7% | 33.9% | 71.1% | 27.8% | 62.4% |
| | 45 | 37.1% | 75.1% | 16.2% | 41.3% | 19.7% | 48.3% |

Similar to Section 5.2, an increase in facility fan power of 100% over corresponding values at idling and 10% loads was considered to be a restriction. As seen in table 6-3, and comparing to results from the previous chapter, the maximum RIT for server utilizations greater than 10% was increased. For 30% load, the upper limit was increased to 44°C. The corresponding value for utilizations of 50% and above was 35°C. Thus, by increasing the target die temperature, the maximum air supply temperature considering the imposed restriction was, in general, increased by 10°C (for utilizations of 30% and above) when compared to the original fan control scheme.

Table 6-3 Ratio of facility fan power to corresponding value at idle load for different rack inlet temperatures and server loads

| Load | $P_{load}/P_{idle}$ | | | |
|---|---|---|---|---|
| | 15°C | 25°C | 35°C | 44°C |
| 10% | 1.00 | 1.01 | 1.01 | 1.01 |
| 30% | 1.02 | 1.03 | 1.02 | 1.67 |
| 50% | 1.03 | 1.04 | 1.54 | 7.70 |
| 70% | 1.02 | 1.04 | 1.54 | 8.29 |
| 98% | 1.02 | 1.05 | 1.93 | 11.16 |

## 6.4 Further Discussion

While the reported discounts were substantial enough to consider modification of server settings; it was also important to consider other effects of implementing such changes. Primarily, a targeted increase in CPU die temperature of 5°C may have noticeable effect on device reliability and life. However, since established correlations between device life (in hours) versus operating temperature for given gate length (32nm) were unavailable, accounting for the same was beyond the scope of this study. Since fan reliability is primarily dependent on bearing life (see Equation 5-2), which is correlated to normalized speed ($n/N_{max}$) and operating temperature, reduction in chassis fan speed may not necessarily increase fan life. This is because, for a given power dissipation at the motherboard, reduction in air flow rate will increase temperature of air approaching the fan and thereby the operating temperature of the bearing. Another advantage of reduced air flow rate per server is during possible failure of facility-level cooling systems. With a fixed amount of cold air in the data hall, servers may stay in operation for longer periods with updated chassis fan control until backup systems are online. Ultimately, the decision to implement said changes is determined by the impact on operating expenditure (OpEx). This involves a tradeoff between savings in facility blower and/or server power consumption, and cost of replacing components (parts and increased personnel hours required) due to a possible rise in server failure rate or downtime. Regardless, an informed decision is imperative before implementing the change in chassis fan control settings.

# CHAPTER 7

## SAVINGS IN COOLING POWER THROUGH DEPLOYMENT OF RACK-LEVEL FANS

Traditional servers are configured to include all sub-systems such as compute, memory, storage, networking and cooling within a single chassis. Common rackmount units generally have a low profile of 1U (U = 1.75 inches), which accommodate small 40mm fans. Manufacturers [97] and standards associations [98] have published data that encourages designers to opt for larger fans to increase their peak total efficiency. However, since fans are generally selected based on server profile, there exist opportunities to instead consolidate larger fans at the rear of a rack to increase savings.

Over the past few years, OEMs (original equipment manufacturers) [99], semiconductor device manufacturers and hyper-scale data center owners [100] have been promoting the concept of “rack disaggregation”. This refers to separation of resources or sub-systems that are traditionally included in a server, into individual modules at the rack. This includes compute, storage, networking, power distribution and cooling, and endeavors to make the rack the fundamental building block of a data center. Increased distance between IT components is countered by introduction of silicon photonics [101, 102]. The primary advantage of such a deployment is the ability to change or refresh subsystems at different frequencies. In addition, disaggregation promotes dematerialization [103], capable of significant environmental impact, through reduction in PCB sizes and sheet metal otherwise used for server chassis. In particular, disaggregation of cooling at the rack is synonymous with the focus of this study.

Preliminary work [104] predicted savings of up to 55% in cooling power by replacing smaller chassis enclosed (60mm) fans with larger rack-mount units for a stack of four web servers. This study advanced this work by experimentally validating maximum possible savings through deployment of 80mm and 120mm fans. In particular, this study previewed a methodology for implementation of a control system that replicated the in-built scheme for modulation of chassis fan speeds. Thus, with minor modifications, row-wise control of rack-level fans was executed with input from each server in the stack. Performance of larger fans under different rack loads and failure conditions was reported and savings in power over the baseline configuration were quantified.

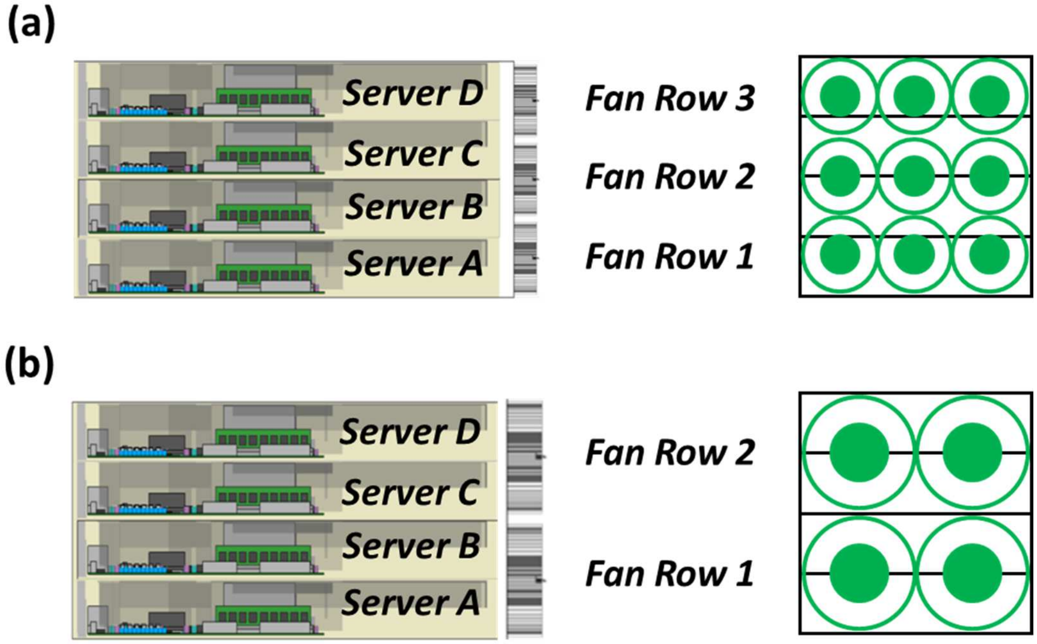


Figure 7-1 Simplified depiction of the fan wall installed behind the stack for both (a) 80mm and, (b) 120mm cases with corresponding names for primary components

## 7.1 Test Setup

As previously discussed in [104], a stack of four servers was considered when evaluating the rack-level solutions. The rear of the stack provided an area of 330mm ×

333mm within which the larger fans were accommodated. Figure 7-1 shows the fan wall installed at a distance of 25.4mm from the rear of the stack for both 80mm (9 units installed in a 3 × 3 array) and 120mm (4 units installed in a 2 × 2 array) cases. Table 7-1 lists specifications of the fans used. A naming scheme was established with servers termed A to D from the bottom of the stack to the top and, each row of fans similarly numbered 1 through 3 (1 and 2 for the 120mm configuration).

Table 7-1 Specifications of fans used

| Setup | Frame (mm) | Max. Air Flow (cfm) | Max. Static Pressure (in. aq.) | Rated Speed (rpm) |
|---|---|---|---|---|
| 60 mm | 60×60×25 | 37.1 | 0.62 | 7600 |
| 80 mm | 80×80×38 | 100.1 | 1.98 | 9500 |
| 120 mm | 120×120×25 | 171.0 | 0.90 | 5100 |

Since the focus of this study was to monitor cooling power consumption, the fans were powered externally as shown in Figure 7-2. However, PWM signals from each server were still used to control the fans and were delivered through a control circuit. This component of the test setup will be explained in detail in a following section. Tachometer output from each fan was logged using a data acquisition (DAQ) unit as well as returned to each server to prevent triggering of a failure scenario (running all remaining fans at full speed to prevent shutdown). It was imperative that the ground (GND) signal from each server and fan be shared between all monitoring and controlling equipment. Since the fans were not powered by the server, a power meter measured the rack (or stack) IT power consumption from a 277VAC source. A workstation communicated with all components in the setup and provided a common timestamp for effective data reduction. An ambient conditions logger recorded air temperature at the inlet to the stack. Over the duration of testing the inlet temperature was found to have a maximum variation of ±1°C with a mean of 25°C.

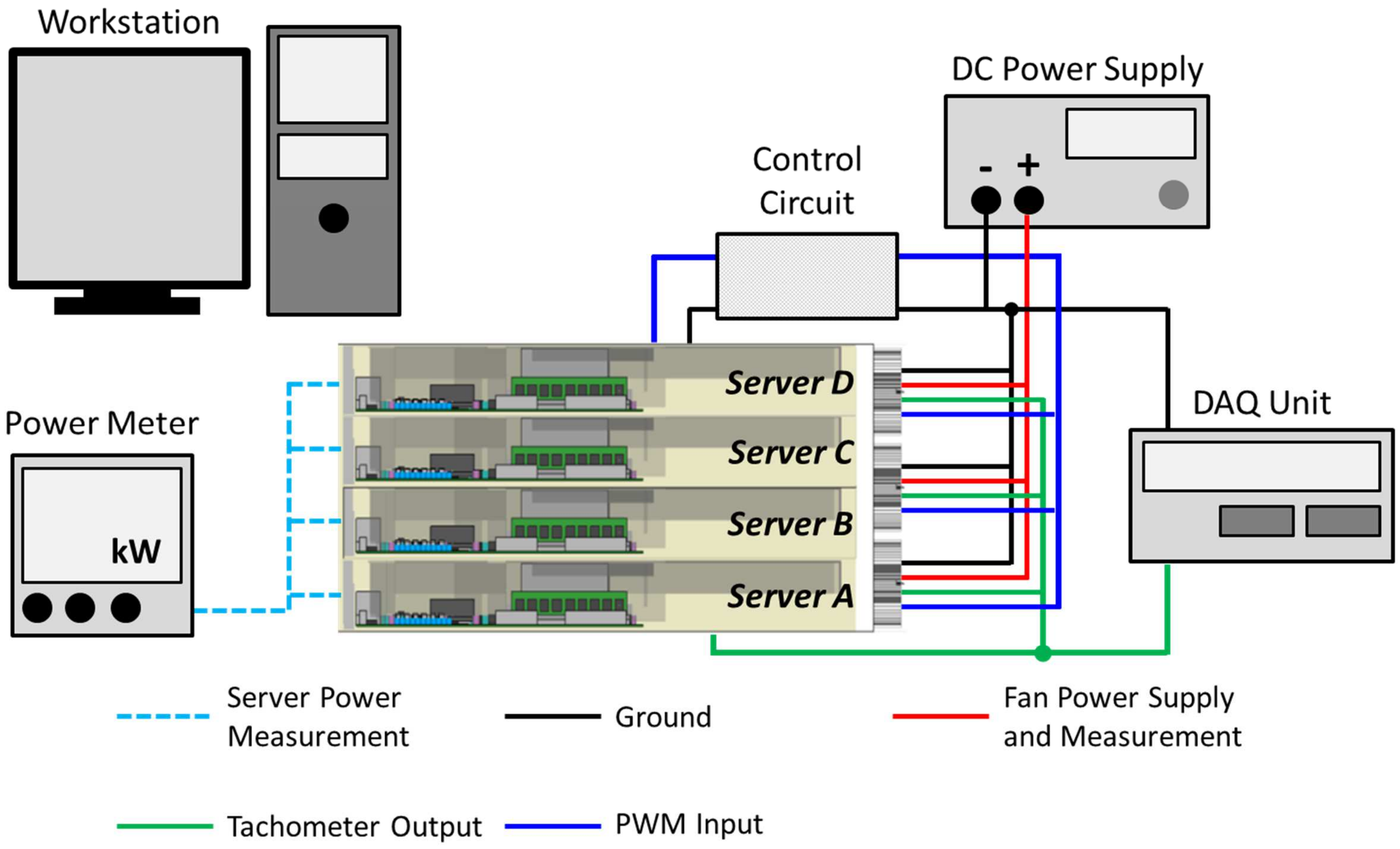


Figure 7-2 Simplified depiction of the test setup with control and data acquisition equipment

## 7.2 Controlling the Fans

### *7.2.1 Row-wise control*

As previously stated, one of the primary objectives was to outline a methodology for row-wise control of larger rack-mount fans. By controlling each row independently based on adjacent server loads, cooling power consumption could be minimized as compared to a scheme where all fan speeds were equally modulated based on the maximum load across the stack. To enable such an arrangement, four PWM signals (one from each server) from the stack needed to be converted to row-wise input; in this case, three signals for the 80mm configuration or two for 120mm counterpart. Contribution of each server's signal to a given row's input needed to be determined to minimize fan power consumption. A simple 'Zone of Influence' test was carried out to determine these parameters.

### *7.2.2 Zone of influence*

Placement of the fan wall behind the stack (with a gap) ensures uniform flow across the rack when all fans are operated at the same speed. This manifests as near-uniform CPU (or operating) temperatures when all servers are subjected to the same load. To study the effect of fan location on cooling across the rack, an individual row was operated at a different speed compared to others. Variation in CPU temperatures would be indicative of influence each row has in terms of proximity. To enable such a comparison, a uniform load was applied to all servers and, in turn, an individual row of fans was operated a higher duty cycle of 30% with the rest running at 10%. The distribution of operating temperatures across the stack for each test in the 80mm configuration is shown in Figure 7-3. For comparison, CPU temperatures for a similar load with all fans operating at 10% were also reported.  As expected, it was clear that fans exert greater influence on servers in higher proximity. These readings were converted to ‘weights’ to meet requirements outlined in the previous section. Difference in operating temperatures between each row-wise test and uniform case (all at 10%) formed the basis to determining the aforementioned parameter.

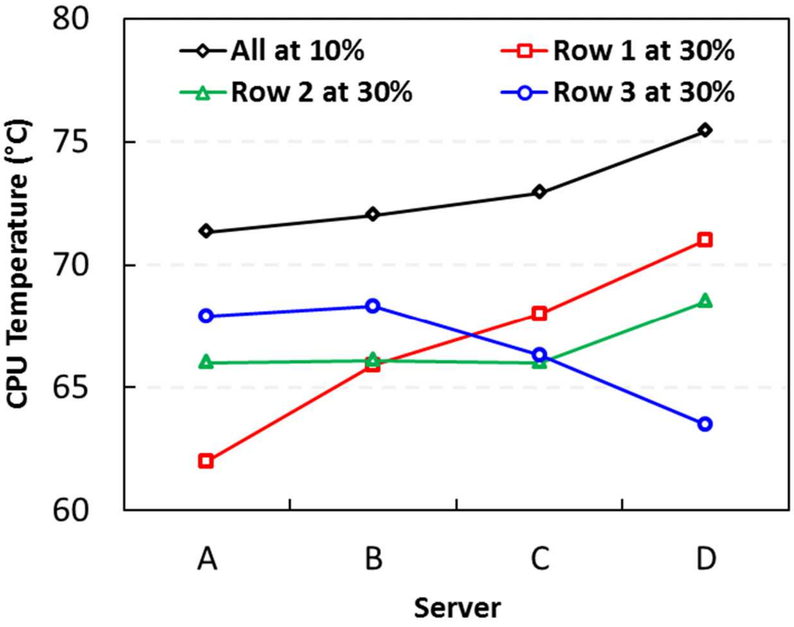


Figure 7-3 Operating temperatures under 98% load when fan rows are individually run at higher speeds

This overcooling term was calculated as follows,

$$\Delta T_{i,j} = T_{i,j} - T_{i,uniform} \qquad (7\text{-}1)$$

Where the subscript ‘i’ corresponds to a given server and ‘j’ to a row of fans running at 30% PWM. The final term in the equation represents CPU temperature for server ‘i’ under uniform fan duty cycles of 10%. The influence (%) was determined by dividing the overcooling terms in each test by the sum as follows,

$$I_i = \frac{\Delta T_{i,j}}{\sum_{i=A}^{D} \Delta T_{i,j}} \qquad (7\text{-}2)$$

Where $I_i$ is the influence on a given server by row 'j' of fans operating at a higher speed. Influence values for all tests are reported in Figure 7-4.

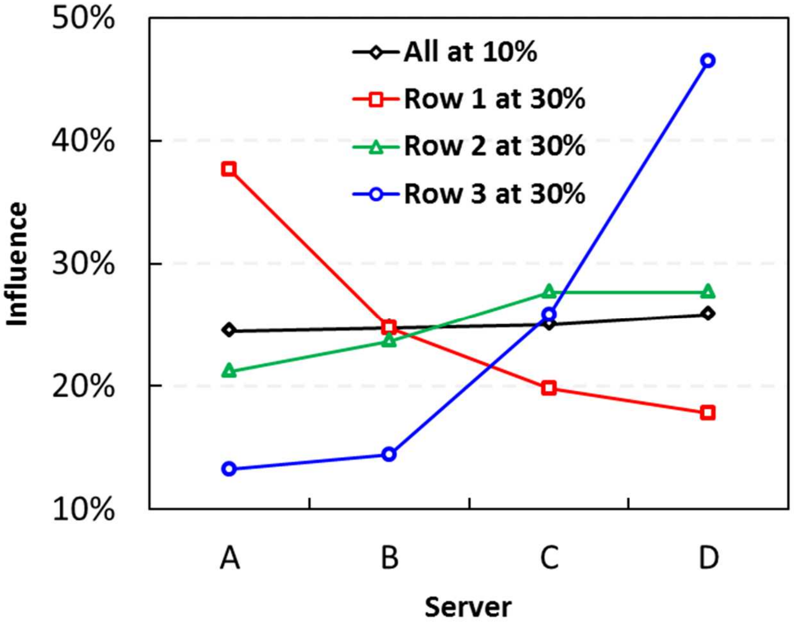
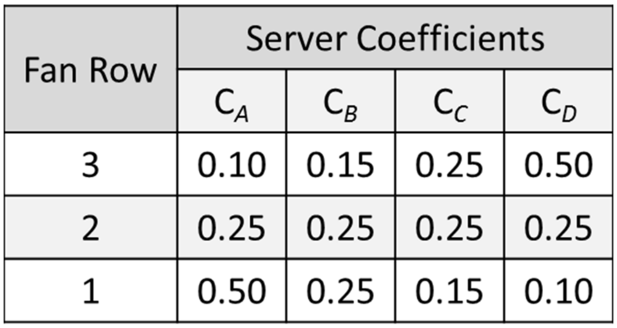


Figure 7-4 Influence of each fan row on server cooling; used to select coefficients for the control scheme

In an ideal state, values for Row 1 and Row 3 would mirror each other and distribution for Row 2 would be uniform (25% across the stack). However, each distribution was skewed towards server D. This was expected, as clearly shown by the CPU temperature variation when all fans were operated at 10% duty cycle in Figure 7-3. These results provided an indication of the tolerances across the stack; specifically in terms of difference in CPU powers. To counter this effect and to provide a scheme that would function across a multitude of servers in a data center, the weight of each server's PWM output with respect to a given fan row was based on the influences reported in

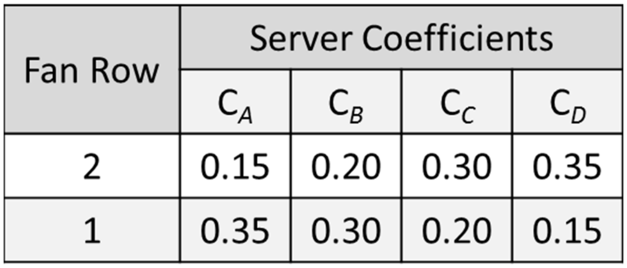
Figure 7-4 and generalized to provide a mirrored form as listed in Tables 7-2 and 7-3. These coefficients were fed to the control circuit to process each PWM input from the stack to a row-wise duty cycle output. The following section will describe the control circuit and system in detail.

Table 7-2 Coefficients (80mm) for each server's duty cycle signal



| Fan Row | Server Coefficients | | | |
|---|---|---|---|---|
| | $C_A$ | $C_B$ | $C_C$ | $C_D$ |
| 3 | 0.10 | 0.15 | 0.25 | 0.50 |
| 2 | 0.25 | 0.25 | 0.25 | 0.25 |
| 1 | 0.50 | 0.25 | 0.15 | 0.10 |

Table 7-3 Coefficients (120mm) for each server's duty cycle signal



| Fan Row | Server Coefficients | | | |
|---|---|---|---|---|
| | $C_A$ | $C_B$ | $C_C$ | $C_D$ |
| 2 | 0.15 | 0.20 | 0.30 | 0.35 |
| 1 | 0.35 | 0.30 | 0.20 | 0.15 |

*7.2.3 Control circuit and operation*

A detailed diagram of the control circuit, using [105], can be seen in Fig. 5. Each server’s output signal was converted to an analog voltage through a low-pass filter. This conversion simplified the circuit as an analog voltage signal was relatively easier to read in comparison to a PWM input. Each filter consisted of a 0.1uF capacitor and two 33kOhm resistors in series to provide a cut-off frequency of around 25Hz. This value was three orders of magnitude lower than the PWM frequency [78] and provided required response as the input frequency did not vary. Analog voltage output ($V_{out,i}$) of the filter was correlated to the duty cycle ($DC_i$, %) of the input signal as follows,

$$V_{out,i} = DC_i \times 5 \tag{7-3}$$

Where 5 VDC corresponds to the logic high of the input signal. Analog input ports in an Arduino microcontroller [64] board were used to measure the processed PWM signal from each server. A LabVIEW [65] program read input to the board and generated PWM output based on the coefficients in Tables 7-2 and 7-3 as follows,

$$DC_j = \sum_{i=A}^{D} C_{i,j} \frac{V_{out,i}}{5} \tag{7-4}$$

Where $DC_j$ is the duty cycle signal to fan row 'j' and $C_{i,j}$ is the weight of server 'i' corresponding to row 'j'. Thus, the control system as described in the previous sections was implemented for both 80mm and 120mm cases.

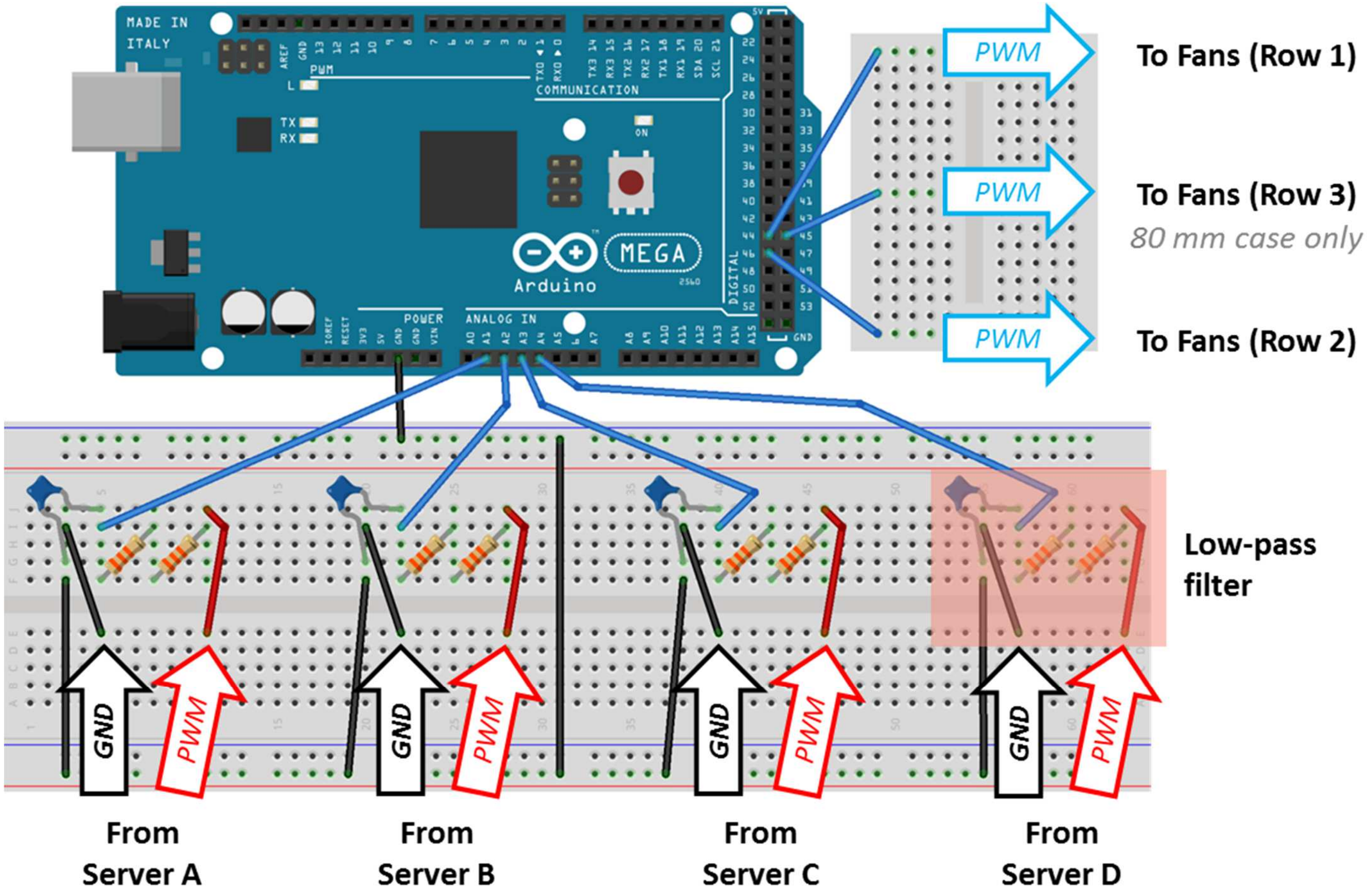


Figure 7-5 PWM signal from each server is converted into an analog voltage by a low-pass filter and read by the microcontroller board. These inputs are processed by a LabView program and a PWM output is sent to each fan row.

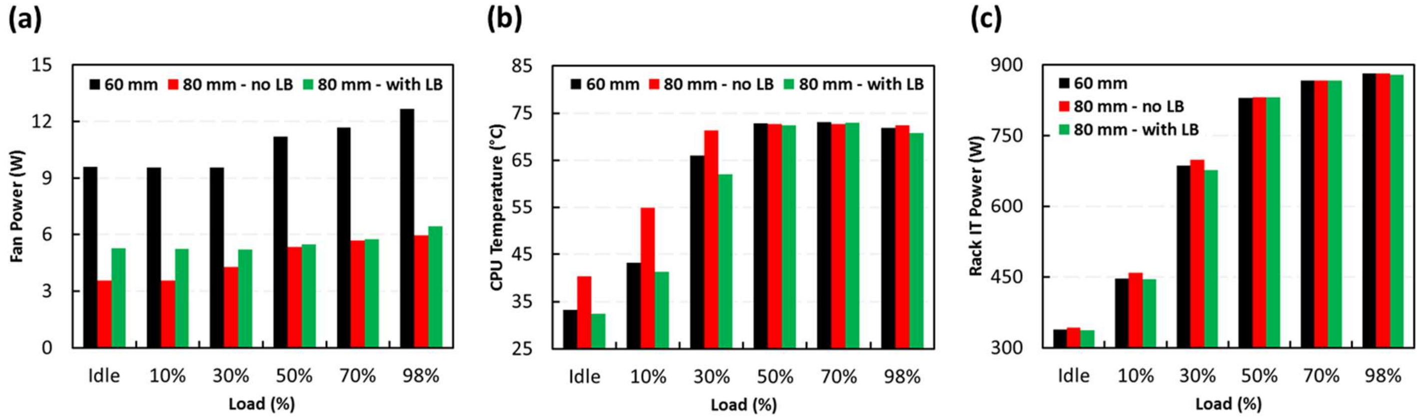


Figure 7-6 Comparisons between (b) operating temperature and, (c) IT power demonstrate that a 7.5% lower bound (LB) in 80 mm fan duty cycle is required

### *7.2.4 Need for a lower bound*

It is critical to remember that, in order to make comparisons between the baseline and rack-level configurations in terms of cooling power consumption, other critical parameters (average CPU temperature and IT power) must be held constant. In an ideal case, maximum available savings for a fixed operating temperature are reported. However, when a control system based on server inputs is implemented, achieving an exact target across all setups is near impossible. Therefore, to provide reasonable comparison between configurations, the 80mm and 120mm control schemes were modified to ensure that CPU temperatures and rack IT power were either equal to or below values reported for the baseline case (60mm). This way, savings in cooling power could be reported without need to discuss remaining critical parameters.

As seen in Figure7-6, for the 80mm configuration, a 7.5% lower bound in the PWM output signals was introduced to meet requirements. The original control scheme outlined in Equation 7-4 provided greater savings in cooling power as compared to the final. However, higher CPU temperatures caused a corresponding increase in IT power through the effect of leakage. The control system was modified to account for the lower bound as follows,

$$DC_j = 0.075 + 0.925 \times \sum_{i=A}^{D} C_{i,j} \frac{V_{out,i}}{5} \tag{7-5}$$

The 120mm fans were inherently configured to produce no change in speed between 10% and idling duty cycles. To account for this and ensure that the output to each row of fans would not lag behind input signals from the rack, the control scheme was modified to account for the lower bound as follows,

$$DC_j = 0.075 + 0.925 \times \sum_{i=A}^{D} C_{i,j} \frac{V_{out,i}}{5} \tag{7-6}$$

No further modification to this system was required as the fans were found to overcool the servers while idling at 10% PWM irrespective of the load as seen in Figures 7-7 to 7-9. Thus, settings for each control system (80mm and 120mm cases) were finalized and deployed for further testing and evaluation.

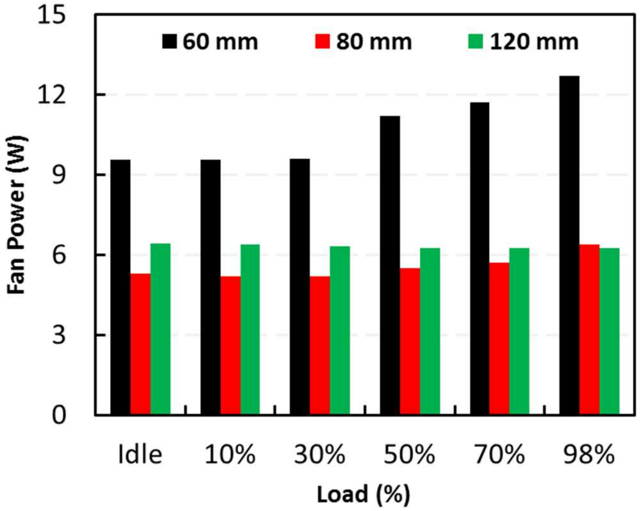


Figure 7-7 Cooling power for all final configurations under uniform load

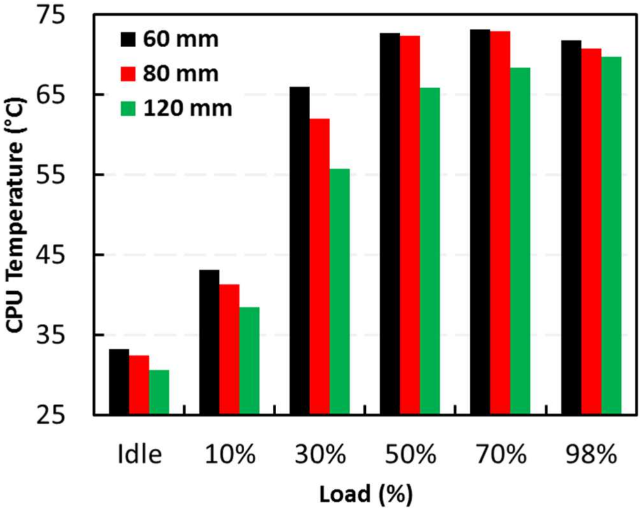


Figure 7-8 Average operating temperature of the stack for all final configurations under uniform load

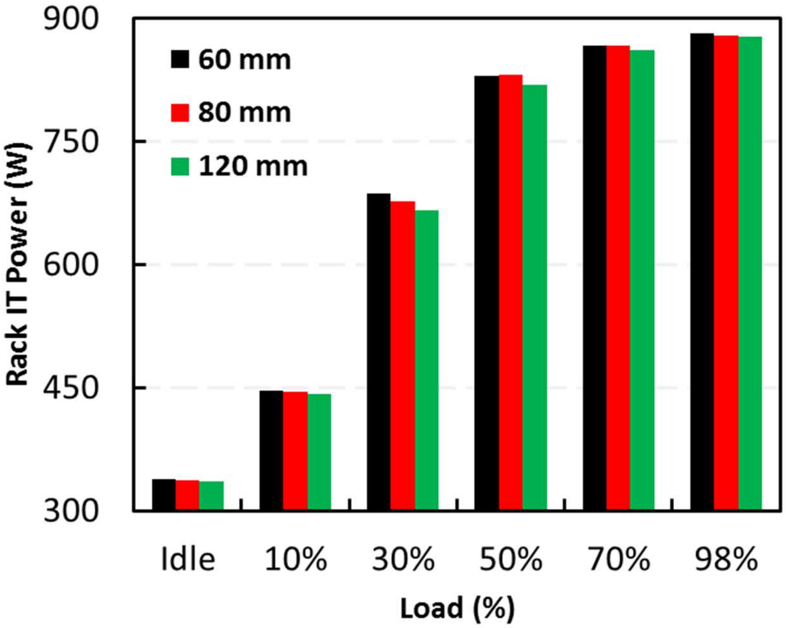


Figure 7-9 IT power consumption for all final configurations under uniform load

## 7.3 Results and Discussion

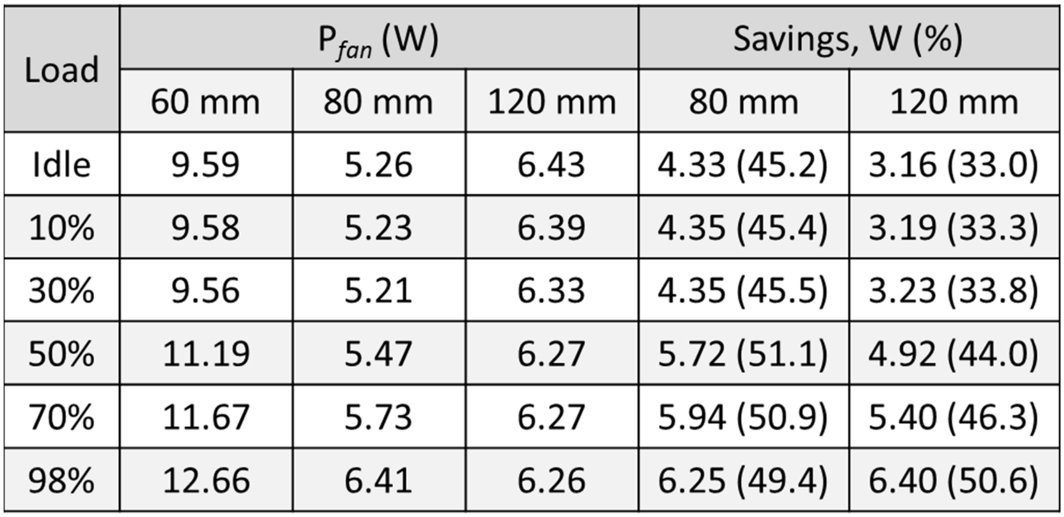
The objective of this study was to confirm that substantial savings are available when 60mm chassis fans are replaced by rack-level configurations with effective control schemes that simulate a product or solution deployed within a data center facility. It was imperative that, in addition to identical utilizations seen in Figures 7-7 to 7-9, realistic conditions such as non-uniform workloads and fan failure scenarios would be investigated to reinforce merits of employing larger fans.

### *7.3.1 Comparison under uniform loading*

It was shown that despite overcooling the servers with rack-level setups, under uniform utilizations, savings in cooling power were available irrespective of the load.

Table 7-4 summarizes the extent of reduction in fan power available through deployment of 80mm and 120mm fans. In either case, it was apparent that savings were maximized when the stack was operated at high workloads. The reason for reduced efficiency at utilizations at or below 30% was because 60mm fans operate at idling speeds in such conditions. In comparison to the baseline, the 80mm case reported savings of around 45% to 51%. Similarly, the 120mm configuration provided reduction in cooling power of the order of 33% to 50%. This setup provided reduced savings at lower loads owing to the fact that these fans operated at idling speeds (10% lower bound) across the entire test spectrum and significantly overcooled the stack for utilizations below 70%.

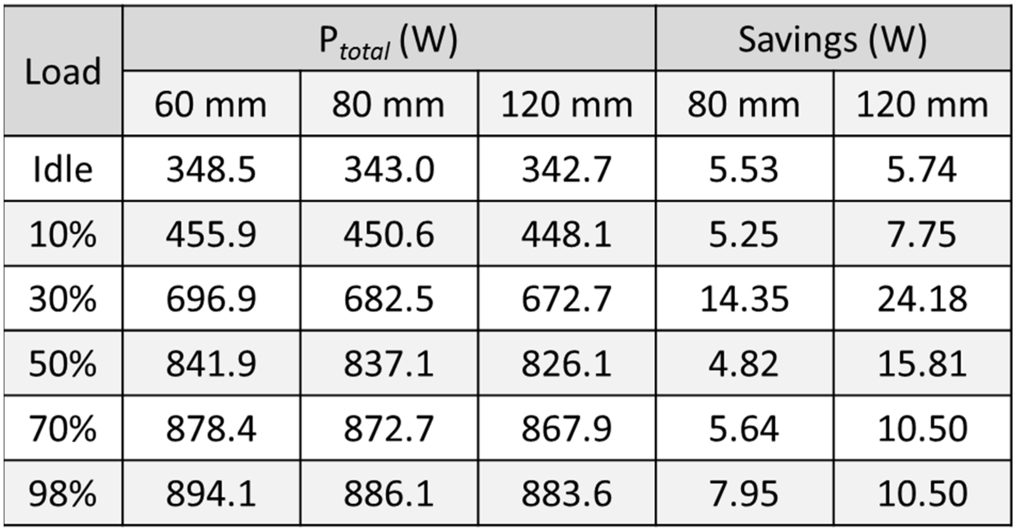
Table 7-4 Savings in cooling power under uniform loads



| Load | $P_{fan}$ (W) | | | Savings, W (%) | |
|---|---|---|---|---|---|
| | 60 mm | 80 mm | 120 mm | 80 mm | 120 mm |
| Idle | 9.59 | 5.26 | 6.43 | 4.33 (45.2) | 3.16 (33.0) |
| 10% | 9.58 | 5.23 | 6.39 | 4.35 (45.4) | 3.19 (33.3) |
| 30% | 9.56 | 5.21 | 6.33 | 4.35 (45.5) | 3.23 (33.8) |
| 50% | 11.19 | 5.47 | 6.27 | 5.72 (51.1) | 4.92 (44.0) |
| 70% | 11.67 | 5.73 | 6.27 | 5.94 (50.9) | 5.40 (46.3) |
| 98% | 12.66 | 6.41 | 6.26 | 6.25 (49.4) | 6.40 (50.6) |

Comparisons were made with ideal conditions (matching operating temperatures across all configurations) to ensure that the deployed control systems did not deviate significantly from reported maximum available savings (by comparing average CPU temperatures). For the 80mm case, a maximum deviation of 5% was observed when compared to projected savings of 50% to 53%. While this reduction could be considered acceptable, it existed due to the effects of overcooling at loads between idling to 30% as seen in Figure 7-8. However, for the 120mm case, a reduction of 15% was observed in comparison to expected savings of 48% to 54%. This was attributed to the fact that in ideal conditions, comparisons between the baseline and 120mm configurations were made with respect to CPU temperatures reported while testing the latter case. It was therefore unfair to draw comparisons between results from matching CPU temperatures and this study for the 120mm setup.

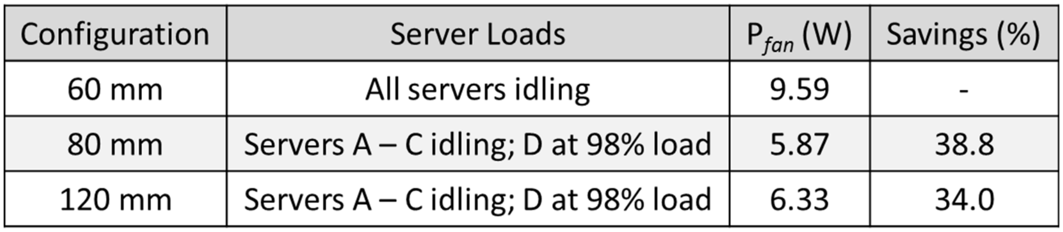
Table 7-5 Total savings under uniform loads



| Load | $P_{total}$ (W) | | | Savings (W) | |
|---|---|---|---|---|---|
| | 60 mm | 80 mm | 120 mm | 80 mm | 120 mm |
| Idle | 348.5 | 343.0 | 342.7 | 5.53 | 5.74 |
| 10% | 455.9 | 450.6 | 448.1 | 5.25 | 7.75 |
| 30% | 696.9 | 682.5 | 672.7 | 14.35 | 24.18 |
| 50% | 841.9 | 837.1 | 826.1 | 4.82 | 15.81 |
| 70% | 878.4 | 872.7 | 867.9 | 5.64 | 10.50 |
| 98% | 894.1 | 886.1 | 883.6 | 7.95 | 10.50 |

When comparing both rack-level configurations from the perspective of cooling power consumption, it would be understandable to claim that the 80mm case was more efficient. However, as seen in Figure 7-9, it was evident that reduction in IT power consumption was available due to overcooling and lower leakage. Therefore, comparisons were made in terms of total power consumption (cooling + IT) as outlined in Table 7-5. In this case, it was apparent that the benefits of reduced IT power more than compensated for lower savings in fan power. However, it must be noted that increased total savings through overcooling was accompanied by an increase in air flow rates through the stack, which would raise cooling power consumption at the facility-level. Therefore, attention was focused at rack-level fan performance only.

Table 7-6 Savings under non-uniform load



| Configuration | Server Loads | $P_{fan}$ (W) | Savings (%) |
|---|---|---|---|
| 60 mm | All servers idling | 9.59 | - |
| 80 mm | Servers A – C idling; D at 98% load | 5.87 | 38.8 |
| 120 mm | Servers A – C idling; D at 98% load | 6.33 | 34.0 |

*7.3.2 Comparison under non-uniform load*

Since it would be time and resource intensive to test each configuration under all possible workloads observed in a data center facility, a simple test was conducted to prove that rack-level fans were more efficient that the baseline under non-uniform loads. This involved comparing the 60mm setup when all servers were idling (lowest fan power consumption) with the larger fan cases when only one server in the stack was operating at maximum utilization. The latter represented an extreme case of non-uniform loading and server D was chosen as it consistently reported CPU temperatures higher than the remaining units (see Figure 7-3). Thus, by reporting savings under such conditions, it would be implicit that rack-level fans under study were superior under non-uniform loads.

Table 7-6 outlines the results from these tests. It was observed that both control schemes provided substantial savings (around 35%) in fan power when compared to the baseline. In addition, these results supported the need for row-wise control as, for the 80mm case, 5% reduction in fan power was observed when compared to a similar test wherein all fans were controlled by server D. Since the fan control did not engage for the 120mm configuration, similar savings were reported in both the aforementioned test and this study.

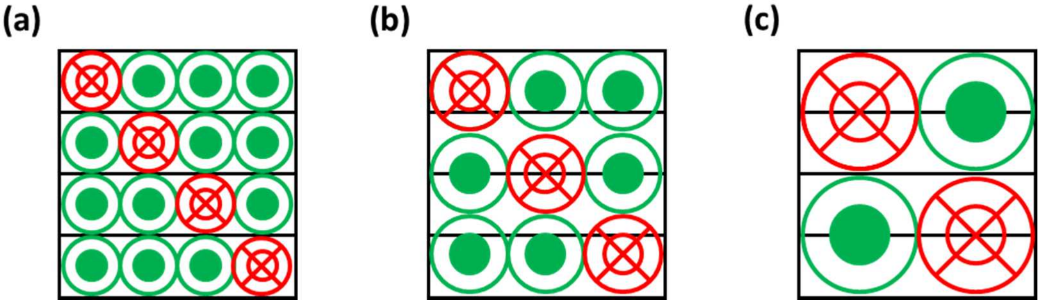


Figure 7-10 Depicting locations of simulated individual fan failures for (a) 60mm, (b) 80mm, and (c) 120mm configurations

### *7.3.3 Fan failure study*

Along with other server components such as hard drives and memory, failure of fans commonly occurs in data center facilities. To account for the same, thermal engineers are required to design cooling systems for servers that ensure uptime even when a fan fails. Therefore, it was imperative that all configurations under study were tested while simulating failure to ensure no detriment to performance of the system such as throttling, increase in power consumption, etc. Figure 7-10 provides an illustration of locations of single fan failures for all configurations. The diagonal pattern was chosen to reduce testing time as power consumption under failure was found to be independent of location [104]. For the rack-level setups, preliminary trials were found to be in agreement as well. For each configuration, results from all tests were averaged and reported in Table 7-7. It was observed that failure of a single fan had a marginal effect on power consumption for the baseline and 80mm cases. However, for the 120mm setup, a 33% increase in fan power was observed under simulated failure. This could be attributed to lower available redundancy caused by the restriction in available area at the rear of the stack that limited the number of units to four. It was important to note that, while the difference in performance between 80mm and 120mm cases was apparent, the frequency of failures and time between failure and replacement were equally important when making a decision between configurations.

Table 7-7 Fan failure testing

| Test Condition | $P_{fan}$ (W) | | |
|---|---|---|---|
| | 60 mm | 80 mm | 120 mm |
| No Failures | 12.66 | 6.41 | 6.26 |
| One Fan Failed | 12.81 | 6.23 | 8.33 |
| Penalty (%) | 1.19 | -2.81 | 33.07 |

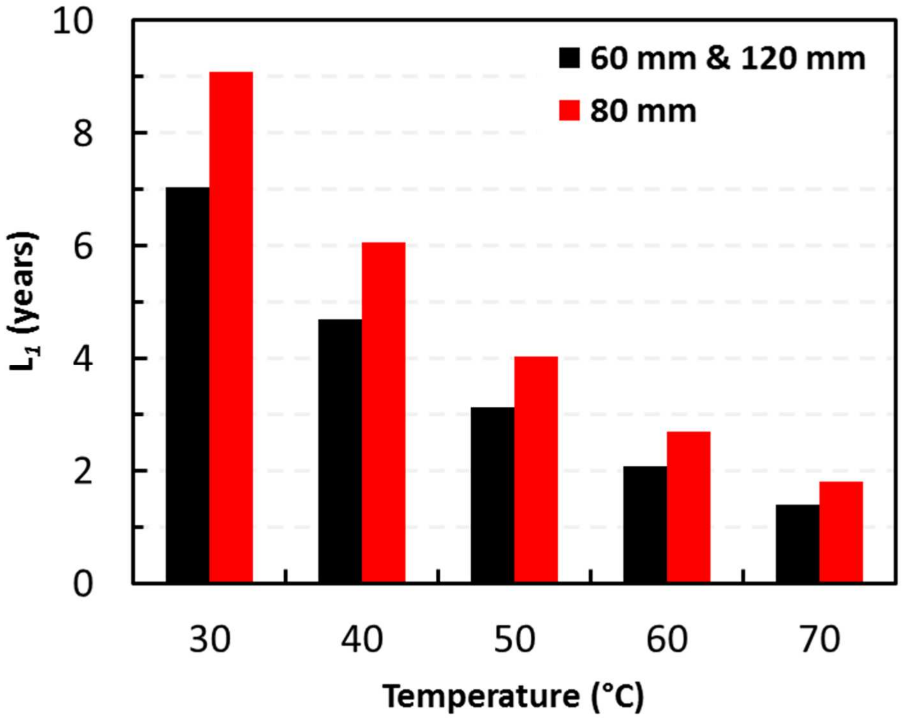


Figure 7-11 Life expectance of chosen fans as a function of operating temperature

*7.3.4 Further discussion*

As previously discussed, multiple factors must be considered before drawing conclusions from a fan failure study. One such statistic is the life expectance of selected fans. Manufacturers publish L10 (time at which 10% of the fans will fail) and MTTF (mean time to failure) along with specifications to aid selection based on product lifecycle. Based on operating conditions of the fan, life expectance can be scaled up or down using an acceleration factor (AF) of 1.5 for every 10°C change in temperature [89].

$$AF = 1.5^{\left[\frac{T_{test}-T_{use}}{10}\right]} \tag{7-7}$$

Where $T_{test}$ is the temperature at which manufacturers conduct tests to report data and $T_{use}$ is the temperature at which the fan will be operated in a system. Data centers are mission critical facilities and, as such, component failure rates of 10% may be considered unacceptable. Therefore, $L_{10}$ values reported by manufacturers were scaled to more stringent requirements such as $L_1$ (time at which 1% of the fans will fail) [106].

$$L_{10} = L_1 \times \left(\frac{ln(0.99)}{ln(0.90)}\right)^{1/\beta} \tag{7-8}$$

Where β is the Weibull shape parameter and chosen to be 3 in this case [89, 107]. $L_1$ values for all three fans as a function of operating temperature are plotted in Fig. 9. 80mm fans have 22.5% greater life expectance than 60mm and 120mm counterparts. However, this does not translate to lower failures than the 120mm units. Considering there are more than twice the number of parts in the 80mm fan wall, more failures would be expected due to the sheer difference in quantity. In addition, since 120mm fans operated at a fixed duty cycle (at ambient temperatures of 25°C or below), and changes in speed are detrimental to fan life, lower failure rates could be expected.

Previous discussions have made valid arguments for choosing either rack-level configuration over the baseline case. However, making a selection between the larger fans is not straightforward. A decision between configurations would require consideration of advantages outlined herein as well as factors that exist beyond the scope of this study. Accommodation of all parameters would necessitate a study of total cost of ownership (TCO). Setup and results from this study would factor in both capital and operational expenditure, based on which a decision could be made. Regardless, based on reported observations, maximizing savings through deployment of rack-level fans could be achieved through a compromise between efficiency, redundancy and cost. These terms are ultimately dependent on the size of fans selected.

# CHAPTER 8

# CONCLUSIONS

To tackle the growth of the data center industry and continued strain on the national electricity grid, there is a need to promote energy-efficiency within such facilities. Specifically, targeting improvements in cooling systems at different levels with the data center is imperative. In the reported series of studies, module, server and rack-level evaluations of single-phase air and liquid cooling solutions are outlined as follows.

## 8.1 Dynamic Cold Plates for High Power Modules

The concept for a dynamic cold plate, which relies on segregated flow control of cooling resources based on requirement for thermal management of high power devices, was introduced. A multichip module was selected as the reference platform for designing such a solution. Performance parameters of the original cold plate were previewed and provided insight regarding the extent of savings available based on predictions from numerical analysis of the dynamic counterpart. Considerations were made to account for limitations of the manufacturing procedure employed to fabricate samples of the proposed solution. Conjugate heat transfer analysis of the cold plate assembled with the MCM reported moderate operating temperatures of components with low impedance to flow in individual sections. Specifically, characteristics of fluid flow in the designed cold plate predicted significant improvements over the original solution. However, concrete conclusions on performance necessitate experimental testing. Specifics for cost-effective design and fabrication of a MCM thermal test vehicle were outlined. Control circuits that enabled automation of testing through modulation of heaters and cooling systems were demonstrated with a simplified test setup. Details of the liquid cooling test bench for evaluation of both original and dynamic cold plates were included. The test matrix and future work were also discussed.

### 8.2 Effects of Elevated Rack Inlet Temperatures on Performance of Web Servers

The influence of raising rack inlet temperature was exhibited in increased server power consumption less through the effect of leakage and to a greater extent through increased fan power consumption. Beyond an RIT of 35°C and for moderate to high server loads (30% or more), around 4% to 14% increase in power with respect to corresponding values at 15°C were seen. It was observed that cooling (or flow rate) was an equally important parameter to consider when deliberating RIT due to its influence on cooling system blower power consumption. Based on server performance, it seemed prohibitive to operate the data hall beyond 30°C for low to medium server utilizations (up to 70%) under the restriction of doubling facility fan power. At maximum load, this upper bound became 25°C. Studying efficiency (pPUE) of the server delivered similar conclusions as those based on power consumption. Increase in equipment noise levels was found to vary between the influence of PSU and chassis fans. Regardless, irrespective of RIT or load, reported values were within regulation standards (85dBA for 8 hours of continuous exposure). Each server's cooling system was determined to operate as intended with critical motherboard components operating at reliable temperatures, well below specified limits. However, reliability of fans and hard disk drives could be a concern only at high utilizations.

Selecting the optimal range of set-point temperatures for a data center requires evaluation of TCO due to the tradeoff between installed capacities (CapEx) and operating conditions (OpEx) to maximizing savings. In summary, it is critical to evaluate IT equipment performance across the intended range of RITs before initiating changes in settings of facility cooling systems.

## 8.3 Understanding the Tradeoff between Cooling Power and Leakage Current

By controlling the chassis fans externally from idling to maximum speeds, a distribution of total server power consumption ($P_{Total}$) versus CPU die temperature was reported. It was observed that a distinct minimum in $P_{Total}$ was seldom observed as fan power consumption exerted greater influence than effects of leakage current. Also, comparing the distribution with corresponding operating points when fans were controlled internally by the chassis fan control scheme, showed that savings in both power and air flow rate could be available by increasing CPU operating temperature. Parameters of the deadband control were increased by 5°C while ensuring that component temperatures were maintained within reliable limits. Primarily, savings or improvements available were due to reduction in chassis fan speeds. Specifically, at higher computational loads (50% or more) and rack inlet temperatures (35°C and 45°C), fan speeds were lowered by around 18% to 53% when compared to the original control scheme. As a result, air flow rate through each server and sound pressure levels were reduced by 16% to 45% and 4% to 27% respectively. Reported savings varied between servers due to difference in chassis fans employed. Reduction in server air flow rate requirements propagated to the facility level with blower power of cooling systems lowered by around 40% to 83%. While reported savings suggested the need to change chassis fan control settings, it was important to recognize corresponding effects on component and system reliability due to higher operating temperatures. An analysis of operating expenditure (OpEx) is required to ensure that improvements in power consumption (server and facility levels) are not negated by increased failure of components.

## 8.4 Savings in Cooling Power through Deployment of Rack-Level Fans

Configurations of larger fans, setup to mirror a product that could be deployed in a data center facility, were shown to provide greater efficiency in cooling web servers traditionally configured to use smaller (60mm) chassis fans. In unison with [104], a detailed methodology that outlined selection of fans, prediction and validation of savings, and setup for server-dependent operation was previewed. A control scheme was implemented for both 80mm and 120mm cases that delivered savings in cooling power consumption of the order of 45% to 51% and 33% to 50% respectively under uniform rack load. Testing both rack-level setups under highly non-uniform load reported savings of around 35% when compared to baseline (60mm) fan power for idling utilization. Simulation of fan failure showed marginal penalties in performance for both 60mm and 80mm cases. However, a 33% increase in cooling power was observed for the 120mm configuration and attributed to lack of available redundancy (fan count). A summary of advantages in support of either rack-level setup is included below.

**Advantages of 80mm configuration**

- Consistent savings in fan power across a spectrum of loads
- No penalty or increase in cooling power for single fan failure
- 22.5% greater life expectance per fan
- Less overcooling corresponds to lower air flow requirement which influences facility-level cooling power consumption

**Advantages of 120mm configuration**

- Greater overcooling gives rise to decrease in IT power consumption due to reduced leakage. These savings more than compensate for relatively higher fan power at lower loads.
- At ambient temperatures of 25°C and below, fans will idle (10% duty cycle). Lack of change in speed will increase life expectance.
- With less than half the number of fans (4) in the wall, lower number of total failures is expected

Analysis of total cost of ownership would be required to make an informed decision between the two rack-level solutions. A final selection would represent a compromise between efficiency, redundancy and cost.

# APPENDIX A

# NOTES AND SUPPLEMENTARY FIGURES

Funding

Work presented here is supported by the National Science Foundation Industry/University Collaborative Research Center for Energy Smart Electronic Systems (ES2) under Grant No. IIP - 1134821. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

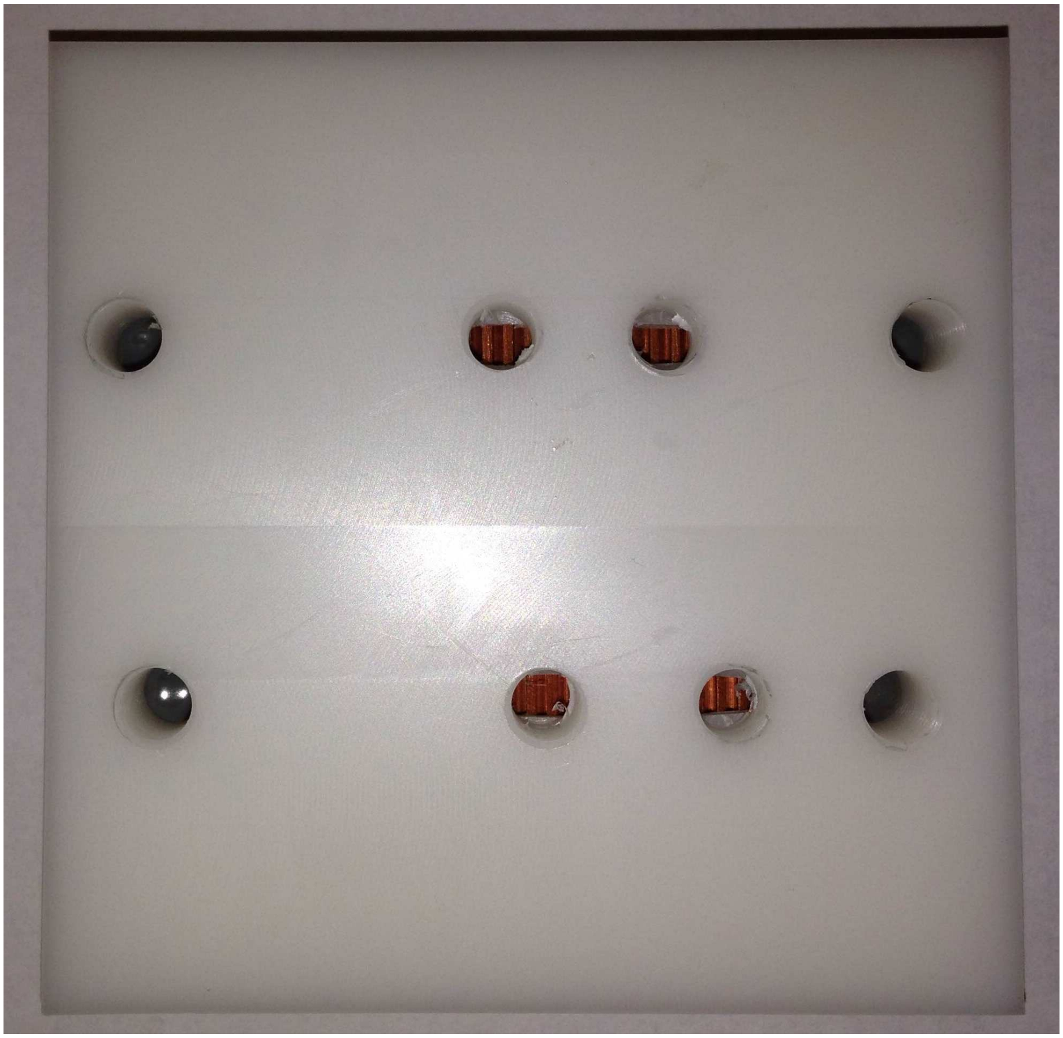


Figure A-1 Fabricated sample of finalized cold plate design

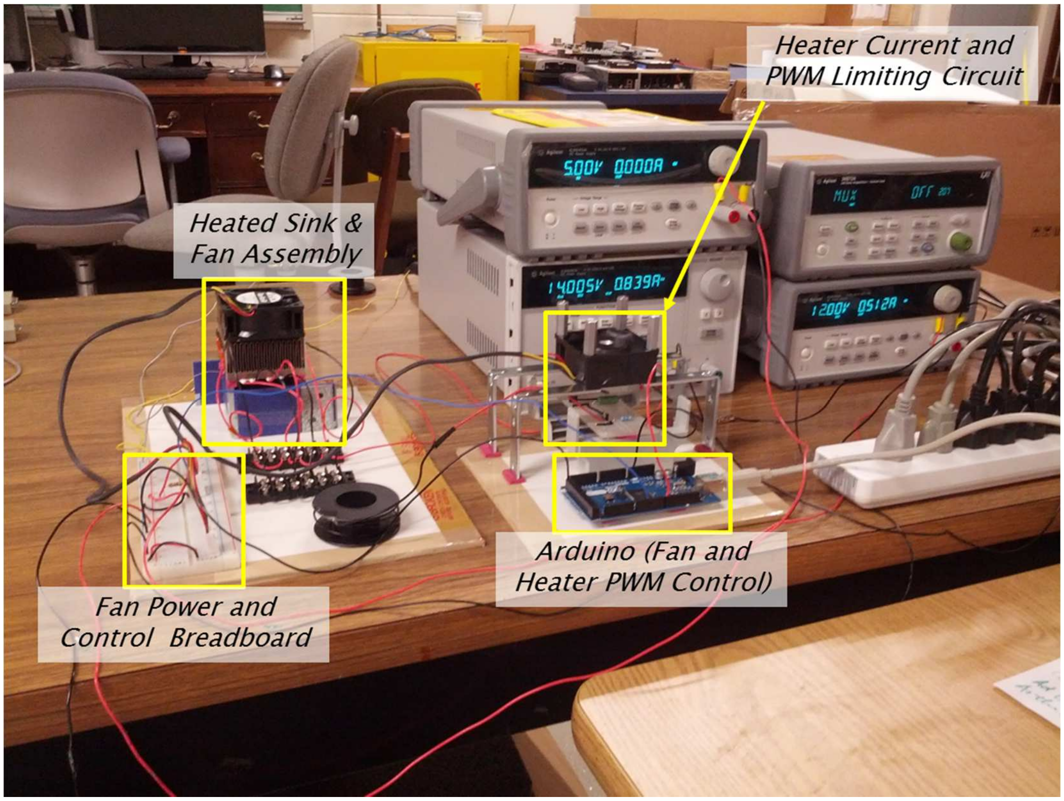
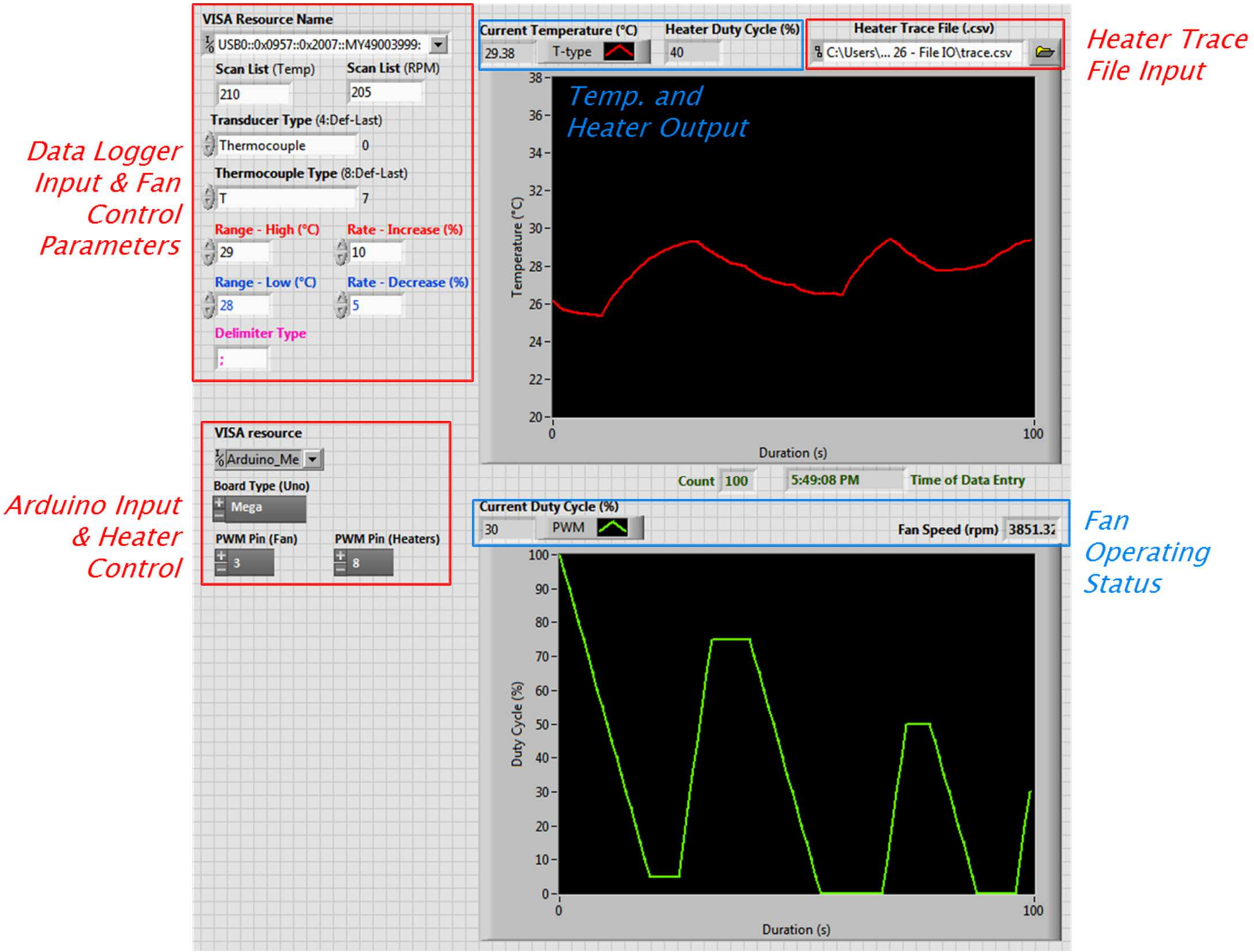


Figure A-2 Setup and components of heater and fan control study



Data Logger Input & Fan Control Parameters

VISA Resource Name: USB0::0x0957::0x2007::MY49003999:

Scan List (Temp): 210

Scan List (RPM): 205

Transducer Type (4:Def-Last): Thermocouple 0

Thermocouple Type (8:Def-Last): T 7

Range - High (°C): 29

Rate - Increase (%): 10

Range - Low (°C): 28

Rate - Decrease (%): 5

Delimiter Type: ;

Arduino Input & Heater Control

VISA resource: Arduino_Me

Board Type (Uno): Mega

PWM Pin (Fan): 3

PWM Pin (Heaters): 8

Current Temperature (°C): 29.38 T-type

Heater Duty Cycle (%): 40

Heater Trace File (.csv): C:\Users\... 26 - File IO\trace.csv

Heater Trace File Input

Temp. and Heater Output

Count 100

5:49:08 PM Time of Data Entry

Current Duty Cycle (%): 30 PWM

Fan Speed (rpm): 3851.3

Fan Operating Status

Figure A-3 Details of LabVIEW front panel

# REFERENCES

[1] ASHRAE TC 9.9, Thermal Guidelines for Data Processing Environments, Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers Inc., 2011.

[2] ASHRAE TC 9.9, Datacom Equipment Power Trends and Cooling Applications, Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers Inc., 2005.

[3] ASHRAE TC 9.9, "IT Equipment Thermal Management and Controls," American Society of Heating, Refrigeration and Air-Conditioning Engineers Inc., Atlanta, 2012.

[4] R. Brown, "Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431," 2008.

[5] J. Koomey, "Growth in data center electricity use 2005 to 2010," Analytical Press, completed at the request of The New York Times, 2011.

[6] DatacenterDynamics Intelligence, "DCD Industry Census 2012: Energy," *DatacenterDynamics Focus,* pp. 38-41, November 2012.

[7] Digital Realty, "N. America Campos Survey Results," January 2013. [Online]. Available: http://content.digitalrealty.com/sfc/servlet.shepherd/version/download/06880000000k573. [Accessed 23 January 2015].

[8] J. Scaramella and M. Eastwood, "Solutions for the Datacenter's Thermal Challenges," January 2007. [Online]. Available: https://www-935.ibm.com/services/fr/igs/pdf/idc_opinion_coolblue_wp.pdf. [Accessed 23 January 2015].

[9] S. Pelley, D. Meisner, T. F. Wenisch and J. W. VanGilder, "Understanding and abstracting total data center power," in *Workshop on Energy-Efficient Design*, 2009.

[10] M. Iyengar, M. David, P. Parida, V. Kamath, B. Kochuparambil, D. Graybill, M. Schultz, M. Gaynes, R. Simons, R. Schmidt and T. Chainer, "Server liquid cooling with chiller-less data center design to enable significant energy savings," in *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, San Jose, CA, USA, 2012.

[11] M. Iyengar and R. R. Schmidt, "Analytical Modeling of Energy Consumption and Thermal Performance of Data Center Cooling Systems: From the Chip to the Environment," in *ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference*, Vancouver, British Columbia, Canada, 2007.

[12] J. Kaiser, J. Bean, T. Harvey, M. Patterson and J. Winiecki, "Survey Results: Data Center Economizer Use," 2011. [Online]. Available: http://www.thegreengrid.org/~/media/WhitePapers/WP41-SurveyResultsDataCenterEconomizerUse.pdf. [Accessed 23 January 2015].

[13] T. Harvey, M. Patterson and J. Bean, "Updated Air-Side Free Cooling Maps: The Impact of ASHRAE 2011 Allowable Ranges," 2012. [Online]. Available: http://www.thegreengrid.org/~/media/WhitePapers/WP46UpdatedAirsideFreeCoolingMapsTheImpactofASHRAE2011AllowableRanges.pdf. [Accessed 23 January 2015].

[14] ASHRAE TC 9.9, Green Tips for Data Centers, Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers Inc., 2011.

[15] S.-C. Lin and K. Banerjee, "Cool chips: opportunities and implications for power and thermal management," *IEEE Transactions on Electron Devices,* vol. 55, no. 1, pp. 245-255, 2008.

[16] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir and V. Narayanan, "Leakage current: Moore's law meets static power," *Computer,* vol. 36, no. 12, pp. 68-75, 2003.

[17] B. Agostini, M. Fabbri, J. E. Park, L. Wojtan, J. R. Thome and B. Michel, "State of the art of high heat flux cooling technologies," *Heat Transfer Engineering,* vol. 28, no. 4, pp. 258-281, 2007.

[18] R. Simons, K. Moran, V. Antonetti and R. Chu, "Thermal Design of the IBM 3081 Computer," in *National Electronic Packaging and Production Conference*, Anaheim, CA, USA, 1982.

[19] R. Chu, U. Hwang and R. Simons, "Conduction Cooling for an LSI Package: A One Dimensional Approach," *IBM Journal of Research and Development,* vol. 26, no. 1, pp. 45-54, 1982.

[20] U. Hwang and K. Moran, "Cold Plate for IBM Thermal Conduction Module Electronic Modules," *Heat Transfer in Electronic and Microelectronic Equipment,*

vol. 29, pp. 495-508, 1990.

[21] D. Delia, T. Gilgert, N. Graham, U. Hwang, P. Ing, J. Kan, R. Kemink, G. Maling, R. Martin, K. Moran, J. Reyes, R. Schmidt and R. Steinbrecher, "System cooling design for the water-cooled IBM Enterprise System/9000 processors," *IBM Journal of Research and Development,* vol. 36, no. 4, pp. 791-803, 1992.

[22] R. R. Schmidt, "Liquid Cooling is Back," 1 August 2005. [Online]. Available: http://www.electronics-cooling.com/2005/08/liquid-cooling-is-back/. [Accessed 23 January 2015].

[23] L. A. Barroso and U. Holzle, "The case for energy-proportional computing," *Computer,* vol. 40, no. 12, pp. 33-37, 2007.

[24] D. Copeland, "Review of Low Profile Cold Plate Technology for High Density Servers," 1 May 2005. [Online]. Available: http://www.electronics-cooling.com/2005/05/review-of-low-profile-cold-plate-technology-for-high-density-servers/. [Accessed 23 January 2015].

[25] J. Fernandes, S. Ghalambor, D. Agonafer, V. Kamath and R. Schmidt, "Mutli-Design Variable Optimization for a Fixed Pumping Power of a Water-Cooled Cold Plate for High Power Electronics Applications," in *IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, San Diego, CA, USA, 2012.

[26] U. Hwang, K. Moran and R. Kemink, "Cold Plate Design for IBM ES/9000 TCM Electronic Modules," *Advances in Electronic Packaging,* pp. 75-81, 1992.

[27] J. Fernandes, S. Ghalambor, A. Docca, C. Aldham, D. Agonafer, E. Chenelly, B. Chan and M. Ellsworth, "Combining Computational Fluid Dynamics (CFD) and Flow Network Modeling (FNM) for Design of a Multi-Chip Module (MCM) Cold Plate," in *ASME International Electronic Packaging Technical Conference and Exhibition*, Burlingame, CA, USA, 2013.

[28] R. Remsburg, "Nonlinear Fin Patterns Keep Cold Plates Cooler," 1 February 2007. [Online]. Available: http://powerelectronics.com/thermal-management/nonlinear-fin-patterns-keep-cold-plates-cooler. [Accessed 23 January 2015].

[29] S. Karajgikar, D. Agonafer, K. Ghose, B. Sammakia, C. Amon and G. Refai-Ahmed, "Multi-Objective Optimization to Improve Both Thermal and Device Performance of a Nonuniformly Powered Micro-Architecture," *Journal of Electronic Packaging,* vol. 132, no. 2, p. 021008 (8 pages), 2010.

[30] A. K. Coskun, T. S. Rosing and K. C. Gross, "Proactive temperature balancing for low cost thermal management in MPSoCs," in *International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, 2008.

[31] A. K. Coskun, J. L. Ayala, D. Atienza and T. S. Rosing, "Modeling and dynamic management of 3D multicore systems with liquid cooling," in *International Conference on Very Large Scale Integration (VLSI-SoC)*, Florianopolis, 2009.

[32] A. K. Coskun, D. Atienza, T. S. Rosing, T. Brunschwiler and B. Michel, "Energy-efficient variable-flow liquid cooling in 3D stacked architectures," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, 2010.

[33] S. Kabbani, R. Beyerle and D. Bachelder, "Active thermal control system with miniature liquid-cooled temperature control device for electronic device testing". United States of America Patent US 7355428 B2, 8 April 2008.

[34] T. J. Chainer, M. P. David, M. K. Iyengar, P. R. Parida, R. R. Schmidt and M. D. Schultz, "Dynamically limiting energy consumed by cooling apparatus". United States of America Patent US 20130138252 A1, 30 May 2013.

[35] T. J. Chainer, M. P. David, M. K. Iyengar, P. R. Parida and R. E. Simons, "Coolant and ambient temperature control for chillerless liquid cooled data centers". United States of America Patent US 20130264046 A1, 10 October 2013.

[36] T. J. Chainer, M. K. Iyengar and P. R. Parida, "Thermally determining flow and/or heat load distribution in parallel paths". United States of America Patent US 20140146845 A1, 29 May 2014.

[37] "CoolIT Systems Inc.," 2014. [Online]. Available: http://www.coolitsystems.com/. [Accessed 23 January 2015].

[38] "Asetek," 2015. [Online]. Available: http://asetek.com/. [Accessed 23 January 2015].

[39] P. Tuma, "The merits of open bath immersion cooling of datacom equipment," in *IEEE Semiconductor Thermal Measurement and Management Symposium*, Santa Clara, CA, USA, 2010.

[40] M. M. Ohadi, S. V. Dessiatoun, K. Choo, M. Pecht and J. V. Lawler, "A comparison analysis of air, liquid, and two-phase cooling of data centers," in *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, Orlando, FL, 2012.

[41] H. Coles, M. Ellsworth and D. J. Martinez, ""Hot" for Warm Water Cooling," in *State of the Practice Reports*, Seattle, WA, 2011.

[42] M. K. Patterson, "The effect of data center temperature on energy efficiency," in *Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*, Orlando, FL, 2008.

[43] N. El-Sayed, I. A. Stefanovici, G. Amvrosiadis, A. A. Hwang and B. Schroeder, "Temperature management in data centers: why some (might) like it hot," in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*, London, 2012.

[44] R. Eiland, J. Fernandes, M. Vallejo, D. Agonafer and V. Mulay, "Flow rate and inlet temperature considerations for direct immersion of a single server in mineral oil," in *IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Orlando, FL, USA, 2014.

[45] A. Addagatla, J. Fernandes, D. Mani, R. Eiland, D. Agonafer and V. Mulay, "Effect of warm water cooling for an isolated hybrid liquid cooled server," in *IMAPS Advanced Technology Workshop and Tabletop Exhibit on Thermal Management*, Los Gatos, CA, 2014.

[46] J. E. Fernandes, M. Sahini, D. Agonafer, V. Mulay, J. Na, P. McGinn, M. Soares and C. Turner, "Evaluating Liquid Cooling at the Rack," in *IMAPS Advanced Technology Workshop and Tabletop Exhibit on Thermal Management*, Los Gatos, CA, 2014.

[47] S. Strutt, C. Kelley, H. Singh and V. Smith, "Data Center Efficiency and IT Equipment Reliability at Wider Operating Temperature and Humidity Ranges," 23 October 2012. [Online]. Available: http://www.thegreengrid.org/en/Global/Content/white-papers/WP50-DataCenterEfficiencyandITEquipmentReliabilityatWiderOperatingTemperatureandHumidityRanges. [Accessed 23 January 2015].

[48] D. Atwood and J. G. Miner, "Reducing Data Center Cost with an Air Economizer," August 2008. [Online]. Available: http://www.intel.com/content/www/us/en/data-center-efficiency/data-center-efficiency-xeon-reducing-data-center-cost-with-air-economizer-brief.html. [Accessed 23 January 2015].

[49] B. Schroeder, E. Pinheiro and W.-D. Weber, "DRAM errors in the wild: a large-scale field study," in *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems (SIGMETRICS)*, Seattle, WA,

2009.

[50] S. Sankar, M. Shaw and K. Vaid, "Impact of temperature on hard disk drive reliability in large datacenters," in *International Conference on Dependable Systems & Networks (DSN)*, Hong Kong, 2011.

[51] B. Schroeder and G. A. Gibson, "Disk failures in the real world: What does an MTTF of 1, 000, 000 hours mean to you?," *FAST,* vol. 7, pp. 1-16, 2007.

[52] B. Schroeder and G. A. Gibson, "A large-scale study of failures in high-performance computing systems," *IEEE Transactions on Dependable and Secure Computing,* vol. 7, no. 4, pp. 337-350, 2010.

[53] S. Sankar, M. Shaw, K. Vaid and S. Gurumurthi, "Datacenter scale evaluation of the impact of temperature on hard disk drive failures," *ACM Transactions on Storage (TOS),* vol. 9, no. 2, 2013.

[54] A. Waseem and Y. W. Wu, "A survey on reliability in distributed systems," *Journal of Computer and System Sciences,* vol. 79, no. 8, pp. 1243-1255, 2013.

[55] R. Mahajan, C.-P. Chiu and G. Chrysler, "Cooling a microprocessor chip," *Proceedings of the IEEE,* vol. 94, no. 8, pp. 1476-1486, 2006.

[56] "i3 Electronics, Inc.," 2014. [Online]. Available: http://www.i3electronics.com/. [Accessed 23 January 2015].

[57] M. Reeves, J. Moreno, P. Beucher, S.-J. Loong and D. Brown, "Investigation on the Impact on Thermal Performances of New Pin and Fin Geometries Applied to Liquid Cooling of Power Electronics," 2011. [Online]. Available: http://www.microcooling.com/images/pdfs/technical_articles/MicroCool_PCIM_2011.pdf. [Accessed 23 January 2015].

[58] "Micro Deformation Technology," [Online]. Available: http://www.microcooling.com/technology/micro-deformation-manufacturing/micro-deformation-technology. [Accessed 23 January 2015].

[59] FLUENT is a product of ANSYS, Inc., 2600 ANSYS Drive, Canonsburg, PA 15317, USA.

[60] Workbench is a product of ANSYS, Inc., 2600 ANSYS Drive, Canonsburg, PA 15317.

[61] M.-C. Lu, B.-C. Yang and C.-C. Wang, "Numerical study of flow mal-distribution on the flow and heat transfer for multi-channel cold-plates," in *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, San Jose, CA, 2004.

[62] 6SigmaET is a product of Future Facilities Ltd., 1 Salamanca Street, London SE1 7HX.

[63] "CPR-500-1 Stripline Chip Resistors," 2015. [Online]. Available: http://www.componentgeneral.com/CPR-500-1-Stripline-Chip-Resistors.html. [Accessed 23 January 2015].

[64] "Arduino Mega 2560," 2015. [Online]. Available: http://arduino.cc/en/Main/ArduinoBoardMega2560. [Accessed 23 January 2015].

[65] LabVIEW is a product of National Instruments Corp., 11500 N Mopac Expwy, Austin, TX 78759-3504.

[66] "MCP50X," 2015. [Online]. Available: http://www.swiftech.com/MCP50X.aspx. [Accessed 23 January 2015].

[67] E. Frachtenberg, A. Heydari, L. Harry, A. Michael, J. Na, A. Nisbet and P. Sarti, "High-efficiency server design," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, Seattle, WA, 2011.

[68] "Intel® Xeon® Processor X5650 (12M Cache, 2.66 GHz, 6.40 GT/s Intel® QPI)," [Online]. Available: http://ark.intel.com/products/47922/Intel-Xeon-Processor-X5650-12M-Cache-2_66-GHz-6_40-GTs-Intel-QPI. [Accessed 23 January 2015].

[69] P. Sarti and F. Frankovsky, "700W-SH/450W-SH Power Supply Hardware v1.0," 24 May 2012. [Online]. Available: http://www.opencompute.org/assets/download/Open_Compute_Project_700W_450W_Power_Supply_v1.0.pdf. [Accessed 23 January 2015].

[70] "H5TQ2G83BFR," [Online]. Available: http://hynix.com/datasheet/eng/computing/details/computing_19_H5TQ2G83BFR.jsp. [Accessed 23 January 2015].

[71] "H5TQ1G83TFR," [Online]. Available: http://www.hynix.com/datasheet/eng/computing/details/computing_19_H5TQ1G83TFR.jsp. [Accessed 23 January 2015].

[72] "Intel® 5500 Chipset (Intel® 5500 I/O Hub)," [Online]. Available: http://ark.intel.com/products/36784/Intel-5500-IO-Hub. [Accessed 23 January 2015].

[73] "Intel® 82801JR I/O Controller," [Online]. Available: http://ark.intel.com/products/34395/Intel-82801JR-IO-Controller. [Accessed 23 January 2015].

[74] S. Furuta, "Server Chassis and Triplet Hardware v1.0," 7 April 2011. [Online]. Available: http://files.opencompute.org/oc/public.php?service=files&t=97a085fe84e16d7d2e950d883838c9ba. [Accessed 23 January 2015].

[75] D. Carraway, "lookbusy -- a synthetic load generator," 22 April 2013. [Online]. Available: https://devin.com/lookbusy/. [Accessed 23 January 2015].

[76] "mpstat(1): Report processors related statistics - Linux man page," [Online]. Available: http://linux.die.net/man/1/mpstat. [Accessed 23 January 2015].

[77] "free(1): amount of free/used memory in system - Linux man page," [Online]. Available: http://linux.die.net/man/1/free. [Accessed 23 January 2015].

[78] Intel Corporation, "4-Wire Pulse Width Modulation (PWM) Controlled Fans," September 2005. [Online]. Available: http://www.formfactors.org/developer%5Cspecs%5C4_Wire_PWM_Spec.pdf. [Accessed 23 January 2015].

[79] R. Eiland, J. Fernandes, B. Gebrehiwot, M. Vallejo, D. Agonafer and V. Mulay, "Air filter effects on data center supply fan power," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, San Diego, CA, 2012.

[80] Airflow Measurement Systems, "Instruction Manual for AMCA 210-99 Airflow Test Chamber," [Online]. Available: http://fantester.com/MAN30.pdf. [Accessed 23 January 2015].

[81] Sound Meter is a product of Smart Tools co. (https://play.google.com/store/apps/details?id=kr.sira.sound).

[82] R. Schmidt and M. Iyengar, "Thermodynamics of information technology data centers," *IBM Journal of Research and Development,* vol. 53, no. 3, pp. 9:1-9:15, 2009.

[83] E. Frachtenberg, D. Lee, M. Magarelli, V. Mulay and J. Park, "Thermal design in the open compute datacenter," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, San Diego, CA, 2012.

[84] V. Mulay, "Learning Lessons at the Prineville Data Center," 17 November 2011. [Online]. Available: http://www.opencompute.org/blog/learning-lessons-at-the-prineville-data-center/. [Accessed 23 January 2015].

[85] M. Iyengar and R. Schmidt, "Analytical modeling for thermodynamic characterization of data center cooling systems," *Journal of Electronic Packaging,* vol. 131, no. 2, p. 021009 (9 pages), 2009.

[86] T. J. Breen, E. J. Walsh, J. Punch, A. J. Shah and C. E. Bash, "From chip to cooling tower data center modeling: Influence of server inlet temperature and temperature rise across cabinet," *Journal of Electronic Packaging,* vol. 133, no. 1, p. 011004 (8 pages), 2011.

[87] Occupational Safety and Health Administration, *Health Standards (29 CFR 1910), Section 1910.95,* US Department of Labor, 1979.

[88] US Department of Health and Human Services, *Criteria for a recommended standard: occupational noise exposure. Revised criteria 1998,* Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 1998.

[89] X. Tian, "Cooling fan reliability: failure criteria, accelerated life testing, modeling and qualification," in *Reliability and Maintainability Symposium (RAMS)*, Newport Beach, CA, 2006.

[90] H. Oh, M. H. Azarian, M. Pecht, C. H. White, R. C. Sohaney and E. Rhem, "Physics-of-failure approach for fan PHM in electronics applications," in *Prognostics and Health Management Conference (PHM)*, Macao, 2010.

[91] X. Jin, M. H. Azarian, C. Lau, L. L. Cheng and M. Pecht, "Physics-of-failure analysis of cooling fans," in *Prognostics and System Health Management Conference (PHM)*, Shenzhen, 2011.

[92] Occupational Safety and Health Administration, *OSHA safety and health standards. 29 CFR 1910.1000,* US Department of Labor, 1983.

[93] S. Mukhopadhyay, A. Raychowdhury and K. Roy, "Accurate estimation of total leakage current in scaled CMOS logic circuits based on compact current modeling," in *Proceedings of the 40th annual Design Automation Conference*,

Anaheim, CA, 2003.

[94] D. Copeland, "64-bit server cooling requirements," in *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, San Jose, CA, 2005.

[95] S. Krishnan, S. V. Garimella, G. M. Chrysler and R. V. Mahajan, "Towards a thermal Moore's law," *IEEE Transactions on Advanced Packaging,* vol. 30, no. 3, pp. 462-474, 2007.

[96] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson and M. V. Fischetti, "Silicon CMOS devices beyond scaling," *IBM Journal of Research and Development,* vol. 50, no. 4.5, pp. 339-361, 2006.

[97] N. D. Strike, "Fan Efficiency, An Increasingly Important Selection Criteria," [Online]. Available: http://www.nmbtc.com/fans/white-papers/fan_efficiency_important_selection_criteria/. [Accessed 23 January 2015].

[98] M. Brendel, "Fan Efficiency Grade Classification for Fans," 2012. [Online]. Available: http://www.amca.org/userfiles/file/ashrae_2012_sa-1_ppt.pdf. [Accessed 23 January 2015].

[99] M. Arlitt, C. Bash, S. Blagodurov, Y. Chen, T. Christian, D. Gmach, C. Hyser, N. Kumari, Z. Liu, M. Marwah, A. McReynolds, C. Patel, A. Shah, Z. Wang and R. Zhou, "Towards the design and operation of net-zero energy data centers," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, San Diego, CA, 2012.

[100] P. Teich and P. Moorhead, "Intel's Disaggregated Server Rack," 21 August 2013. [Online]. Available: http://www.moorinsightsstrategy.com/wp-content/uploads/2013/08/Intels-Disagggregated-Server-Rack-by-Moor-Insights-Strategy.pdf. [Accessed 23 January 2015].

[101] S. Han, N. Egi, A. Panda, S. Ratnasamy, G. Shi and S. Shenker, "Network support for resource disaggregation in next-generation datacenters," in *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks (HotNets)*, New York, NY, 2013.

[102] J. Weiss, R. Dangel, J. Hofrichter, F. Horst, D. Jubin, N. Meier, A. La Porta and B. J. Offrein, "Optical interconnects for disaggregated resources in future datacenters," in *European Conference on Optical Communication (ECOC)*, Cannes, France, 2014.

[103] J. Chang, J. Meza, P. Ranganathan, A. Shah, R. Shih and C. Bash, "Totally green: evaluating and designing servers for lifecycle environmental impact," in *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, London, 2012.

[104] B. Nagendran, S. Nagaraj, J. Fernandes, R. Eiland, D. Agonafer and V. Mulay, "Improving cooling efficiency of servers by replacing smaller chassis enclosed fans with larger rack-mount fans," in *Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Orlando, FL, 2014.

[105] A. Knorig and B. Howell, "Advanced prototyping with fritzing," in *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, Cambridge, MA, 2010.

[106] S. Sidharth and S. Sundaram, "A methodology to assess microprocessor fan reliability," in *Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Las Vegas, NV, 2004.

[107] S. Kim, C. Vallarino and A. Claasen, "Review of fan life evaluation procedures," *International Journal of Reliability, Quality and Safety Engineering,* vol. 3, no. 1, pp. 75-96, 1996.

# BIOGRAPHICAL INFORMATION

John Edward Fernandes received his bachelor's degree (BE) in Mechanical Engineering from the University of Mumbai in 2006. In September 2006, he began his graduate studies at the University of Texas at Arlington. His master's research focused on characterization of airflow over a mannequin using stereo particle image velocimetry. John served as both a teaching and research assistant and earned his MSc in Mechanical Engineering in December 2008. He began his doctoral degree in January 2010 and joined the EMNSPC to conduct research on thermal management of data centers and IT equipment. His presentation on energy-efficient air-cooling of web servers using rack-level fans earned him a 'best poster' award in the thermal track at ITherm 2014. John received his PhD degree in Mechanical Engineering from the University of Texas at Arlington in May 2015.