COMPARATIVE CFD ANALYSIS OF THE EFFECT OF AIR AND LIQUID COOLING ON THE FORM FACTOR OF GPU SERVER

by

ANKIT SUTARIA

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

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2019

ACKNOWLEDGEMENTS

I would like to thank Dr. Dereje Agonafer who helped throughout my work. His continuous guidance and support over the last two years of my study and research at The University of Texas at Arlington helped me to cross all the obstacles.

I would like to thank Dr. Haji-Sheikh and Dr. Andrey Beyle for evaluating my work as committee members. I would like to give a special thanks to Mr. Uschas Chowdhury and Mr. Rajesh Kasukurthy for their immense help throughout my stay at UT Arlington. I would also like to thank all members at EMNSPC.

Finally, I would like to thank my parents who has always been the beacon of hope in my life. I thank Almighty God for providing me the strength and inspiration.

ABSTRACT

COMPARATIVE CFD ANALYSIS OF THE EFFECT OF AIR AND LIQUID COOLING ON THE FORM FACTOR OF GPU SERVER

Ankit Sutaria, MS

The University of Texas at Arlington, 2019

Supervising Professor: Dereje Agonafer

Artificial intelligence (AI) has becoming an important area which may significantly impact everyone's daily life. This means a lot of high-performance chips such as high-performance CPU, GPU, FPGA (Field-programmable gate array), ASIC (application specific integrated circuit) devices may need. The thermal design power (TDP) of these chips are high, and now it is very common to see a processor TDP reaches as high as 300Watts[1]. When these high-power density processors are packaged together, the power density of the servers and compute boxes are significantly high. Traditional air-cooling is approaching its cooling capability limitation, especially when a cluster of racks is fully populated with computing-intensive nodes. The air cooling may still capable for high density thermal management, however there will be a significant amount of cost associate with it. Liquid cooling may provide a higher cooling efficiency, which can easily reject larger amount of heat generated from processors. This lowers the cooling PUE. In this work, the server used for study is FACEBOOK BIGBASIN which has 3OU chassis, consisting 8 Nvidia Volta V100 GPU, 4 PCIe cards, and 8 hot swappable fans. On the GPUs, there are 8 heatsinks on in the air-cooled model and 8 impingement type cold plates in the liquid cooled model. This work is proposed to identify, evaluate and develop effective liquid cooling by using cold plate which gives high performance compute with compared to air cooled solution. The GPU chips are assumed to be running at maximum Utilization. The cold plate is optimized for better thermal performance in 6sigma ET software.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

ρ	Density (kg/m^3)
k	Thermal Conductivity (W/m-K)
υ	Velocity (m/s)
μ	Viscosity (N/m2S)
ε	Kinematic Rate of Dissipation (m^2/s^3)
$\dot{m}$	Mass Flow Rate (kg/sec)
Q	Heat Load (KW)
P	Power (W)
ϑ	Volumetric Flow Rate (cfm)
p	Pressure (Pa)
T	Temperature (K)
C_p	Specific Heat Capacity (J/kg-k)
R_e	Reynolds Number
l	Characteristic Length (m)

CHAPTER 1

INTRODUCTION

1.1 Data Center

There was a time when our information needs were simpler. We had tv shows with handful of channels, using paper for information transfer and backup, used wired phones and even cell phones were only for making calls. But with the dawn of the internet, high-bandwidth broadband, smartphones and all the new technologies now we stay connected all time, demanding more data on our computers, gaming systems, TVs and smartphones. Simpler things got digitized. With this requirement of fast data delivery came the need for computer and networking equipment. Thus, the modern Data center was born. Data centers are simply centralized locations where computing and networking equipment is concentrated for the purpose of collecting, storing, processing, distributing or allowing access to large amounts of data[2].

Figure 1. 1 Facebook's Data Center at Lulea, Sweden[3]

The main component of data center is the servers. Servers are classified based on their applications. Platform servers, Application servers, mail servers, proxy servers, web servers and communication servers are few of the types. The servers can be of different shape and size according to their chassis design. The enclosure which holds the multiple servers are rack. Rack mounted servers are of standard sizes termed as 1U servers (1U=44 mm). Which means if a server is 2U size, it has height of 2.5 inches. However, some big companies use different measure for server size for example Facebook uses its 1OU (1OU=48mm) as standard size. Here are some types of servers and racks used commonly now a days in data centers. Figure 1.2 to 1.6 shows

Figure 1. 2 IBM HS20 blade servers[4]

Figure 1. 3 cisco UCS C4200 rack server[5]

Figure 1. 4 google liquid cooled rack for Machine learning[6]

Figure 1. 5 HP C7000 rack[7]

Different types of server and racks used now days commonly in data centers.

1.2 Data Center Energy Consumption

Today data centers are widely used for cloud-based computing, virtualization, online data storage and access, which is increasing day by day. That means large data centers, high performance computing equipment, more power and density. Adding to this is the cost of energy and maintenance for providing cooling and prevent failures. Many engineers are trying to optimize and improve the efficiency of our data centers by considering new server designs and cooling methods. In 2014, data centers in the U.S. consumed an estimated 70 billion kWh, representing about 1.8% of total U.S. electricity consumption[8]. The report estimates 73 billion kWh of energy usage by 2020. The trend is expected to rise further every year.

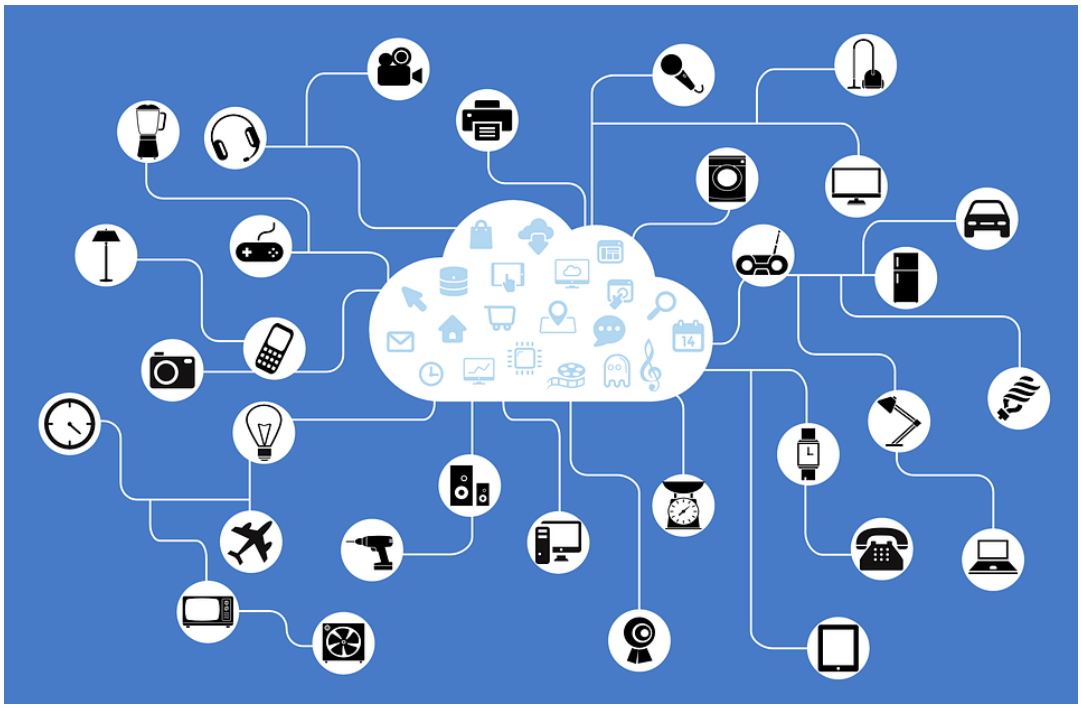

Figure 1. 6 Representation of Internet of Things linked to data center[9]

1.3 Thermal Management of Data Center

To store, process and distribute large amounts of data, various IT equipment consisting servers, storage subsystems, networking switches, routers are housed data center. which needs cabling and physical rack units to organize and interconnect throughout the facility. The facility also consists of power supplies, backup power supplies, environment controls (such as air-

conditioning) and security measures. The servers along with other IT equipment generate large amount of heat which require the proper thermal management at the device, rack and room level. Substantial amount of energy is required to maintain ta certain operating temperature for the facility at IT equipment. Thus, more measures are needed to improve the data center efficiency. The most common metric used to measure the efficiency of data center is Power Usage Effectiveness (PUE). PUE is a ratio of the total energy used to the energy consumed by IT equipment. For example, if the PUE of a certain data center is 2.0 means that for every watt used by IT equipment, another watt is used by cooling, power distribution, and lighting equipment. The global average PUE of data centers is between 1.8 and 1.89[10].

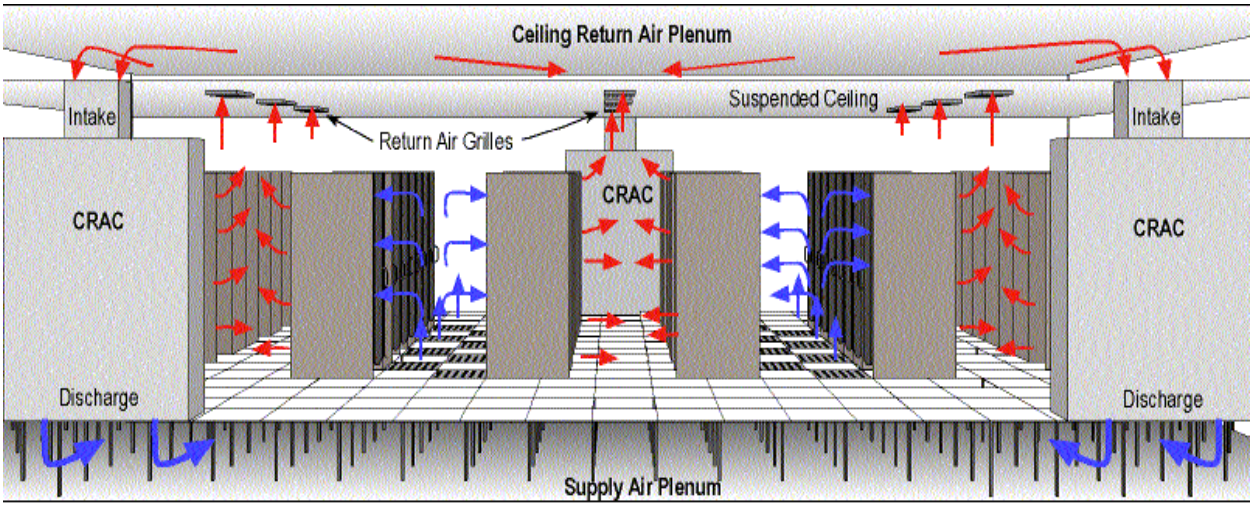

Figure 1. 7 Typical thermal layout of data center[11]

The typical data center layout is shown in figure 1.7. The computer room air conditioner (CRAC) unit provides cold air through the raised floor. The cold air gets hot from IT equipment and rise to the suspended ceiling then again it returns to the CRAC unit. The airspace is divided into cold aisle and hot aisle. American Society for Heating Refrigeration and Air Conditioning Engineers (ASHRAE) committee has developed a guideline for design, operation, maintenance, and efficient energy usage of data centers[12]. According to ASHRAE guidelines the recommended temperature zones are from 18ºC (64.4º F) to 27ºC (80.6º F) and less than 60% humidity level.

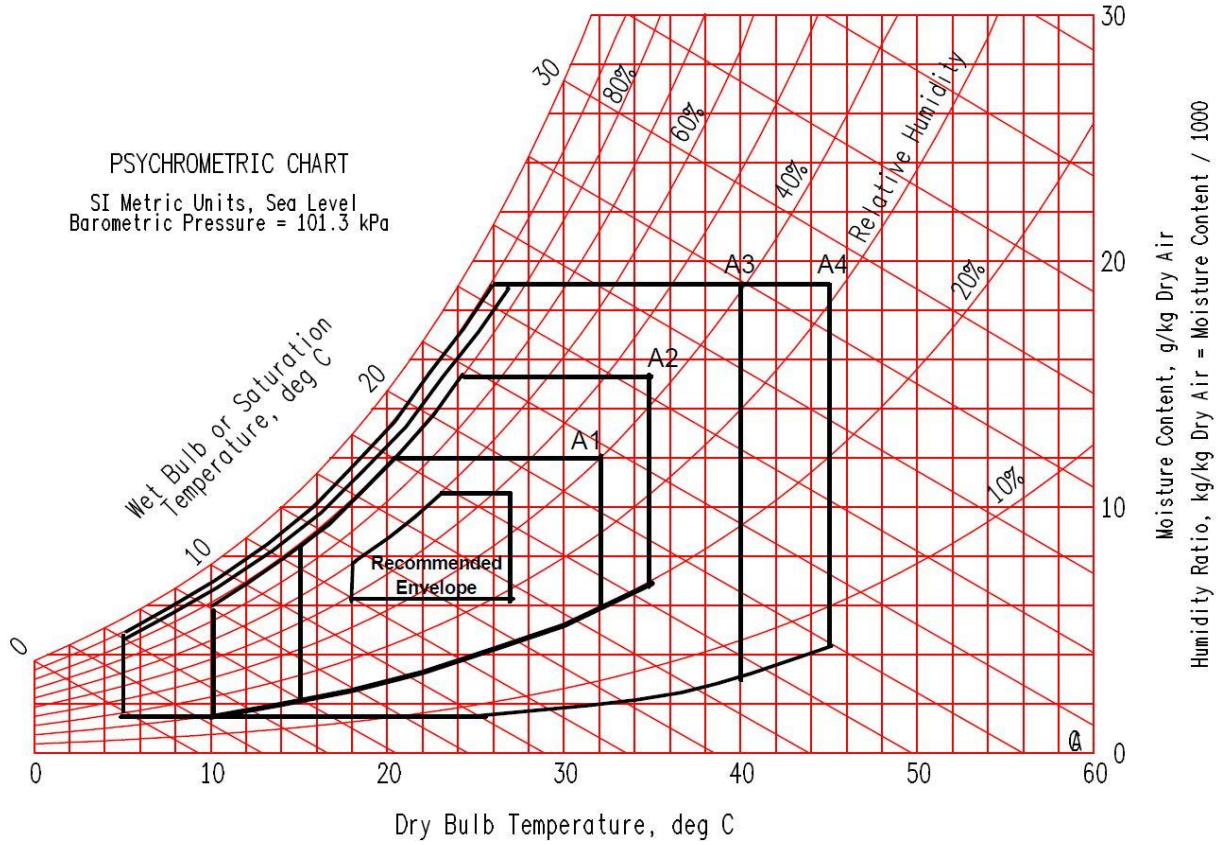

Figure 1. 8 ASHRAE Environmental Classes for Data Centers[10]

The main heat generative component in data center are computing chips in the server. The process of heat removal from chip to the ambient air includes design of server chassis, design of heat sinks, use of thermal interface materials (TIM) and heat spreader, proper fan installation and precise ducting for air delivery. In a rack, to improve the cooling efficiency, various arrangements like perforated tiles, hot and cold aisle containment, rear door heat exchangers, and even direct liquid cooling for high power density servers. For high powered microelectronic devices liquid cooling is also used[13]. The cold plates are used for cooling high watt microelectronic chips. Liquid cooling can provide higher efficiency in rejecting large amount of heat generated by processors[1]. This lowers the PUE.

This study is devoted to understanding the impact of both reduced air and liquid flow rate on the thermal performance, energy consumption and form factor of a high-power server. Next chapter will discuss about the server taken into this study and its major components.

CHAPTER 2

SERVER DISCRIPTION

2.1 Introduction

From scientific discovery to artificial intelligence, high performance computing (HPC) is an important pillar that fuels the progress of humanity[14]. The traditional CPUs are no longer delivering the performance gain they used to for HPC data centers. The increasing need for deep learning, Machine learning and big data calculations require significant amount of computing power and compute resources. Which means more devices such as CPU, GPU, FPGA and ASIC chips will require for HPC data centers. Now it is very common to see thermal design power of these chips reaches as high as 300 Watts[1]. When such processors are packaged together, the power density of the servers and racks become an important thermal challenge.

The Facebook has been deploying various AI hardware infrastructure with high machine learning capability to provide better communication experience for their customers. The server under study is facebook's "Big basin". It is a 3OU rack mount GPU based server. It has 8 Nvidia Tesla V100 sxm2 GPUs.

Figure 2. 1 Facebook's Big basin Open compute server[15]

Figure 2. shows front view of the big basin server. This chapter will further describe each components of the server. Here is the list of some important components of the server.

Table 2. 1 Big basin list of important items[16]

Item	Description/Major Components
GPU Baseboard	8 Nvidia Tesla V100 SXM2 GPUs x16 PCIe slot x4 8 x16 PCIE Gen3 slots for GPU PEX 8780 PCIe Switch x2 PEX 8764 PCIe Switch x2 2 x16 Standard PCIe Gen3 Slots for AICs
GPU Module	NVIDIA Tesla series – Volta V100, 300 W TDP PCIe x16
Middle Plane Board	BMC and CPLD Power connectors from Rack Bus bar Power/Signal connectors for Baseboard Power/Signal connectors for IO board
Host Retimer Card	4 mini-SAS HD connectors x16 PCIE Gen3 retimer
Bridge card W/O Retimer	4 mini-SAS HD connectors x16 PCIE Gen3
Fan	Fan Quantity: 4 Fan size (L x W x H in mm): 9276 (76 x 92 x 92)
Power Supply	Through Open Rack Power Bus Bar, Open Rack V2 support
Chassis	3 OU BOX for Open Rack V2

2.1.1 GPU baseboard

The server under investigation uses 8 Nvidia Tesla series socket motherboard with 4 PCIe slots. The motherboard uses Nvidia GPUs in SXM2 form factor with a TDP (thermal design power) Up to 300 W each. It supports NVLINK interconnection across all GPUs running at 25.7812 Gbps[16].

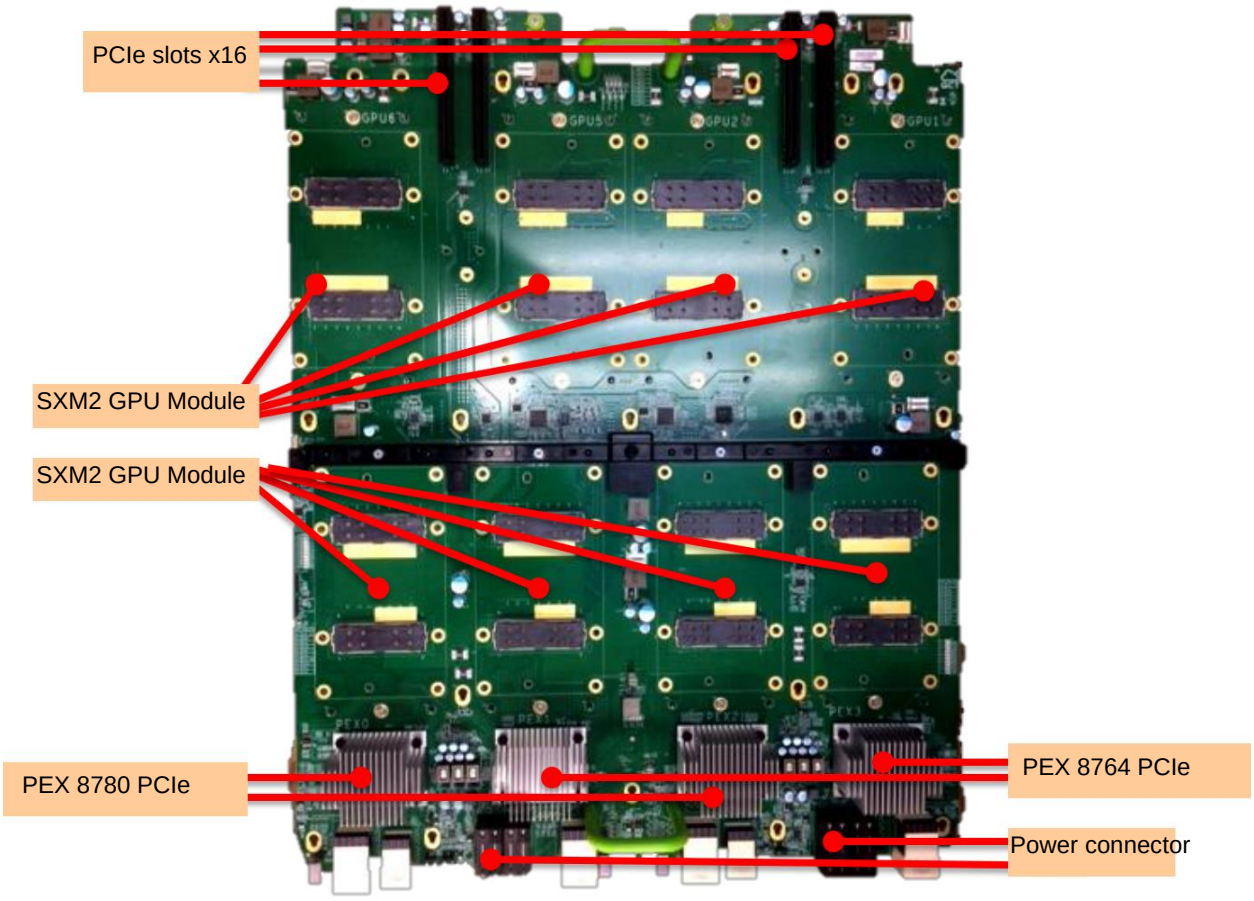

Figure 2. 2 GPU Baseboard[17]

2.1.2 Middle plane board

The middle plane board is placed rear of the baseboard PCB. Middle plane board has power connectors for the baseboard PCB and other components. The main power from the open rack comes from the bus bas connectors. Middle plane board pc has 4 fan connectors at the back of the server. Figure 2.3 shows the top view of the middle plane board PCB.

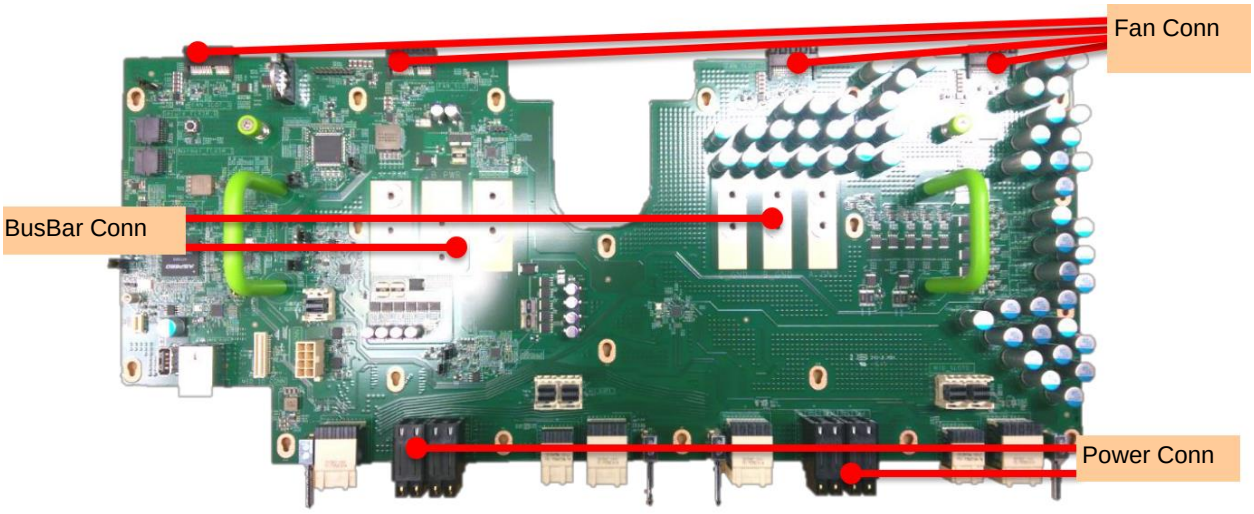

Figure 2. 3 Middle plane Board[17]

2.1.3 Host Retimer card

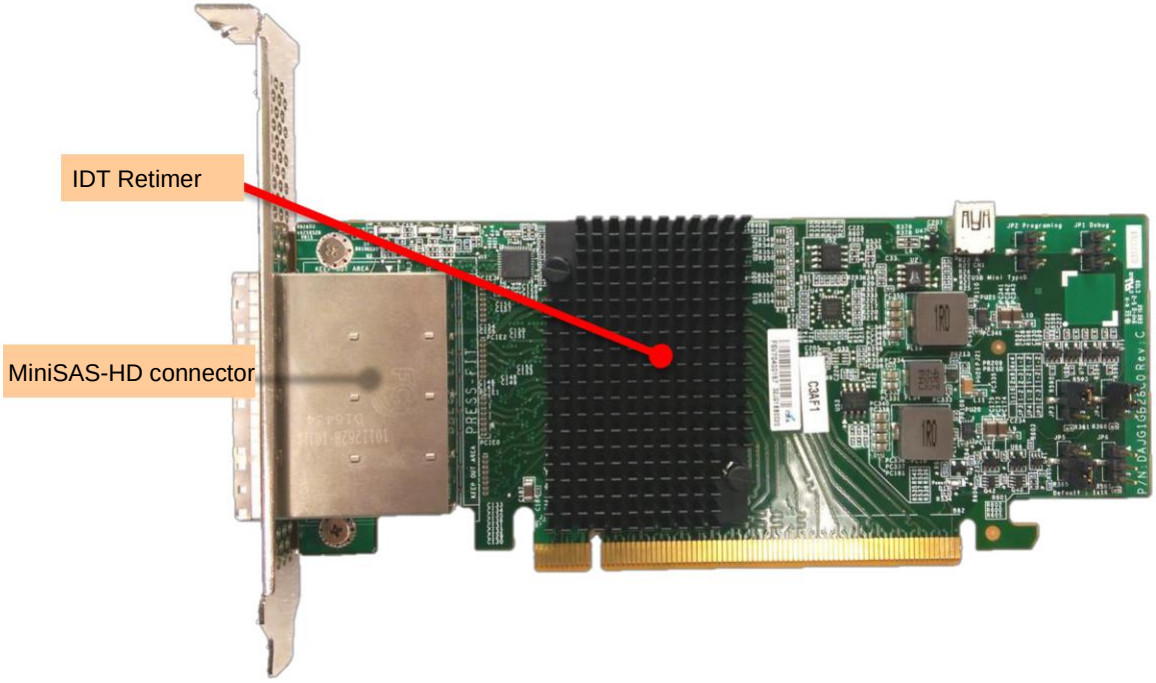

Figure 2. 4 Host Retimer Card[17]

2.1.4 Bridge card W/O Retimer

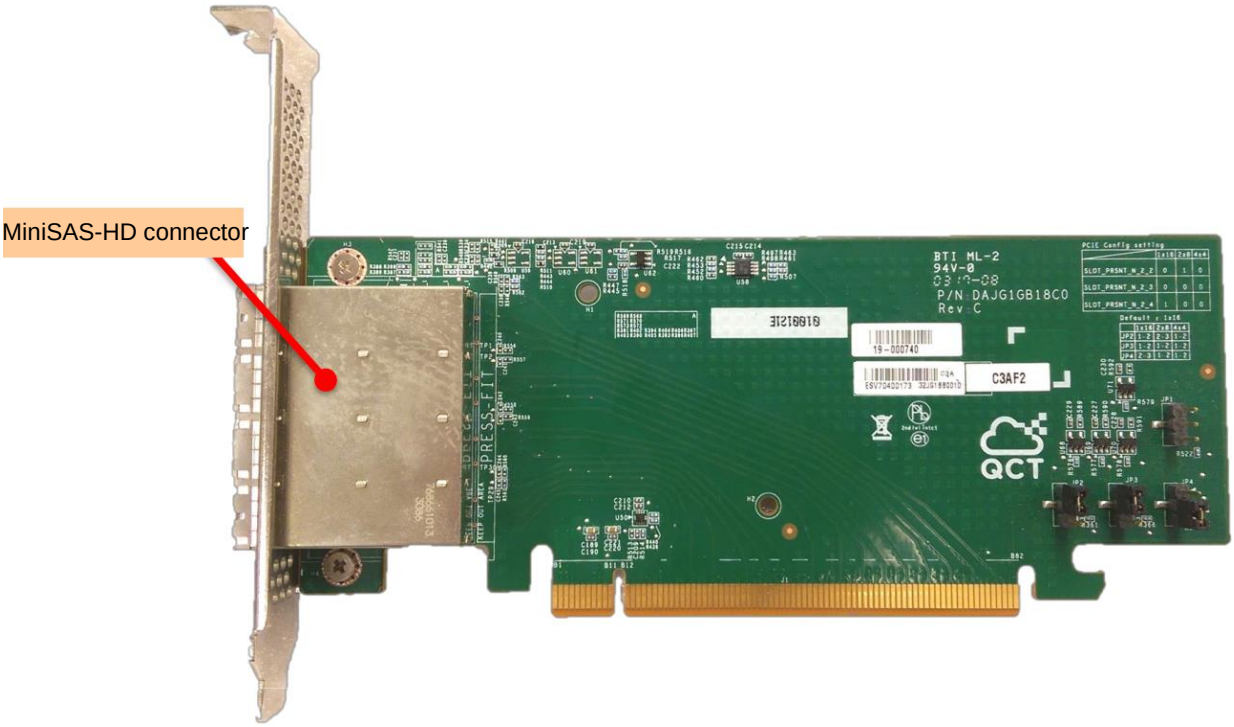

Figure 2. 5 Bridge card W/O retimer Placement[17]

2.1.2 Server Chassis

The chassis is 141.5 mm tall which houses 3 OU space in an open rack (OR) form factor. The chassis is 537.2 mm (21” inch) wide and 931 mm (36.8” inch) long. It has 4 fans located at the rear. The front the chassis has four vents with honeycomb type holes. The GPU module of the server is mounted on a sliding tray for ease of maintenance.

Figure 2. 6 baseboard with sliding tray[18]

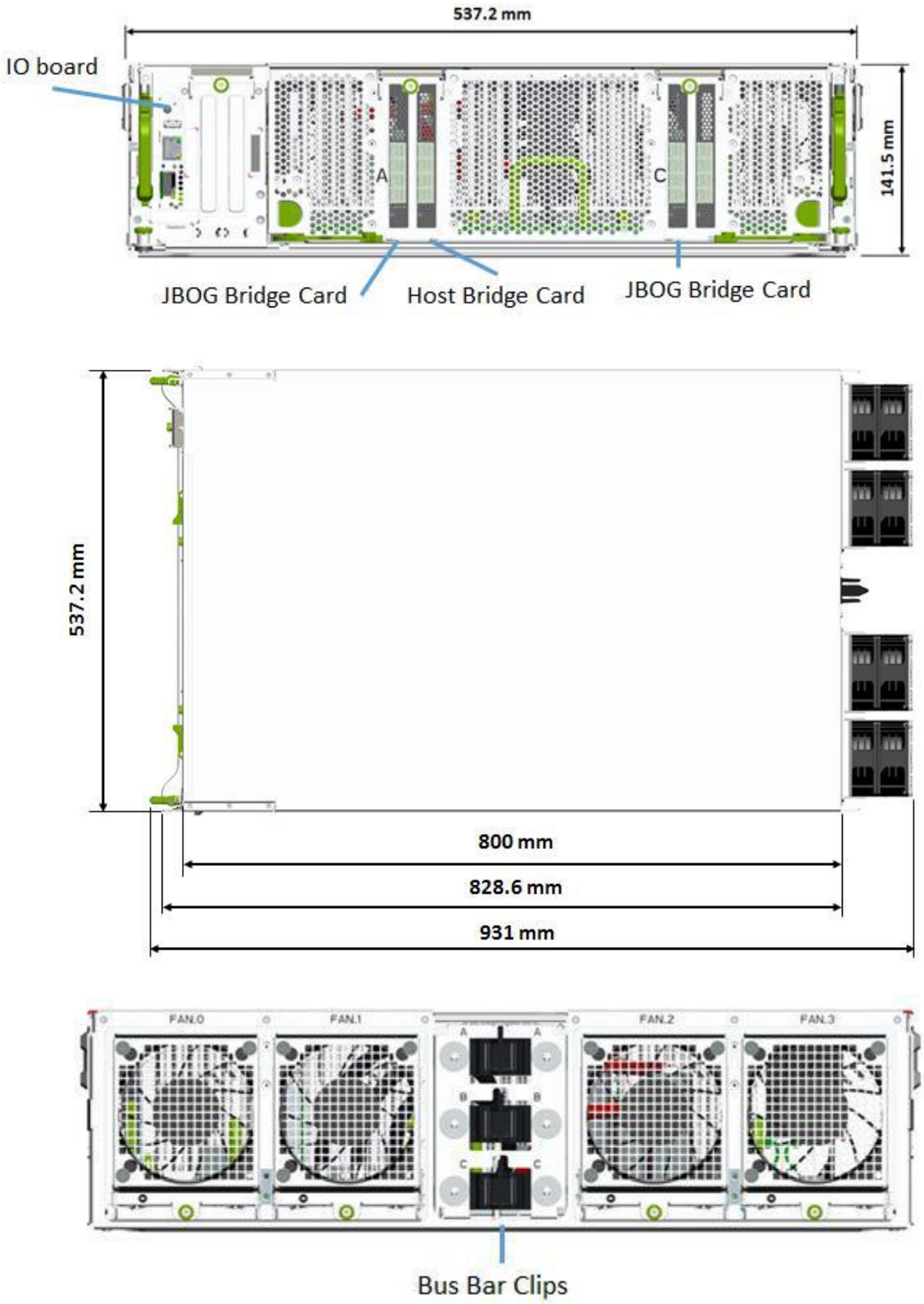

Figure 2. 7 Front, Top and Rear view of chassis[16]

CHAPTER 3

COMPUTATIONAL FLUID DYNAMICS (CFD)

3.1 Introduction to CFD

Computational fluid dynamics (CFD) is the numerical simulation of fluid flow. Fluid flow can be rain, wind, air pollution or contaminants, heating, ventilation and air conditioning, complex flow in furnaces, heat exchangers, blood flow and breathing of human body and so on and so forth. Computational fluid dynamics provides qualitative and quantitative prediction of fluid flow by means of mathematical modeling (partial differential equations), numerical methods (discretization and solution techniques), software tools (solvers, pre- and postprocessing utilities). CFD enables engineers to perform numerical experiments (i.e. computer simulations) in a virtual flow laboratory[19]. The outcoming of the CFD simulations can be used to analyze the design and optimize the design process and products. Computer processors are used for performing the calculations required to simulate the fluid interactions with surfaces based on the boundary conditions. Nowadays data centers also use CFD to simulate air and liquid flow, temperature, pressure and other variables throughout the calculation domain.

CFD gives an insight into highly detailed flow patterns that are difficult, expensive or impossible to study using experimental techniques. It is a link between pure theory and pure experiment. CFD works by discretizing forms of partial differential equations (PDEs) for fluid flow and heat transfer and gives approximate solutions of the governing equations at a predetermined number of points that are specified by a grid of elements (Mesh) formed within a geometric boundary. Compared to conducting experiment, CFD provides faster results and does parallel simulations to solve different problem scenario. Thus, it is very less expensive for every change made to optimize the design. CFD tools offers various features like designing, mesh generating, fluid flow animations, pre-installed libraries of material data with all the properties and several other features.

3.2 Governing Equations

The Navier-stokes equations are the base of computational fluid dynamics codes. The numerical solutions of fluid dynamics problems are obtained by solving series of three differential equations. These equations are conservation of mass, conservation of momentum and conservation of energy. The generalized form of these equations is given by

- Conservation of mass,

$$\frac{\partial \rho}{\partial x} + \nabla \cdot (\rho u) = 0$$

- Conservation of momentum

$$\frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u u) = \nabla \cdot (\mu \, grad \, u) - \frac{\partial p}{\partial x} + B_x + V_x$$

- Conservation of energy for a steady low velocity flow

$$\nabla \cdot (\rho u h) = \nabla \cdot (k \, grad \, T) + S_h$$

3.3 Global Computational Domain

In general flow field, when considering a closed volume inside a finite region of flow which is defined as control volume. Governing equations for mass momentum and energy are solved in control volume. The control volume may me fixed in space or maybe moving along with the fluid. For most computational problems the external ambient temperature, velocity, pressure, mass flow at inlet and outlet, fluid viscosity, thermal conductivity, specific heat and other environmental conditions are included in boundary conditions. Boundary conditions depends on the type of heat transfer such as conduction or convection and any radiation factors. By fixing the boundary conditions the solution of those governing equations are obtained. In addition to boundary conditions the computational domain wall also need to be specified if they are open, closed (adiabatic) or symmetrical in nature.

CFD analysis process includes steps such as Problem statement, mathematical model, mesh generations, space discretization, time discretization, iterative solver, CFD software,

simulations run, post processing, verification. Discretization in CFD process is important as it converts differential equations into algebraic equations. Depending on the discretization method, the governing equations as expressed in integral or differential form. The three main methods of discretization are finite difference method (FDM), the finite volume method (FVM) and finite element method (FEM). The FVM and FEM are integral schemes, while FDM is differential scheme and is based on Taylor series expansion. The FVM solves a problem by dividing it into small volume and integrate around the mesh elements. In the FDM method the differential terms are discretized into series of small grid points.

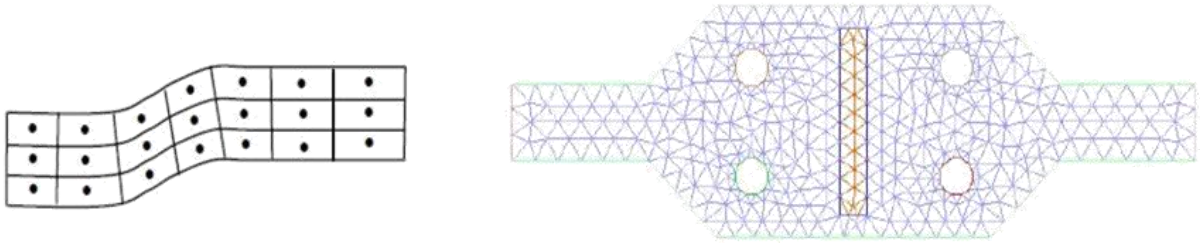

Figure 3. 1 Graphical representation of Discretization by FVM[20]

The present problem is solved by using CFD model based on finite volume method, using the future facility's 6sigmaET software. Thus, the governing equations are solved by integrating over the control volume and applying divergence theorem. The variable to be calculated are located at the centroid of the finite volume. FVM is locally conservative as it is based on balance approach.

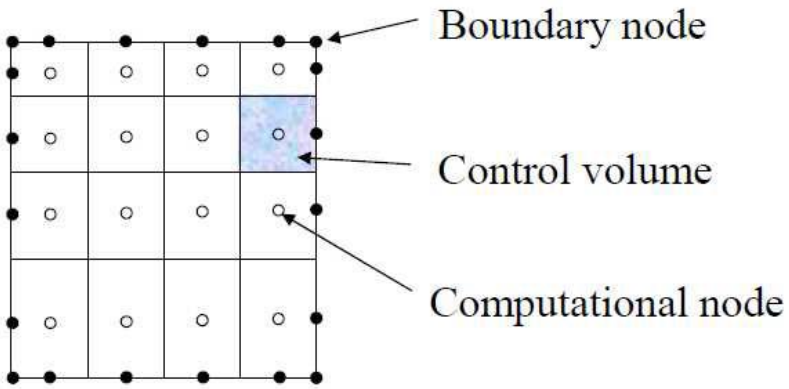

Figure 3. 2 Graphical representation of 2D grid

3.4 Turbulence Modeling

All flow become unstable and random in every direction above certain Reynolds number. The velocity and all other properties of flow vary in a random and chaotic way, This regime is called turbulent flow[21]. When the Reynolds number of a fluid flow gets above critical Reynolds number Re_{crit}, the flow becomes turbulent. The fluid flow of a Reynolds number below Re_{crit} are considered as laminar flow.

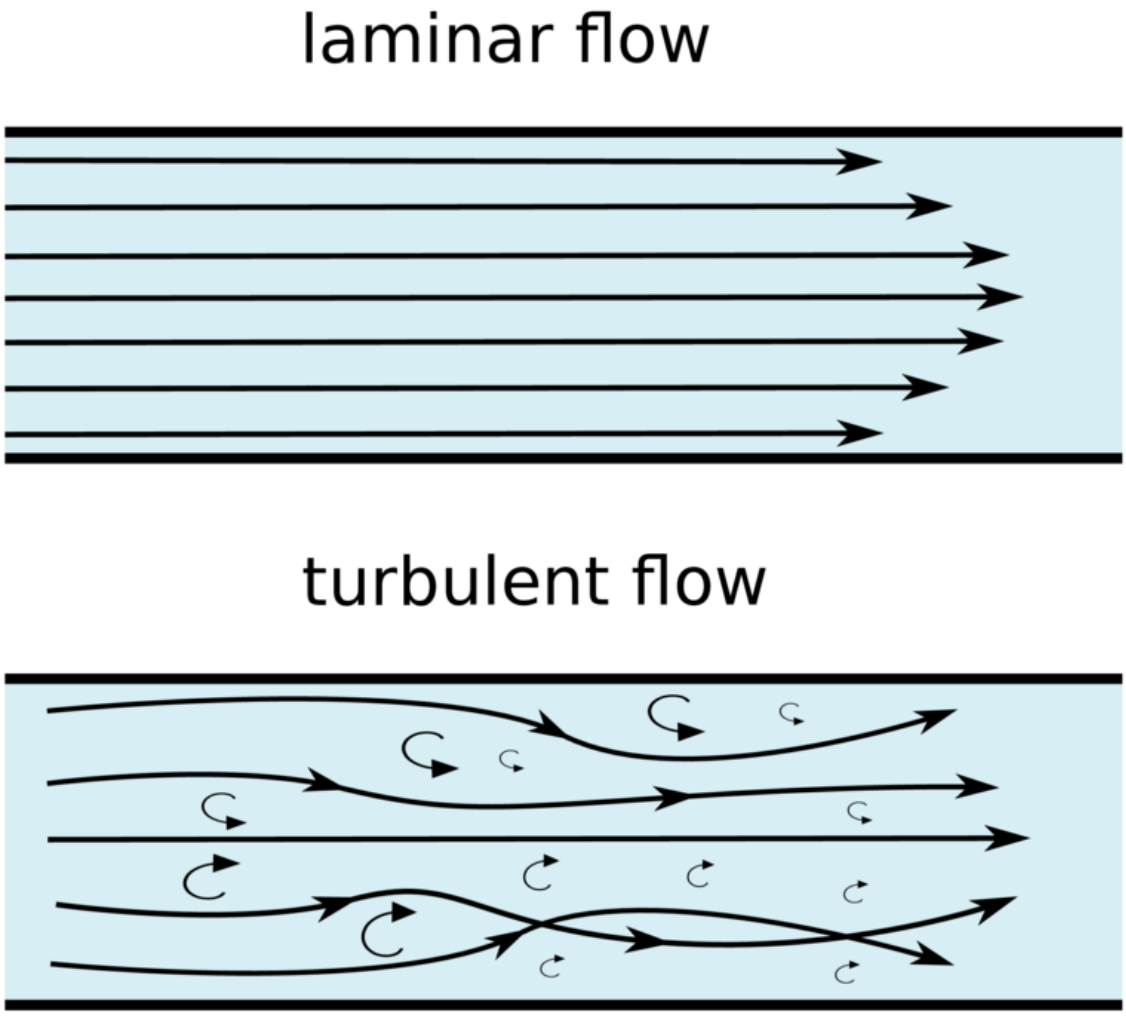

Figure 3. 3 Graphical representation of laminar vs turbulent flow[22]

3.4.1 K-Epsilon Turbulence Model

K-epsilon model is the most common model used in CFD to simulate characteristics of turbulent flow. It is a two-equation model which represents the turbulent properties of the flow. The first transport variable is turbulent kinetic energy and other one is turbulent dissipation. The variable of turbulent dissipation determines the scale of the turbulence whereas the variable of turbulent kinetic energy determines the energy in the turbulence. The following are the transport equation of K-Epsilon model[23].

- For turbulent kinetic energy

$$\frac{\partial(\rho k)}{\partial t} + \frac{\partial \rho k u_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_i}\right] + G_k + G_b - \rho\varepsilon$$

- For turbulent dissipation

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \frac{\partial p\varepsilon u_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial \varepsilon}{\partial x_i}\right] + C_{1s}\frac{\varepsilon}{k}(G_k + C_{3s}G_b) - C_{2s}\rho\frac{\varepsilon^2}{k}$$

3.5 Grid constraints and Meshing

In 6sigmaET grid constraints are used for specifying limits of maximum number of cells across the geometry.6sigmaET uses Cartesian grid and object-based gridding. Which means it has predefined grid control settings for specific objects such as PCB, chassis, heatsinks, cold plates, chip socket, capacitors, power supply unit, vents, fans, hard drives etc.[24]

3.6 Objects in 6sigmaET

As mentioned above 6sigma has pre-defined objects used in electronics industry. The entities are parametrically defined in 6sigmaET with their characteristics and function. This feature of 6sigmaET helps building and analyzing the model with ease. It includes objects such as test chamber, chassis, heatsink, chip socket, component, duct, PCB and fans etc. instead of solid block for representing. Which also can be modified based on our requirements.

3.6.1 Test chamber

A test chamber is a virtual wind tunnel where you can set up your model and apply the air flow within it in several ways. It is works as an environment in 6sigmaET. The walls of the test chamber can be defined as open or any specific environment like constant fluid flow. It can also be disabled or uninstalled if not require.

3.6.2 Chassis

A chassis is the default solution domain in 6sigmaET. All objects get placed inside a chassis unless a test chamber is used. If the physical structure of the chassis is not required, the chassis sides can be uninstalled. The modification can be done with the parameters of the chassis such as material, thickness, dimensions etc. The objects directly attached to chassis are organized under chassis node. Chassis objects are:

- Cooling - fans, blowers and heatsinks
- Electronics - all hard disk drives and bays, PCBs and the components mounted on them (including all their associated hardware)
- Fluid Cooling - includes internal ducting for fluid cooling system and pumps
- Obstructions - used for modeling non heat producing items such as internal ironmongery, DIP switches and any heat creating items such as transformers and large inductors
- Power – includes simple power supplies and their attached fans
- Sensors – consists of only the sensors that are attached to chassis. Sensors can be attached anywhere inside the test chamber and chassis for pressure, temperature, velocity, density etc.

3.6.3 PCB

Printed circuit board (PCB) is combinations of layers of copper and non-conductive materials. PCB consists several conductive paths to transfer data and leaving other area non-conductive. The components are mounted on PCB and are connected through copper paths.

3.6.4 Fan

Fans are used for cooling the electronics equipment by providing an air flow throughout the device. The fan generated air flow by converting supplied torque energy into kinetic energy and doing so fans increases pressure difference across fan rotor.

3.6.5 Heatsink

Heatsinks are passive heat exchangers that are used to transfer heat away from the electronic component into fluid medium (air or liquid). The heatsinks are mounted onto the heat generative components such as CPU and GPU.

3.6.6 PAC

In 6sigmaET PAC stands for Parameterize, Analyze, and Compare. PAC is used for performing parametric analysis for alternative cases. It practically allows us to change the variables for different cases and can solve multiple cases parallelly. It reduces the time to compare the different cases by giving results of those cases in one tabular format. In this study PAC is used to obtain mesh sensitivity analysis, optimization of cold plate, system resistance and thermal resistance of the system.

CHAPTER 4

CFD MODELING AND FLOW ANALYSIS

4.1 Detailed modeling of server

Server modeling was done using 6SigmaET. The CFD model is generated from the CAD model of the server. It is very difficult to include all the components of the motherboard into the CFD model as they do not contribute in heat transfer or system resistance. Such components are negligible compared to certain high-power components of the server that generates large amount of heat. In this model several major flow restricting and heat generating components are chosen. It is done based on dimensions and TDP of each component. The components considered for modeling are baseboard PCB, Middle plane PCB, 8 Nvidia GPU chips, 8 Heatsinks, 4 PCIe cards, IO board, 4 fans and a middle obstruction for flow direction. Later, the heat sinks are replaced by the cold plates for liquid cooling purposes. The modeling process of whole server is carried out in several small segments. All the detail of server is obtained from Open compute project website which includes dimensions, geometry and TDP of each component[16].

4.1.1 GPU

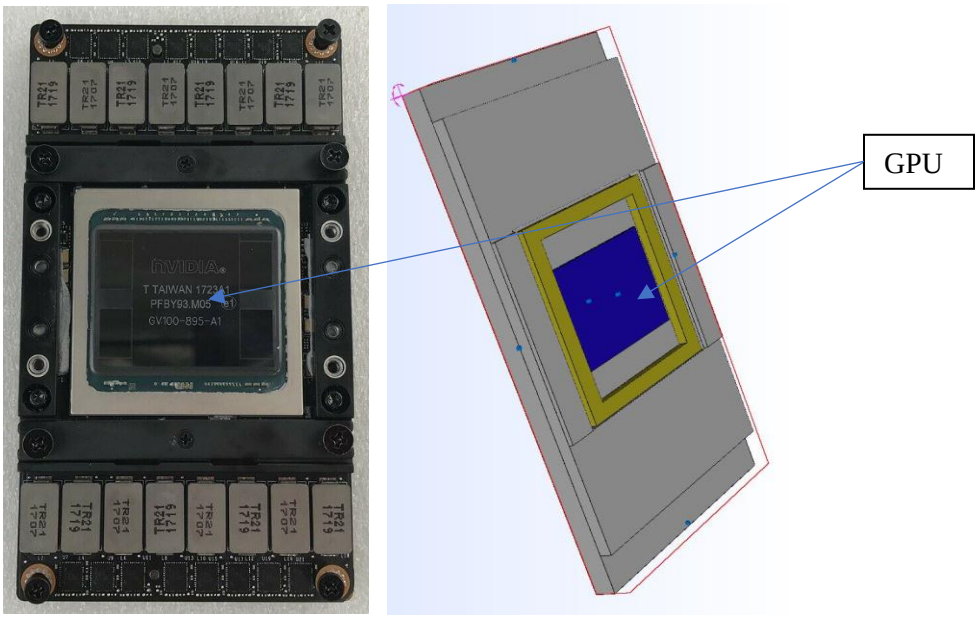

Figure 4. 1 GPU actual and 6sigmaET model

4.1.2 PCIe card

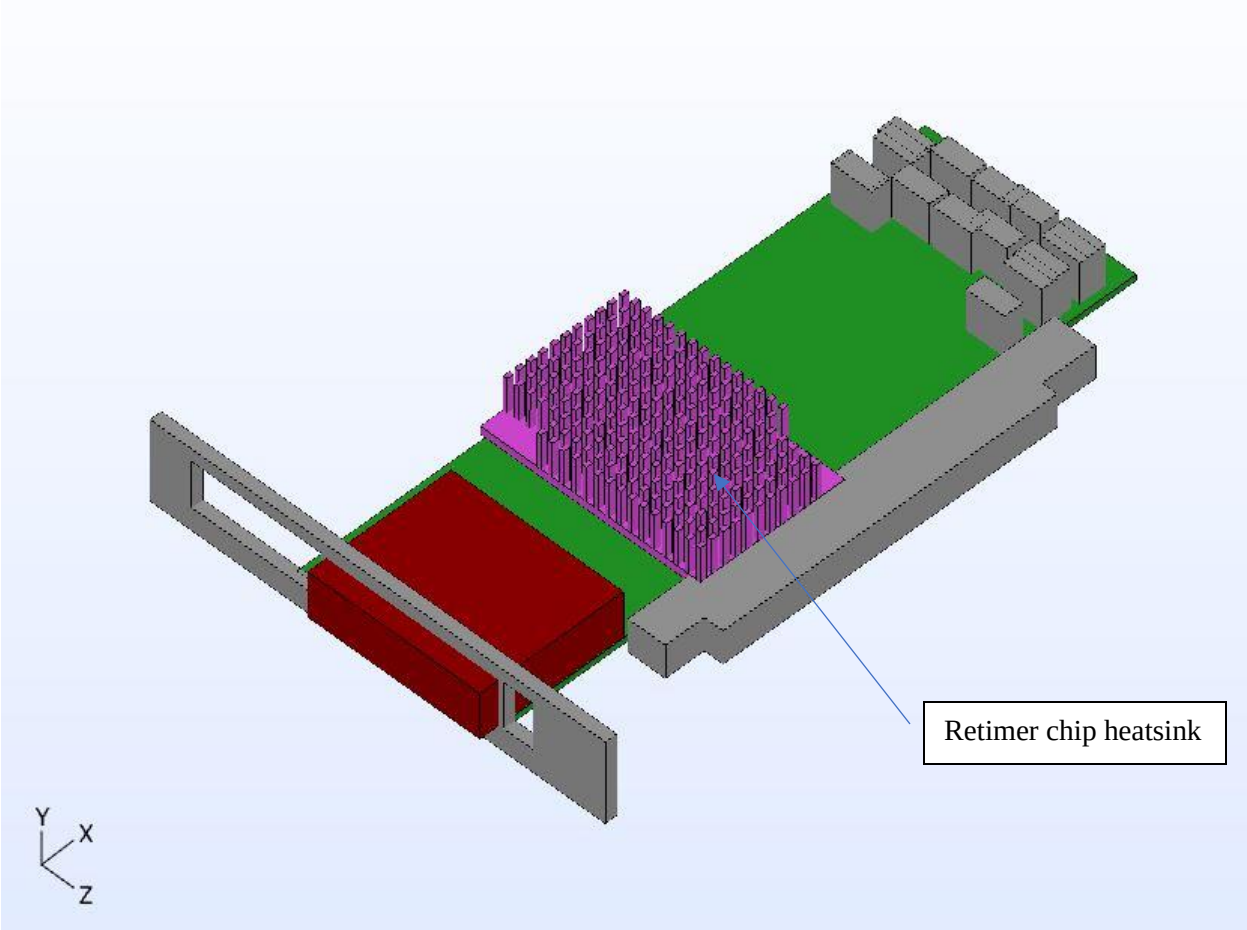

Figure 4. 2 PCIe card

4.1.3 PEX switch heatsink

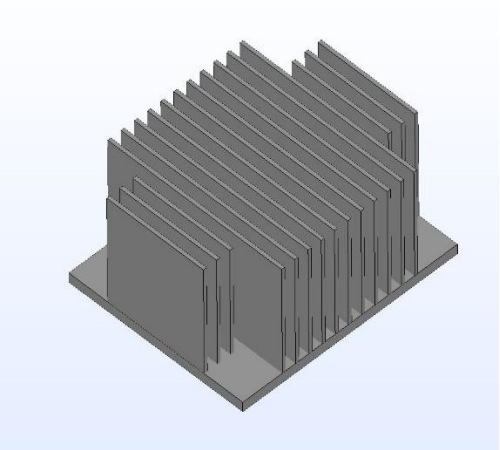

Figure 4. 3 PEX heatsink

4.1.4 Baseboard PCB

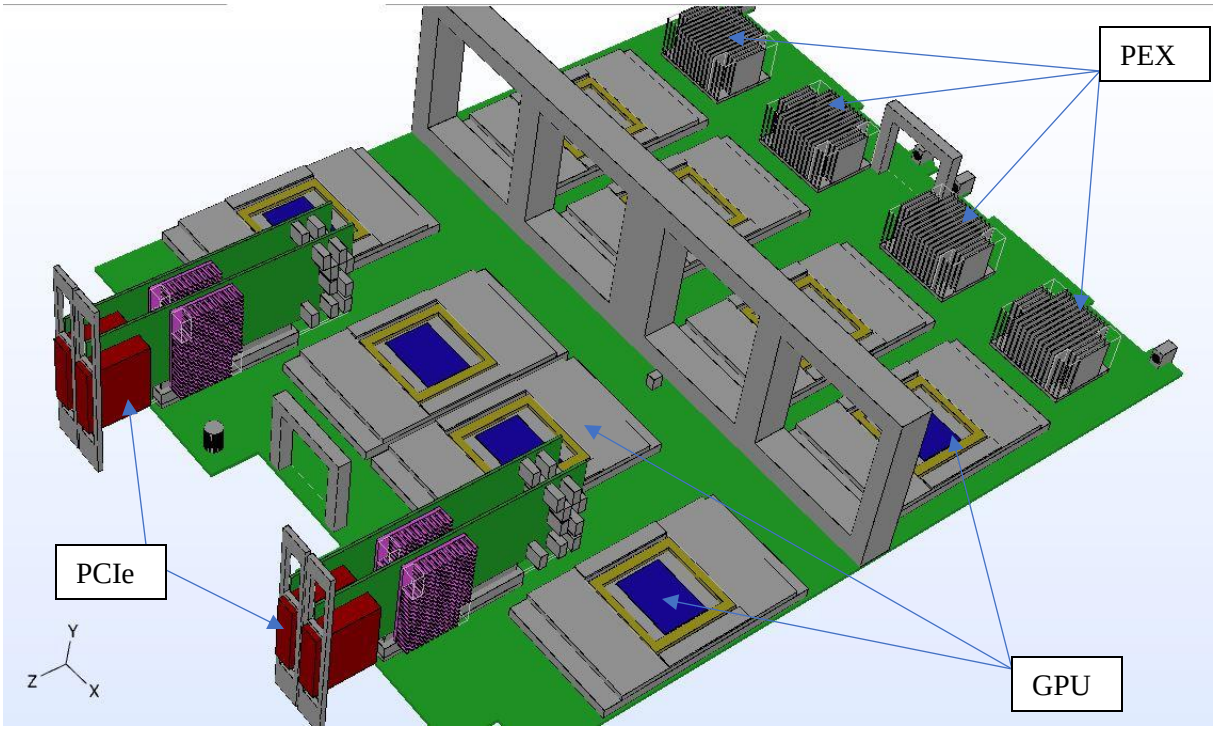

Figure 4. 4 Baseboard PCB with GPU, PCIe card and PEX switch

4.1.5 Middle plane PCB

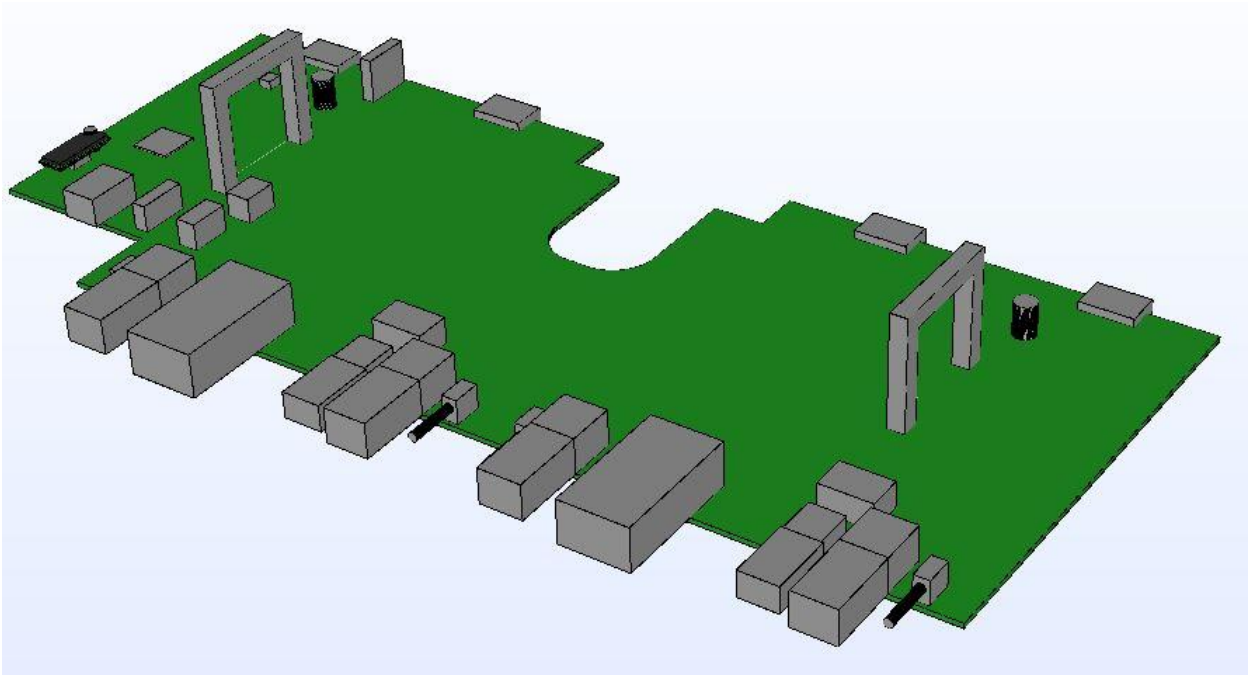

Figure 4. 5 Middle plane PCB

4.1.6 GPU Heatsink

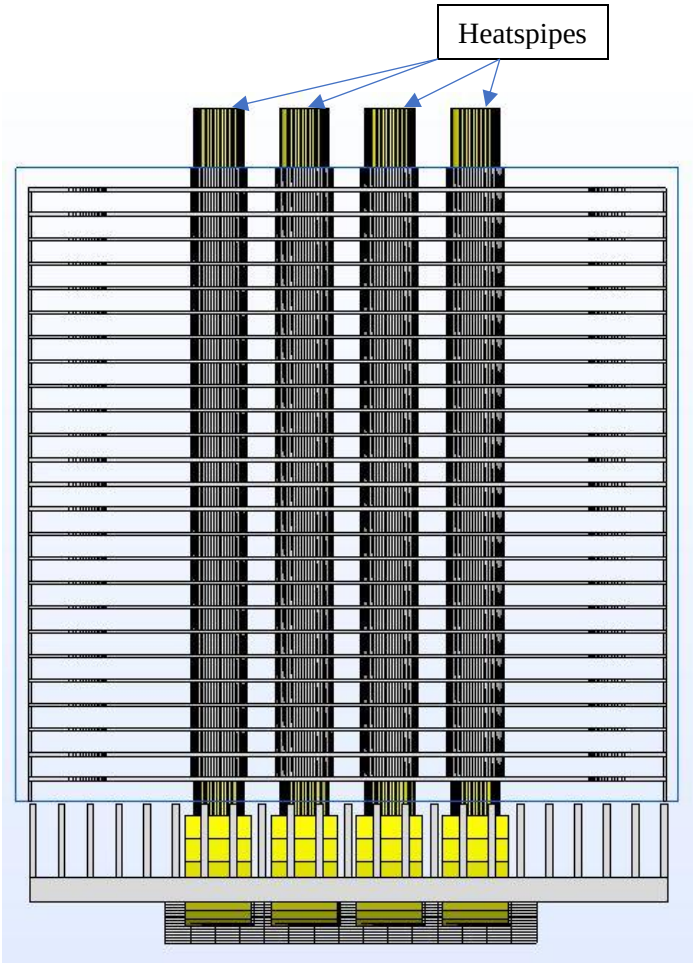

Figure 4. 6 Heatsink front view

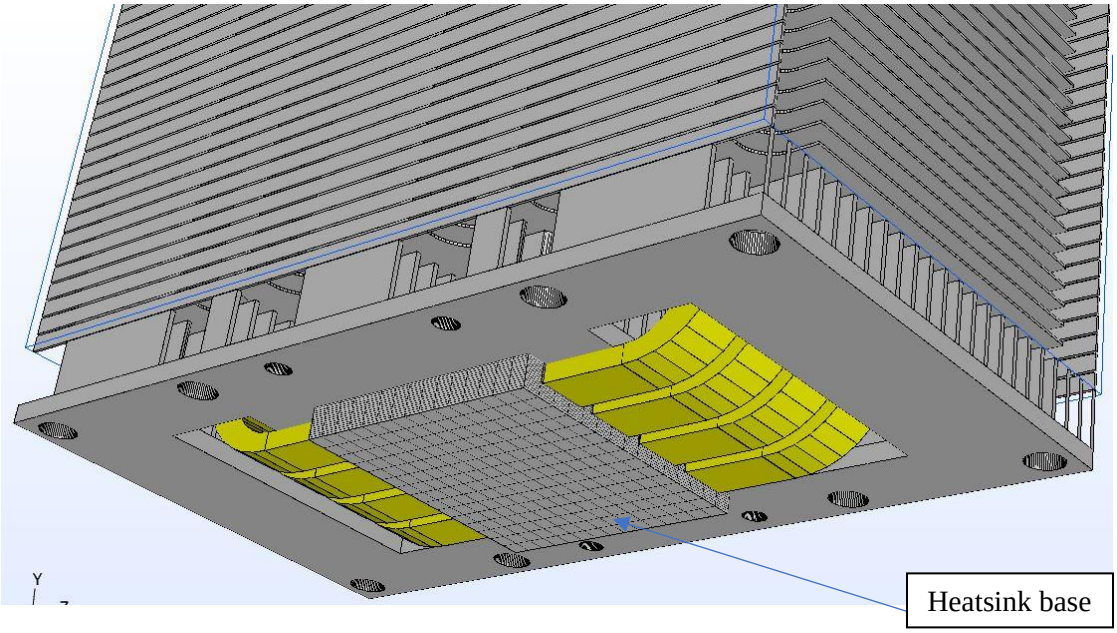

Figure 4. 7 GPU heatsink front view and bottom view

4.1.7 Server assembly

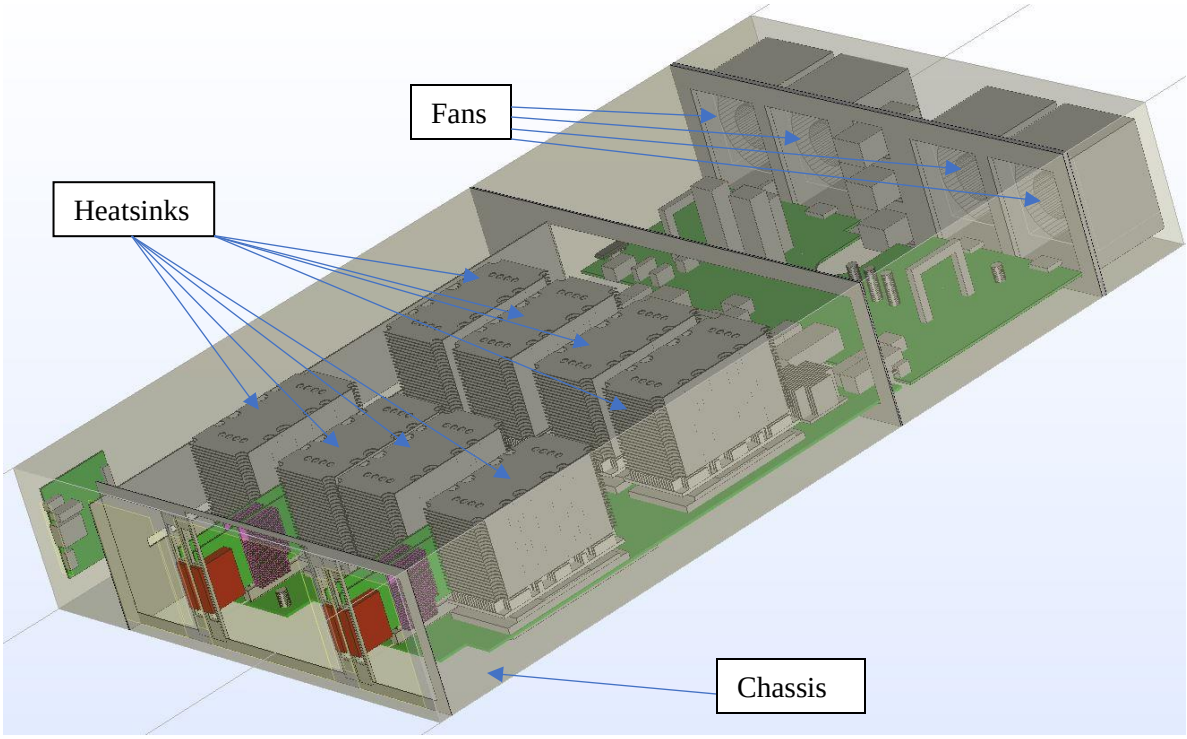

Figure 4. 8 Isometric view

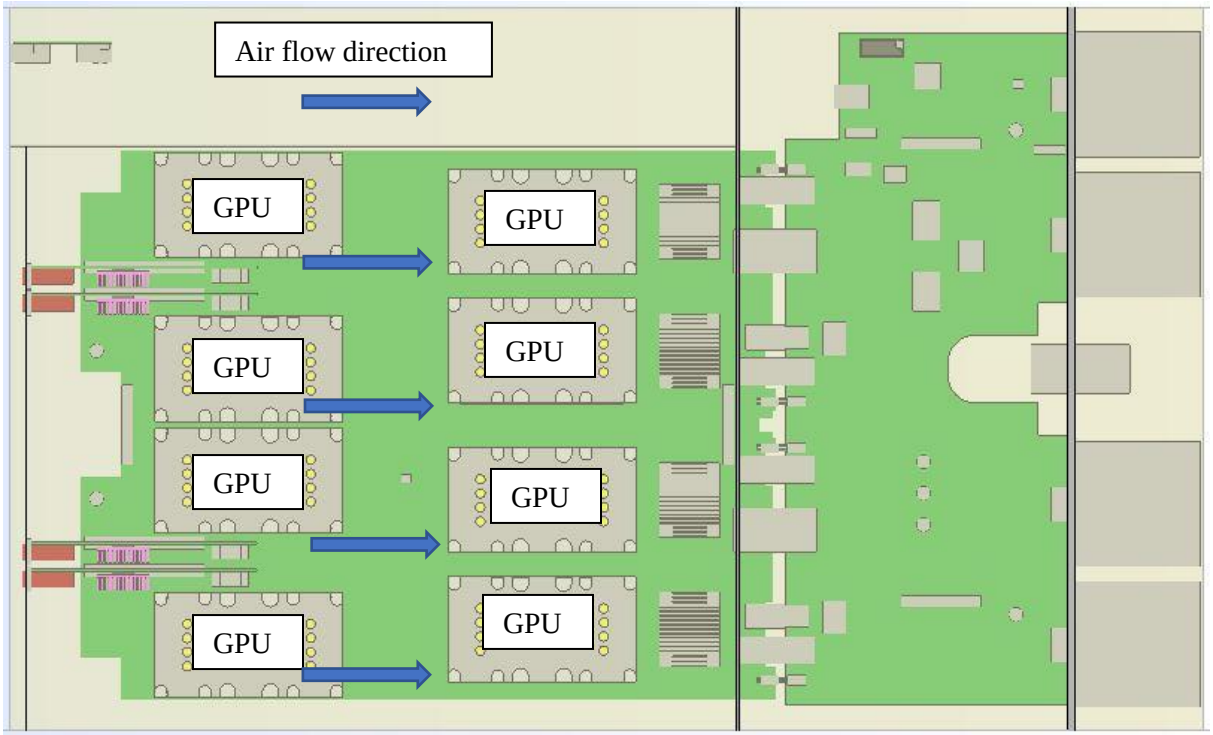

Figure 4. 9 Top and iso view of server assembly

4.2 Compact heatsink model

Figure 4. 6 and Figure 4. 7 shows the geometry of the heatsink used in this GPU server. The overall mesh count of the server reaches to 93 million. It will be very difficult for the computer to solve this problem with that much of higher mesh count and will take several days to solve one case with this mesh count. After investigation of each component of the server it is found that the heatsink contributes up to 80 million mesh count out of 93 million. After mesh sensitivity analysis of the heatsink each heatsinks gridding took up to 9.6 million gridding.

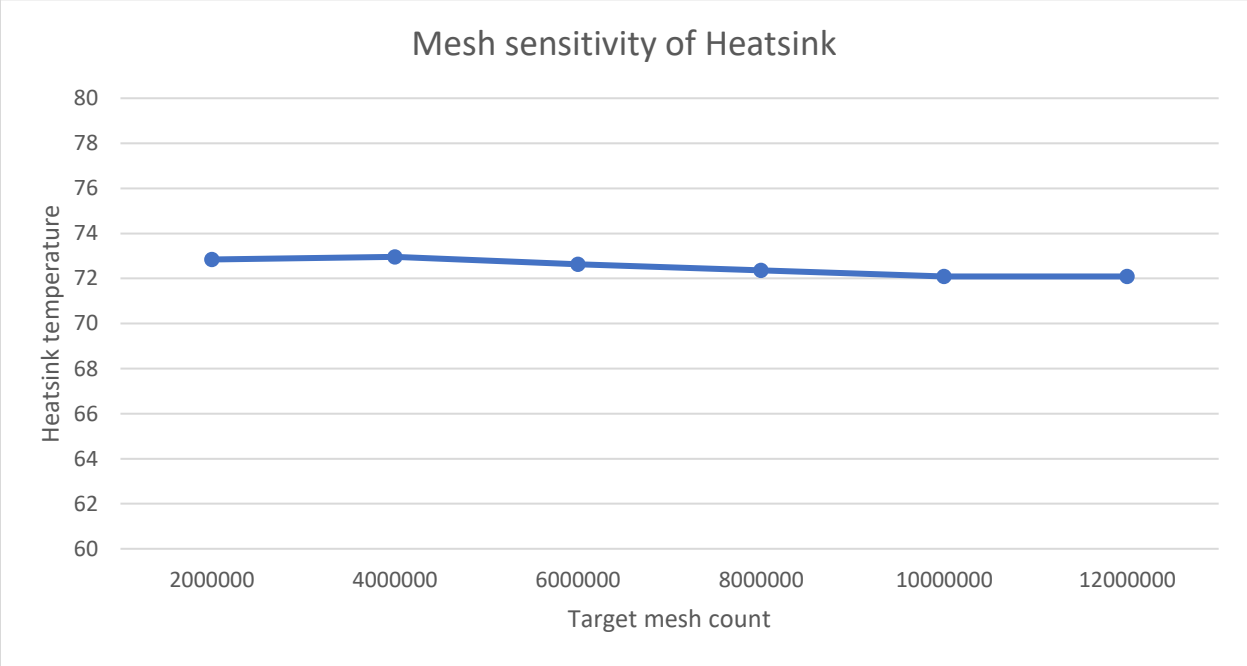

Figure 4.10 Max heatsink temp vs Target mesh count

To make a compact heatsink 6sigmaET requires the thermal resistance curve data and system resistance curve of the actual heatsink. Which can be obtained by placing actual heatsink into a test chamber of similar size. System resistance curve is pressure difference across heatsink vs air flow provided to the heatsink. Thus, four different air flow of 1,3,5 and 7 m/s velocity is considered. The obtained result of the system resistance is shown in Figure 4. 12. similarly, thermal resistance curve is thermal resistance vs air flow. Thermal resistance is a ratio of

temperature difference between inlet and outlet of chamber vs total power supplied which is 300 Watts.

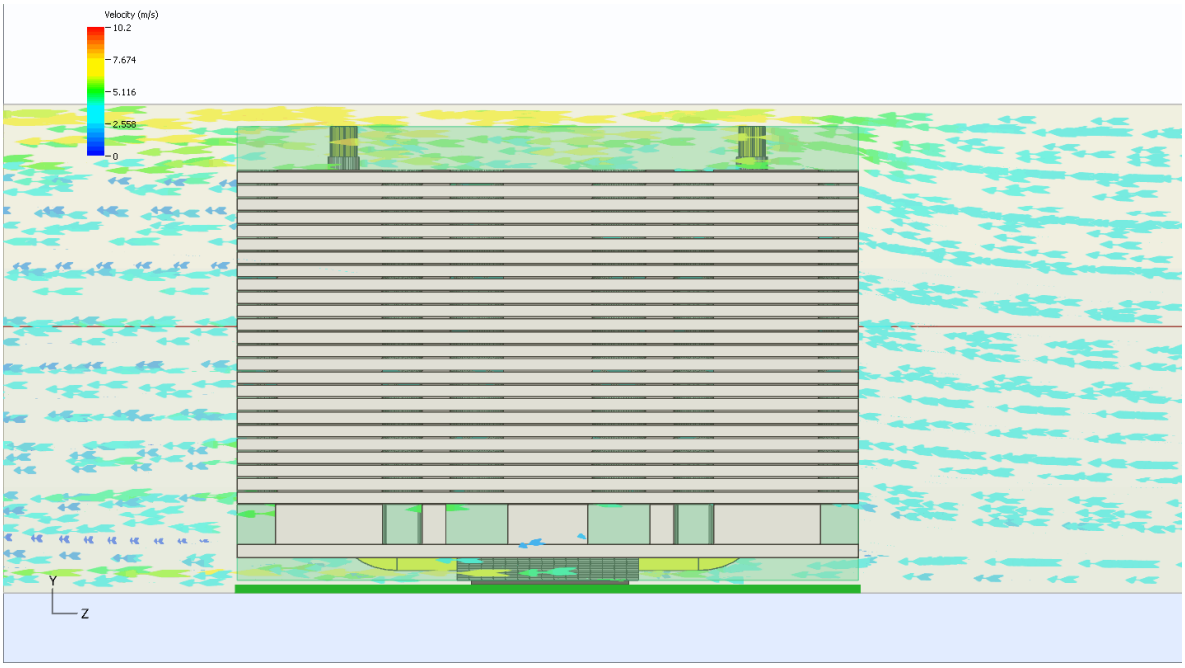

Figure 4. 11 Actual heatsink in test chamber

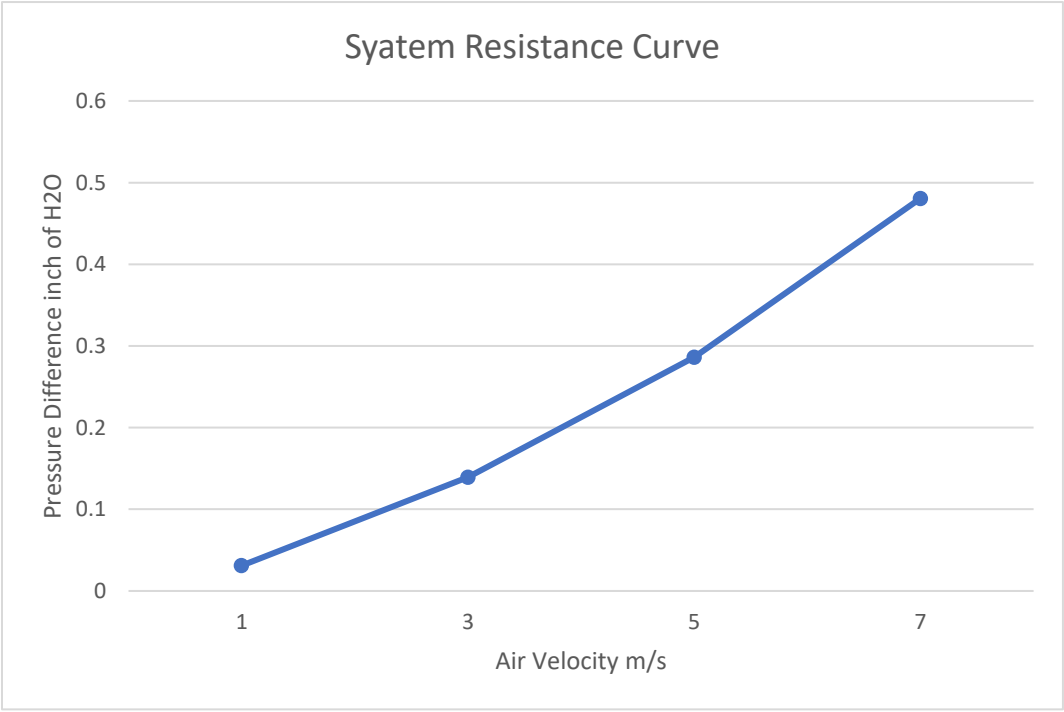

Figure 4. 12 System resistance curve of heatsink

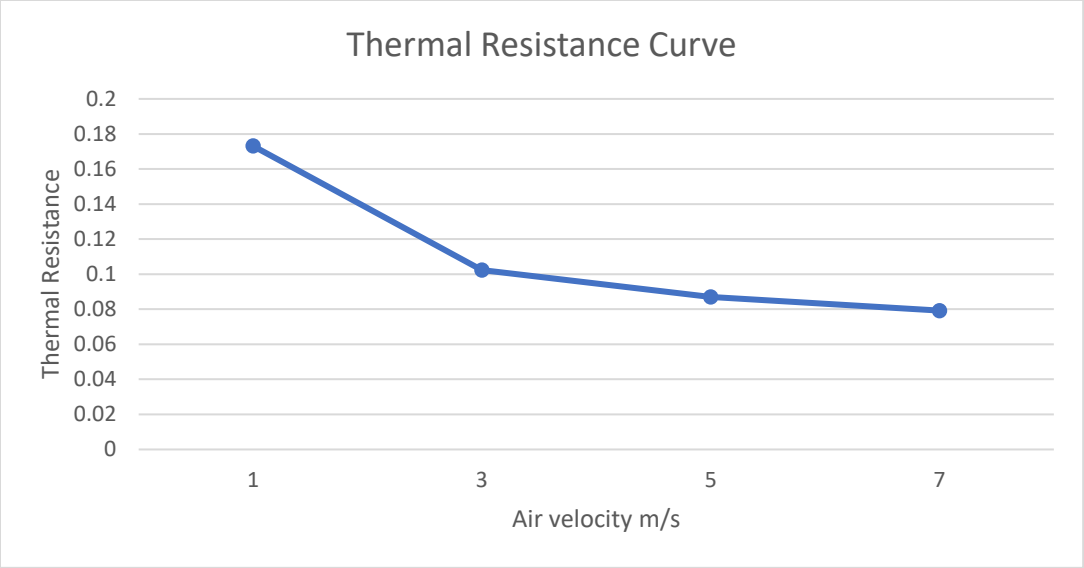

Figure 4. 13 thermal resistance curve of heatsink

The compact heatsink is a feature of heatsink that 6sigmaET provides for easy and fast computing with accuracy. The compact heatsink is the library component of the 6sigmaET in which whole heatsink is designed from the base plate. The base plate of the compact heatsink works as heat transfer medium. Which means the heat transfer or heat conduction occurs from the baseplate only. The area cover by fins in actual heatsink design is act as a flow resistance region and does not contributes in to heat transfer. This helps generating less mesh than actual heatsink.

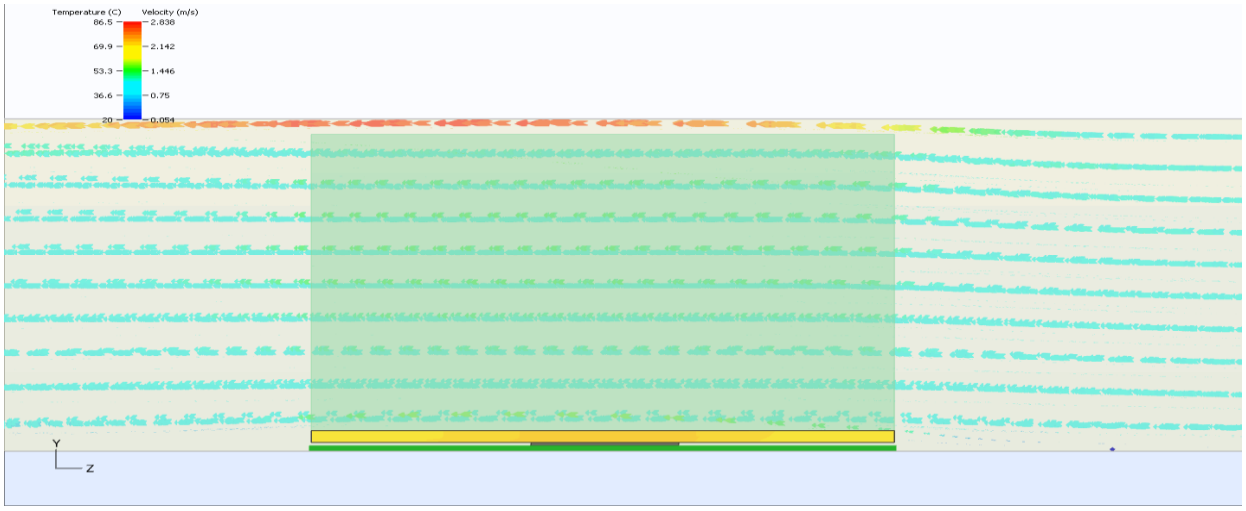

Figure 4. 14 Compact heatsink model

Here is the comparison of thermal resistance curve and system resistance curve of actual heatsink vs compact heatsink.

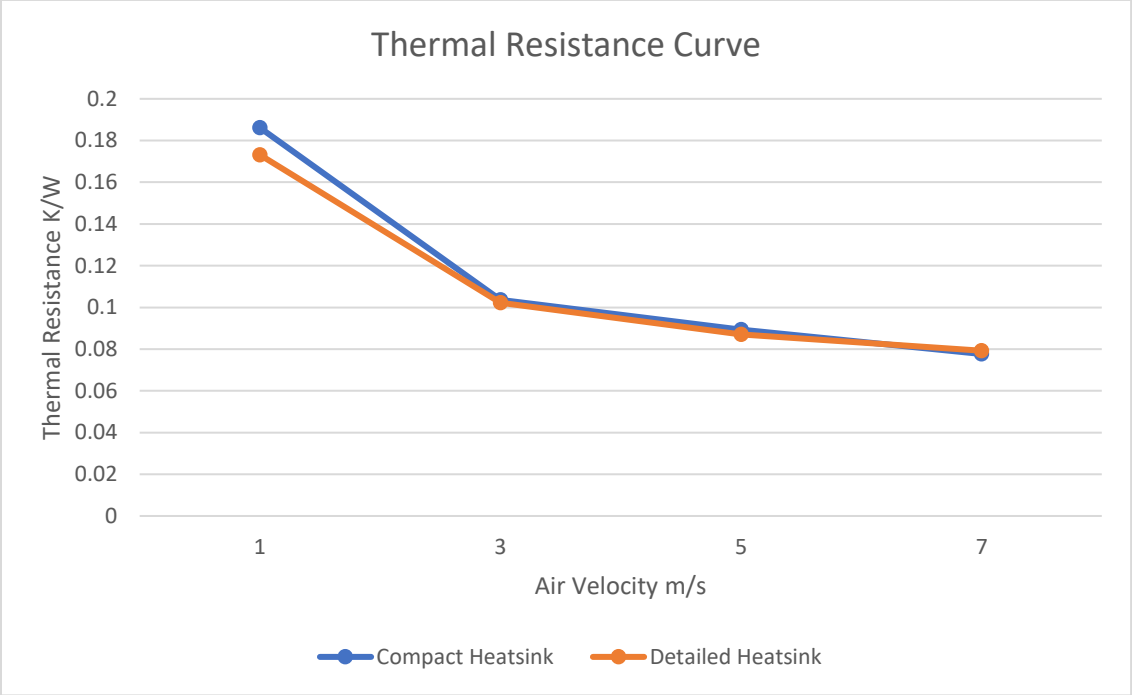

Figure 4. 15 Comparison of Thermal resistance curve

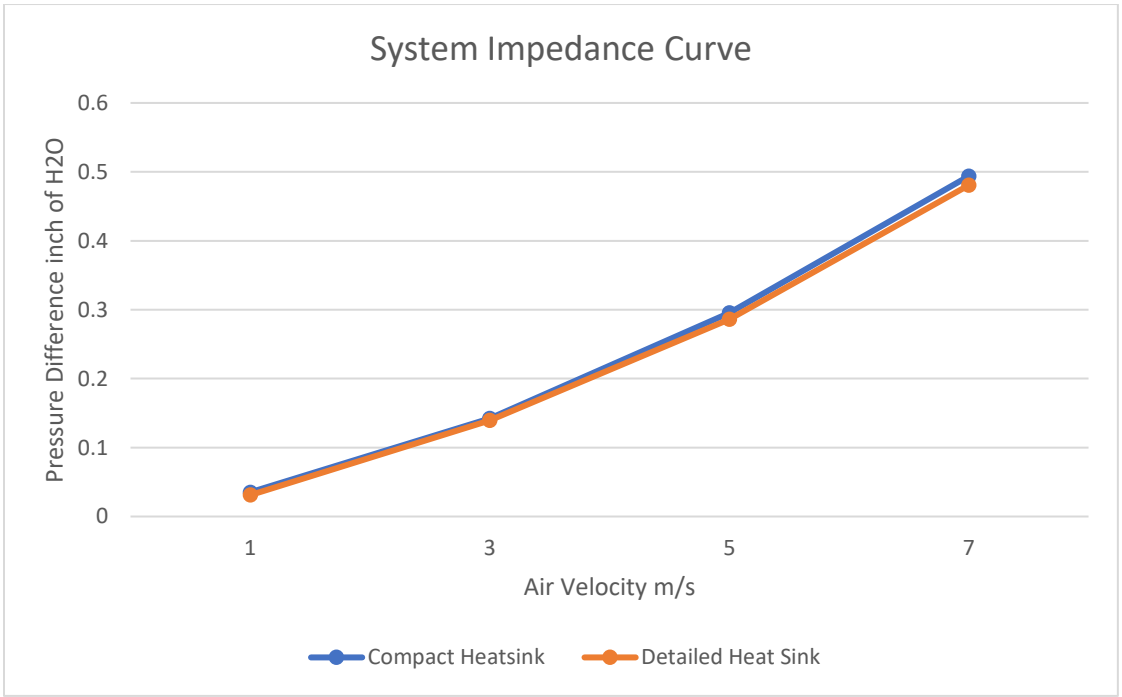

Figure 4. 16 comparison of actual vs compact heatsink

From figure 4.15 and 4.16 it can be said that both the curves of actual heatsink almost matches with compact heatsink. Hence, compact heatsink was applied into the server model. The mesh count reduces from 93 million to 15 million for whole server assembly after introducing the compact heatsink. The server model with 15 million mesh count was used for CFD simulations.

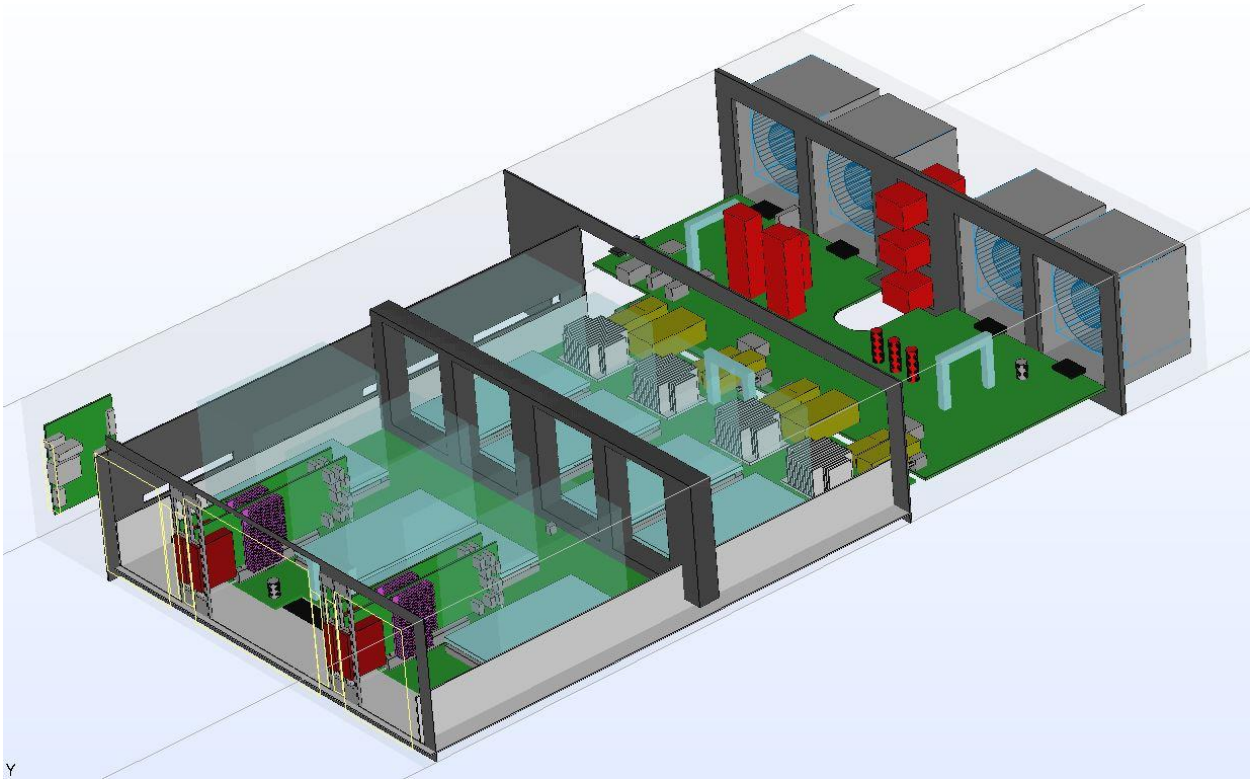

Figure 4. 17 Server with compact heatsink

4.3 Modeling and characterization of cold plate

All the above component ‘s details are available at the Open compute project website. This Particular server is 3OU size. Which means it uses 3 slots in an open rack. On top of that the server uses 4 fans that provides significant amount of CFM to cool the GPUs. The main factor for increasing the height of the server is TDP of the GPU therefore the required heatsinks to properly cool it. This study uses liquid cold plate as an alternative approach for cooling GPUs.

Cold plates are heat transfer devices that uses liquid flow to cool the microprocessors chips. In this case it was used for cooling the 8 GPUs. In this attempt the cold plated used for the server is jet flow impingement type, which impinge the cooling fluid into the fins in the form of

small jets. Water was selected as cooling fluid. The cooling fluid then passes through the fins of the cold plate and carry the heat out of the component. The graphical design and characterization of cold plate Is shown in figure 4.18.

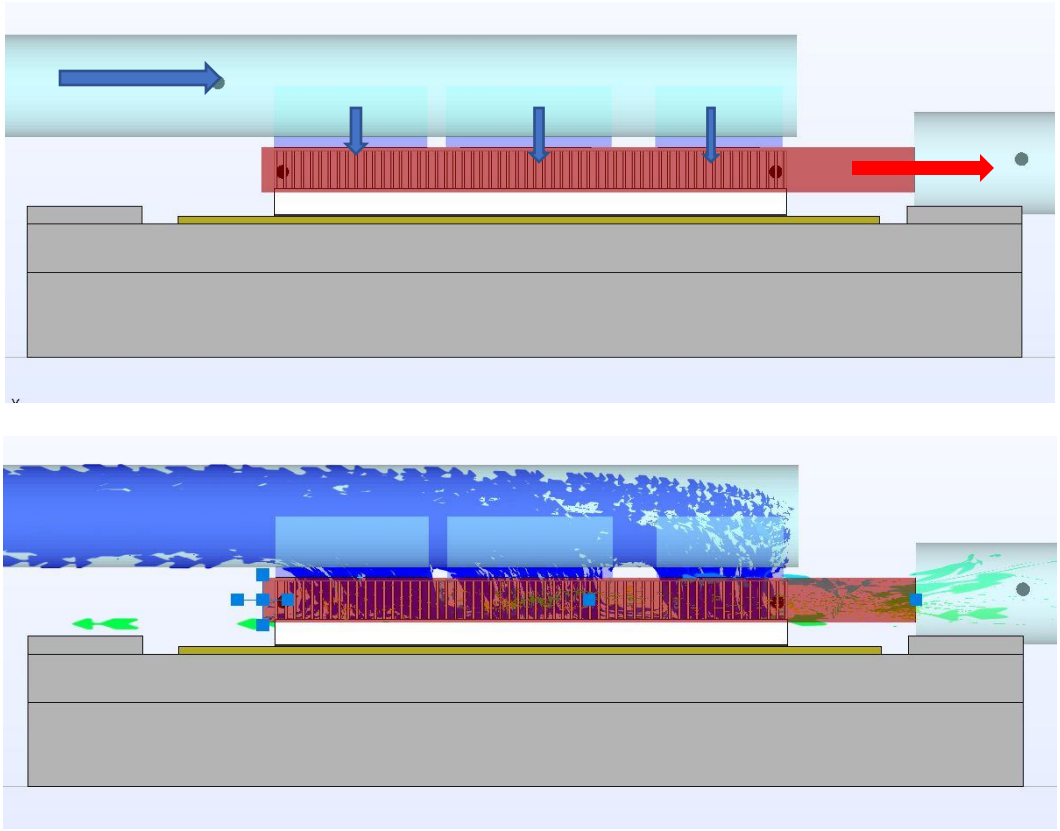

Figure 4. 18 Cold plate side view

Figure 4.15 shows the side view of the cold plate. The fluid flow at 15^0C comes from the supply pipe and passes through three ducts such that the liquid impinges in the middle of the cold plate. The three opening are created such a way that the pressure and distribution of fluid remain equivalent throughout the cold plate chamber. The cold plate chamber has two paths at the end of the fins through which it allows liquid to flow to the return pipe. This prearticular cold plate is designed for cooling high power chip with TDP up to 500Watts.

The characterization of cold plate is carried out by changing parameters such as fin thickness, fin separation therefore number of fins, and fin height inside the cold plate. The outcome of this characterization process is shown below in tabular format.

Table 4. 1 characterization by fin thickness and sin separation

Power W	Fin thickness mm	Number of fins	Fin separation mm	Flow rate L/min	Supply temperature ^{0}C	Component temp ^{0}C	Pressure Difference pa	Temp difference ^{0}C
500	***0.1***	***101***	***0.3004***	***1***	***15***	***52.81***	***1035.9***	***7.92***
500	0.1	51	0.7008	1	15	58.95	919.81	7.32
500	0.1	37	1.0122	1	15	64.22	1114.8	7.45
500	***0.3***	***67***	***0.3036***	***1***	***15***	***52.32***	***997.18***	***7.26***
500	0.3	40	0.7215	1	15	61.14	945.4	7.29
500	0.3	31	1.028	1	15	62.8	1059.4	7.3
500	0.5	50	0.309	1	15	53.32	1057.51	7.26
500	0.5	34	0.7012	1	15	62.33	933.2	7.29
500	0.5	27	1.0246	1	15	71.76	952.3	7.27

The characterization by fin thickness and fin separation is shown in table 4.1. the fin thickness is chosen as 0.1,0.3 and 0.5 mm. and separation of fins chosen are 0.3, 0.7 and 1 mm. the flow rate of fluid is 1 L/min. height of the fins, for this stage was 3 mm default. The supply temperature of the water was 15^{0}C. The TDP of the chip was 500Watts. As we can see from the results that two cases of fin geometry are highlighted. These two cases have relatively low component (chip) temperature. Pressure difference shown is between inlet and outlet of the cold plate which contributes in pumping power. The temperature difference between inlet to outlet is remain almost similar for all the cases. Temperature difference of 7.92^{0}C means that there is an increase of 7.92^{0}C in the fluid temperature compared to inlet temperature.

Now case of fin thickness 0.1 and fin thickness 0.3 with component temperature 52.81^{0}C and 52.32^{0}C are to be characterized further by using different fin height.

Table 4. 2 Characterization by fin height

Fin height mm	Fin thickness mm	Number of fins	Component temperature ^{0}C	Pressure Difference Pa	Temperature difference ^{0}C	Return temperature ^{0}C
3	0.1	101	52.81	919.81	7.26	22.26
4	0.1	101	52.3	964	7.3	22.3
5	0.1	101	58.8	816.4	7.3	22.3
3	0.3	67	52.32	997.18	7.92	22.92
4	***0.3***	***67***	***51.4***	***914.5***	***7.3***	***22.3***
5	0.3	67	54.3	788.4	7.4	22.4

As it can be seen from the obtained results from table 4.2 that when the fin height is 4 mm, fin thickness is 0.3 mm and number of fins is 67 mm we get the best possible temperature for the GPU (component). Which is 51.4^{0}C. the cold plate with this geometry is used for simulations.

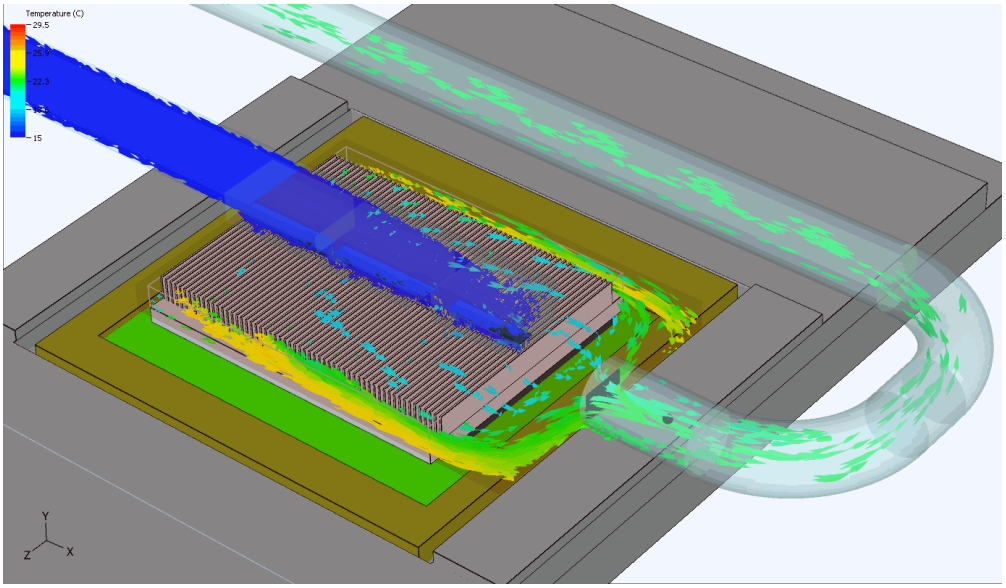

Figure 4. 19 Iso view of characterized cold plate with GPU

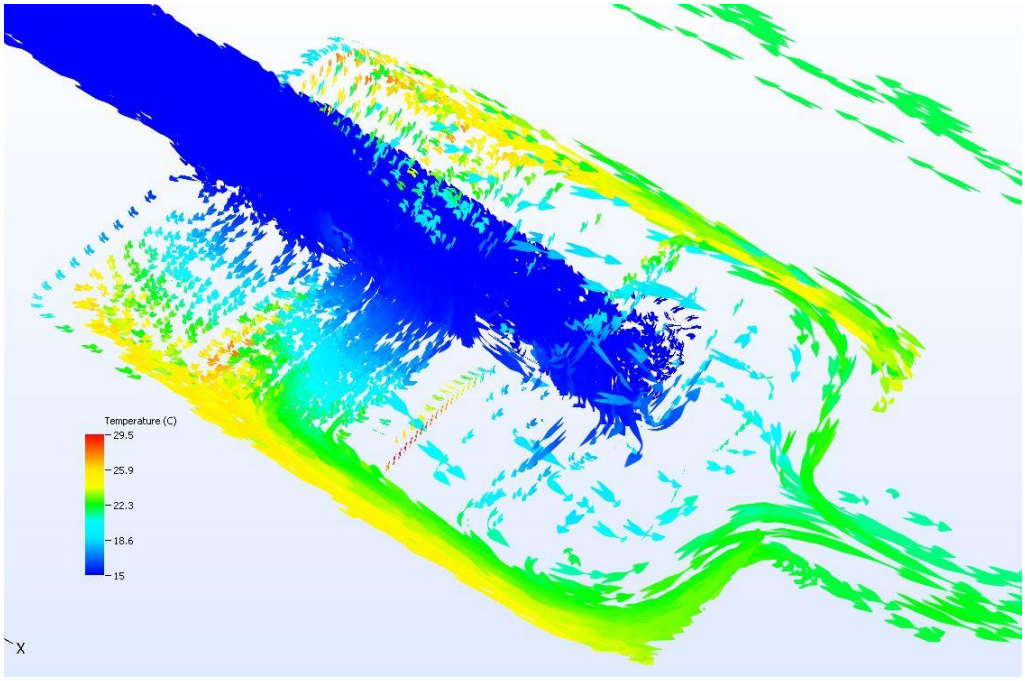

Figure 4. 20 Fluid flow of characterized cold plate

The cold plate obtained after the characterization it is now ready to introduce into the GPU server. The server has 8 GPUs therefore 8 cold plate will be placed on top of GPUs. The 6sigmaET software allows the single-phase liquid cooling. It has features to provide fluid control system in the user interface. In this study the fluid control devices such as pump is located outside of the server. The fluid supply inlet and outlet are located at the front of the server. the ducting for the cold plates was done parallelly as it gives an individual inlet supply for each cold plate from the supply duct. The fluid return from the cold plates are carried out the same way with fluid return duct situated below the supply duct. This gives each cold plate an equal amount of fluid supply so that no single GPU chip get significantly overheated than the rest of GPUs.

The parallel connection of cold plate is shown in figure 4.21. Fluid supply duct is shown in blue color and fluid return duct are in greenish yellow color. The color of the ducts is based on its liquid temperature.

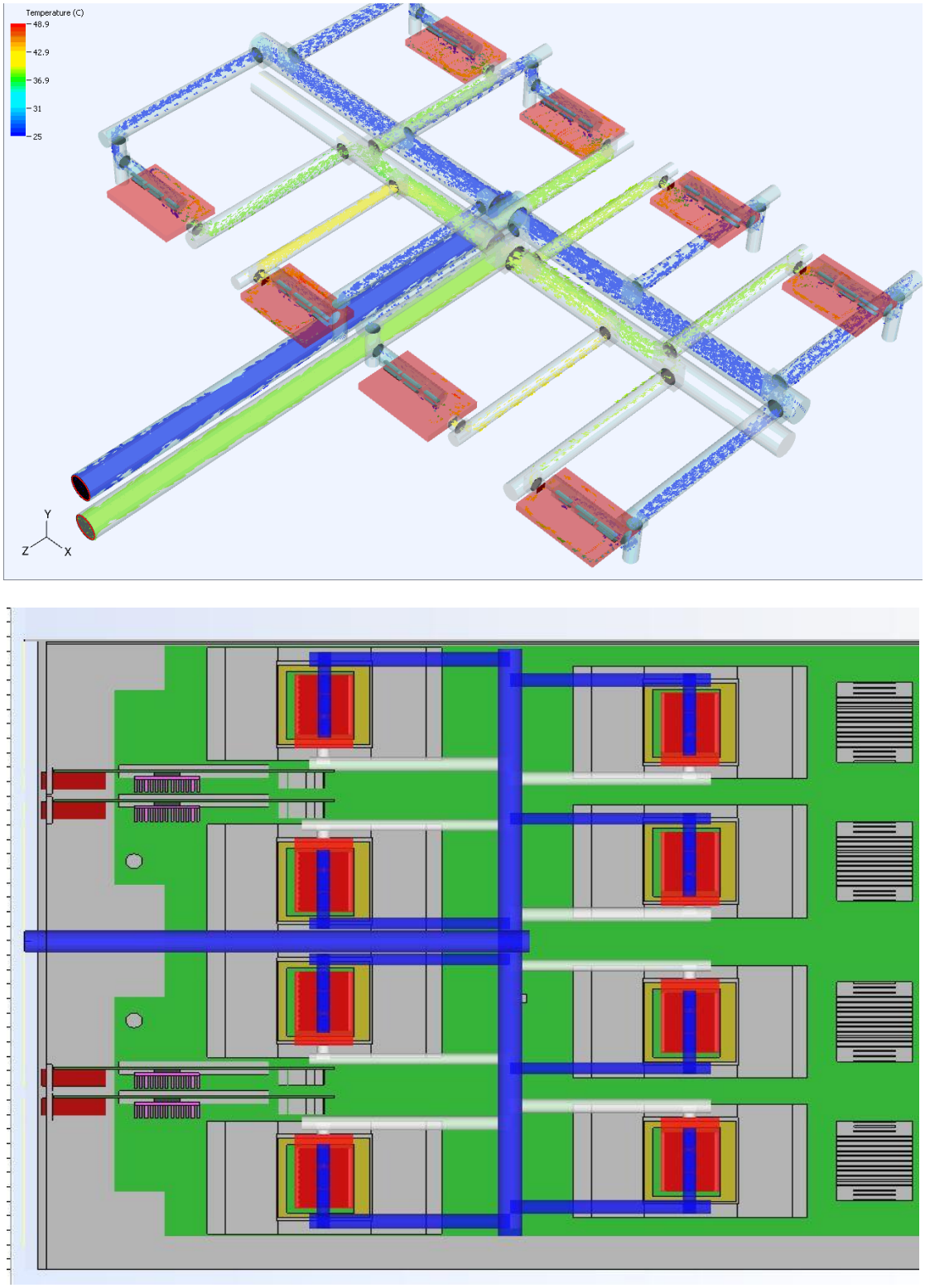

Figure 4. 21 Ducting of cold plates

CHAPTER 5

RESULTS AND COMPARISION

5.1 Air cooing Results

The GPU server was put under virtual test chamber with new compact heatsinks on top of each GPU for air cooling. The air flow was provided at the front of the test chamber. Two pressure sensors were placed at the front and the end of the server chassis to measure the pressure difference. The system resistance curve was obtained from pressure difference data and air flow rate.

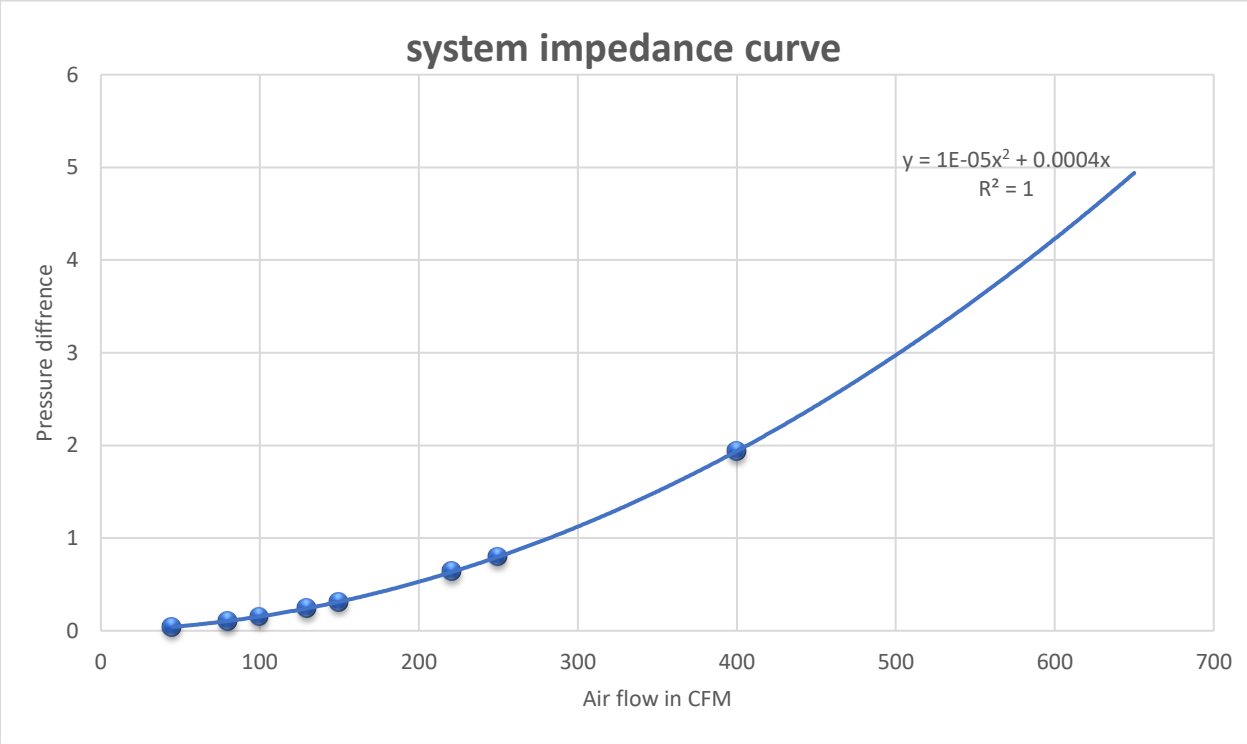

Figure 5. 1 system impedance curve – air cooling

Figure 5.2 shows the trend of the pressure difference vs supplied air flow rate. The system resistance curve follows the equation shown in the figure 5.1.

The GPUs are set to run at 100% utilization. Constant air flow from 40 CFM to 400 CFM was provided to the GPU server and the temperature of the GPUs were measured from the simulation results. The air was supplied at 20^0C. The temperature of all the GPU were measured over the different CFM and are expressed below in figure 5.1.

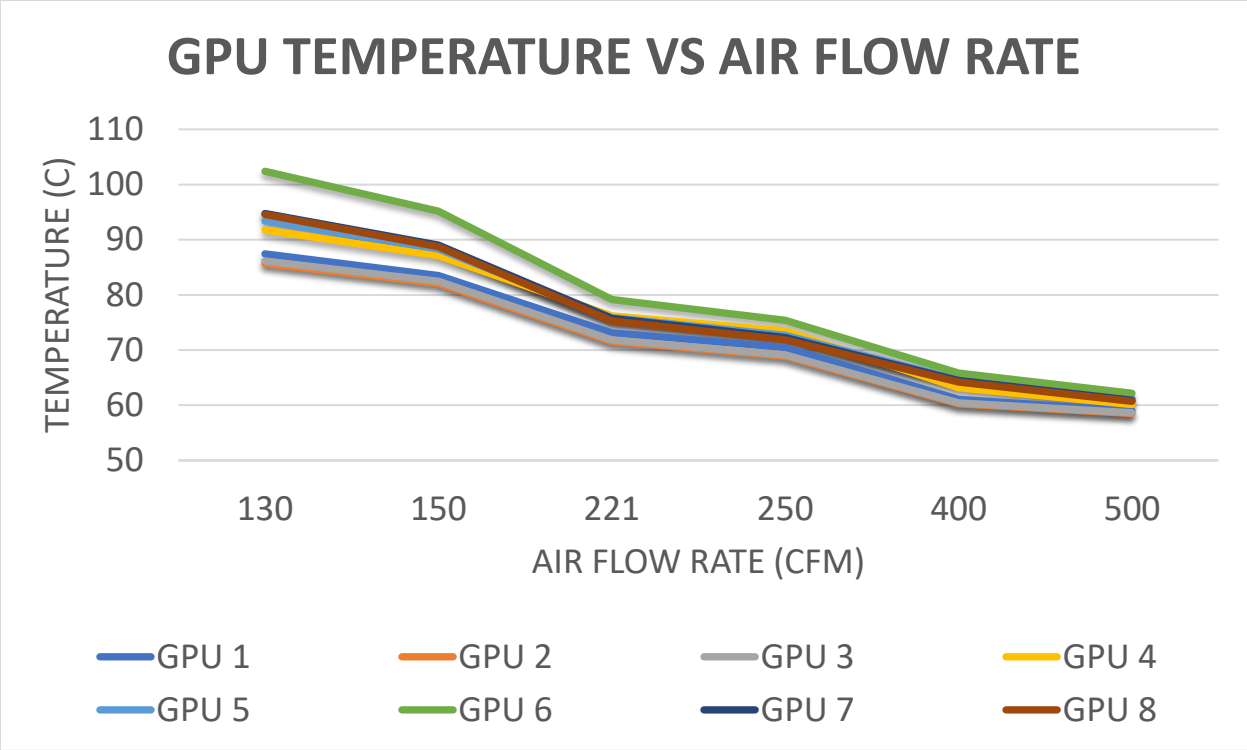

Figure 5. 2 GPU temp vs Air flow rate

It is observed from the figure 5.2 that the temperature of the GPU decreases with increase in air flow. The GPUs operational limit is 80^0C[16]. The CFD from the data obtained, it was clear that the temperature of all the GPU reaches below 80 around 221 CFM. CFD simulation ran for six air flow rates 130, 150, 221, 250, 400 and 500 CFM. After 221 CFM the temperature of all GPUs stay in operational limit. Hence, the air flow provided by the fans should be higher than at least 250 CFM for safer side. It can be seen from the figure 5.2 that for 500 CFM the temperature reaches as low as 60^0C.

From the simulation, the data regarding the pumping power or power required for cooling the whole server were calculated. table 5.1 gives the important details of pumping power at each CFM for air cooling the GPU server.

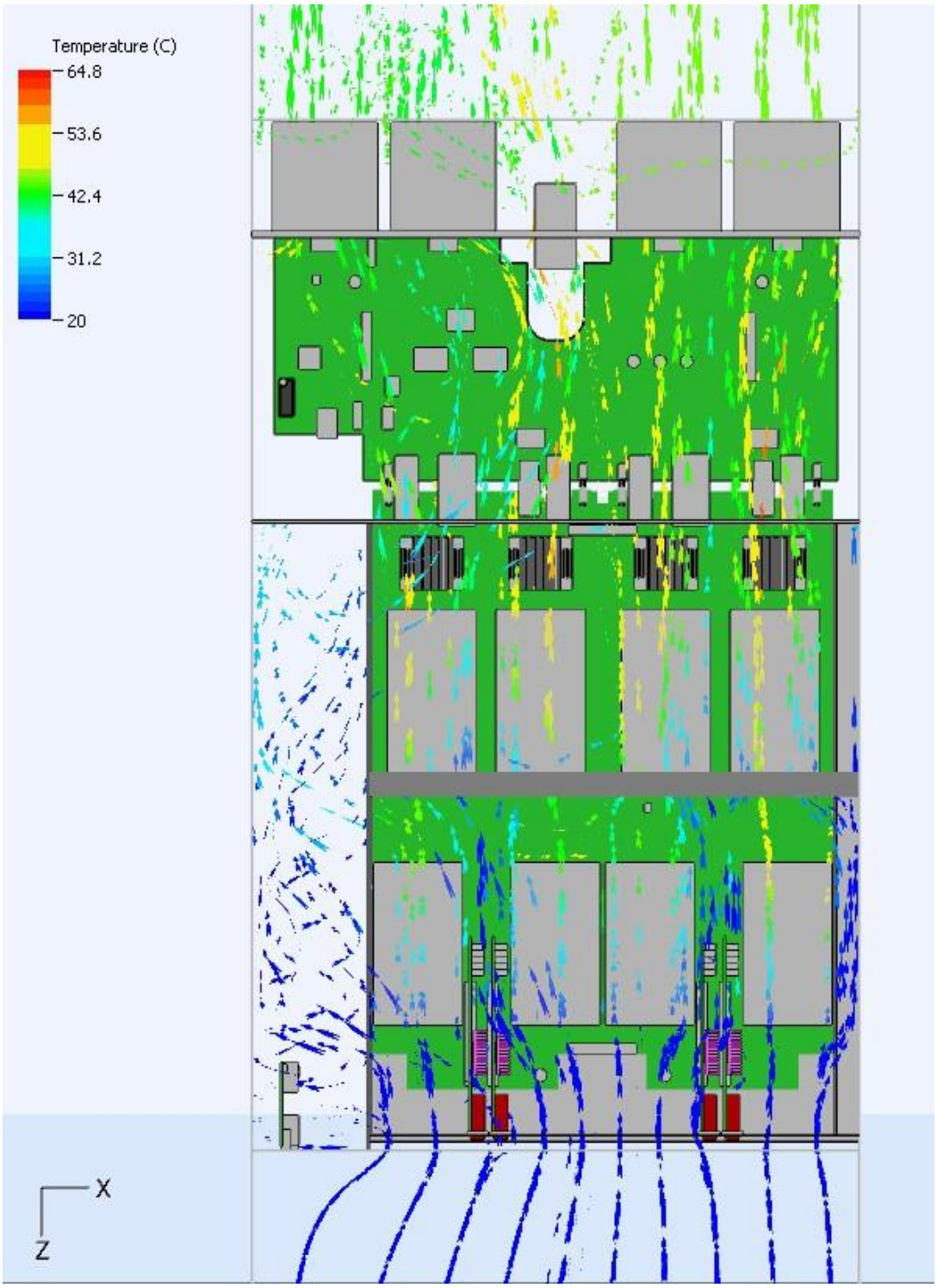

Figure 5. 3 Flow pattern of air-cooled server

Figure 5.3 shows the top view of the GPU server with air flow pattern across all the components.

Table 5. 1 pumping power for air cooling

Flow rate Cfm	Flow rate m^3/hr	Pressure Difference Inch of H_2O	Pressure Difference Kpa	Pumping power Watts
130	220.871	0.241	0.060	3.683
150	254.851	0.311	0.077	5.492
221	375.481	0.642	0.159	16.681
250	424.752	0.800	0.199	23.489
400	679.604	1.936	0.481	90.983
500	849.505	2.962	0.737	173.968

It is clear from the table 5.1 that with increasing the supplied air flow rate the pumping power also increases significantly to keep GPU under thermal operational conditions.

5.2 Liquid cooling of Server

The GPUs are of 300 watts each which are the primary and most heat generative component of this server. To obtain better cooling with air cooling method the server would only require more pumping power. To reach lower temperature than air cooling the new designed cold plate was implemented inside the server.

To obtain system impedance curve of the liquid cooled system the pressure sensors were placed inside inlet duct and outlet duct of the fluid supply system. The cooling liquid was water at 25^0C temperature. Four different flow rates were taken for the simulations.

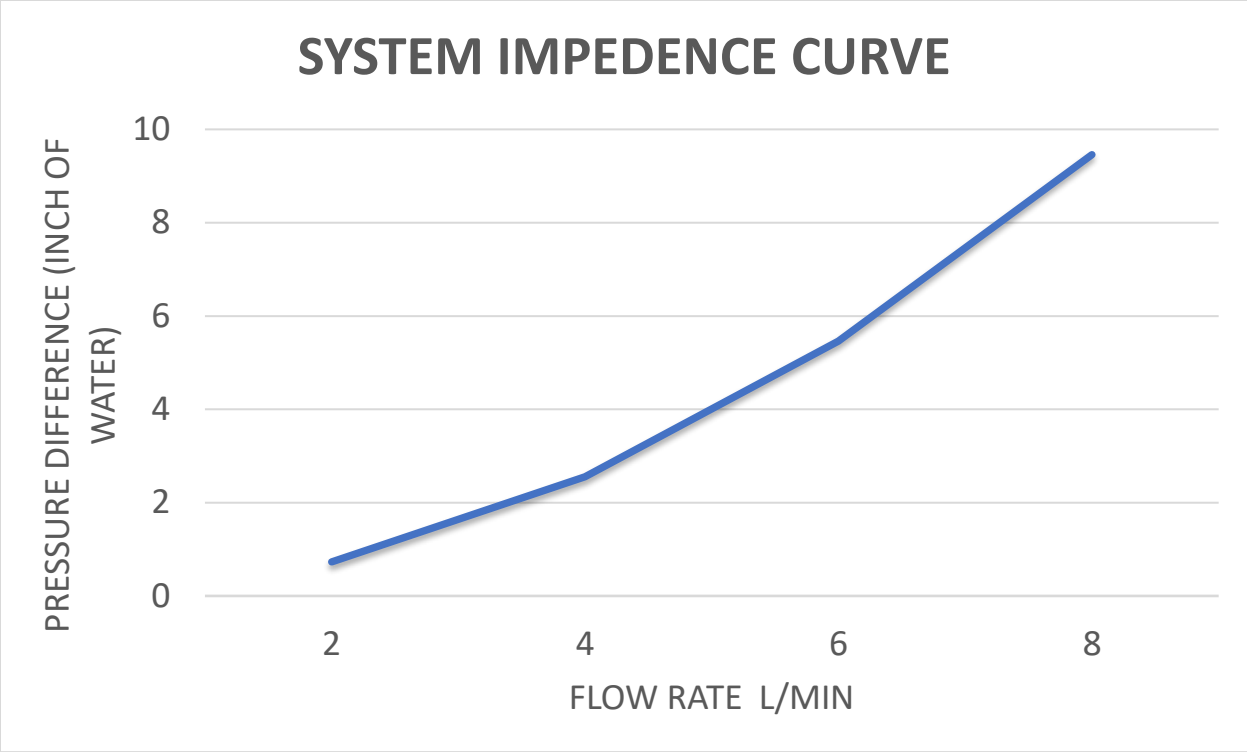

Figure 5. 4 system impedance curve – liquid cooling

Figure 5.4 shows the pressure difference of the fluid supply system vs supplied liquid flow rate. From the curve, the pressure difference between liquid supply duct and liquid return duct increases with increase in liquid flow rate.

By applying liquid cooling, the temperature of the GPUs drops down significantly. Figure 5.5 shows the temperature profile of all the GPUs with respect to supplied liquid flow rate. For flow rate of 2 L/min all the GPUs remain between temperature 60^0C to 65^0C. the temperature of al GPU gets much lower when introducing 8 L/min flow rate to the system.

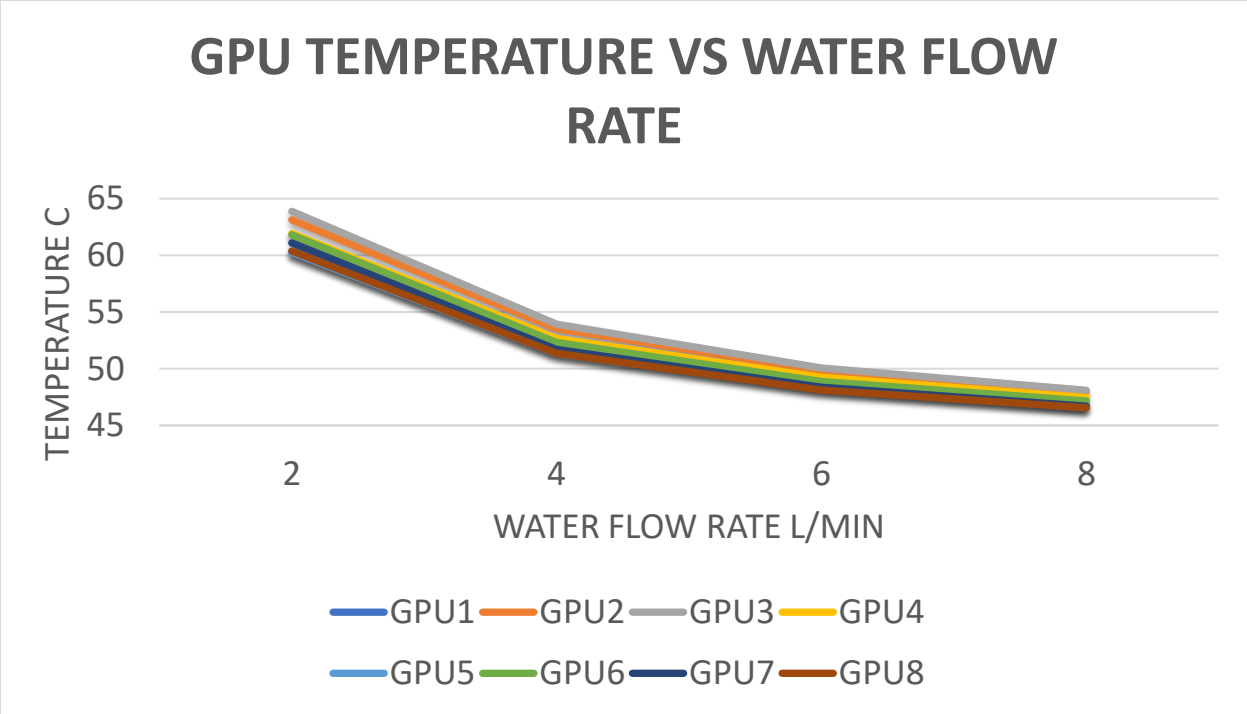

Figure 5. 5 GPU temp vs liquid flow rate

As observed from the figure 5.5 data that liquid cooling gives much better temperature compared to air cooling for all GPUs. The pumping power required for liquid cooling the GPUs are given in table 5.2. The required pumping power for liquid cooling also gives outstanding results compared to air cooling. It is clearly seen from the data that even for 8 L/min liquid supply it requires much less power.

Table 5. 2 pumping power for liquid cooling

Water low Rate L/min	Pressure Difference Inch of H_2O	Pressure Difference kpa	Pumping power Watts
2	0.729	0.181	0.006
4	2.553	0.635	0.042
6	5.464	1.359	0.135
8	9.458	2.353	0.313

However, the server still requires a certain amount of air flow rate to cool other small components such as PEX switches which are 12 to 16 watts. The temperature of these small

components was below their maximum thermal limits by providing a small amount of air flow at only 20 CFM across the server.

Table 5. 3 Total Pumping power (liquid + air)

Air flow rate cfm	Air flow rate m^3/hr	Pressure difference kpa	pumping power Watts	total pumping power(liquid+air) watts
20	33.980	0.698	6.592	6.906

Table 5.3 shows the total pumping power required for (liquid + air) cooled server. from the data obtained from air and liquid cooling methods it was clear that the liquid cooling provides much better outcomes than the air cooling with respect to GPU temperature and pumping power. Also, cold plates use less vertical space than the heatsinks. Assuming the 4 PCIe card can be mounted horizontally, the server height can also be reduced by significant margin.

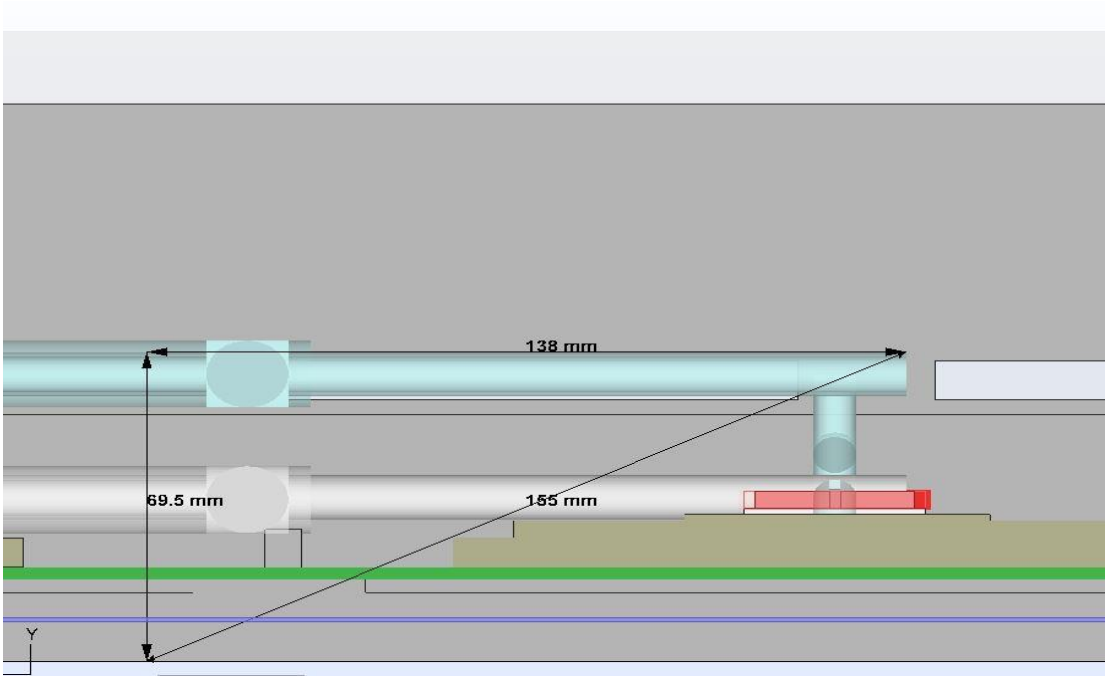

Figure 5. 6 Form factor of GPU server

The figure 5.6 shows the height required for the cold plates including ducting. The server is of 3OU size. And its actual height is about 142 mm and the figure shows the required height of the ducting and cold plates is about 70mm. It can be said that the server height can be reduced from 3OU to 1.5 OU that is 50% reduction in height.

CHAPTER 6

CONCLUSION

In this study, the two different cooling approaches for high powered GPU server has implemented and simulated to obtain better temperature. For air cooling, the server was subjected to six different air flow. The GPU temperature and pumping power was calculated from the CFD simulation of air-cooling approach. It was found out that to stay within the thermal limits of the GPU, the server needed air flow of more than 250 CFM. The maximum temperature of GPUs was about 75^0C at 250 CFM which require minimum of 23 watts theoretically. To reach GPU temperature lower than that, required more pumping power and air flow.

To apply liquid cooling technology into the server, specific high cooling capacity cold plate was needed. Therefore, to cool the GPUs with liquid cooling method, new cold plate was designed and implemented into the server model. The results of the liquid cooling approach were proved to be better than air cooling. The temperature obtained from applying liquid cooling to GPUs was as low as 48^0C at 8 L/min. which was well below thermal limitation of GPUs. The pumping power of fluid supply system was calculated from the CFD simulations. Also considering small air flow of 20 CFM for cooling of small components on the motherboard and 8L/min liquid flow rate the (liquid + air) cooling approach required about 7 watts of pumping power theoretically.

Both air cooling and Liquid cooling method was performed on GPU server. Significant reduction in pumping power obtained by liquid cooling approach compared to air cooling. Which means significant improvement in power usage per server. Assuming the 4 PCIe card can be mounted horizontally, the liquid cooling approach my reduce the height of the server by almost 50%. Considering a 42OU rack, instead of 14 of these GPU servers, 28 can be accommodated in one rack. Hence significant reduction in size of overall facility with less pumping power than before.

REFRENCES

[1] T. Gao, H. Tang, Y. Cui, and Z. Luo, “A Test Study of Technology Cooling Loop in a Liquid Cooling System,” *Proc. 17th Intersoc. Conf. Therm. Thermomechanical Phenom. Electron. Syst. ITherm 2018*, pp. 740–747, 2018.

[2] “What is a data center? - How Data Centers Work | HowStuffWorks.” https://computer.howstuffworks.com/data-centers1.htm.

[3] sweden facebook datacenter at lulea, “Inside the Facebook Data Center in Sweden (Video) | Data Center Knowledge.” https://www.datacenterknowledge.com/archives/2016/05/24/inside-facebook-data-center-sweden-video.

[4] “File:IBM HS20 blade server.jpg - Wikipedia.” https://en.wikipedia.org/wiki/File:IBM_HS20_blade_server.jpg.

[5] “Cisco UCS C-Series Multinode Rack Servers - Cisco.” https://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-c4200-series-rack-server-chassis/index.html.

[6] “Google’s scalable supercomputers for machine learning, Cloud TPU Pods, are now publicly available in beta | Google Cloud Blog.” https://cloud.google.com/blog/products/ai-machine-learning/googles-scalable-supercomputers-for-machine-learning-cloud-tpu-pods-are-now-publicly-available-in-beta.

[7] “HP BladeSystem.” https://en.wikipedia.org/wiki/HP_BladeSystem.

[8] A. Shehabi *et al.*, “United States Data Center Energy Usage Report,” 2016.

[9] pixabay, “Network Iot Internet Of Things - Free vector graphic on Pixabay.” https://pixabay.com/vectors/network-iot-internet-of-things-782707/.

[10] “2011 Thermal Guidelines for Data Processing Environments-Expanded Data Center Classes and Usage Guidance Whitepaper prepared by ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment,” 2011.

[11] hp, “Optimizing data centers for high-density computing, 2nd edition.”

[12] “ASHRAE TC9.9 Data Center Networking Equipment-Issues and Best Practices.”

[13] 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 Transf. Eng.*, vol. 28, no. 4, pp. 258–281, Apr. 2007.

[14] “THREE REASONS TO DEPLOY NVIDIA TESLA V100 IN YOUR DATA CENTER,” 2018.

[15] K. Y. Thoon, “Big Basin-Flexible GPU Expander.”

[16] “Big Basin-JBOG Specifications Rev 1.0,” 2016.

[17] “QCT Rackgo X Big Basin Contribution Following Big Basin Specification.” http://files.opencompute.org/oc/public.php?service=files&t=2d705157b11cca46279d18312a6b80d1&download.

[18] “The next step in Facebook’s AI hardware infrastructure - Facebook Code.” [Online]. Available: https://code.fb.com/ml-applications/the-next-step-in-facebook-s-ai-hardware-infrastructure/.

[19] D. Kuzmin, “Introduction to Computational Fluid Dynamics.”

[20] “Lecture 5-Solution Methods Applied Computational Fluid Dynamics Instructor: André Bakker.”

[21] H. K. Versteeg and W. Malalasekera, “An Introduction to Computational Fluid Dynamics Second Edition.”

[22] “Laminar vs. Turbulent Flow.” https://www.cfdsupport.com/OpenFOAM-Training-by-CFD-Support/node334.html.

[23] K. Dhinsa, C. Bailey, and K. Pericleous, “Low Reynolds number turbulence models for accurate thermal simulations of electronic components,” in *5th International Conference on Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems, 2004. EuroSimE 2004. Proceedings of the*, pp. 483–490.

[24] “Features | 6SigmaET by Future Facilities.” https://www.6sigmaet.info/software/features/.

BIOGRAPHICAL INFORMATION

Ankit Sutaria has received his Bachelor of Engineering degree in Mechanical Engineering from the Gujarat Technological University (GTU), Gujarat, India. He has completed his Master's in Mechanical Engineering from the University of Texas at Arlington, Texas, USA in August 2019.

During his master's, Ankit worked on projects under Dr Dereje Agonafer in the University of Texas at Arlington. Thermal management and optimization in data center is his main area of research. He associated himself with significant industry collaborated research projects and studied areas of air cooling, liquid cooling and direct/indirect evaporative cooling etc. He has gained fundamental knowledge of CFD and worked on data center specific codes like 6sigmaE, 6sigmaRoom, ANSYS icepack and FloTherm. He has worked experimentally on cisco servers with air flow bench for thermal and system resistance study. Ankit was a member of three NSF and IUCRS funded projects collaborated with university of Texas at Arlington and one of his projects research paper on modular data center was accepted in ASME interpack conference.