

**ADVANCED THERMAL MANAGEMENT STRATEGIES FOR  
ENERGY EFFICIENT DATA CENTERS**

A Thesis  
Presented to  
The Academic Faculty

by

Ankit Somani

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in the  
School of G.W. Woodruff School of Mechanical Engineering

Georgia Institute of Technology  
December 2008

**(C) COPYRIGHT 2008 BY ANKIT SOMANI**

# **ADVANCED THERMAL STRATEGIES FOR ENERGY EFFICIENT DATA CENTERS**

Approved by:

Dr. Yogendra Joshi, Advisor  
School of Mechanical Engineering  
*Georgia Institute of Technology*

Dr. Seyed Ghiaasiaan  
School of Mechanical Engineering  
*Georgia Institute of Technology*

Dr. Karsten Schwan  
College of Computing  
*Georgia Institute of Technology*

Date Approved: [Date Approved by Committee]

[To my loving family]

## **ACKNOWLEDGEMENTS**

I would like to sincerely thank my advisor, Dr. Yogendra Joshi, for his guidance over the past year and the confidence he showed in every step of my work and the decisions I made. I also want to take this opportunity to thank my committee members for their valuable suggestions

Amongst my friends, I would like to mention two people in particular, Shawn and Nikhil. Shawn has been a friend and colleague with whom I have had several interesting discussions, both academic and otherwise, and who helped me right from building the experimental setup to critiquing my work. Nikhil has been a source of endless support and has encouraged me whenever I was bogged down. I would also like to thank the other METTL members for their valuable feedback.

I want to express my deepest love and gratitude for my family for their continued encouragement and support during all these years.

# TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF SYMBOLS AND ABBREVIATIONS	xiv
SUMMARY	xv
 <u>CHAPTER</u>	
1. INTRODUCTION	1
1.1. Historical Data and Trends in Data Centers	4
1.2. Power Flow in a Data Center – Area of Focus	6
1.3. Importance of Air Management in a Data Center	9
1.4. Literature Survey	10
1.4.1. Government and professional organizations and regulatory bodies	10
1.4.2. Industry	16
1.4.3. Academic work	19
1.5. Objective of Research	20
1.5.1. Objective 1	21
1.5.2. Objective 2	21
1.5.3. Objective 3	21
2. AMBIENT INTELLIGENCE BASED LOAD MANAGEMENT	23
2.1. Modeling	23
2.1.1. Geometry	23

2.1.2.	Boundary Conditions	24
2.1.3.	Meshing	26
2.1.4.	Simulation Facts/Conditions	27
2.2.	Simulation Results	28
2.3.	Requirements for a load management algorithm	33
2.3.1.	Extrinsic Requirements	35
2.4.	Assumptions and Simplifications	37
2.4.1.	Effect of Natural Convection	38
2.4.2.	Humidity	39
2.4.3.	Linear Variation in Temperature	39
2.4.4.	Linear transformation between power dissipated and % CPU utilization.	40
2.4.5.	Server fan velocities constant	40
2.5.	Formulation	40
2.5.1.	Calibration Phase	41
2.5.2.	Operational Phase	43
2.6.	AILM Results and Discussions	46
2.6.1.	Calibrating the Data Center	46
2.6.2.	Comparison with the “Unmanaged” Data Center	46
2.6.3.	Case study 1: Load Distribution variation with $V_{CRAC}$	53
2.6.4.	Heterogeneous Data Center	57
2.6.5.	Case study 2: Heterogeneous Air-Cooled Data Center	58
2.6.6.	Case study 3: Design of Data Center for Liquid Cooled Racks	65
2.6.7.	Case study 4: Sensitivity of DCHDC with $T_C$	68

2.7.	Experimental Setup	69
2.7.1.	APC Server Simulators	69
2.7.2.	Temperature Measurements	70
2.7.3.	Uncertainty Analysis	71
2.7.4.	Anemometer	72
2.7.5.	Control Volume	72
2.7.6.	Placements of measurement devices:	75
2.8.	Experimental Results and Discussions	79
2.8.1.	Conditions simulated	79
2.8.2.	Repeatability	80
2.8.3.	Setting up AILM – Calibration of the Data Center.	80
2.8.4.	Case Study 5: $T_C = 32.2^\circ\text{C}$	82
2.8.5.	Case Study 6: $T_C = 28^\circ\text{C}$	86
3.	SCALABLE PODS BASED CABINET ARRANGEMENT AND AIR DELIVERY	88
3.1.	Introduction	88
3.2.	Computational Model	91
3.2.1.	Boundary conditions	92
3.2.2.	Solution and convergence	94
3.3.	Results and Discussions	95
3.3.1.	Case 1: $V_{\text{CRAC}} = 7 \text{ m/s}$ , $q'''$ (for each rack) = $16161.61 \text{ W/m}^3$	95
3.3.2.	Case 2: $V_{\text{CRAC}} = 4 \text{ m/s}$ , $q'''$ (for each rack) = $8888.89 \text{ W/m}^3$	96
3.3.3.	Case 3: Optimal velocity for the S-POD case	101
3.3.4.	Case 4: Possible modifications in the S-POD layout.	102

4.	Quantifying Inefficiency – Data Center Metrics	108
4.1.	Metrics	109
4.1.1.	Issues in defining metrics	109
4.1.2.	What is a Good Metric?	111
4.1.3.	General Metrics	122
4.1.4.	Component Specific Metrics	123
4.1.5.	Server-Level Metrics	124
4.1.6.	IT Metrics	128
4.1.7.	Facilities Metrics	130
4.1.8.	Air-management metrics	131
4.1.9.	Panoptic Metrics	132
4.2.	Control Parameters	135
5.	CONCLUDING REMARKS	138
5.1.	Summary	138
5.2.	Future Work	139
	APPENDIX A: Computational fluid dynamics-heat transfer	141
	REFERENCES	145

## LIST OF TABLES

	Page
Table 2.1 : Grid Independence Study .....	27
Table 2.2. Rack specifications for cases .....	59
Table 2.3. Rack configurations for Case 2.....	61
Table 2.4. Server heat load map for repeatability experiments .....	80
Table 2.5. The "A" matrix of AILM .....	82
Table 2.6. Server heat load map for AILM, Case Study 5.....	83
Table 2.7. Server heat load map for "Uniform", Case study 5 .....	84
Table 2.8. Server heat load map for AILM, Case Study 6.....	86
Table 2.9. Server heat load map for Uniform, Case Study 6 .....	86
Table 3.1: Mass Flow rates and other features for S-Pod Layout.....	101
Table 4.1. List of Power metering devices, location and applications .....	115
Table 4.2. List of Power metering devices, location and applications (Contd.).....	116
Table 4.3. List of Power metering devices, location and applications (Contd.).....	117
Table 4.4. List of Temperature sensors, location and applications.....	118
Table 4.5. List of Temperature sensors, location and applications (Contd.) .....	119
Table 4.6. List of velocity sensors/flow meters, location and applications .....	120
Table 4.7. List of Pressure sensors, location and applications.....	121

## LIST OF FIGURES

	Page
Figure 1.1. A typical data center [1] .....	2
Figure 1.2 A Hot Aisle-Cold Aisle (HACA) arrangement of racks in a data center .....	3
Figure 1.3. Trends for heat load dissipations for IT systems [7] .....	4
Figure 1.4. EPA Prediction of energy use in data centers [2].....	6
Figure 1.5. Power flow diagram of a typical data center .....	8
Figure 1.6. Variation in data center energy usage [7].....	8
Figure 1.7. Roadmap for the Thesis.....	22
Figure 2.1. Computational model simulated; Nomenclature of servers .....	25
Figure 2.2. Temperature variations (K) in the streamline path originating from perforated tiles ( $V_{CRAC} = 10$ m/s).....	29
Figure 2.3. Temperature variations (K) in the streamline path originating from perforated tiles ( $V_{CRAC} = 5$ m/s).....	30
Figure 2.4. Streamlines generated from back of $A_1$ and $B_1$ racks (i) $V_{CRAC} = 5$ m/s, (iii) $V_{CRAC} = 10$ m/s; Back of $A_{11}$ - $A_{41}$ and $B_{11}$ - $B_{41}$ (ii) $V_{CRAC} = 5$ m/s, (iv) $V_{CRAC} = 10$ m/s .	32
Figure 2.5. Temperature contours at rack inlets (K).....	33
Figure 2.6. Algorithm for AILM.....	41
Figure 2.7. Various cases possible for DCHDC with VCRAC .....	45
Figure 2.8. Temperature contours at inlet face of racks for loading according to (a) Uniform case, (b) AILM case. ( $V_{CRAC} = 5$ m/s) .....	48
Figure 2.9. Server inlet temperatures versus server no. (please not nomenclature on top left).....	50

Figure 2.10. Data center Heat Dissipation Capacities (DCHDC) with $V_{CRAC}$ for "Uniform" and AILM cases .....	52
Figure 2.11. $V_{CRAC} = 3$ m/s, Power Dissipation Map: (a) B Rack, (b) A Rack .....	54
Figure 2.12. B Rack Power Dissipation Map: (a) $V_{CRAC} = 4$ m/s, (b) $V_{CRAC} = 5$ m/s, .....	56
Figure 2.13. Power Dissipation Map - (a) $V_{CRAC} = 7$ m/s, B Rack, (b) $V_{CRAC} = 7$ m/s, A Rack, (c) $V_{CRAC} = 8$ m/s, A Rack.....	57
Figure 2.14. Different rack layouts .....	59
Figure 2.15. Data center Heat Dissipation Capacity with $V_{CRAC}$ for Table 2.1 .....	60
Figure 2.16. Data center Heat Load Efficiency with $V_{CRAC}$ for different configurations. 62	
Figure 2.17. Data center Workload efficiency with $V_{CRAC}$ for different configurations .. 63	
Figure 2.18 Heat load and Workload efficiency for objectives being maximum workload .....	64
Figure 2.19. Rack inlet temperature contours for $V_{CRAC} = 5$ m/s (a) Uniform case, (b) Heterogeneous Case.....	67
Figure 2.20. DCHDC with $T_c$ for the considered model. ....	68
Figure 2.21 Connection of temperature measuring instruments.....	71
Figure 2.22 The control volume for experiments (a) Elevation, (b) Near CRAC inlet (Rack on left) (c) the experimental rack (d) Underfloor plenum (e) collapsible back. ....	73
Figure 2.23 (a) Schematic of server simulator rack with thermocouple placement, (b) Schematic of CRAC with thermocouple and anemometer placement.....	77
Figure 2.24. (a) Wires hanging in a facility [9], (b) wire obstruction in the current control volume.....	78

Figure 2.25. Thermocouple base units outside control volume .....	78
Figure 2.26. Temperature stability plot for experiments for a 30 min time interval .....	81
Figure 2.27. Comparison of temperature with server inlet for experimental and computational values for Case Study 5.....	84
Figure 2.28. Comparison of temperature with server inlet for experimental and computational values for Case Study 6.....	87
Figure 3.1 Various modifications to the Hot Aisle-Cold Aisle layout [6].....	90
Figure 3.2. Scalable Pods (S-Pod) based layout with maintenance options .....	91
Figure 3.3. Symmetrical model simulated for the HACA layout .....	93
Figure 3.4. Symmetrical model simulated for the S-Pod layout.....	94
Figure 3.5. Rack inlet temperature contours for HACA and S-Pod layouts, $V_{CRAC} = 7$ m/s .....	96
Figure 3.6. Rack inlet temperature contours for HACA and S-Pod layouts, $V_{CRAC} = 4$ m/s .....	98
Figure 3.7. Recirculation patterns in S-Pod layout .....	99
Figure 3.8. Temperature differential between rack inlet and CRAC outlet for various cases .....	100
Figure 3.9. Barrier placement for S-Pod layout .....	103
Figure 3.10. Location of Liquid Cooled racks in a primarily air-cooled datacenter .....	104
Figure 3.11. Temperature Contours and rack inlets for heterogeneous case .....	106
Figure 3.12. Temperature differential between rack inlet and CRAC outlet for S-Pod and Heterogeneous S-Pod cases .....	107

Figure 4.1. Power Flow Diagram of a data center; Type and location of instrumentation required to measure metrics .....	114
Figure 4.2. Control Scheme for a data center .....	136

## LIST OF SYMBOLS AND ABBREVIATIONS

$A_{ij}$	Server in A row, i rack and j height
A	The basic AILM matrix
ACR	Air-Cooled Racks
AILM	Ambient Intelligence based Load Management
CFD	Computational Fluid Dynamics
CFM	Cubic Feet per Minute (unit for volume flow rate)
CRAC	Computer Room Air-Conditioning unit
DCHDC	Data center Heat Dissipation Capacity
FPM	Foot per Minute (velocity unit)
HACA	Hot Aisle-Cold Aisle
i, j	Servers
IT	Information Technology
LCR	Liquid-Cooled Racks
$I_{on}$	Server heat load in just “on” state
$I_{max}$	Server heat load in 100% CPU utilization state
n	Number of Servers
$q'''$	Volumetric Heat generation rate ( $W/m^3$ )
Ra	Rayleigh Number
Re	Reynolds Number
SLA	Service Level Agreements
S-Pods	Scalable Pods Layout
UPS	Uninterrupted Power Supply
$V_{CRAC}$	CRAC inlet velocity to plenum (m/s)

## SUMMARY

A simplified computational fluid dynamics/heat transfer (CFD/HT) model for a unit cell of a data center with a hot aisle-cold aisle (HACA) layout is simulated. Inefficiencies dealing with the mixing of hot air present in the room, with the cold inlet air leading to a loss of cooling potential are identified. For existing facilities, an algorithm called the *Ambient Intelligence* based *Load Management* (AILM) is developed which enhances the net data center heat dissipation capacity for given energy consumption at the facilities end. It gives a scheme to determine how much and where the computer loads should be allocated, based on the differential loss in cooling potential per unit increase in server workload. While the gains predicted are validated numerically initially, experimental validation is conducted using server simulators. For new facilities, a novel layout of the data center is designed, which uses scalable pods (S-Pod) based cabinet arrangement and air delivery. For the same floor space, the S-Pod and HACA facilities are simulated for different velocities, and the results are compared. An approach to incorporate heterogeneity in data centers, both for lower heat dissipation and liquid cooled racks has been established. Various performance metrics for data centers have been analyzed and sorted on the basis of their applicability. Finally, a roadmap for the transformation of the existing facilities to a state of higher cognizance of Facilities/IT performance is laid out.

# 1. INTRODUCTION

Data centers are centralized facilities for processing, storing and managing digital data/information. They house equipments, commonly referred to as data processing equipments, which include servers, telecommunication devices and storage media. While the above mentioned hardware corresponds to the Information Technology (IT) aspect of it, there are another set of equipments, commonly referred to as Facilities, which support and ensure the proper functioning of the IT infrastructure. Fig. 1.1 shows a typical air cooled datacenter. The servers and storage media are stacked in racks which are arranged in rows. The layout shown in Fig. 1.1 is a typical Hot Aisle-Cold Aisle (HACA) arrangement, which has been recommended as one of the best practices [1]. The expected airflow pattern is portrayed in Fig 1.2. The cold aisle provides air to the servers, which is then exhausted to the hot aisle. This air is picked up by the Computer Room Air Conditioning (CRAC) units located generally near the walls of the room. They cool this air and then supply it to an underground or over floor plenum. This helps in transporting air from the CRAC inlet to the perforated tiles in the cold aisle, where the cold air is released into the room. While, this is one of the general arrangements, other variations of it are mentioned in Chapter 3. The networking devices are sometimes stored in a separate rack, or they are affixed to walls from where lines are drawn to the server racks.

Data centers can consume 25 to 50 times more electric power than a standard office space of the same footprint [1]. Servers and data centers consumed 61 billion kWh (1.5% of total US Energy) in 2006 [2] which is equivalent to energy consumption by 5.8 million average US households every year. With the advent of internet and higher

computational and database requirements over the last decade, the number and size of these facilities have increased tremendously [3, 4]. While leading chipset manufacturers have researched on better and faster chips, another aspect of improvement is making the chip size smaller or equivalently achieving higher computation for the same size. This has translated to a higher density of computing power in a data center, proportionately increasing the power supply needs. While high power needs are generated at this end, another aspect of the data center is to dissipate this power effectively and keep the electronics below a certain threshold temperature to ensure reliable and safe working conditions.



Figure 1.1. A typical data center [1]

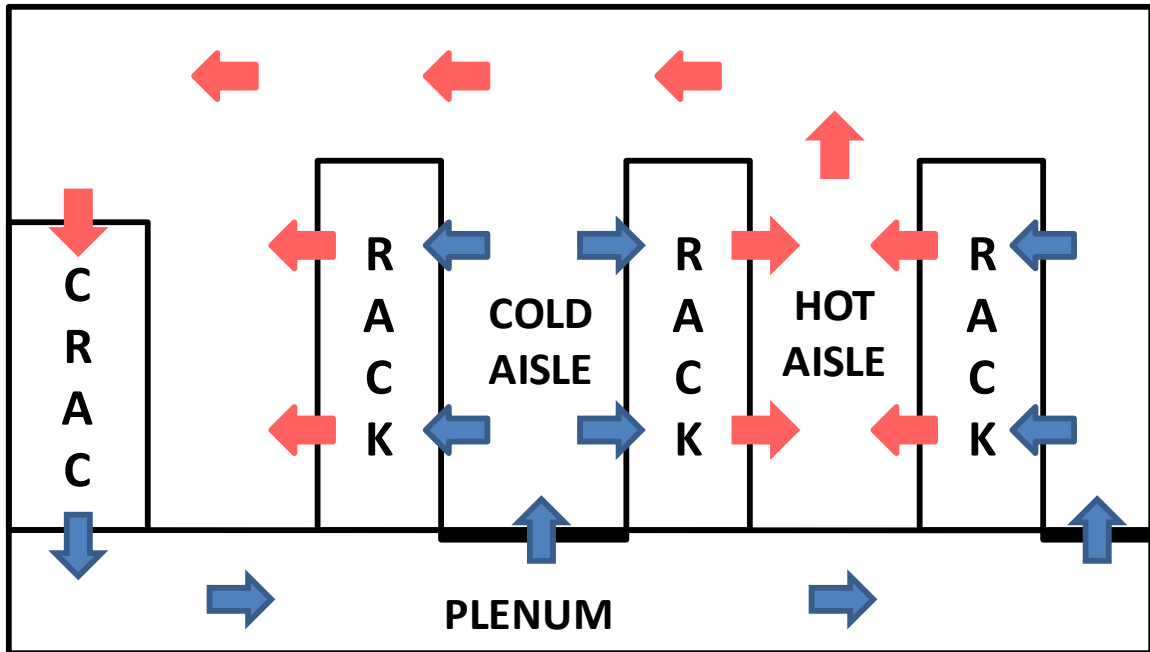


Figure 1.2 A Hot Aisle-Cold Aisle (HACA) arrangement of racks in a data center

The growing concern for achieving optimized cooling leading to lower net cost is the chief motivation behind this study. It has been identified that the current systems are so ineffective that for every 1 kW that enters a datacenter facility, an additional 0.5 kW is required for cooling equipments [5]. This chapter deals with understanding the basics of these issues. Initially the historic trends of datacenter power consumption are laid out, followed by the importance of cooling in it. Further, the power flow in a data center is presented and the scope of this study is defined. At this stage, the literature published by government, regulatory bodies, industry, academic institutions and consulting firms has been reviewed. Finally, the objective of the study is presented.

## 1.1. Historical Data and Trends in Data Centers

A comprehensive study by the American Society for Heating, Refrigerating and Air-Conditioning Engineers breaks down the IT assets into its constituent parts, and studies the trends for each [6].

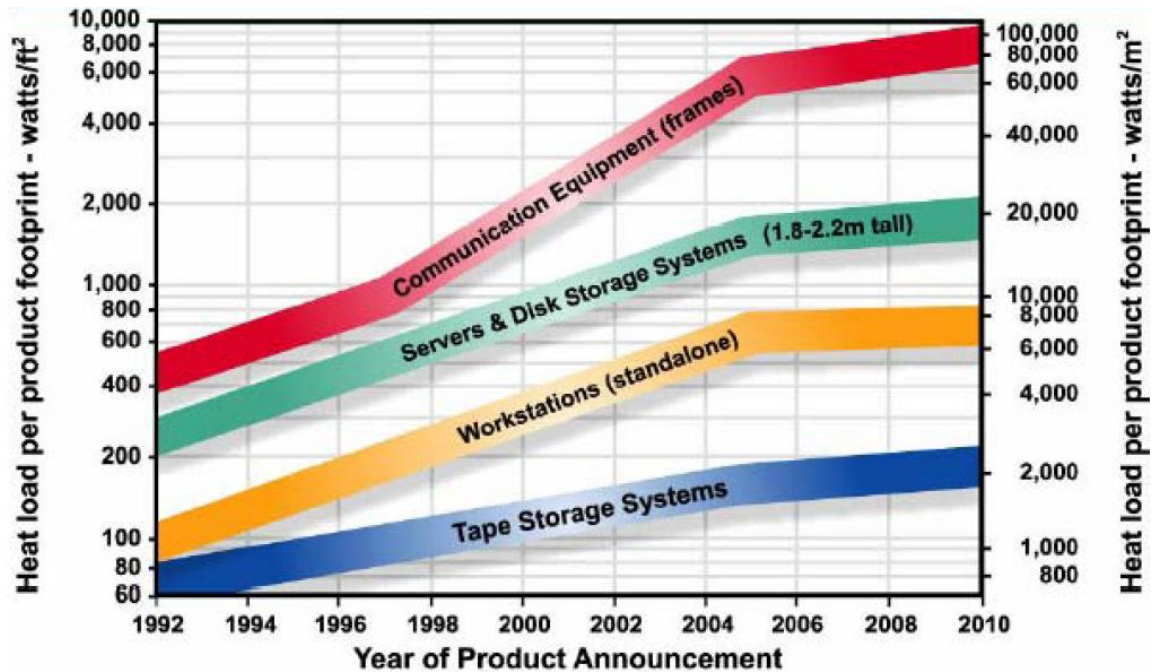


Figure 1.3. Trends for heat load dissipations for IT systems [7]

The data taken from hardware manufacturers typically provides the nameplate power dissipation values, i.e. maximum power the server can dissipate including the derating factor [6]. This is often conservative since most equipment dissipate only up to 40-50% of this value, at least in the first year where a potential of upgrades is kept in equipments. However, Fig 1.3 is drawn based on measured data for a fully configured system in the first year. ASHRAE has converted the comparison metric to  $W/ft^2$  equipment area chart, and have extrapolated it to the future. It is noticed that the power

consumption has more than doubled since 2000, and it is estimated that the annual energy usage could reach over 100 billion kWh by 2011 (2.5% of U.S. energy) [8].

The increase in the electric power requirement of a datacenter is attributed to the following reasons:

1. There are increasing number of chips produced with decreasing dimensions and thus increasing the load densities.
2. The Internet facilities demand increasing amounts of power for supporting their operations
3. There was an over prediction of the increase in energy density for data centers.

The variations in prediction of energy increase can be noticed in Fig. 1.4 which gives a plethora of scenarios predicting the increase in energy from 0.5 to 2 times. Thus, considering designers consider conservative estimates for ensuring functionality, this lead to over-sizing of data center, leading in-turn to more inefficient operations.

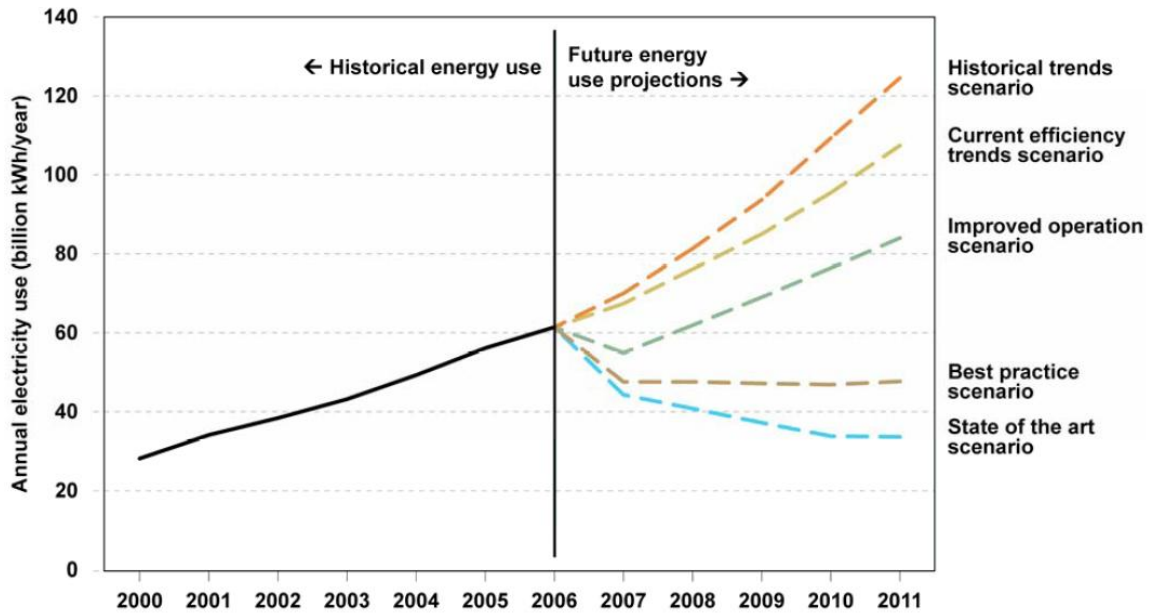


Figure 1.4. EPA Prediction of energy use in data centers [2]

## 1.2. Power Flow in a Data Center – Area of Focus

Figure 1.5 gives the power flow diagram for a datacenter. The main power coming from the utility or in-house power generator feeds into the Switch Gear. This is the point where power is divided amongst the IT equipment and the facilities and support systems. On the IT branch, power goes to the Uninterrupted Power Supply (UPS). UPS is also provided power from any power storage media available, such as a backup battery, in case of power failure from the main source. The UPS provides power to various IT units inside the rack and the networking systems. The IT units consist of servers and the data storage units. In this diagram, the CRAC unit, though being on the facilities end, is powered via the UPS for damage mitigation during power failures, based on the feedback on industry experts. The switch gear also provides for the lighting systems inside the datacenter. On the facilities end, power lines are routed from the switch gear to the cooling tower, chiller and chiller pumps.

Lawrence Berkeley National Laboratory (LBNL) had performed benchmarking studies on 23 different facilities. Figure 1.6 gives the energy consumption by each of the power train units, described above, as percentages of the total energy, based on real life case studies conducted in various facilities [7]. According to the figure, though the servers consume most of the power, the total HVAC loading accounts for the single largest component of power consumed on the facilities end. It was found that the average percentage of power consumed just by IT equipments was just 57%. This leads to the conclusion that a considerable optimization on the facilities end can be done to achieve higher net data center efficiency. The current study focuses on optimizing energy consumed within the datacenter. This is because lower power consumption at the CRAC unit will have a trickling down effect on the power consumption of the chiller and subsequently chiller pumps and cooling tower.

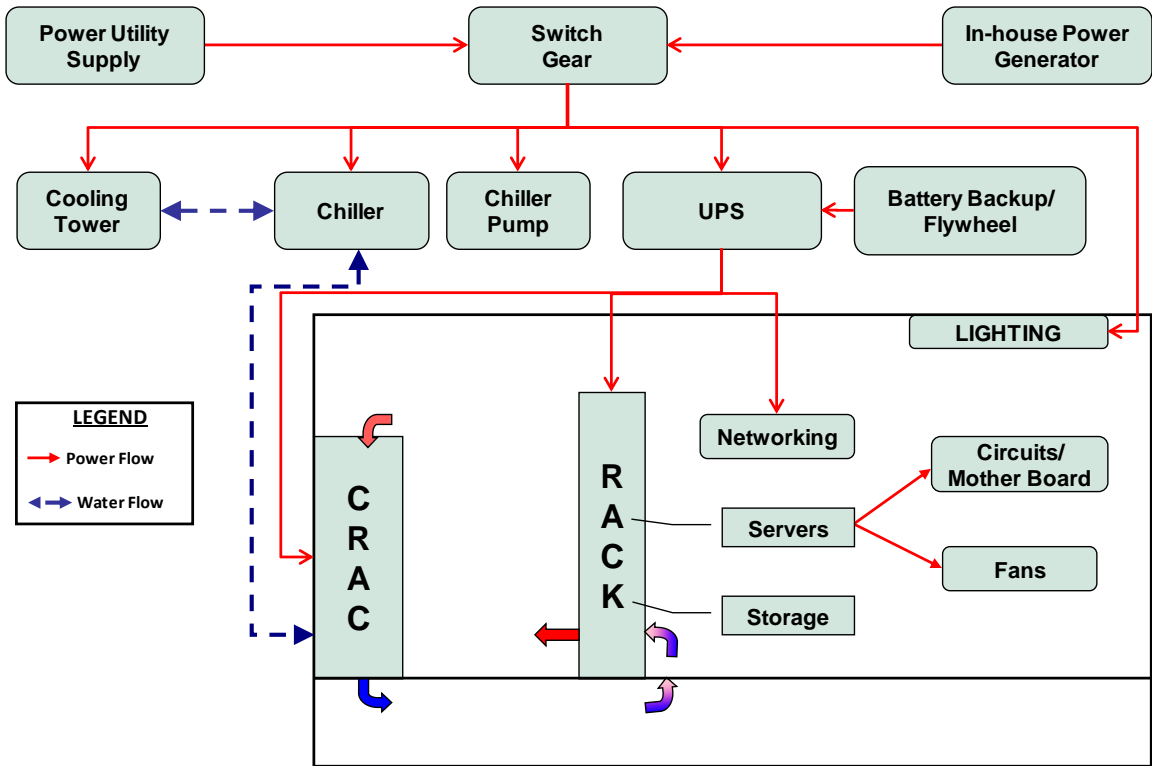


Figure 1.5. Power flow diagram of a typical data center

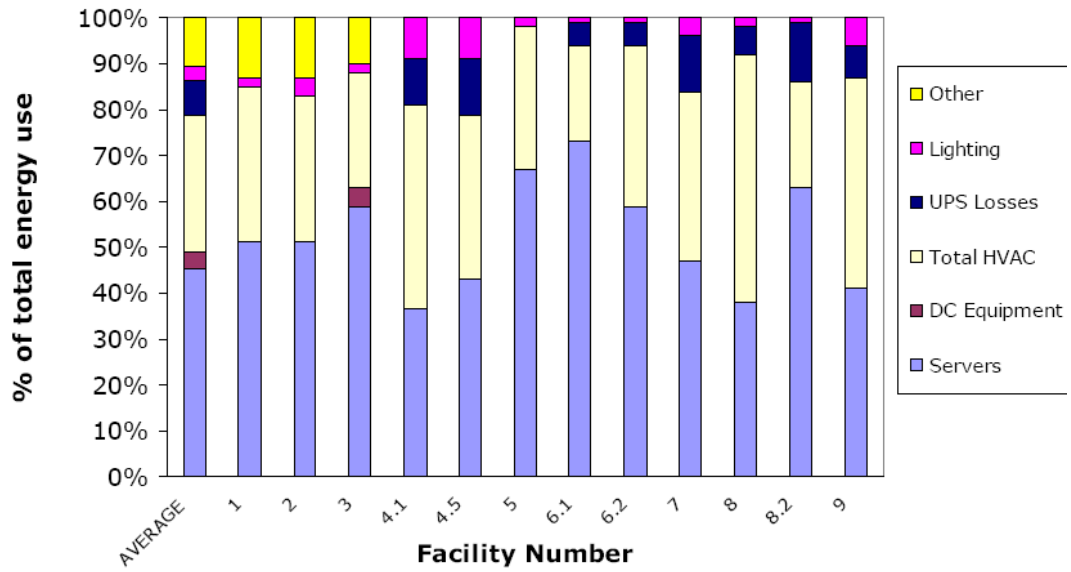


Figure 1.6. Variation in data center energy usage [7]

### **1.3. Importance of Air Management in a Data Center**

As explained before, the cooling system is at the heart of a functional data center. This study focuses on a primarily air-cooled data center utilizing a HACA layout (Fig. 1.2). According to [2], a well-designed air management system can reduce operating costs, first cost of equipments and thermally related processing interruptions or failures. Further, it may increase datacenter's power density ( $\text{W}/\text{ft}^2$ ) capacity. High return side temperature of cooling units will increase their efficiency. This is because these units are typically designed to serve office spaces and thus can control humidity loads (latent). Since the number of occupants is low in a data center, increasing the return air temperatures can convert some of the latent heat load capacity into sensible cooling capacity, thus extracting more out of it. The fan power consumption of the HVAC units for a  $100 \text{ W}/\text{ft}^2$  design data center can vary from  $11 \text{ W}/\text{ft}^2$  to over  $22 \text{ W}/\text{ft}^2$  (ratio of total fan power to the total footprint of the facility). Also, reducing the supply air flow rate leads to a large, non-linear reduction of this fan energy use. For example, a 20% reduction in airflow volume can lead to reduction of power consumption by 45-50%. Fan speeds can be reduced if there is no mixing of the cool air with the hot air [2]. Thus a better airflow management can help achieve higher return air temperature to the CRAC unit, lower fan speeds of CRAC/CRAH units and lower overall maximum temperatures at the rack inlets.

## 1.4. Literature Survey

Since the chief area of interest has been identified, following studies help understand the ongoing efforts in this field. The studies have been tiered with the genre of institutions undertaking them.

### 1.4.1. Government and professional organizations and regulatory bodies

Due to the issues existing in the data centers, as explained in the previous sections, a number of government bodies have shown an interest in identifying the major inefficiencies, laying out a set of Best Practices and even giving data center Managers incentives to go green.

Most notable, is the effort by the American Society for Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). They have published several reports and reviews of actual datacenter based case studies to explain the importance of maintaining a safe and healthy datacenter. In [6], different aspects necessary to be considered to design any Datacom facility are defined as:

1. Knowledge of division of data center with various applications based on floor area/white space and power consumption. This helps one identify the critical equipment and the bottle-necks for new designs.
2. The estimated growth in performance for the same footprint area of the facility. It was found that, for the same space, it is generally in the range of 20-25%, but if a new facility is planned, it can jump to 100%. Thus planning for a set of years, enough leeway must be maintained to retain the growth rate required.

3. Processing capability compared to storage capability: This is workload and application dependent. Based on historical data, on average, storage runs at 50% capacity and servers run at 70% capacity.

4. Estimated growth rate of applications: While this is a tough parameter to determine given the fast changes this industry is experiencing, it has been found that there is a growth rate of at least 15-20% in applications.

5. Asset Turnover/Tradeoff: This helps data center Managers to determine the life of equipments based on the value it is bringing to a company. While poor equipment will have lower processing capacity with lower power dissipation, newer equipment can have higher processing capacity but also having higher power dissipation density. The newer equipment will also have an added initial cost. Thus the idea is to get an effective usage plan for the existing footprint.

The manufacturers are recommended by ASHRAE to provide 2 different working conditions for operation, high reliability and performance. The intake temperature should be below 32°C for operation, and below 25°C for high reliability. The relative humidity should be between 35% – 80%.

ASHRAE has also listed the various datacenter layouts and configurations currently being used in the industry along with the advantages/disadvantages of the most prominent HACA layout. This will be explained in details in Chapter 3.

Lawrence Berkeley National Laboratories (LBNL) has laid out a research roadmap for data centers in [9]. The objective was to identify various short term and long term strategies to achieve higher efficiencies. They benchmarked processes in 23 data centers on common parameters, thus identifying various ideas that work best at different

stages in the system at the least effort. For increasing efficiency from a long term perspective, both monitoring and updating technology based on cost benefit analysis is required. Monitoring and control tools are needed for evaluating data center's performance, controlling equipments and thus improving energy performance and shutting down equipments in events of emergency. For monitoring, various metrics were defined, which will be discussed in details in Chapter 4. Also, advanced analysis/design tools are required for data centers. The current usage of Computational Fluid Dynamics-Heat Transfer (CFD/HT) tools is low because of complexity, technical limitations, perceived inaccuracy and high cost. A desire for simplified yet accurate approach for the industry is expressed. The non-existence of such tools leads to very conservative estimates when the data center is designed. Often equipments, especially support systems such as electrical facilities are over-sized. The actual power demand based on metered values is around 30-55 W/ft<sup>2</sup> [10], while the cooling infrastructure is designed for 100-300 W/ft<sup>2</sup> [11]. This is an issue since over-sized equipments running at part load waste more power. Moreover, this leads to increased initial costs and delay in getting power, owing to higher requirements. Also, excess investment in transmission and distribution infrastructure, can lead to soliciting power from other states and thus overall delay. Further, lack of planning and management leads to reliability issues and poor environmental designs [5, 12].

Since the life cycles of electronic components are short because of increasing number of advanced chipsets in market, there is little study on initial vs. life cycle cost and thus empirical designs are common based on what has worked in the past [1].

Environment Protection Agency (EPA) has also reported to the congress on server and data center efficiency [2]. Improved airflow management can lead to up to 30% improvement in infrastructure energy efficiency. If “Best Practices” for the data centers are followed, this number can increase up to 70%. This could be achieved by using higher efficiency and more aptly sized facilities equipments, using free cooling when appropriate, and by employing improved transformers and UPS for power distribution. An additional 10% can be obtained by using liquid cooling techniques and optimizing server level power consumption. The identified barriers to the promotion of energy-efficient techniques were:

1. Lack of metrics to quantify data center efficiency: Metrics are currently nonexistent because it is tough to quantify the output of a server as it is server and application dependent.
2. The absence of personnel with sufficient knowledge of both IT and Facilities equipments. Thus, being conservative on both ends and not having a common optimizing objective can lead to increased energy wastage.
3. As the data center is primarily an IT asset, facilities are just support systems. Thus, in order to achieve the IT goals and reliability standards, the facilities end is often neglected and a “good enough” system is put in place.
4. Rapidly changing technology leads to reluctance of incorporating energy efficient measures and the initial investment is not recovered in the short life cycle.

Considering the above issues, following are the measures leading to higher adoption of energy efficient technologies:

1. Labeling the products with higher efficiencies can help consumer better decide the tradeoff between the initial/operating costs and processing power and can help increase competitiveness amongst the vendors in designing better systems. For example, Energy Star program by EPA and DOE for labeling and benchmarking energy consumption.
2. Different stages of energy consumption in a data center can be benchmarked. This will help develop a set of best practices making a designer/consumer more literate about the issues and possible solutions.
3. With the energy costs sky-rocketing, there are financial incentives for the data centers owners to look into higher efficiency products and incorporating strategies leading to lower operational costs.

At the cabinet level, the issues existing are related to heat removal from the servers. For increasing efficiency at “room level”, following recommendations are proposed:

1. Improving airflow management by using standardized configurations like Hot Aisle-Cold Aisle and incorporating retrofits for penetration sealing.
2. Adjusting environmental conditions (temperature and humidity set points) in CRAC units to allow wider range while still meeting manufacturer specifications
3. Optimizing data center airflow configuration using better visualization tools and CFD/HT modeling.

The Uptime Institute, a consortium of companies devoted towards increased energy-efficiency in data centers, claims that data center power density has increased more than 300% between 1992 and 2002 [13]. The existing cooling techniques work well

with around 40 W/ft<sup>2</sup>. A breakup of the space utilization of each kind of equipment is provided, and equipment rearranging techniques have been discussed which could help achieve higher footprint/gross area of server equipment without adversely impacting cooling. There have been a number of studies discussing various metrics quantifying energy-efficiency and “greenness” of a data center, which have been discussed in details in Chapter 4. The Institute also found that 10% of racks have air intake conditions above the critical environmental conditions specified by the manufacturer and recommended by ASHRAE [14]. Also surprisingly, highest % of hot spots were found in rooms with light loads even though they had 3.2 to 14.7 times more cooling than required. This has been attributed to poor air-management.

Based on case studies of existing facilities owned by member companies, it has been concluded that 60% of the air coming out of tiles is directly short circuiting and mixing with the already present hot aisle air and then gets pulled in to the return side of the CRAC unit. Thus, a more important issue than providing more cooling is optimizing the existing setup. There is a high percentage of bypass flow from unsealed cable cutouts and misplaced perforated tiles. Based on best-practices, the bypass flow was reduced from 43% to 10% in a facility. Some computer room hot spots were removed by shutting down 11 of the 24 cooling units considered in the case study. Predictions from CFD/HT models did not match the experimental studies in one case because of excess bypass flow being not modeled in numerical simulations. It was concluded that CFD/HT modeling should be acceptable for bypass flow below 10% of the total flow coming out of the perforated tiles. The hot spots at room level have been identified to be:

1. **Zone Hot Spots:** These are seen at the rack inlets. They typically occur in a row of equipments and can be monitored using temperature sensors at the rack inlets.

2. Vertical Hot Spots: This can be seen within a rack itself. There might be a temperature variation of more than 5°C over short vertical distances. The lower the distance for which this phenomenon is observed, the more serious the problem.

The Green Grid is another global consortium which focuses on quantifying issues in a data center and then suggesting mitigation measures. As quantification is the focus, their work concentrates on developing various metrics and is talked about in detail in Chapter 4. Also, one of their studies finds a linear dependence of power consumption to % CPU utilization with a  $\pm 5\%$  error. Although, power consumption in just “on” state was found to be around 66% of that in 100% CPU utilization state [15]. They have given a set of best practices, power distribution standards, floor layout options and location of vented floor tiles using CFD/HT models in [16].

#### 1.4.2. Industry

Hewlett Packard (HP) were the proponents of sensing and control and formulated the Dynamic Thermal Management (DTM) [17] concept, which suggests change in processor power consumption based on feedback from temperature sensors located at strategic locations. This can be used to reduce the net power consumption by optimizing the facility [18, 19]. Also in [17], power that can be safely allowed for a server, is considered inversely proportional to the difference between the plenum temperature and the server exhaust temperature. This study claims up to 25% reduction in cooling costs.

A collaboration of HP with Duke University has reported on Dynamic Smart Cooling (DSC) technology [20]. This work concentrates on developing software for data center management which has lowest cost as the chief guiding principle. A utility

computing infrastructure is created which provisions isolation of blocks of servers from a shared cluster (*Splice*©) [21]. Sensors are deployed and the data center is metered. Then machine learning techniques are applied to understand the relation between workload, internal and ambient temperatures (*Consil*©) [22]. These analytical tools help in determining the cost associated with each move of the data center manager (*Weatherman*©) [23]. Different cost aware schemes like Zone based Discretization, minimizing heat recirculation and *Weatherman* based placement are formulated and compared with each other. They report around 165% savings from the worst case scenario [24]. There has been past work addressing these issues by focusing on power-aware solutions that optimize the power consumption of the compute equipment in the server clusters within a data center [25-27].

The International Business Machines (IBM) Corp. has reported on the chilled water production being optimized in one of the IBM plants using a thermal energy storage (TES) [28]. A load leveling approach has been applied to minimize the required equipment and storage capacities used in the operation. Greenhouse gas (GHG) emissions were observed to be reduced by 35%. Also, energy consumption was found to decrease by 5300 MWh per year.

They have identified the “4 R’s” for a Green data center [29]. These are regaining power and cooling capacity, recapturing resiliency, reducing energy costs and recycling end-of-life equipments. They have also developed a Rear Door Heat Exchanger which removes heat from the server exhaust air before it enters the hot aisle. This way, even if recirculation is not controlled, the effect of it on the incoming cold air could be reduced [30]. On the IT end, they have developed the Blade Centers which reduced energy

consumption by 20% for the same compute load [31]. From the monitoring perspective, IBM has developed a “Cool Blue” portfolio for addressing power and cooling issues. It provides a software, PowerXecutive to measure power consumption by x servers. The need for measuring, analyzing and relocating power and cooling loads to achieve maximum efficiency is stressed [32].

Intel Corporation has studied data centers from a Total Cost of Ownership (TCO) perspective [33]. They concluded that higher density data centers are a better choice for lowering the total cost of ownership. Densities of workload equivalent to 1,000 W/ft<sup>2</sup> are achievable with high efficiencies. They also suggest that the design and construction phase of the data center should make greater use of CFD/HT tools to better understand the airflow management. Another interesting conclusion from the paper is that uniform server loading makes the task of airflow management easier and better. This has been observed to be **not true** from the present study. They have developed a platform resident Policy Manager (PM) which monitors thermal and power data and then enforces built-in policies related to workload management [34].

Intel along with Microsoft Corporation has also concentrated on monitoring and controlling the various features of a data center in real time [35]. Some of the common best practices mentioned are having CRAC units in front of the hot aisles, and providing airflow equivalent to the net required by the racks to avoid recirculation and short circuiting.

### 1.4.3. Academic work

The Consortium for Energy Efficient Thermal Management (CEETHERM) has contributed towards CFD/HT modeling of data centers [36]. The full-field CFD/HT analyses have very large degrees of freedom. It can take several days to simulate a typical data center layout. An alternate is to use reduced order models. The proper orthogonal decomposition (POD) to develop such models uses snapshots of fewer numbers of simulations to generate the flow and temperature fields for any given initial conditions within a range. This has successfully decreased the degrees of freedom by several orders of magnitude, making such tools applicable for design. Also,  $Ra/Re^2$  estimations are shown to be low such that effects of natural convection can be neglected for the scales considered. In future, as the velocities increase to dissipate more heat, the ratio will reduce even further. The requirements of a future “Open” data center have been laid out in [37]. Here, design for lowest cost has not been considered as the chief goal. It also considers reliability, robustness, concurrency, flexibility, customization and leeway for future growth. The reduced order model based on POD [36] has been used as an optimum heat load distribution determination tool [38]. A Compromise Decision Support Problem (CDSP) architecture is employed and the net improvements are estimated to be \$47,000-\$189,000 annually for real world data centers.

Research on thermal aware job allocation has been performed at Arizona State University’s IMPACT Laboratory [39]. Initially, server power consumption is shown to be linear to CPU utilization. The optimization framework uses genetic algorithms and minimizes the increase of peak inlet temperature. Three different workload assignment

schemes are tried and the concept of minimum computing energy is shown to be the best option. Cooling savings of 30% are predicted numerically.

There has been a joint work performed by McKinsey and Co. and The Uptime Institute on increasing data center Efficiency [40]. They predict that the GHG emissions by data centers will quadruple by 2020. They have identified the issues leading to inefficiency as poor capacity and demand planning, and failings in asset management - 6% average server utilization and 56% facility utilization. They have suggested means to double IT efficiency by 2012 by checks and measures by appointed personnel. This includes integrating asset management capabilities to remove underutilization issues. Also, another all-encompassing metric called Corporate Average Data Efficiency (CADE) has been formulated.

### **1.5. Objective of Research**

The above literature survey proves that data center energy issues have been an active area of research for the past 6-8 years. The inefficiencies have been coarsely identified and some mitigation measures along with a standard set of best practices have been provided. The need for sensing and control over the existing units has been expressed and work is underway in displaying useful metrics to better understand exact inefficiency causes. While the importance of controlling has been mentioned, there have been very few attempts in this direction. All the previous studies focus on CFD based model development which, as explained, can take months to solve a scenario and thus not useful as a design tool. Also, CFD simulations for existing facilities can be misleading because of the leakage issues. Moreover, the layout of a data center has been

set to HACA, which as shown in the present study, is not the most optimal setup. Thus, based on these issues, following are the objectives of the present study:

1.5.1. Objective 1

For existing data centers, develop a control and load allocation strategy which is self controlling with minimal manual intervention.

1.5.2. Objective 2

For new data centers, study different layouts which could achieve better airflow management than the HACA arrangement.

1.5.3. Objective 3

For existing and new data centers, enlist the various metrics in the literature and compare them with each other, also giving guidelines to best measure parameters for defining the metrics.

The roadmap of the thesis is presented in Fig. 1.7.

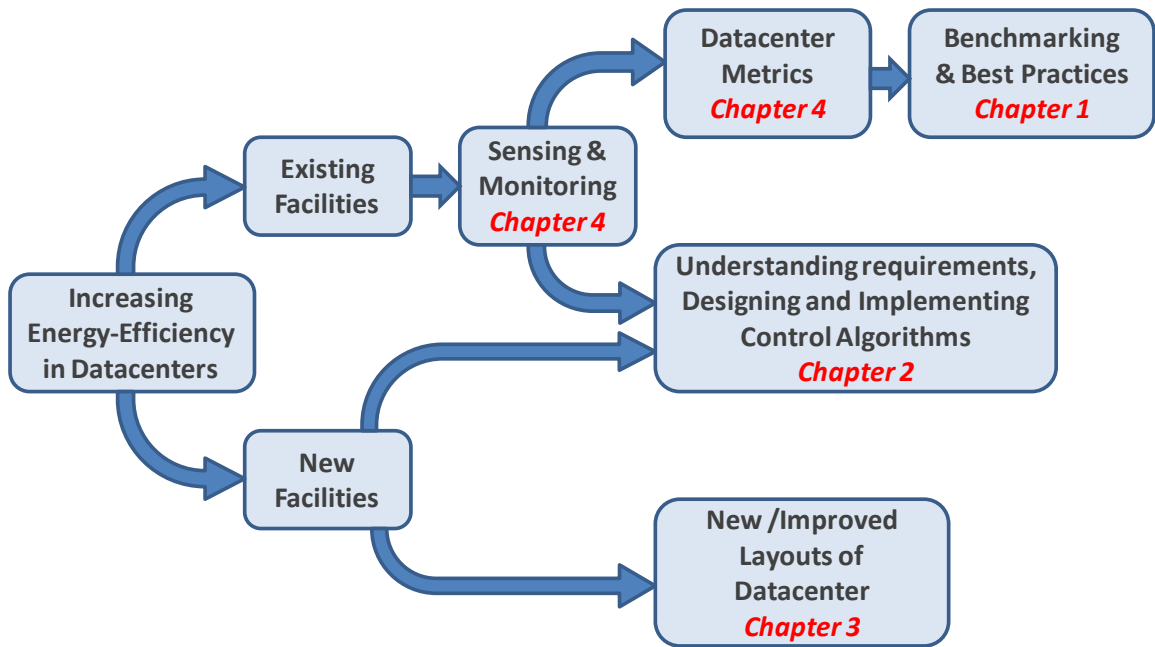


Figure 1.7. Roadmap for the Thesis

## **2. AMBIENT INTELLIGENCE BASED LOAD MANAGEMENT**

This chapter deals with the development of a thermal aware job scheduling algorithm called Ambient Intelligence Based Load (AILM). Initially the modeling, meshing, conditions simulated, and results are presented for a model facility which helps the determination of the exact issues as listed in the literature and their causes. This helps in the development of the algorithm. The numerical results are presented, comparing it to a standard case. Several case studies are undertaken to understand its performance in heterogeneous data centers, the tradeoff of workload with heat load, and its sensitivity to the threshold rack inlet temperature. Finally, an experimental validation is presented which gives the setup, instruments and parameters considered, along with comparisons with the numerical results and causes for any discrepancies.

### **2.1. Modeling**

#### **2.1.1. Geometry**

The computational model is shown in Fig. 2.1. This model is in accordance with [38]. It is one-fourth of the total facility with symmetry walls being the two “transparent planes” through which the inner view of the data center is possible. Thus, this model is also referred to as the unit cell in this study. The under-floor plenum is 0.86 m deep and facilitates the passage of the cool coolant coming out of the Computer Room Air Conditioning (CRAC) unit to the room. The racks are arranged in 2 rows, A and B. The arrangement is according to the ASHRAE recommended HACA layout [6]. The floor between the two racks is perforated tiles which allow the cool coolant to enter the space

between the two rows of racks, referred to as the cold aisle. The other two aisles are referred to as the hot aisle where the hot coolant is exhausted from the back of the servers. This hot coolant is subsequently taken in by the CRAC unit which cools it and resupplies it to the plenum. The commonly used coolant is air with limitations on relative humidity and temperatures. The roof is 3 m above the raised floor. The CRAC unit's volume is not included in the model, which includes only its inlet and outlet surfaces, and side walls. The racks are modeled as boxes with 0.6 m x 1.1 m footprint and 2 m height. They are divided in 6 compartments to accommodate 6 servers.

#### 2.1.2. Boundary Conditions

The 4 walls, including the ceiling and floor, and two other walls are modeled as symmetry boundary condition. For the wall boundary condition, the velocities and gradients perpendicular to the surface are set to zero. The plenum has similar boundary conditions. It also has an inlet surface from the CRAC bottom which is modeled as velocity inlet.

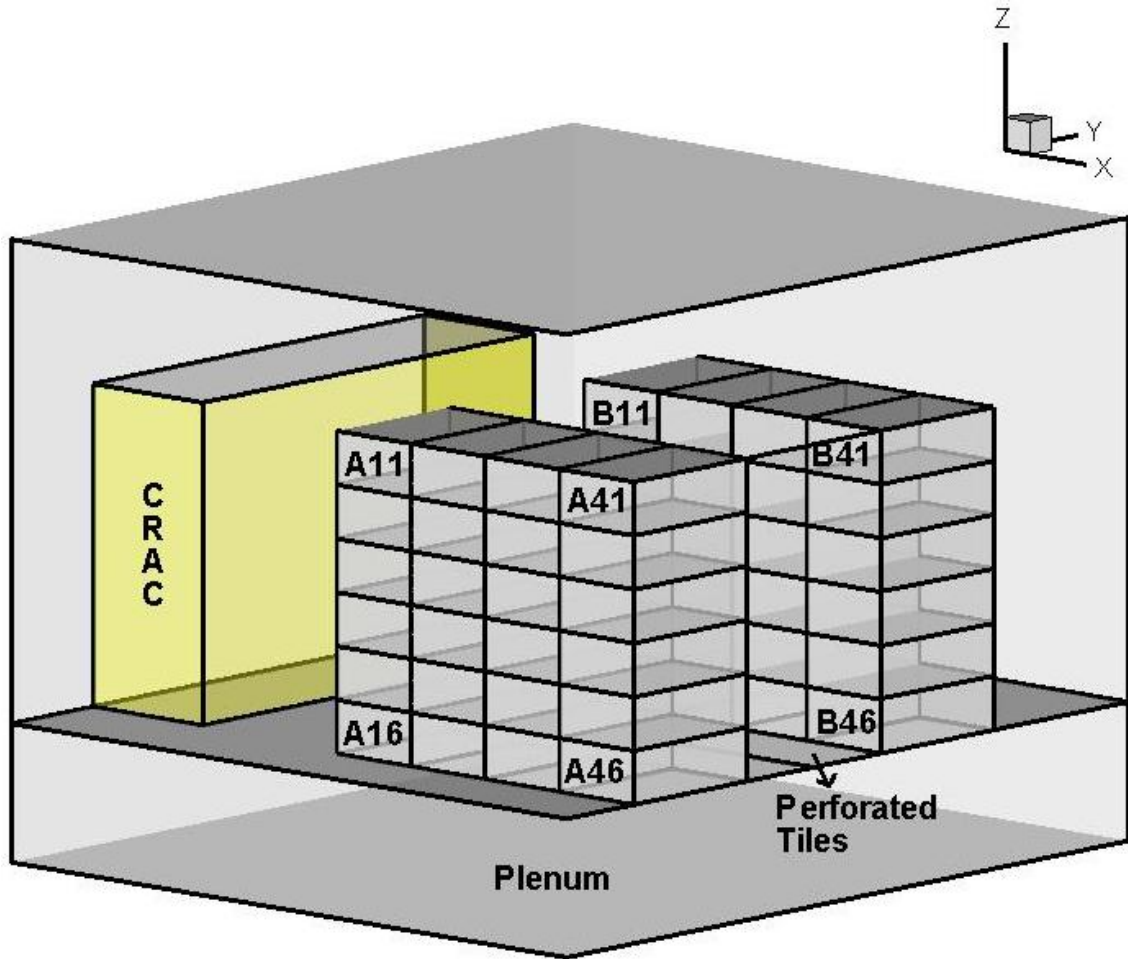


Figure 2.1. Computational model simulated; Nomenclature of servers

The temperature and velocity of the inlet air has been specified. The temperature for all simulations is fixed to 288.15K, while velocity is a parameter which is varied. The CRAC's top surface is a velocity outlet with the same velocity specified as for CRAC bottom. To calculate the top surface's temperature, an energy balance is applied between the total heat dissipated from racks and the cooling provided by the air coming from the CRAC. The racks are modeled as black boxes with constant volumetric heat generation to account for the heat dissipated in the servers. Modeling it in this simplified way leads to a

lower mesh size requirement for the final model and helps concentrate on the inefficiencies at the room level scales. The nomenclature of the servers is explained in Fig. 2.1. The face of the rack facing the cold aisle is assigned a specified porosity to simulate server inlets. While actual server inlets constitute only 20% of server front face, their location varies with servers. Thus, for this model, the entire face is an inlet to simplify the conditions. These porous inlets are specified with a thickness of 0.035 m, over which the pressure drop is a combination of Darcy's Law and an additional inertial loss term, which varies with velocity according to the equation:

$$\Delta p = - \left( \frac{\mu}{\alpha} v + C_2 \frac{1}{2} \rho v^2 \right) \Delta m \quad (2.1)$$

The values of various parameters in the above equation are fixed according to [38]. The  $C_2$  value is fixed at 13809.93 /m. The perforated tiles are also governed by the same boundary condition, 20% opening and a different  $C_2$  value of 36787.7 /m. The face of the rack facing the hot aisle is modeled as fan with polynomial pressure drop-velocity curves described as:

$$\Delta p = \sum_{n=1}^N f_n v^{n-1} \quad (2.2)$$

Where,  $N=4$   $f_1= 744.6$   $f_2= -439.41$   $f_3= 99.784$   $f_4= -57.961$

These values are in accordance with the industry standard fans for servers [38].

### 2.1.3. Meshing

Gambit 2.4.6 has been used for constructing and meshing the model with a mesh size of 0.12 m. Due to the block structure of the geometry, quad elements are chosen for surface meshes, and hexahedral elements are chosen for volume meshing. The total number of nodes is 630,000. A grid convergence study has been performed with  $V_{CRAC}$  4

m/s and constant heat generation per rack 8888.89 W/m<sup>3</sup>. The details are provided in the following table:

Table 2.1 : Grid Independence Study

No. of Modes	Temperature		Velocity	
	Max. T (K)	Error in T (%)	Max. V (m/s)	Error in V (%)
1,260,000	321	-	4.26	-
945,000	320.3	2.1	4.15	2.6
630,000	320.1	2.5	4.12	3.1
315,000	319.8	3.6	4.05	5.1

#### 2.1.4. Simulation Facts/Conditions

FLUENT 6.3.26 has been used for the computations. A steady state formulation has been applied. Standard k-  $\epsilon$  turbulence model has been used to model viscous flow. A first order upwind scheme has been chosen for initial set of iterations on momentum equations and is then switched to a second order upwind to achieve a faster convergence. For all the other equations, a second order upwind scheme has been used. The governing equations for mass, momentum and energy are presented in Appendix A.

Accounting for natural convection, using the Boussinesq approximation leads to a maximum error of less than 1.2% in the temperature difference between the fan outlets and CRAC supply. Thus the effect of this has been neglected. Also, since no economizers are used for this model, the same air is recycled back to the data center. Thus humidity variations inside the room are neglected for this study, as done in [37].

## 2.2. Simulation Results

Figure 2.2 shows a virtual experiment where a set of particles are being released at the perforated tiles. The view is from behind the B Rack with CRAC unit on the right side. The released particles follow the streamlines starting from this point. Thus, this will help us understand the dynamics of flow in the cold aisle. For the case of  $V_{CRAC} = 10$  m/s, the momentum of the flow coming out of the tiles is very high. The particles rising in the cold aisle are forced away from the  $A_1$  and  $B_1$  racks. This can be explained by the pushing effect of the recirculated flow already existing in the room. This flow originates from the fans of the servers, especially in the  $A_1$  and  $B_1$  racks. Thus, a direct entry of the cold air is partially not allowed for  $A_1$  and  $B_1$  racks. Also, because of the interaction of the cold air with the hot air, the temperature in the cold air stream gradually increases, as noticed in the temperature gradient across the streamlines in Fig. 2.2.

This combined effect of obviation of the intake of cold air in the servers, along with increase in the inlet air temperature for rest of the servers is referred to as recirculation. Also, cold air leaving the cold aisle from the top without being used up by the servers is called short circuiting. Since this air will ultimately mix with the hot air and exit through the CRAC top surface, it will reduce the effective temperature of the return flow. This leads to a reduced load on the CRAC unit since it has to cool through a lower differential. While, this will have a cascading effect on the energy consumption of rest of the facilities units like Chiller, Chiller Pump etc., it'll also come with a tradeoff of increased CRAC velocity. The final effect on the net energy consumed in the facilities end is uncertain since this study focuses on energy usage inside the data center room.

Increase in CRAC velocity leads to a superlinear increase in CRAC power consumption. Thus the decrease in power consumption due to decrease in temperature differential for cooling cannot overcome the added tradeoff of increased power consumption due to higher velocity requirements for the same.

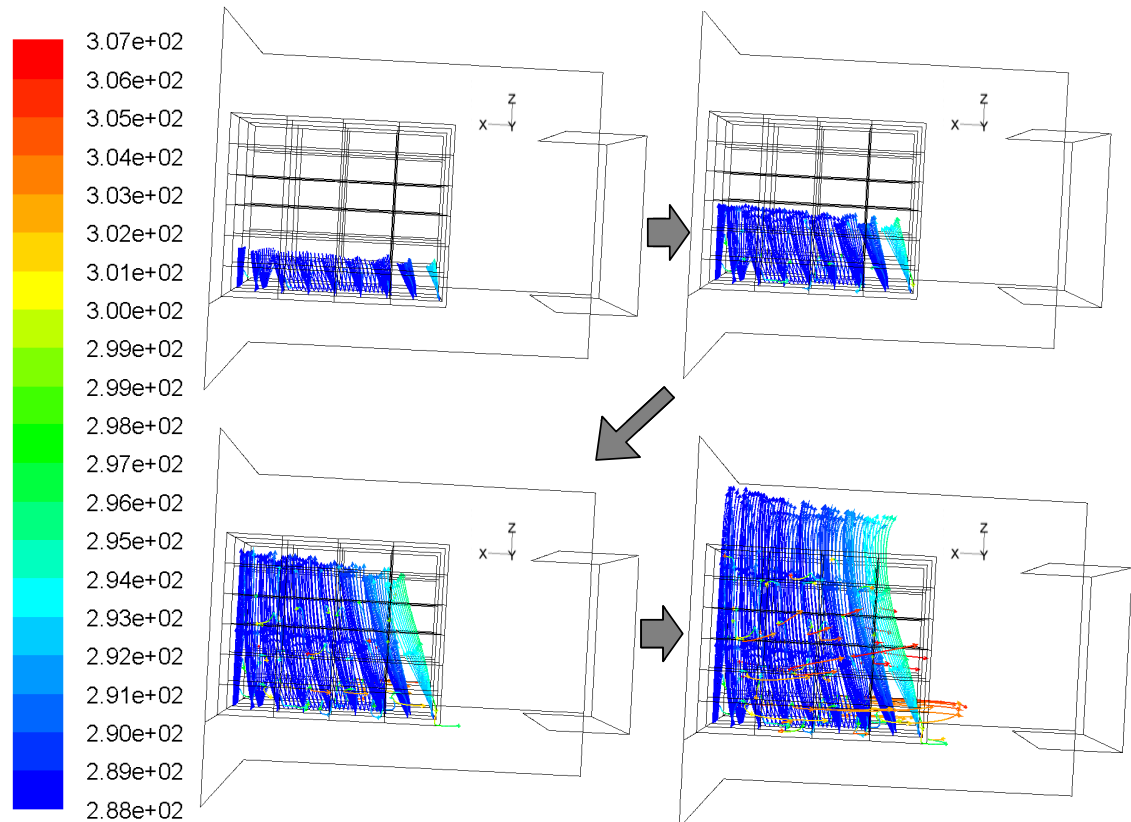


Figure 2.2. Temperature variations (K) in the streamline path originating from perforated tiles ( $V_{CRAC} = 10$  m/s)

Fig. 2.3 shows the same analysis for a lower  $V_{CRAC}$ , 5 m/s. For this case, two important features are noticed as compared to the Fig. 2.2. Firstly, the effect of recirculation has amplified. Now, servers  $A_{11}$ ,  $A_{12}$ ,  $B_{11}$  and  $B_{12}$  are completely devoid of direct cold aisle flow and receive only recirculated air. Here the recirculated air can be

seen in the fourth quadrant of Fig 2.3 as the red arrows coming out of the servers. They are red since they have passed through the constant heat generation servers already and are seen to be moving towards the cold aisle. Secondly, the effect of short circuiting has reduced because of the lower mass flow rate coming out of the tiles. This also supports the explanation given for Fig 2.2.

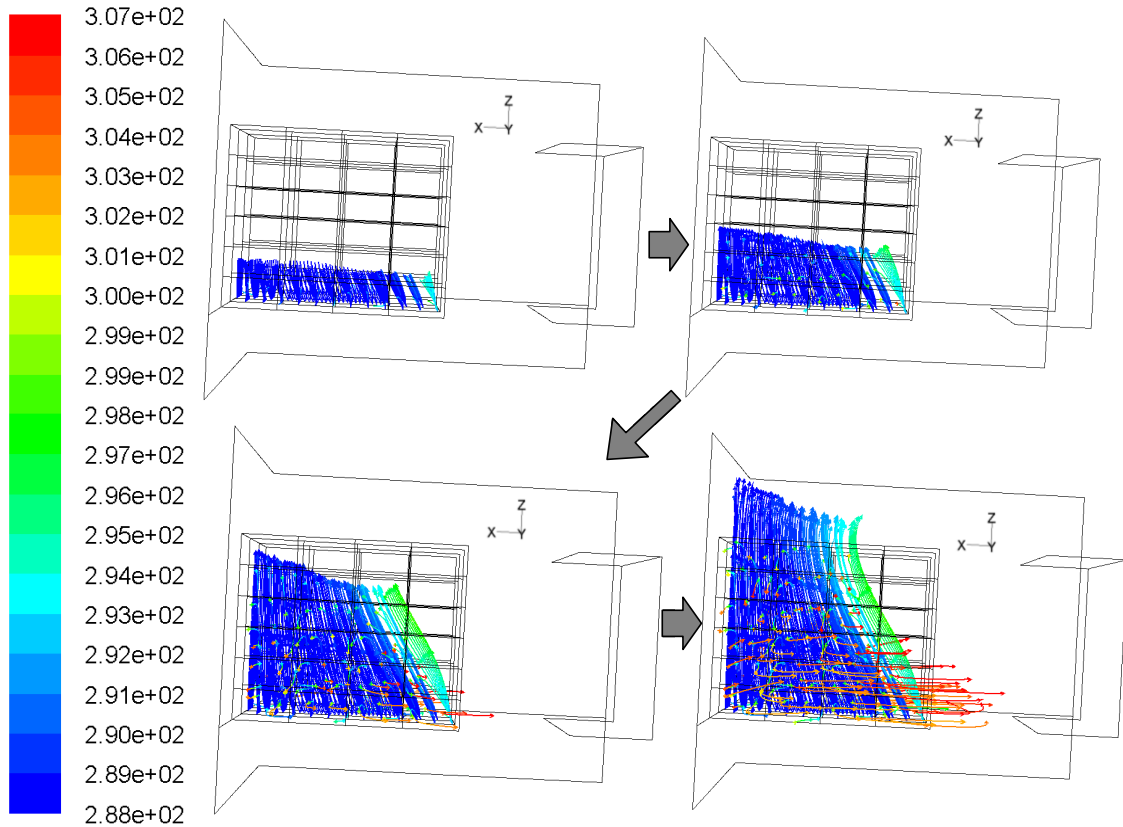


Figure 2.3. Temperature variations (K) in the streamline path originating from perforated tiles ( $V_{CRAC} = 5 \text{ m/s}$ )

With the explanation of recirculation conditions given for the previous figures, one tends to question the predictability of these inefficiencies. Whether bottleneck

servers, which lead to the above explained flow conditions, can in general be identified. The explanation for this is given with the help of the Fig. 2.4.

Fig. 2.4 considers a similar virtual experiment as explained before with particles emanating from the fans of A1 and B1 racks. Fig 2.4(i) and 2.4(ii) are for  $V_{\text{CRAC}} = 5$  m/s and Fig. 2.4(iii) and 2.4(iv) are for  $V_{\text{CRAC}} = 10$  m/s. In Fig. 2.4(iii), it can be seen that the flow coming out of the back of A racks is recirculating to the A rack inlet. Fig. 2.4(iv) shows that the flow on the top of the aisle, though vortical in nature, is not mixing with the cold aisle fluid. This can be also be explained by the fact that the short circuited air shown in Fig. 2.3, is preventing the hot aisle air to enter from the top. In Fig 2.4(i), the recirculation effects reverse. Here most of the flow from A rack tends to go straight to the CRAC top surface, while the flow from the back of B rack is heavily recirculated to both A and B rack inlets. In this case too, the recirculation from top side is missing. Thus from these two studies, we conclude the following:

1. Recirculation effects, while present for both the velocity cases, do not always have the same source and thus produce varied effects. Even if we can find a threshold velocity for which the recirculation effects from both the racks balance out, which will be between 5 m/s and 10 m/s for this model, in a bigger facility with interaction between different CRAC flows beneath a plenum, it'll be difficult to generate such a CRAC flow which will balance out recirculation always.
2. The top side recirculation is not present for the chosen set of velocities because of higher momentum of the cold aisle flow and subsequent short circuiting. But, if the velocity is decreased further, we will notice top side recirculation which will affect inlets of  $A_{11}$ ,  $A_{21}$ ,  $A_{31}$ ,  $A_{41}$  and the corresponding B rack servers.

Such studies haven't been performed for other server fans because apart from a few servers on the bottom of racks, these are not the sources of recirculation, and exhausted air is directed towards the CRAC top surface.

Fig. 2.5 shows the temperature contours of rack inlets for  $V_{CRAC} = 5$  m/s and uniform volumetric heat generation ( $q'''$ ) of  $12121.21$  W/m<sup>3</sup> for all racks. The temperature contours clearly show that the inlet temperature for the racks affected by recirculation, is higher. The ASHRAE specified upper limit of the rack inlet temperature is 305.15K [6]. Since the heat generation is kept constant for all racks, racks A<sub>1</sub> and B<sub>1</sub> act as bottlenecks for the upper limit of maximum volumetric heat generation.

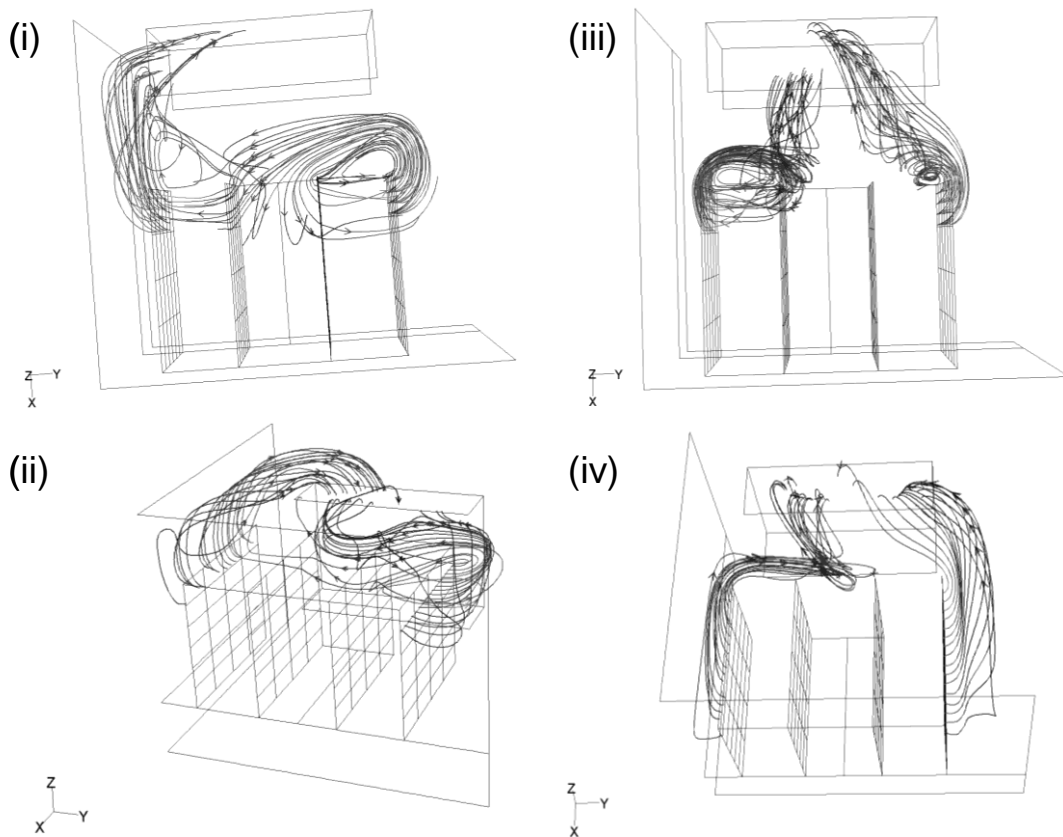


Figure 2.4. Streamlines generated from back of A<sub>1</sub> and B<sub>1</sub> racks (i)  $V_{CRAC} = 5$  m/s, (iii)  $V_{CRAC} = 10$  m/s; Back of A<sub>11</sub>-A<sub>41</sub> and B<sub>11</sub>-B<sub>41</sub> (ii)  $V_{CRAC} = 5$  m/s, (iv)  $V_{CRAC} = 10$  m/s

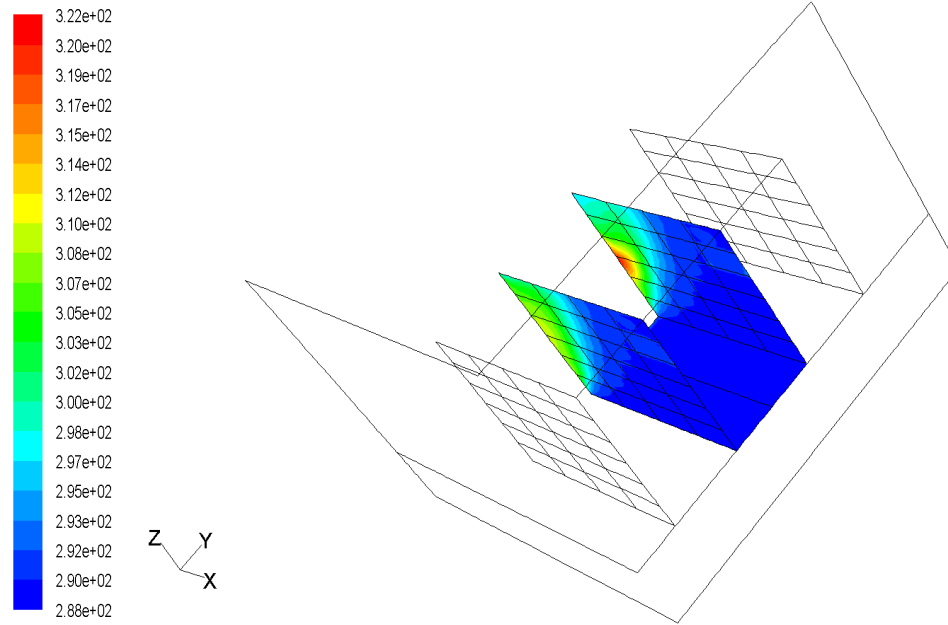


Figure 2.5. Temperature contours at rack inlets (K)

### 2.3. Requirements for a load management algorithm

As can be seen from previous section's results, if a uniform heat load is maintained for all racks, then the  $A_1$  and  $B_1$  racks become the bottleneck in deciding the net heat load capacity of the data center. If on the other hand, these loads can be re-distributed such that the  $A_1$  and  $B_1$  racks can take lower loads and the remaining load is distributed amongst other servers, the data center might have a higher total heat load potential. This might also be in accordance with the current state of data center in which a wide range of servers are present. Since, the server technology is rapidly advancing [4], we could expect frequent arrivals of new servers. This will lead to a more heterogeneous data center. For such a facility, different servers will have different workload capacities and heat dissipation values. Thus apart from reallocation of load during operational phase, one should also determine the optimal arrangement of these servers to maximize

the highest workload potential. While, redistribution, if it could be achieved, should be a plausible solution for the workload already existing in the data center, another crucial question is about the allocation of every unit of new workload at the doorstep of the data center. An ad-hoc approach for determining this might not be efficient enough for larger data centers. Thus a pre-calibrated facility with a defined algorithm can also guide towards the right allocation of this extra workload, such that the net work done in achieving an optimal setup can be minimized. This is required since both the determination of the correct placement of workload and the subsequent reallocation has to be done in real time. This is to make sure that the data center can meet the Service Level Agreements (SLA) within the turnaround time bounds. While the goal now is set to achieve such kind of redistribution, following issues arise:

1. *Determining Bottlenecks*: In the  $V_{CRAC} = 5$  m/s case,  $A_1$  was causing recirculation while for  $V_{CRAC} = 10$  m/s case,  $B_1$  rack was causing this effect. One issue is how to set up a parameter which helps the data center Manager decide, how the bottlenecks vary with velocity. Secondly, in this analysis only streamlines from  $A_1$  and  $B_1$  have been plotted. The exhausts from other servers might also contribute, however little, to the net recirculation, in some cases just directing the  $A_1$  and  $B_1$  rack flows towards the cold aisle, in other, there might be a direct effect. Considering the inherent non-linearity of velocity field based on the input velocity, such parameters are tougher to decide analytically.

2. *Determining the final destination*: Now, as the workload is taken off from the servers near the racks, the next issue is determining where these loads should be relocated. There can be three different scenarios here.

- Relocating the load equally amongst the rest of the servers.

- Listing out the servers on the basis of increasing values of inlet temperature. Allocating maximum load possible to the “coolest” server. Then, allocating the remaining load in the next coolest server until its maximum capacity is reached and doing so until we completely allocate all the extra workload.
- Listing out the servers on the basis of increasing values of exhaust temperature. Allocating maximum load possible to the “coolest” server, since lower exhaust temperature values could indicate unutilized potential of a server. Then, allocating the remaining load in the next coolest server until its maximum capacity is reached, and doing so until we completely allocate all the extra workload.

Thus, on the basis of above, a set of requirements of a potential algorithm for the data center can be listed out. The above requirements are however intrinsic to the system and are more related to a specific layout. Based on the literature survey presented and the analysis done on our numerical test bed, a set of extrinsic requirements which take a more panoramic view of the problem can also be laid out. Following are the requirements which the author believes covers all the dimensions of the issue at hand:

### 2.3.1. Extrinsic Requirements

1. *Flexibility*: The algorithm should not be specific to any particular server or its characteristics, since servers with more capacity and computational speed will be added frequently to most data centers. Some future racks might even be Liquid Cooled Racks (LCR). Thus the algorithm should be flexible enough to incorporate all the heterogeneities.
2. *Scalability*: The range of scale of data centers is very large. There exist data centers in typical office facilities with ground area  $100 \text{ m}^2$ , to ones built for very large facilities

stretching up to 10 acres. Thus any approach has to be valid for the entire range of facilities.

3. *Cost:* The initial cost of setting up the algorithm and subsequent calibration should be low. This cost also refers to the operational computational needs of the algorithm. The benefits achieved by integrating this approach should be much higher than the loss of the computational power of the data center. For example, consider a 50 server data center with 5 servers being used up by a complex CFD based approach. This directly reduces the total capacity of the data center by 10%. Now the benefits achieved by this new algorithm should be at least greater than 10% for the remaining servers to justify its usage.
4. *Time:* Time taken for initial set up should be low. This disqualifies all only CFD based algorithms, as simulating the entire facility even once could take days. Designing using only CFD based approaches will require multitude of these simulations, increasing the set up time enormously. Also, with new changes incorporated in data centers periodically, the calibration time during operational phase is very high. While the setting up time is considered, another aspect is time for results. Both monitoring and taking an action based on it should be achievable in real time.
5. *Global Optimum:* Data Center Managers are IT oriented as the foremost purpose of a data center is to perform an operation within a set turnaround time. Thus often if one of the bottleneck computers are the fastest, workload will be directed towards them. As has been pointed out in earlier results, this might not be a wise decision thermally. Also, allocation only through thermal guidance might lead to more loads in an

inefficient server. Thus the tradeoffs have to be maintained in the model through internally set metrics. The constraints from both the IT and Facilities ends should be considered and a global optimum should be determined. If an algorithm is primarily based on any one of these set of constraints, it should at least have the capability to be integrated with the other end of the approach.

6. *Operational Ease*: The ease of handling and varying the exogenous parameters for the end user is important. This requirement is more from a practical application perspective. Thus for successful operation, it should be a self controlling mechanism with minimum manual intervention.

#### 2.4. Assumptions and Simplifications

The energy equation is presented in Eq. 2.3:

$$\frac{\partial}{\partial t} (\rho E) + \nabla \cdot (\vec{v}(\rho E + p)) = \nabla \cdot \left[ k_{eff} \nabla T - \sum_j h_j \vec{J}_j + (\bar{\tau}_{eff} \cdot \vec{v}) \right] + S_h \quad (2.3)$$

Where,  $k_{eff}$  is defined as the effective thermal conductivity of the region, and it is the sum of the thermal conductivity of the fluid in the region ( $k$ ) and the turbulent thermal conductivity ( $k_t$ ), estimated by various turbulent models.  $J$  is the diffusion flux of the species  $j$ ,  $S_h$  is volumetric heat generation term (For example, in our case the heat dissipated by the chip). The variable  $E$  used in the L.H.S. is defined as:

$$E = h - \frac{p}{\rho} + \frac{v^2}{2} \quad (2.4)$$

Where,  $h$  is the sensible enthalpy.

#### 2.4.1. Effect of Natural Convection

Due to the temperature variations existing in the field, there can be variations in density. These variations in density coupled with the acceleration due to gravity can generate buoyancy forces and the heat transfer caused by such effects is called natural convection. To check whether the natural convection effects are important, the ratio of Reynolds number (Re) and Rayleigh number (Ra) have to be calculated. Rayleigh number is expressed as:

$$Ra = \frac{g\Delta\rho L^3}{\rho_0\nu\alpha} \quad (2.5)$$

Where,  $\vartheta$  is kinematic viscosity and  $\alpha$  is the thermal diffusivity.

For  $Re/Ra > 10^4$ , this effect can usually be neglected [36]. But since in this case, a correct length scale is difficult to determine, we incorporate this effect using Boussinesq's approximation and check for errors.

*Boussinesq Approximation:* The fluid properties are generally functions of temperature. The temperature differences will lead to variations in density. To account for this, the density is kept constant in the unsteady and convection terms of the energy equation and is treated as a variable only in the body force term in Eq. 2.3. This is part of the Boussinesq approximation which further assumes a linear variation of the density with temperature [41]. The equivalent buoyant force per unit volume term is expressed as:

$$(\rho - \rho_0)g_i = -\rho_0g_i\beta(T - T_0) \quad (2.4)$$

Where,  $\rho$  is the density,  $g_i$  is acceleration due to gravity and  $\beta$  is the coefficient of volumetric expansion. It has been found that the maximum variation in air temperature for simulations run within ASHRAE specified conditions is 20K. This leads to maximum errors of 1% for the maximum temperature in the field. This proves that the effect of

natural convection can be neglected. Also, it is mentioned in [37] that these effects are likely to reduce in future data centers because of increase in power density leading to higher CRAC velocities.

#### 2.4.2. Humidity

The air which comes out of the CRAC unit passes through the servers in the racks and is resupplied to the CRAC unit where it is cooled again. In this cooling process, no water vapor is picked up by the air (The study does not model humans in the data center). Also, since this study does not use any air side economizers, there is no interaction with outside air having low/high humidity levels. Thus in essence, the same air is recirculated and the relative humidity levels change only because of the variation in temperature (within 20K). Thus the humidity effects are neglected in this study.

#### 2.4.3. Linear Variation in Temperature

This forms the basis of the algorithm laid out in the next section. The only non-linear terms presented in Eq 2.3 are non-linear because of the temperature scalar multiplied with the velocity. Since the variations in temperature are within 20K, the velocity equations are first solved in the CFD/HT software assuming an average density. Then, the energy equation is solved. Thus, while solving the energy equation, the velocity field is constant. Also, radiation from all components is neglected in this study. This makes the energy equation linear with respect to temperature. With this we conclude that, if  $\Delta Q_i$  change in the volumetric heat generation of server  $i$  produces a temperature change of  $\Delta T_i$  at a point of interest, and  $\Delta Q_j$  change in the volumetric heat generation of server  $j$  produces a temperature change of  $\Delta T_j$  at the same point, then a simulation run with both

the changes  $\Delta Q_i$  and  $\Delta Q_j$  simultaneously in servers  $i$  and  $j$  respectively will lead to a temperature variation of  $(\Delta T_i + \Delta T_j)$  at the same point considered.

#### 2.4.4. Linear transformation between power dissipated and % CPU utilization.

This has been established and used in various studies [13]. It assumes that change in % CPU utilization in any server is proportional to the power dissipated by it.

#### 2.4.5. Server fan velocities constant

The velocities for server fans are assumed to be constant with change in CPU utilization.

### **2.5. Formulation**

Based on the temperature field linearity concept laid out in the previous section, the following is the guiding principle of the Ambient Intelligence Based Load Management (AILM) algorithm [42]. Change in the volumetric heat generation of server  $i$  ( $i \in \{1, 2, \dots, n\}$  for the given facility) present in the room contributes towards the change in the inlet temperature of server  $j$  ( $j \in \{1, 2, \dots, n\}$  for the given facility). Thus a data center can be calibrated based on how much a unit change in volumetric heat generation at server  $i$  can alter the inlet temperature of server  $j$ . The algorithm is shown in Fig. 2.6 and following is the explanation.

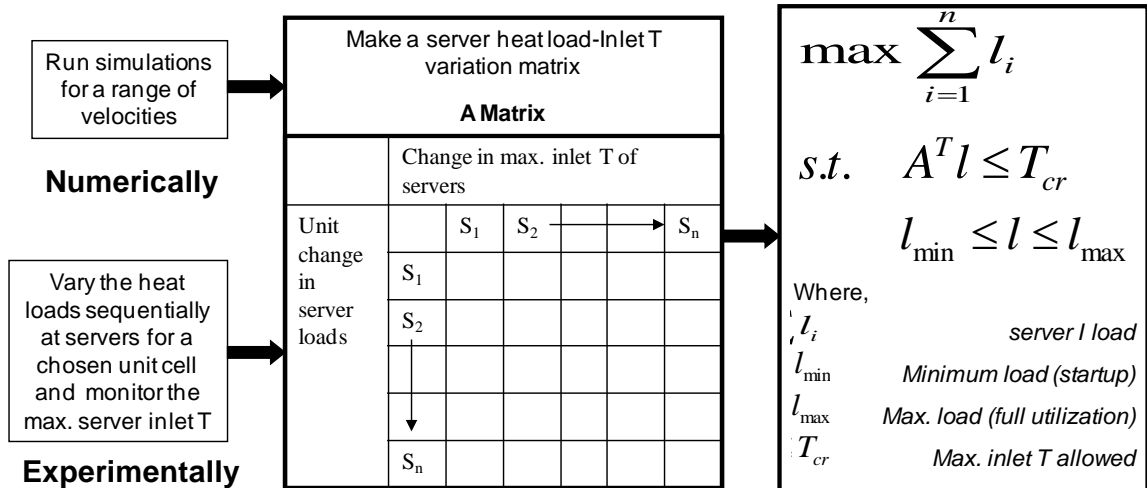


Figure 2.6. Algorithm for AILM

### 2.5.1. Calibration Phase

- For a given set of velocities of the Computer Room Air-Conditioning (CRAC) unit, the servers are run at their baseline loads (chosen to be 0.83 KW), defined as the minimum power dissipation the servers can achieve when they are running with minimum applications. For this condition, the maximum temperature at the inlet of each server is noted with the help of temperature sensors.
- The power dissipated is increased by a unit amount (noted by power sensors and guided by % CPU utilization), and the system is left to reach steady state without changing the CRAC velocity. The maximum temperatures at the server inlet are noted again and the difference from the baseline case is recorded. This difference when calculated for all the servers gives an estimate of how sensitive is server  $i$ 's inlet T with respect to server  $j$ 's heat load for a given CRAC velocity. This process is repeated sequentially for all the servers. Thus we get an  $n \times n$  matrix of values for a

CRAC velocity. Again, this matrix of values is generated for different CRAC velocities within the feasible range of CRAC velocities.

- c) For each CRAC velocity, the desired output is the maximum power dissipation the data center can take. To calculate this and the respective power dissipations for each server, we need to optimize the server loads within the constraints of the maximum and minimum loads of servers and the critical inlet server temperature (as specified by ASHRAE) below which the inlet temperatures should lie. The optimization formulation is:

$$\begin{aligned} & \max. \quad c \cdot l \\ & \text{s.t.} \quad A^T l \leq T_{cr} \\ & \quad \quad l_{min} \leq l \leq l_{max} \end{aligned}$$

Where,

$l \in \mathbb{R}^{n \times 1}$	The solution heat load vector
$l_{min} \in \mathbb{R}^{n \times 1}$	The minimum heat load (0 kW or $l_{on}$ )
$l_{max} \in \mathbb{R}^{n \times 1}$	The maximum heat load (can vary depending on server)
$T_{cr} \in \mathbb{R}^{n \times 1}$	The critical temperature specified by ASHRAE (32°C)
$c \in \mathbb{R}^{n \times 1}$	The cost vector (all entries 1 to calculate total maximum heat load capacity of the data center)
$A \in \mathbb{R}^{n \times n}$	Incremental server load against change in inlet temperature matrix. This is provided in the Fig. 2.6.

The above given formulation finds out the respective heat loads of all the servers in the data center for the maximized condition of net heat load capacity. The first constraint keeps a check on the maximum inlet temperature and keeps it below the critical limit. The second constraint defines bounds on the heat load for each server. For

the lower bound,  $l_{on}$  is the condition used. This is based on the fact that the lowest a server can go is to a just “on-state” where no useful applications are run on it. This is not 0% CPU utilization, but it is the minimum possible based on the operating system and start up programs running. This value will likely vary for different generation of servers. The rationale for using this is that if the servers are completely switched off, then it will be more time consuming to get them operational (switching on time is between 3-5 minutes). The other lower bound option is 0 kW. This considers that if a server is not running any useful application, it could be shut off. Since it is a linear programming framework, for this bound we might get values between 0 kW and  $l_{on}$  for some servers. While this value is not realistic, such servers can be set to 0 kW and their contribution to the net heat load capacity can be subtracted, giving a conservative estimate of the maxima. This is done to avoid a step jump between 0 kW and  $l_{on}$  and thereafter a continuous curve till  $l_{max}$ , as that will require a Mixed-Integer Programming (MIP) formulation. MIP’s are generally more time/memory intensive [43]. The difference between the LP solution and the approximate solution is later found out to be less than 0.5% and thus invalidates the use of an expensive solver.  $l_{max}$  is also a parameter specific to a server. Different values have been assumed for running simulations and will be discussed later. The optimization model is solved in MATLAB using the optimization toolbox with the Revised Simplex Algorithm

### 2.5.2. Operational Phase

The above calibration is performed once at the commencement of the data center operations, and subsequently every time new servers are added. While this gives us the base values, the implementation during operation phase is different. This implementation

can be explained with the help of Fig. 2.7. Here, 6 different  $V_{CRAC} : V_1, V_2, \dots, V_6$  are assumed. The corresponding maximum heat load dissipation of the data center are  $D_1, D_2, \dots, D_6$  with load vectors  $l_1, l_2, \dots, l_6$  respectively. There are five different scenarios of workloads considered:

1. **W<sub>1</sub>**: This workload is higher than the maximum workload the Data center can take. Thus either some of the workload will be discarded or the algorithm will show an error.
2. **W<sub>2</sub>**: While this corresponds to the range of workload the Data center can take, it also is possible for only one velocity, i.e.,  $V_3$ . Thus, the load distribution will take place on the basis of  $l_3$ .
3. **W<sub>3</sub>**: While, the workload lies in the range, there are 2 different velocities where it is exactly equal to the maximum heat load dissipation corresponding to those velocities. In such an instance, the lower velocity,  $V_2$ , will be chosen with the corresponding distribution  $l_2$  as that will lead to lower power consumption.
4. **W<sub>4</sub>**: This is the most interesting case. The workload is within range, but is not equal to any  $D_i$ 's. It lies between  $D_1$  and  $D_2$ ,  $D_4$  and  $D_5$ ,  $D_5$  and  $D_6$ . Now these ranges are compared internally and the highest amongst both are chosen, i.e.  $D_2$ ,  $D_4$  and  $D_6$ . Now within these three, the lowest dissipation is chosen, i.e.  $D_2$ . Further the load vector corresponding to  $D_2$  is scaled down by the scaling factor (s) such that:

$$s \sum_{i=1}^n l_i = W_4$$

Here,  $s = W_4/D_2$ . By the first step, a dissipation value is chosen which could be scaled down to achieve a more conservative solution. During the second step, the dissipation corresponding to the lowest velocity is chosen for using least amount of facilities power. Finally, according to this velocity, the optimal load vector is scaled such that the required workload capacity could be achieved. Thus the approximation in the solution which leads to loss in energy though keeping the servers within safe limits, can be minimized by calibrating for more number of velocities, especially in the range where the solid line in Fig. 2.7 is steep.

5. **W<sub>5</sub>**: This workload can be achieved through any velocity. Thus the lowest velocity,  $V_1$ , is chosen and a corresponding scaling as explained in the previous section is carried out.

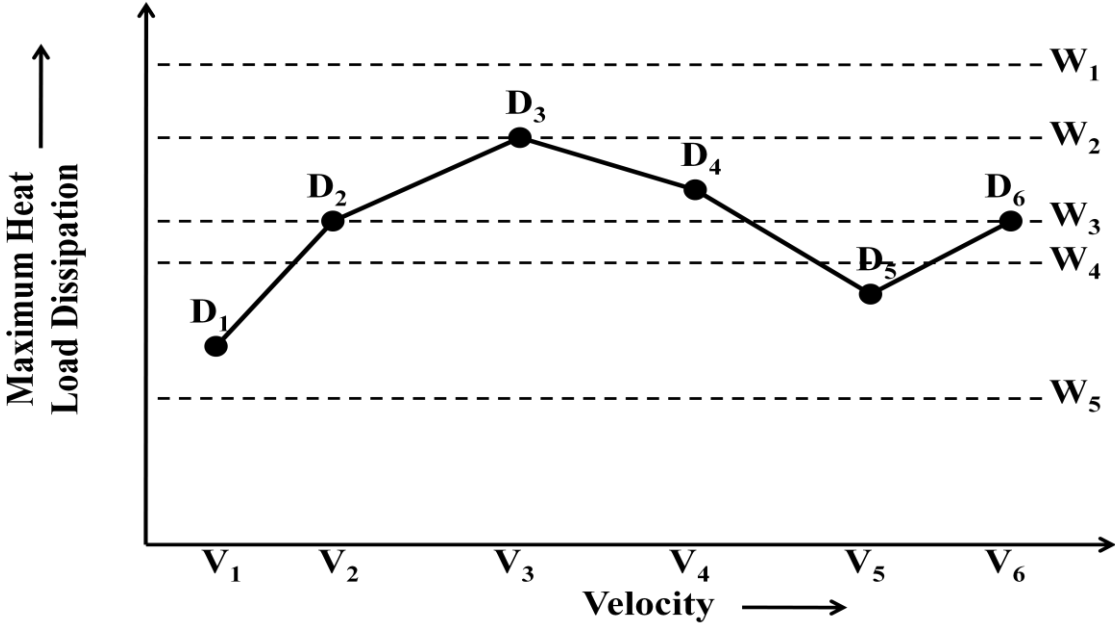


Figure 2.7. Various cases possible for DCHDC with VCRAC

## 2.6. AILM Results and Discussions

### 2.6.1. Calibrating the Data Center

Following are the steps for calibration:

1. For  $V_{CRAC} = 3$  m/s, a uniform heat load of  $l_{on} = 0.83$  kW (note the final results are **independent** of the choice of values here) is given to all the servers. The CFD/HT simulation is run and the maximum rack inlet temperature values are recorded under the tab “Base Values”.
2. The  $A_{11}$  server is now given twice the previous load, i.e. 1.66 kW and all other servers are kept at the same value. The simulation is run and all the maximum rack inlet temperatures are recorded again. The difference in the current and Base Values are listed under the tab “Delta  $A_{11}$ ”. This gives the variation in the inlet temperature of all the servers for a 0.83 kW increase in the heat load of  $A_{11}$ . This process is repeated for all the servers. The values are the different rows of the A Matrix defined in the previous section.
3. All the vectors defined in the optimization formulation are defined. Now, these vectors are input arguments to the `linprog()` function in MATLAB R2007a. This function uses the Simplex optimization framework. The output is the load vector. The optimization takes 0.7s to give a converged solution.
4. This process is repeated for 4, 5, 6, 7, 8 m/s  $V_{CRAC}$ .

### 2.6.2. Comparison with the “Unmanaged” Data Center

Figure 2.8 shows the comparison between the “unmanaged” and AILM cases. The assumptions of the unmanaged case are:

1. The Data Center Manager (DCM) is unaware of thermal issues.
2. All the servers in the data center have the same characteristics.

This makes it clear that the DCM is unbiased about the load distribution in the data center. Thus uniform heat loads are given to all the servers. The AILM case, as explained in the formulation, goes through the algorithm steps and calibrates the data center. Then in the operational phase, it distributes the heat loads based on the previous characterization. The current case is for  $V_{CRAC} = 5$  m/s. For the AILM case, it is found out that the maximum heat load capacity for this velocity could be 220 kW. This is distributed according to the algorithm and the FLUENT simulation is run. To compare the uniform case with this, 220 kW of heat load is distributed equally in all eight the racks, i.e. 27.5 kW per rack, and simulation is run. Fig. 2.8(a) presents the uniform case while Fig. 2.8(b) presents the AILM case.

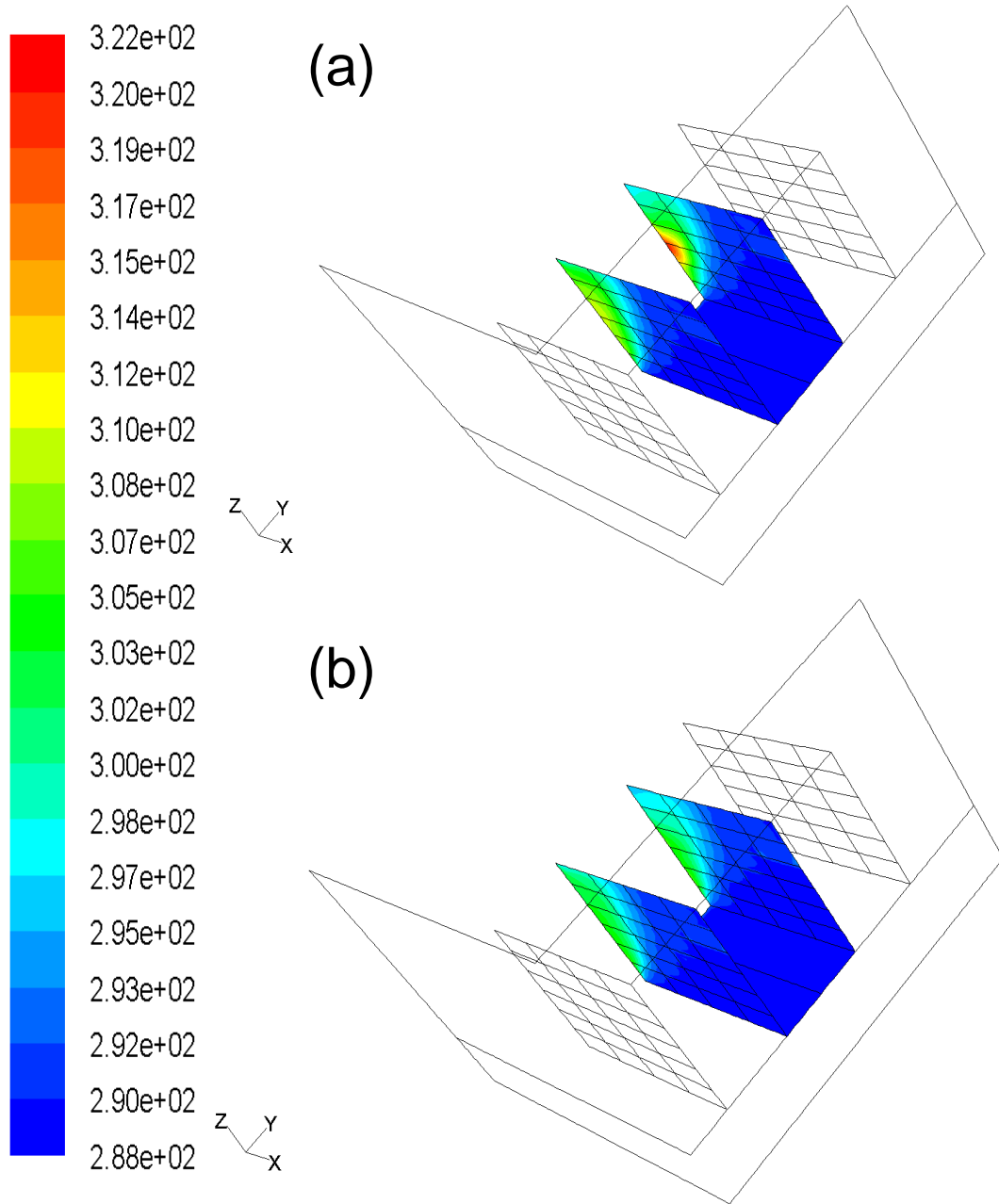


Figure 2.8. Temperature contours at inlet face of racks for loading according to (a) Uniform case, (b) AILM case. ( $V_{CRAC} = 5$  m/s)

The following characteristics are noticed:

- The uniform case has a maximum rack inlet temperature of 322K.

- The AILM case has a maximum rack inlet temperature of 305.3K. This is just above the critical temperature limit of 305.2 K for rack inlets as defined before.
- The shape of contours for the both the cases is almost the same with racks  $A_1$  and  $B_1$  being the hottest. The extent of contours is larger for the uniform case with temperatures above 298K present in  $B_2$  rack. The hottest server is  $B_{14}$  for both the cases.
- Recirculation is not just affecting the topmost servers in a rack.  $B_{14}$  is hotter than  $B_{11}$  for both the cases. This proves our previous conclusion that side recirculation is predominant.
- B racks are hotter than their corresponding A racks. This can be explained by Fig. 2.4(i) where recirculation in B racks is higher than in A racks.

Figure 2.9 gives the plot for the maximum server inlet temperature for both “Uniform” and AILM cases. Here the range for AILM is 0.8 – 7.5 kW per server and a uniform load of 5 kW per server. The simulation is for  $V_{CRAC}$  of 5 m/s. The heat dissipation capacity for AILM case is found to be 298 kW while for the uniform case, 240 kW is simulated. The x-axis is the location of servers and is explained in the left top matrix. The servers 11-16 are closer to the CRAC unit. While the uniform case clearly overshoots the safe temperature limits, the AILM, brings down these temperatures and allows higher heat load dissipation for servers which do not affect the other servers a lot. The maximum temperature for AILM is within range, thus confirming the satisfaction of the optimization model constraints.

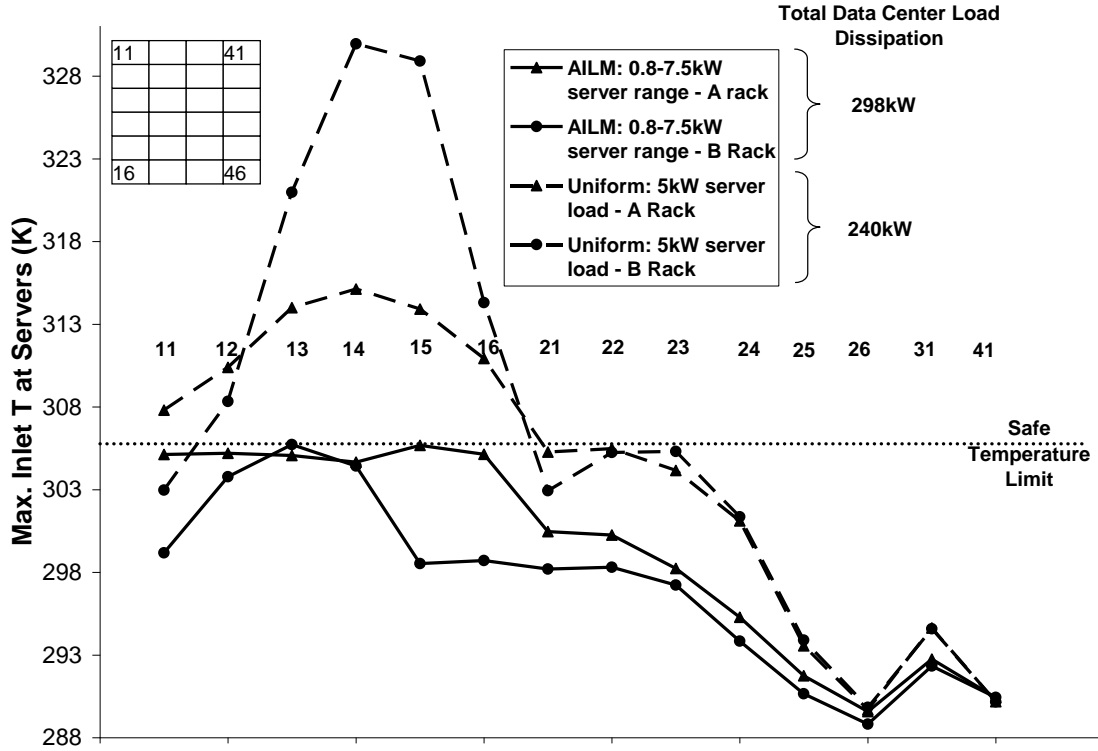


Figure 2.9. Server inlet temperatures versus server no. (please not nomenclature on top left)

Fig. 2.10 gives a qualitative comparison between the “Uniform” and AILM cases. The initial condition for this setup is that each server can vary its loads between 0 – 5 kW. Thus  $l_{max}$  is 5 kW. For this, the simulations are run for 6 different velocities:  $V_{CRAC} \in \{3, 4, 5, 6, 7, 8\}$ . While the algorithm gives the best distribution, to determine the uniform load in racks which does not violate the bound on maximum inlet temperature, following is done:

1. The same A matrix is used which is derived for the AILM. Since now the loads have to be uniform, the  $l$  vector is a constant ( $c$ ) multiplied by a vector of ones. Thus  $A^T l$  is sum of the columns of A multiplied to  $c$ .
2. Thus for each  $i$  ( $i \in \{1, 2, \dots, 48\}$ ),  $c$  can be determined by calculating  $b_i / (A^T l)_i$ .

3. Finally, the value of constant chosen is  $c = \min\{ \min\{b_i/(A^T l)_i\}, l_{max} \}$ .

Thus the net heat load capacity of data center becomes  $48 \times c$ . This value is plotted alongside the AILM values for each velocity in Fig. 2.10. The following trends are noticed:

1. The AILM case data center Heat Dissipation Capacity (DCHDC) doesn't vary significantly. The minimum is for 3 m/s (216 kW) and maximum is for 8 m/s (238 kW). Thus for a 166% increase in velocity, the capacity expansion is just 10.2%. Another way to look at it is that if the required DCHDC is 10% lower, then significant energy can be saved. Also, while the extremes are minimum and maximum, the variation for the in-between values is not apparent. From 4 to 5 m/s, the DCHDC decreases and from 6 to 7 m/s, there is no increase in DCHDC.
2. For the uniform case, the minimum DCHDC is 142 kW for 3, 4 and 5 m/s and the maximum is 228 kW for 8 m/s. In this, for a 166% increase in velocity, the increase in DCHDC is substantial (60.6%). Again, the trend for the in between velocities is not monotonically increasing. For, the first 3 velocities, it is a constant, after which it dips for 6 m/s and then sharply increases for the next two velocities. This can be explained by the fact that the bottleneck server for the first 4 servers received substantial direct cold air supply for higher velocities.
3. The trend followed by the two curves is not the same. From, 4 to 5 m/s, DCHDC for AILM shows a decrease while uniform remains constant. From, 5 to 6 m/s, it increases for AILM while it decreases for uniform case. For 6 to 7 m/s, it remains constant for AILM but increases by 20.6% for the uniform case.

4. For lower velocities, DCHDC for uniform case is significantly lower than the AILM. Thus the improvements achieved can be of the order of 52.1%. The maximum benefit is obtained for 6 m/s and is 64.7%. The % increase in DCHDC decreases for higher velocities. For 8 m/s, it is 5.3% only. This can be explained by the fact that side recirculation decreases drastically for higher velocities, as explained in Fig. 2.4(i) and 2.4(iii).

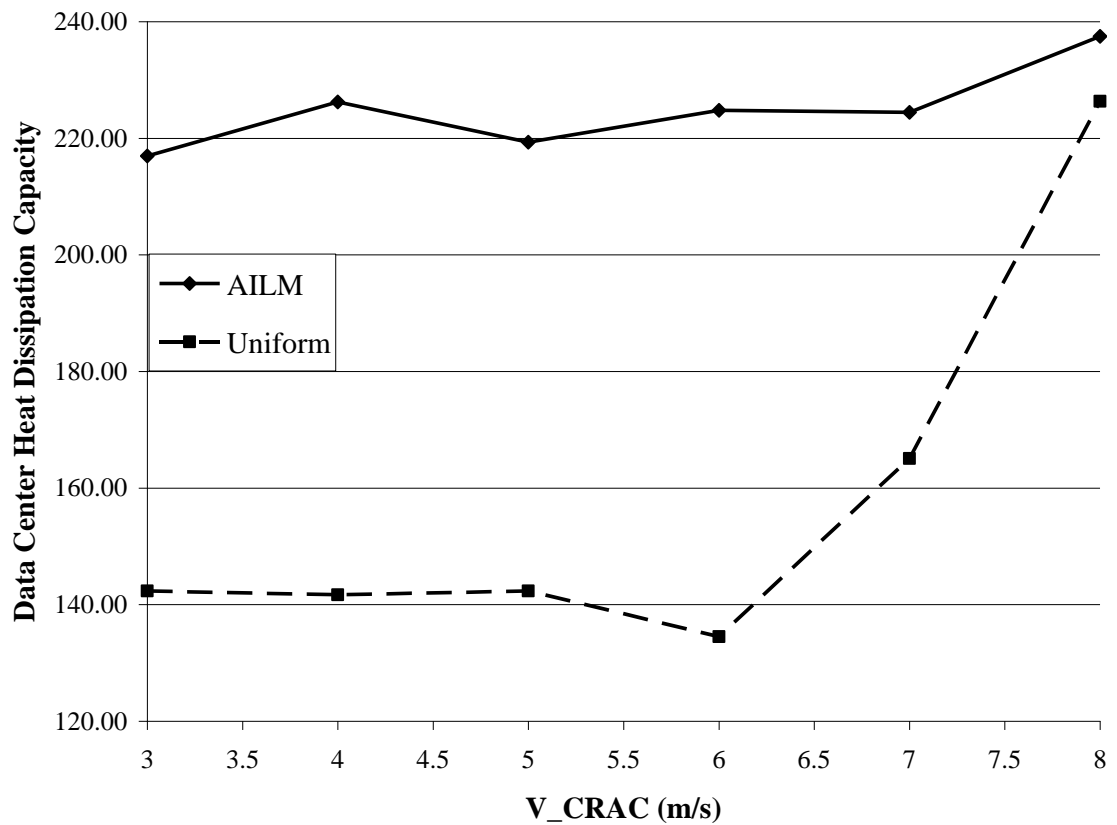


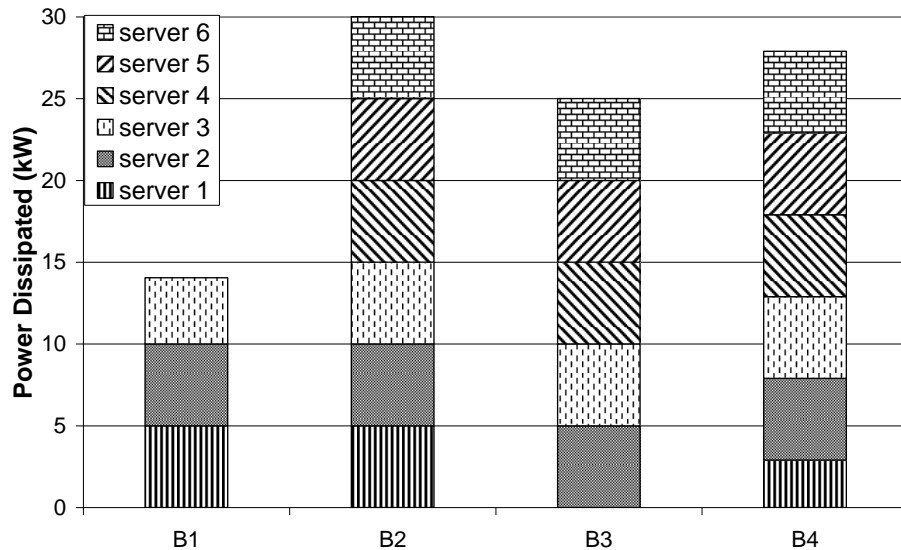
Figure 2.10. Data center Heat Dissipation Capacities (DCHDC) with  $V_{CRAC}$  for "Uniform" and AILM cases

Different case studies are used to understand various aspects of AILM.

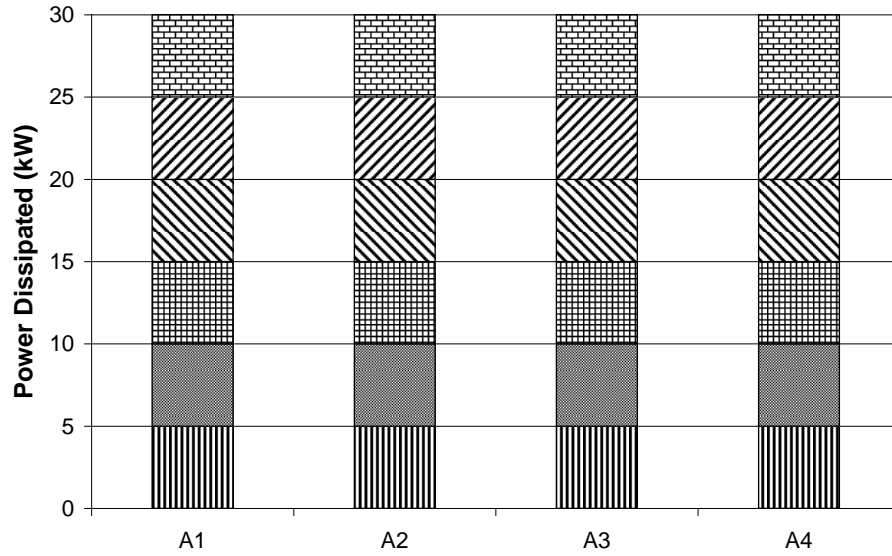
### 2.6.3. Case study 1: Load Distribution variation with $V_{CRAC}$

As explained in the previous section, simulations are run for racks with similar servers. The AILM algorithm is applied and a load distribution is determined. This section deals with the study of this distribution. For these simulations, the server load can vary between 0 – 5 kW. Thus a rack maximum load can be 30 kW. For 3 m/s (Fig. 2.11(a), 2.11(b)), it is noticed that A racks have maximum loads.

This is at the expense of the loss of load in B rack. From previous sections it was understood that for lower velocities, racks were causing the recirculation. Since AILM depends on the inlet temperature increasing characteristic of a server, it is expected that it will push the loads on these servers down. The lowest is in the  $B_1$  rack because of the highest effect of recirculation existing there. One interesting feature is the  $B_2$  rack carrying full load as compared to  $B_3$  and  $B_4$ , even though  $B_2$  causes more recirculation than the others.



(a)



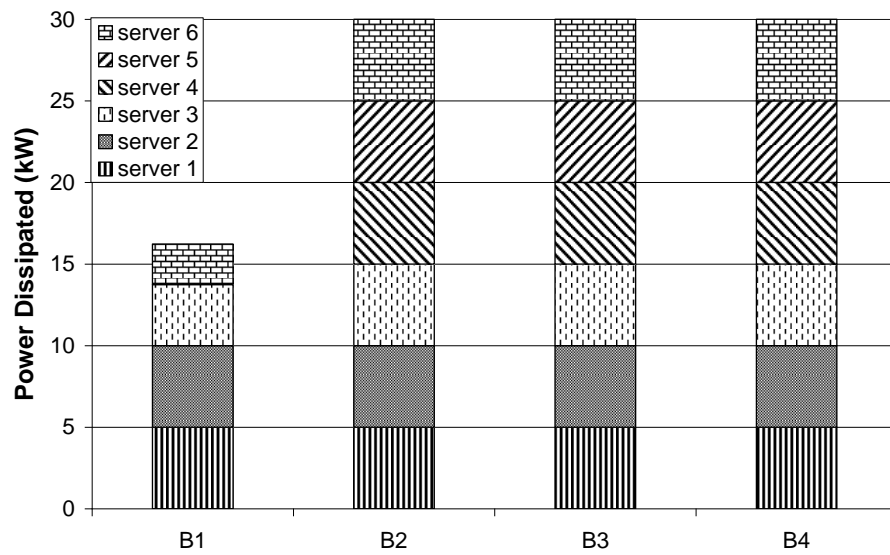
(b)

Figure 2.11.  $V_{CRAC} = 3$  m/s, Power Dissipation Map: (a) B Rack, (b) A Rack

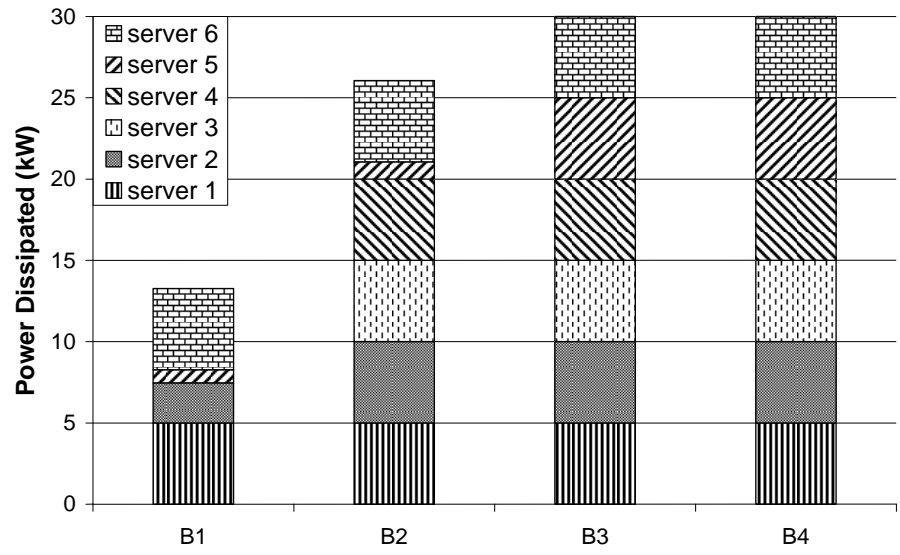
For 4 m/s (Fig. 2.12(a)), the  $B_1$  rack has lower load as expected. It is interesting to notice that full load in all other racks is compensated by the decrease in the load of  $B_1$  only. This increases the data center capacity from 3 m/s case. On the other hand, for 5 m/s (Fig. 2.12(b)) the decrease in  $B_1$  rack load cannot compensate the full load by others. Thus some load is taken off from  $B_2$  rack too. This decreases the new data center capacity from 4 m/s to 5 m/s case. It has been observed that some higher velocities accentuate the effect of recirculation with the transfer of momentum from their cold air flow to the hot air existing in the room.

For 7 m/s (Fig. 2.13(a), 2.11(b)), the flow conditions inside the room reverse. The recirculation is now dominated by the A racks and they are also the ones which are most affected by it. Thus in the B racks, there is only a slight decrease in load for  $B_1$ , rest all staying at maximum. This is counterbalanced by the decrease in load for the A1 rack. It is

noticed that for these higher velocities, the decrease in load is only for the first rack. This is because, firstly, the recirculation on the exhaust aisle is dominated by the first racks primarily. Secondly, the momentum of the flow coming out of the tiles is high enough to drive away the recirculated flow from the hot aisle to the cold aisle to adulterate the inlet flow for other racks. Again, for 8 m/s (Fig. 2.13(c)), due to the high inlet velocity, the load is affected only for A<sub>14</sub> server.

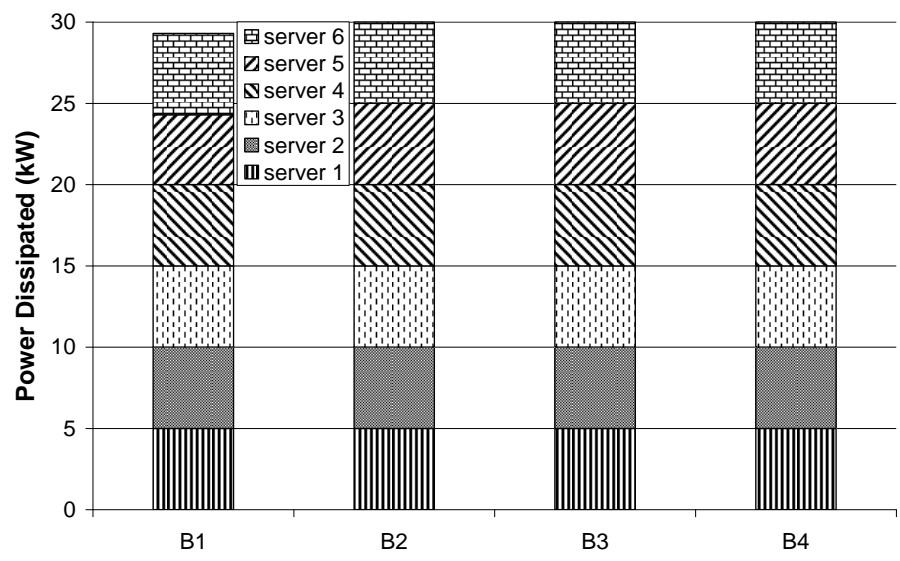


(a)

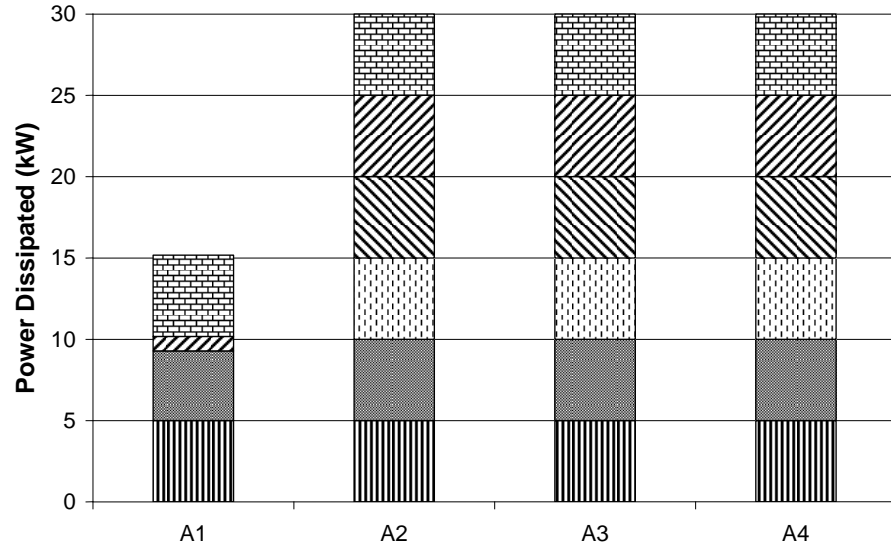


(b)

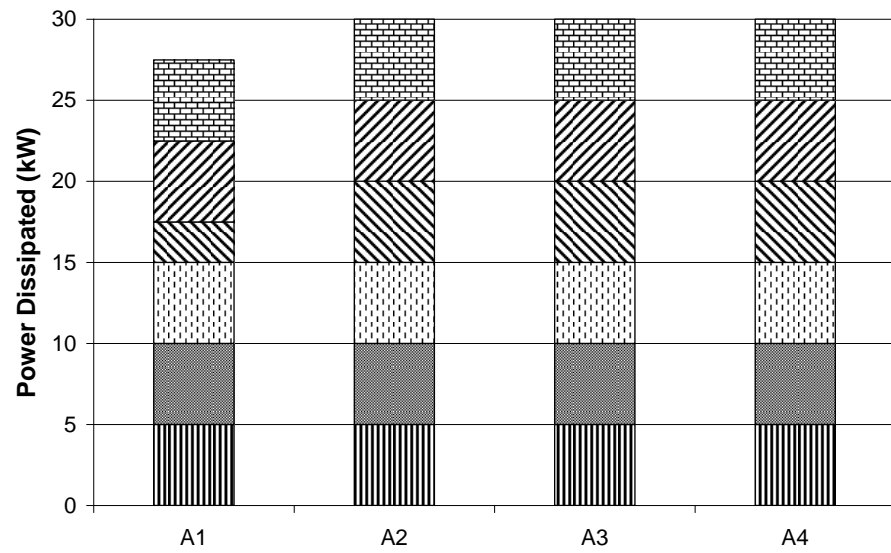
Figure 2.12. B Rack Power Dissipation Map: (a) VCRAC = 4 m/s, (b) VCRAC = 5 m/s,



(a)



(b)



(c)

Figure 2.13. Power Dissipation Map - (a)  $V_{CRAC} = 7$  m/s, B Rack, (b)  $V_{CRAC} = 7$  m/s, A Rack, (c)  $V_{CRAC} = 8$  m/s, A Rack

#### 2.6.4. Heterogeneous Data Center

The previous section had considered all servers to have similar characteristics. In this study, the DCM will be indifferent of the IT aspect, as all the servers perform equally

well. Thus in this case, the biggest guiding factor would be the thermal condition of the room. In actual data centers, a myriad of server types are present and thus while thermal condition is an important constraint, it is not the sole guiding factor. Also, AILM could be potentially used to understand the correct placement of servers in the first place, so that after the optimization, the load distribution achieved can give net higher capacity than any other arbitrary configuration. To understand the characteristics of such conditions, two different cases are considered. One considers two different kind of air-cooled racks and the subsequent distribution of load. The other considers the question of where to place a liquid cooled rack if one is introduced in a primarily air-cooled Data center.

#### 2.6.5. Case study 2: Heterogeneous Air-Cooled Data Center

Figure 2.14 shows 4 different configurations considered. Here the B racks are the more advanced racks with servers giving high performance and lower heat dissipation rates. A racks are the older racks with servers having low performance with higher dissipation rates (For example, A Rack servers are P4 and the B Rack servers are the new Dual Cores from Intel Corporation). Following are their characteristics:

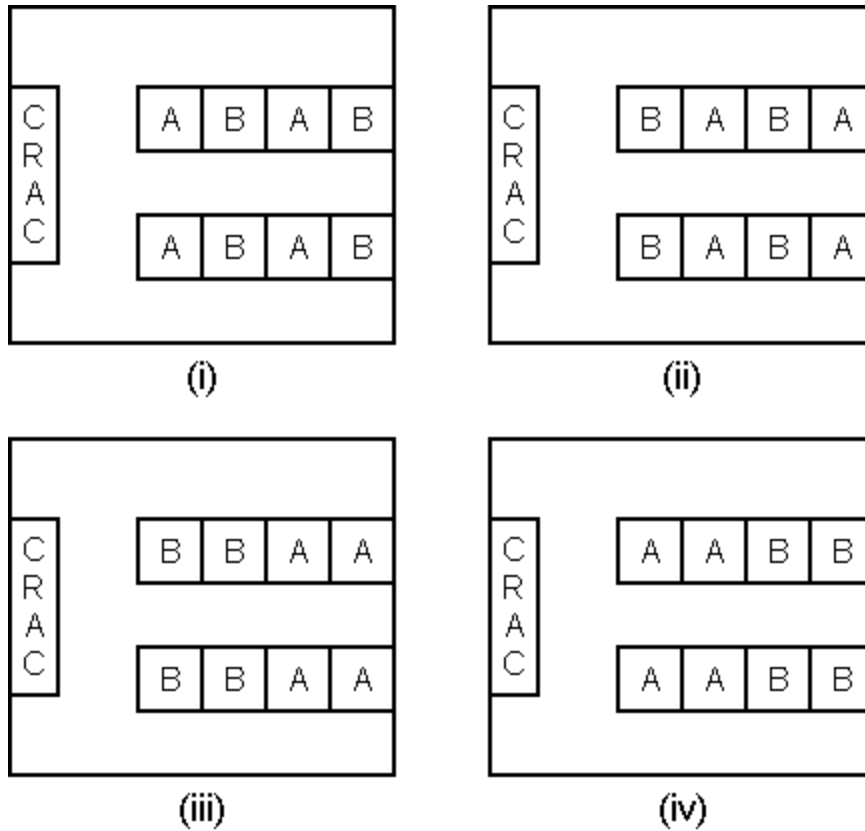


Figure 2.14. Different rack layouts

Table 2.2. Rack specifications for cases

	<b>A Rack</b>	<b>B Rack</b>
<b>Heat Dissipation</b>	0 – 5 kW per server	0 – 3 kW per server
<b>Work Unit</b>	0.5	1

AILM is performed on this configuration to understand which velocity will work the best for an ABAB configuration (Fig. 2.14(i)). The maximum heat load dissipation potential for the Data center is now 192 KW. For each velocity, the DCHDC is calculated using AILM and the efficiency of the Data center is calculated based on the potential.

The graph is plotted in Fig. 2.15. It is seen that the efficiency reduces from 3 to 5 m/s VCRAC. This however increases after that to a 100% for 8 m/s.

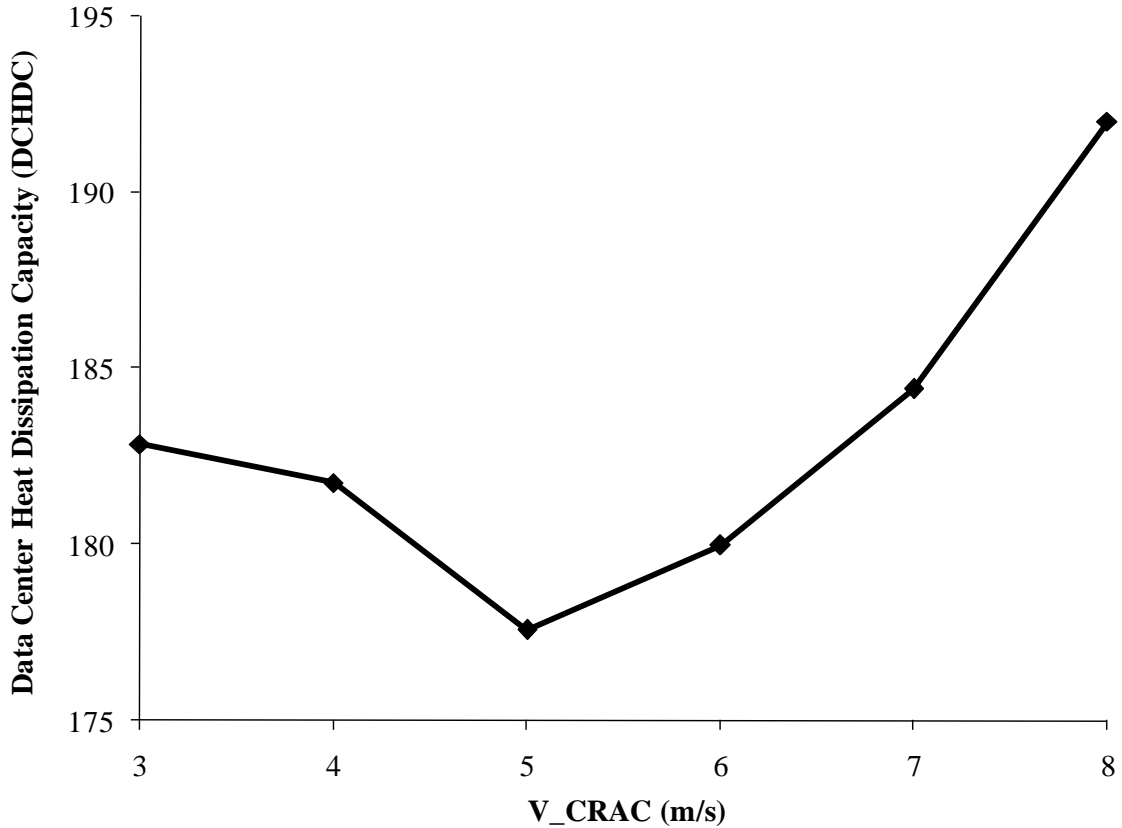


Figure 2.15. Data center Heat Dissipation Capacity with  $V_{CRAC}$  for Table 2.1

This is however not in accordance with Fig. 2.10 where the DCHDC increases from 3 to 4 m/s. Also, the difference between the least and the maximum DCHDC is around 12 kW and thus all efficiencies are greater than 90%. The current Blade centers dissipate up to 6 kW of power. Thus to check for higher heat load variations, the following are the characteristics assumed:

Table 2.3. Rack configurations for Case 2

	A Rack	B Rack
Heat Dissipation	0 – 10 kW per server	0 – 6 kW per server
Work Unit	0.5	1

The work units are still kept to be 0.5 and 1 since they are relative to each other. For these characteristics, configurations Fig. 2.14(i) - 2.14(iv) are simulated. The AILM results are plotted in Fig. 2.15. The efficiency is calculated by dividing the DCHDC by the maximum workload possible, i.e. 192 kW. Following are the trends noticed:

- ABAB configuration has the lowest DCHDC/efficiency for all velocities except 3 m/s, where it's the highest.
- BBAA is the best configuration for all velocities except 3 m/s, where it is worst. This is intuitive since it allows B to be the first 2 racks, which as explained in Case study 1, tend to have not maximum heat loads. Thus the loss of potential will be more if the first rack is an A racks instead of a B rack.
- For all configurations, 4 m/s velocity achieves the maximum efficiency. This was also noticed in Case Study 1
- All the configurations reach similar performance for 8 m/s velocity. Even though the performance is the same, it is not the maximum. This similarity is because of decreased recirculation as explained before. The decreased performance is because due to high z momentum of the flow coming out of the tiles, the bottom servers are not able to draw the required air effectively. Thus the cooling achieved is not optimal.

- The trends in between the minimum and maximum velocities are different for different configurations. For the BBAA configuration, the efficiency increases for increase in velocity from 3 to 4 m/s and then monotonically decreases with increase in velocity. For BABA, there is a slight increase from 7 to 8 m/s. For ABAB and AABB, there is an increase from 5 to 6 m/s and 7 to 8 m/s.

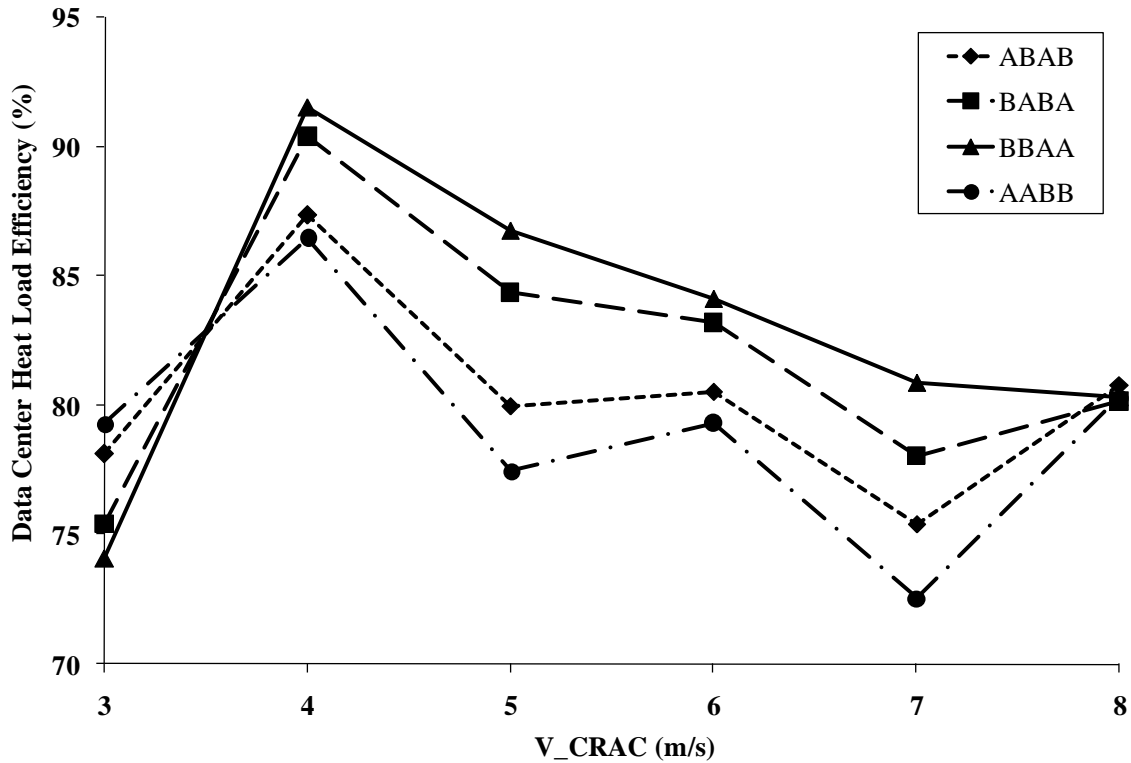


Figure 2.16. Data center Heat Load Efficiency with  $V_{CRAC}$  for different configurations

A 0.5 work unit for A rack would be achieved if it is allowed to dissipate full 10 kW. From 10 to 0, it is scaled linearly between 0.5 and 0. Similarly for 6 to 0 kW for B rack, the work units are scaled linearly between 1 and 0. The efficiency is then calculated for the allowable units with the reference of the maximum work units potential. This will give a thermally guided computer load allocation. In Fig. 2.17, this efficiency is plotted for each configuration and velocity. It is noticed that AABB and ABAB configurations

perform better than the BBAA and BABA. Thus while if only AILM is performed without considering the work unit aspect of it, then BBAA would be the best contender for server placement. But, from this plot, depending on velocity, both AABB and ABAB, can perform well. If one choice has to be made, the author would recommend ABAB since it is within the top two contenders in both the cases.

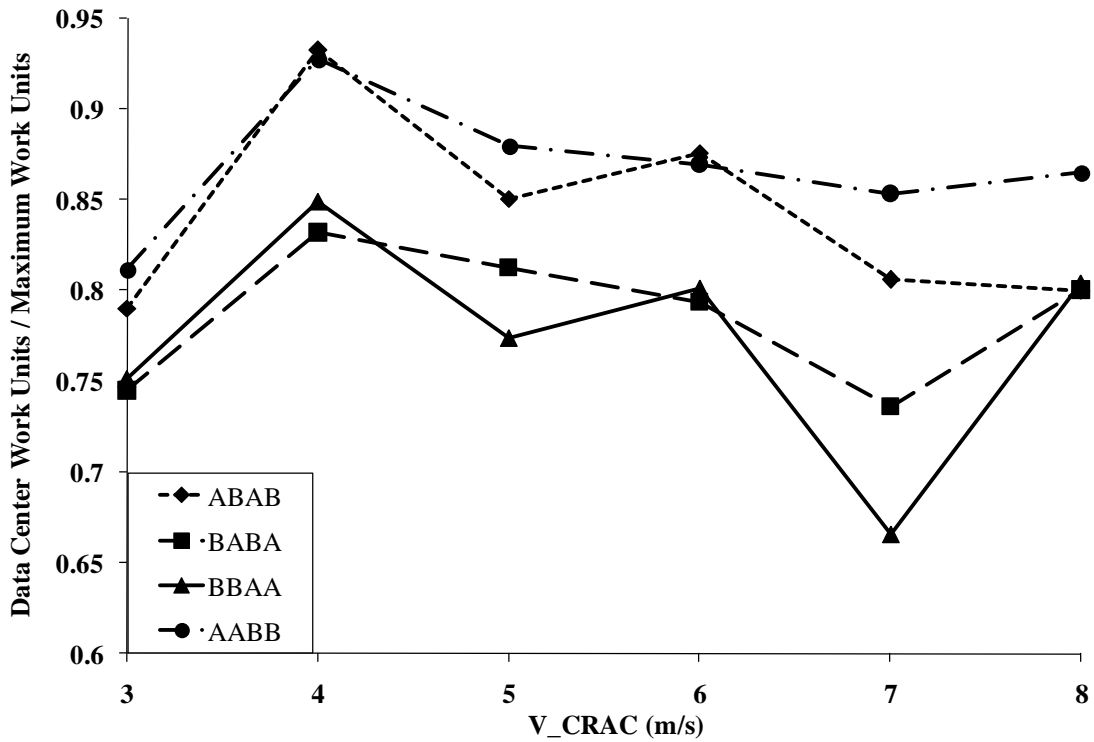


Figure 2.17. Data center Workload efficiency with  $V_{CRAC}$  for different configurations

An important conclusion from Figs. 2.16 and 2.17 is that the best configuration according to heat load maximization is not the best configuration for workload maximization. The main purpose of a data center is to perform maximum (computer) work possible. Thus, while the objective function of maximization being total heat load capacity gives a good insight as to which configurations might work well, it fails to give higher preference to the computer workload and thus might not give the optimized results

for maximum workload the data center can take. The objective function for the case study presented in Fig. 2.18 is changed to maximizing the work units/load, keeping the same constraints as the earlier model.

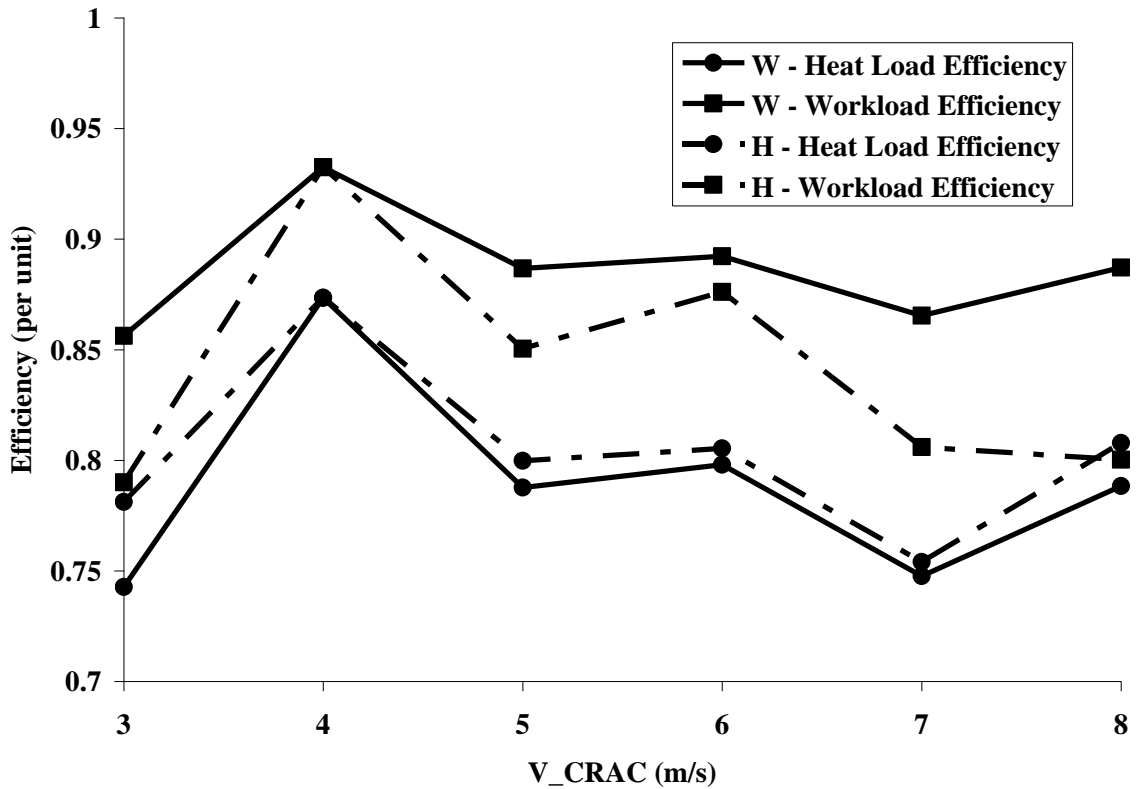


Figure 2.18 Heat load and Workload efficiency for objectives being maximum workload

The results presented in Fig. 2.18 have prefixes *H* and *W*. *H* represents the case where maximizing heat load was the objective and *W* represents the case maximizing work units was the objective. For each velocity, the *W* simulation gives greater than equal to workload than *H* simulation. Also, the *W* simulation gives lesser than equal to heat load capacity than *H* simulation, thus being a bad optimization model, as far as heat load is concerned. This model while giving better results, gives an interesting approach where the IT objective is considered with Facilities constraints.

### 2.6.6. Case study 3: Design of Data Center for Liquid Cooled Racks

The LCRs are increasingly becoming more popular in data centers. Heterogeneous data center with both liquid and air cooled racks will become a common practice in near future. Since, liquid cooled racks will replace the existing air cooled racks in some cases, a systematic way of determining the correct positioning, keeping in mind the infrastructural constraints has to be devised. The following has to be considered:

1. The positioning should be such that it can replace the air cooled rack which is most constrained in providing higher work units due to either inherent load limitations or the characteristic flow for a particular layout of a data center.
2. The positioning should also ensure, if possible, that other air cooled racks benefit out of the placement of this liquid cooled rack, for example by: diverting cold air flow towards the other racks, or decreasing recirculation.

Let us assume a case in which two new liquid cooled racks are brought in a data center (the model described above) and two air cooled racks have to be displaced. The question now is: Which ones should be replaced?

The previously derived  $A$  matrix during the formulation phase of AILM gives how differential change in heat load for a particular server can affect the inlet temperature of another server. Thus it provides the recirculation characteristics (location and extent). Based on the  $A$  matrix, for all the velocities, the two most vulnerable racks to recirculation are the  $A_1$  and  $B_1$  racks. This has also been observed in the temperature contours in Fig. 2.8 which shows the inlet temperatures are highest for the first racks in each row for both the AILM and uniform case. Thus the two ways to determine the most vulnerable racks are:

1. Highest recirculation potential based on the  $A$  matrix.
2. Highest inlet temperatures based on the temperature contours. In an actual data center, this will be represented by the thermocouples placed at the inlet of the server.

Computationally, this replacement can be modeled as two solid blocks of rack dimensions in place of  $A_1$  and  $B_1$  which do not dissipate any heat (since in liquid cooled racks, heat generated is transferred internally to circulating coolant). This block is then simulated for  $V_{CRAC} = 5$  m/s. The net load in the Data center is assumed to be 220 kW and for the uniform case, it has been distributed equally between the racks at 27.5 kW. In the hetero case, the heat load is distributed uniformly to the 6 air cooled racks at 36.67 kW. The results are presented in Fig. 2.19. The results show that, while for the uniform case in Fig. 2.19(a), the inlet temperatures have reached 322K, for the hetero case, the inlet temperatures are just 307K, even though the same load is distributed in 6 racks instead of 8. What this also suggests, is that the load carried by the liquid cooled racks will be additional and the Data center, in essence, could support higher loads. This gives an example of a case where the right placement of liquid cooled racks can increase the capacity of the remaining data center, as well as give additional capacity due to the inclusion of the new racks.

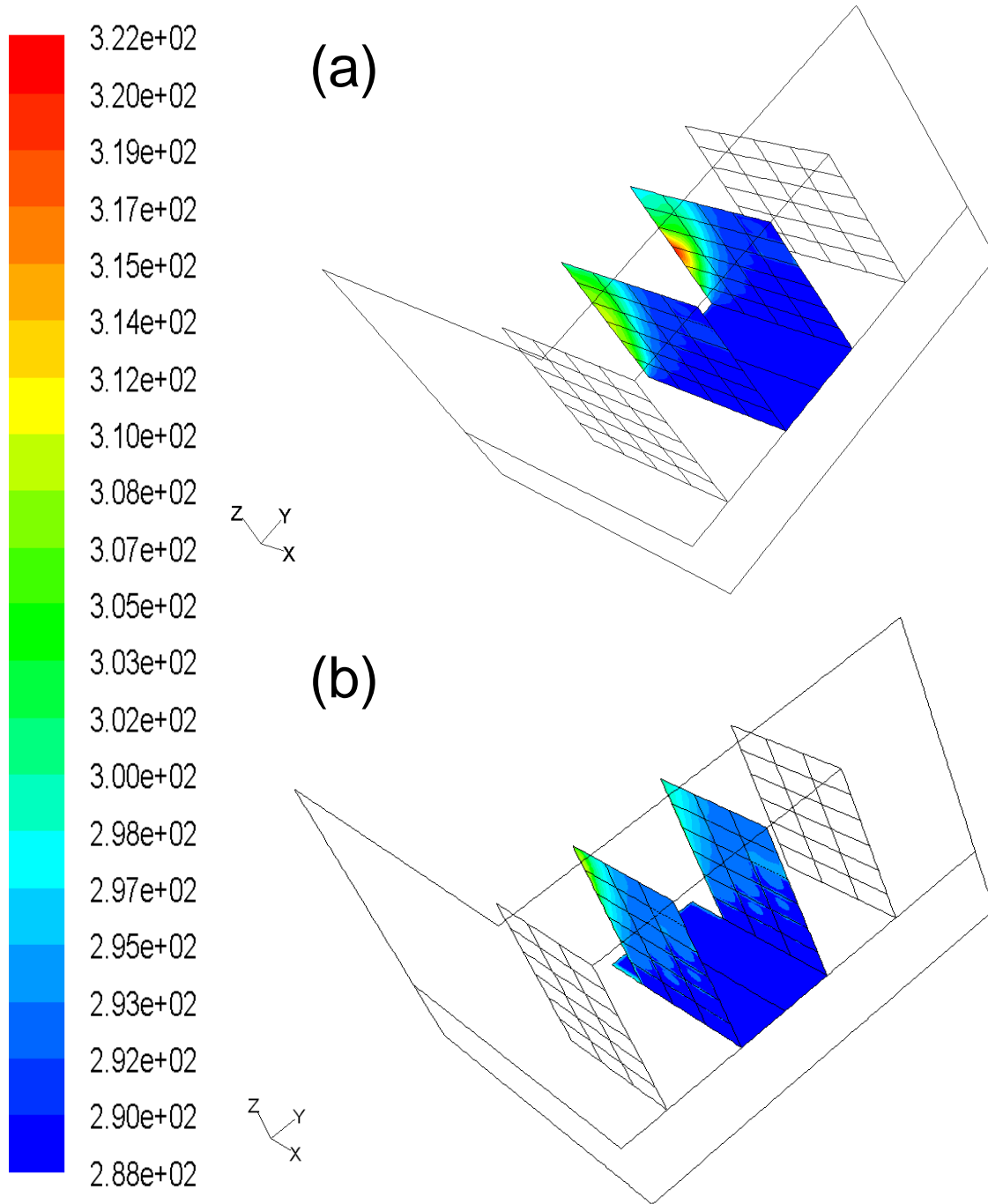


Figure 2.19. Rack inlet temperature contours for  $V_{CRAC} = 5$  m/s (a) Uniform case, (b) Heterogeneous Case

### 2.6.7. Case study 4: Sensitivity of DCHDC with $T_c$

This case study helps to understand the sensitivity of DCHDC as calculated by AILM with the critical temperature. This can help determine that if in future, the threshold temperatures as specified by regulatory bodies changes, how will the net DCHDC get affected. Simulations are run for  $V_{CRAC} = 4$  m/s with the ABAB arrangement as specified in Table 2.2. The results show, that the variation in DCHDC is almost linear with  $T_c$ . An approximate linear curve has also been drawn in Fig. 2.20 along with the actual results. The  $R^2$  value of the line is 0.988 and thus the variation is fairly linear.

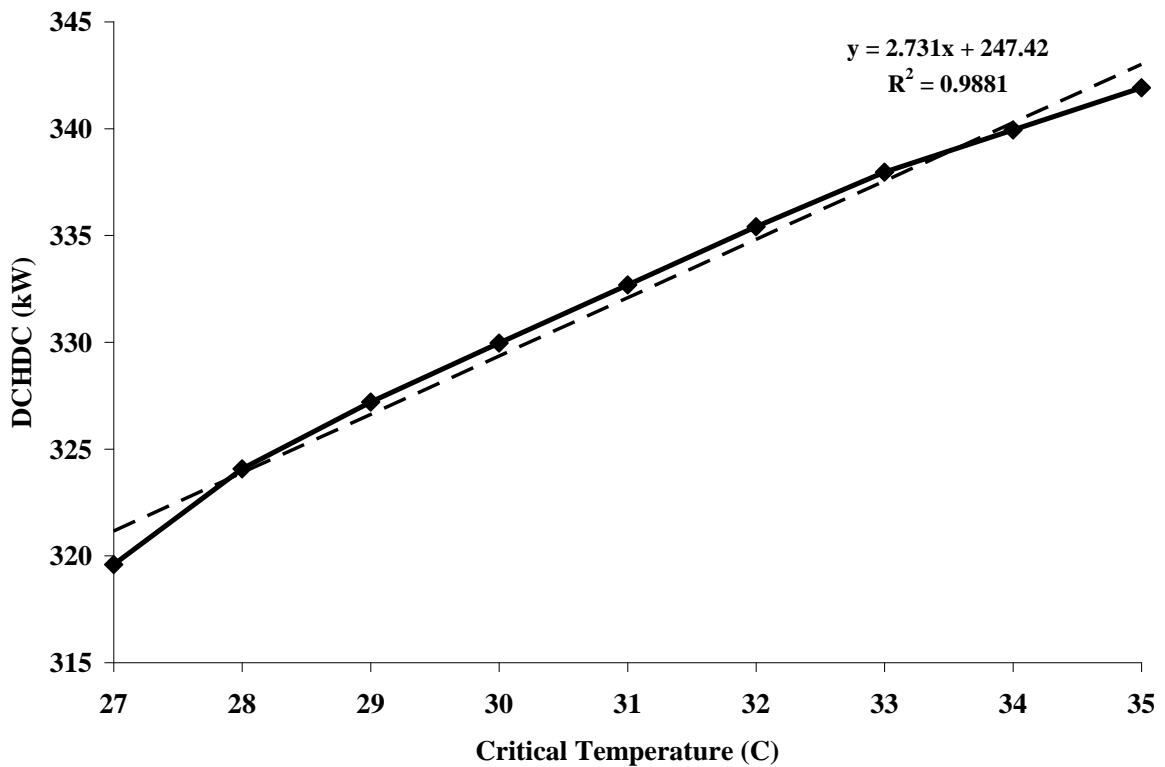


Figure 2.20. DCHDC with  $T_c$  for the considered model.

As the numerical simulations have been performed, experiments have to be performed on an actual test bed to determine the validity of assumptions in the numerical

simulations and understand the issues which will affect the practical application of such a concept.

## **2.7. Experimental Setup**

In this section, the experimental validation of the AILM concept is presented. Initially, the control volume setup is described. A server simulator is used instead of a rack of servers. The advantages and disadvantages of this are listed out. The instruments used for measuring various quantities are also explained in details. Following the setup descriptions, results for this analysis are presented. A comparison between the assumptions present in the CFD/HT model and the experimental analysis is laid out and cause for variations in the same are explained.

### **2.7.1. APC Server Simulators**

American Power Conversion (APC) server simulators are a replacement to actual computing servers. They include an array of heat generators inside them allowing for better quantification and control of parameters of interest. These parameters are:

1. *Heat Dissipation Rates*: The power dissipated in an operating server is tough to quantify. This is an important parameter from a thermal perspective and used in modeling as an equivalent parameter to % CPU usage, and thus the workload in a server. In a server simulator utilized in this study, the power dissipated can be varied discretely between 0 and 5,750 kW and can be selected by the user.
2. *Fan speeds*: As explained before, the AILM approach considers the same fan speed of the server for different power dissipation, as an inherent assumption. As most new servers have fan speeds varying with % CPU consumption, flow conditions inside the

Data center changes even for constant CRAC velocity. The server simulator allows the fan speeds to vary, using a turning knob and can be altered by the user. This also helps in better achieving steady state, as real servers tend to display unsteadiness due to change in power consumption. The knob is a freely rotating one and has a calibration from 1-10. For the currents set of experiments, the knob is maintained at the same calibration at all times.

The details of the APC server simulator, including the fan curves and the discrete heat dissipation values achievable can be found in [44].

### 2.7.2. Temperature Measurements

For temperature measurements throughout the experimental setup, copper-constantan Type T thermocouples were used. The wire thickness was 0.127 mm. The uncertainty estimate in temperature was  $\pm 0.3$  °C based on calibration in and over the range 10 °C to 90 °C.

One end of the thermocouple wire junction is fixed near the point of interest. The other end is wired to National Instruments FieldPoint 8-Channel Thermocouple Input Modules, model FP-TC-120. The connection is shown in Fig. 2.21. These modules were attached to FieldPoint Network Interface Model FP-1601. For the current case, 9 units of FP-TC-120 were attached to one FP-1601 model, which is also its maximum capacity. The FP-1601 also requires a power supply of 12V DC to power up the devices.

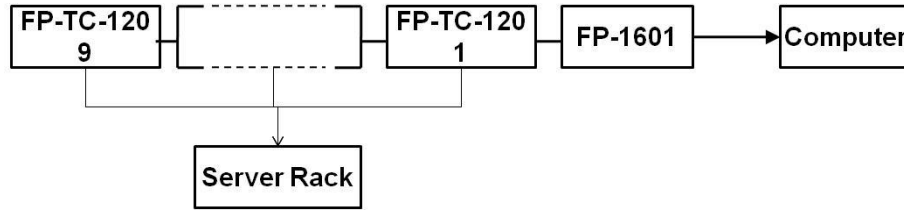


Figure 2.21 Connection of temperature measuring instruments.

The temperature of the thermocouples is a function of the voltage produced. Thus it is important to note that the bases should be stationed in a location where the temperature changes are minimal. The measurements are then transferred using an Ethernet cord to a laptop where it is linked to the LABORatory Virtual Instrumentation Experimental Workbench (LABVIEW).

### 2.7.3. Uncertainty Analysis

The uncertainties include gain and offset errors, differential and integral nonlinearity, quantization errors, noise errors, errors in linearization algorithms, and errors in the cold-junction temperature measurements. The errors in cold junction temperature arise due to the base experiencing thermal gradients. The error estimation is done in three steps:

1. Estimation of error theoretically: It has been performed previously for the thermocouples which are used in this experiment in [44]. The thermal gradient error is reported to be 0.225 °C. The cold-junction error, as listed by the manufacturer as 0.15°C, and the gain error is expected to be between 0.01 and 0.03%. Thus the total error comes out to be 0.4 °C.
2. Validation of error reported: The thermocouples were placed in a bath within a thermocouple calibrator. A Resistance Temperature Detector (RTD) is used to

measure the temperature. This device has a digital display and an error of 0.1°C. The maximum temperature difference between the actual and measured values was noted to be  $\pm 0.3$  °C, thus within the predicted range.

3. Repeatability of Experiments: This experiment is explained in the results section.

#### 2.7.4. Anemometer

A hand held, battery operated TSI VelociCalc 8350 a constant temperature thermal Anemometer [45] is utilized. It has a velocity sensor and a temperature compensation sensor. The velocity sensor is heated using internal electronics, while the temperature sensor reads the ambient air temperature. This is done to maintain the velocity sensor at a constant “overheat”. These two sensors are the two branches of a Wheatstone bridge and are arranged such that the voltage generated is related to the air velocity. The velocity given by the anemometer is for standard conditions and is converted to the actual velocity. The uncertainty in anemometer measurement is mentioned later.

#### 2.7.5. Control Volume

A representative diagram for the control volume setup is shown in Fig. 2.23. The control volume’s aim is to isolate a server simulator from the entire data center. The CRAC supplying cool air to this server is also separated. There are two regions in this control volume divided by the floor tiles.

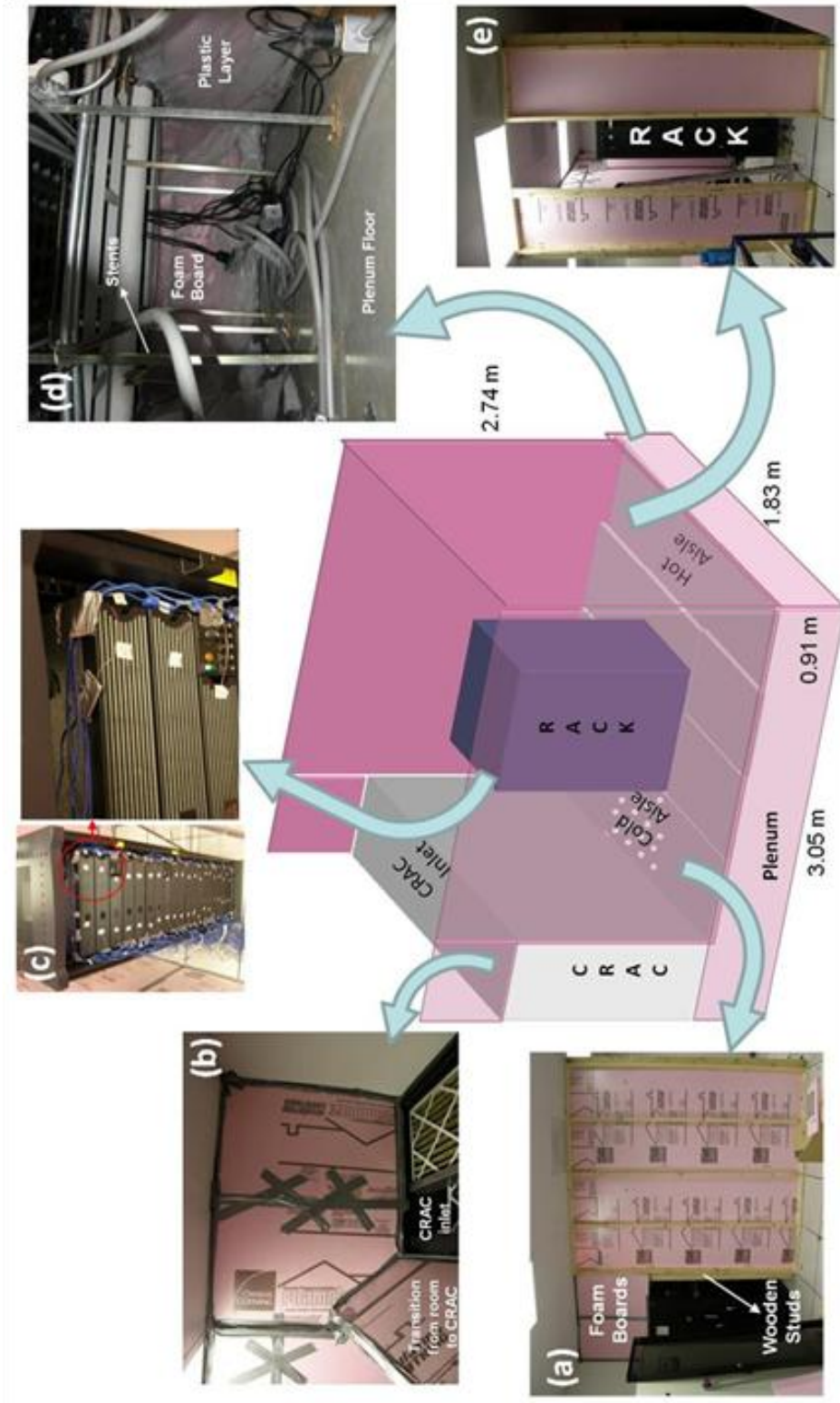


Figure 2.22 The control volume for experiments (a) Elevation, (b) Near CRAC inlet (Rack on left) (c) the experimental rack (d) Underfloor plenum (e) collapsible back.

*Above Floor:* The major dimensions of the room are shown in Fig. 2.22. The floor tiles are squares with each side being 0.6 m. The available area in the room above floor is 3 floor tiles wide (1.83 m) and 4 floor tiles long (2.4 m). The separating boards/room walls are shown in Fig. 2.22 (a). First wooden studs are cut to the height of the room (2.74 m) and are cross nailed with perpendicular studs. This makes the frame ready for each of the sides. The foam boards are fixed over them using broad head nails. The back side of the room is made collapsible (Fig. 2.22 (e)). For this, three separate faces are made and can be removed by hammering them out of their current place. This has been done to allow mobility in the room.

There is another plane created on the CRAC end of the room of height equivalent to a rack. This is done to reduce heat loss due to conduction from the CRAC front face, as it is metallic. Also, this plane is placed at a distance of 0.5 feet from the CRAC surface (Fig. 2.22 (b)). This is to create enough space to place the temperature measurement base units (). The remaining part of the setup is the transition from this plane to the CRAC unit and sealing of the CRAC unit from all sides (Fig. 2.22(e)). The transition is created by placing a foam board in an inclined position with ends being on the above mentioned plane and CRAC inlet.

Two of the sides, above the CRAC unit, are sealed using foam boards without wooden studs (one side shown in Fig. 2.22(e)). They are reinforced with cardboard and pieces of wood. This is done because once the CRAC unit operates, it creates negative gauge pressures leading to sucking in of the boards. An initial failed experiment led to the conclusion that these boards should be reinforced. The other two sides of the CRAC units are walls. For sealing purposes, the boundaries between CRAC units and walls are

lined with duct tapes. To prevent leakage of air from the control volume, the setup is sealed with duct tape at all locations.

*Plenum:* The plenum of the current data center contains internal pipes, power chords, stents and chord benches. Thus a foam board across the entire height was not a plausible solution. Till 0.3 m from the plenum floor, a foam board is erected. Above that, pieces of foam board are fit across the pipes and are joined using duct tapes (Fig. 2.22(d)). An additional sealing is provided here using plastic sheeting as the chances of leakage from the plenum were very high due to divided foam boards.

#### 2.7.6. Placements of measurement devices:

The sensors include thermocouple and anemometer. The placement of the thermocouples has 3 stages:

1. *Fixing the thermocouples in the right locations:* The placement of thermocouples on the server simulator rack is shown in Fig. 2.22(c). Each server has 5 grills (see Fig. 2.23(a)) and each grill has two thermocouples on it. While this gives a coarse picture of the thermal map, the idea in this experiment is to find the point of maximum temperature in each server. It has been found computationally that maximum temperatures occurs closer to the edges and thus the thermocouples are arranged in the way its shown in Figs. 2.22(c) (actual) and 2.23(a) and 2.23(b) (schematic). Four thermocouples are placed on the CRAC unit (Fig. 2.23 (b)). The two placed above are above and below the filter. The two placed in the plenum are placed very close to the anemometer to cause minimum obstruction to flow.

2. *Placement of the Base units:* As explained in the previous section, the base units have to be placed in a region of low thermal gradients. Thus they are placed outside the control volume. The placement of the nine FP-TC-120s and the FP-1601 is shown in Fig. 2.25. The region shown in the figure is between the wall of the control volume and CRAC's front face. An ethernet cable is run from the FP-1601 to a laptop on the other side of the control volume.
3. *Connecting the base units with the thermocouples:* One end of the thermocouples is connected to the base units (Fig. 2.25) and the other is on the servers (Fig. 2.22 (c)). A hole is made in the control volume, in front of the base units to facilitate the above mentioned connection. This hole is sealed using dampers. The wires are shown in Fig. 2.30(b). Fig. 2.30(a) shows a comparison of a similar facility with networking wires occupying a lot of space. In the current setup, no effort has been made to remove the effect of wires because firstly, real conditions are attempted to be simulated. Secondly, if the wire placement is not changed, the AILM approach is independent of them and thus portrays the robustness of the algorithm. The repeatability of the experiment is however dependent on the wire arrangement as it affects the flow inside the control volume. Thus for these set of experiments, the wire connections are not changed to maintain a fixed flow condition.

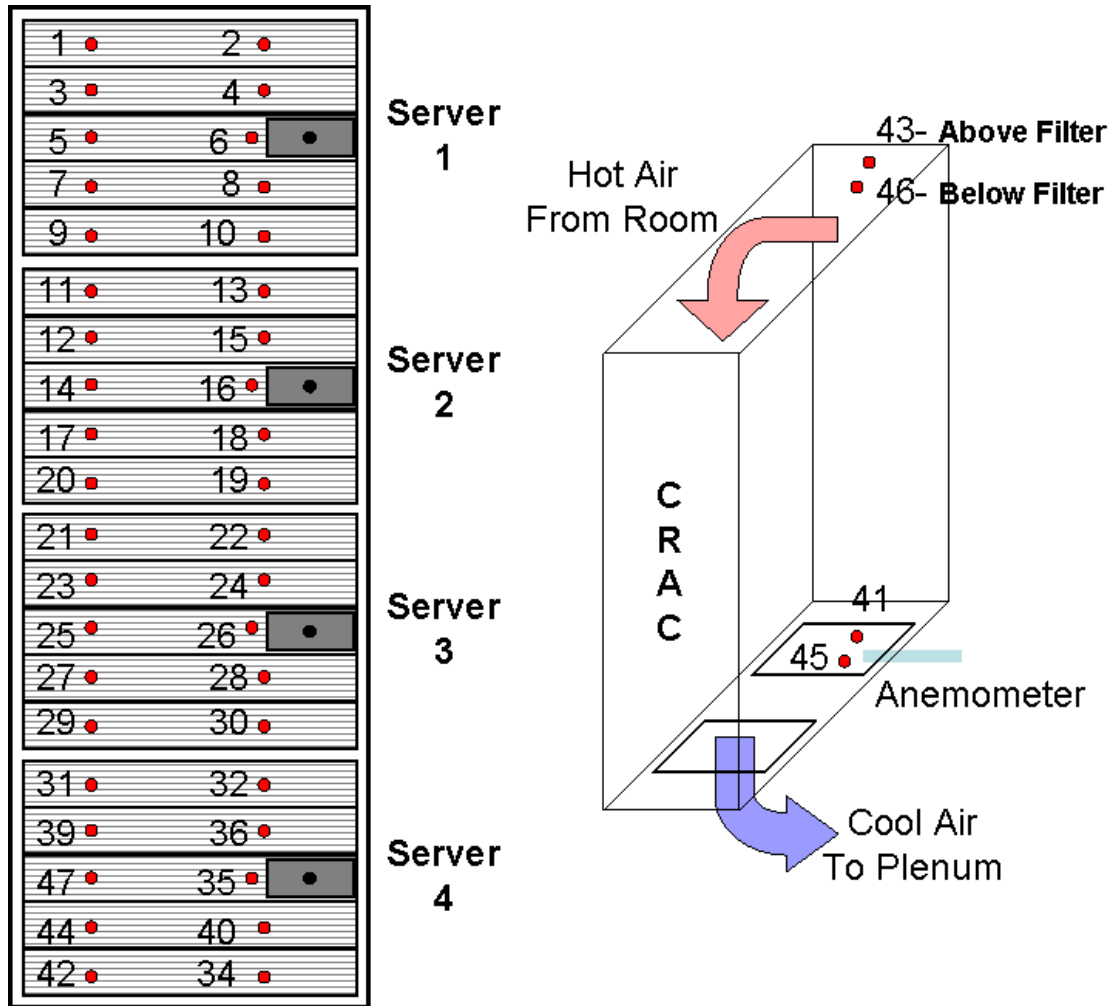


Figure 2.23 (a) Schematic of server simulator rack with thermocouple placement, (b) Schematic of CRAC with thermocouple and anemometer placement

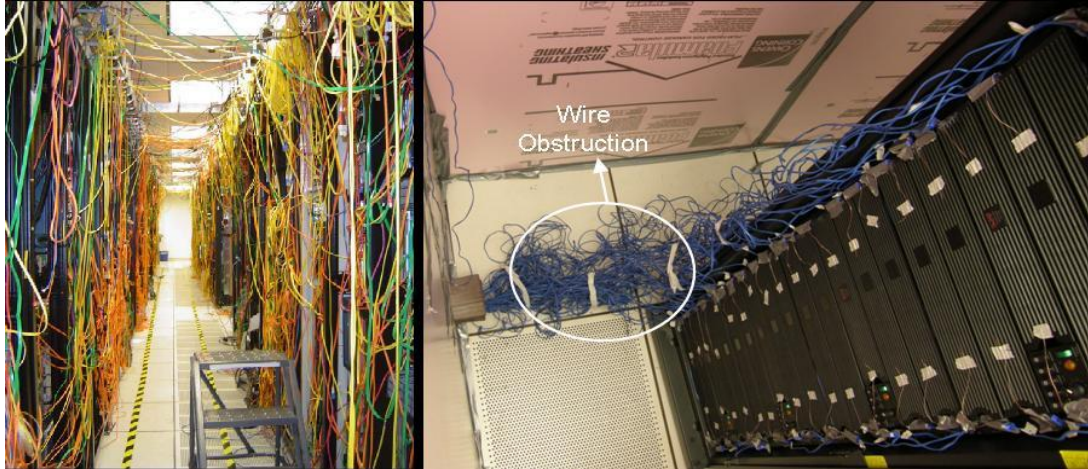


Figure 2.24. (a) Wires hanging in a facility [9], (b) wire obstruction in the current control volume

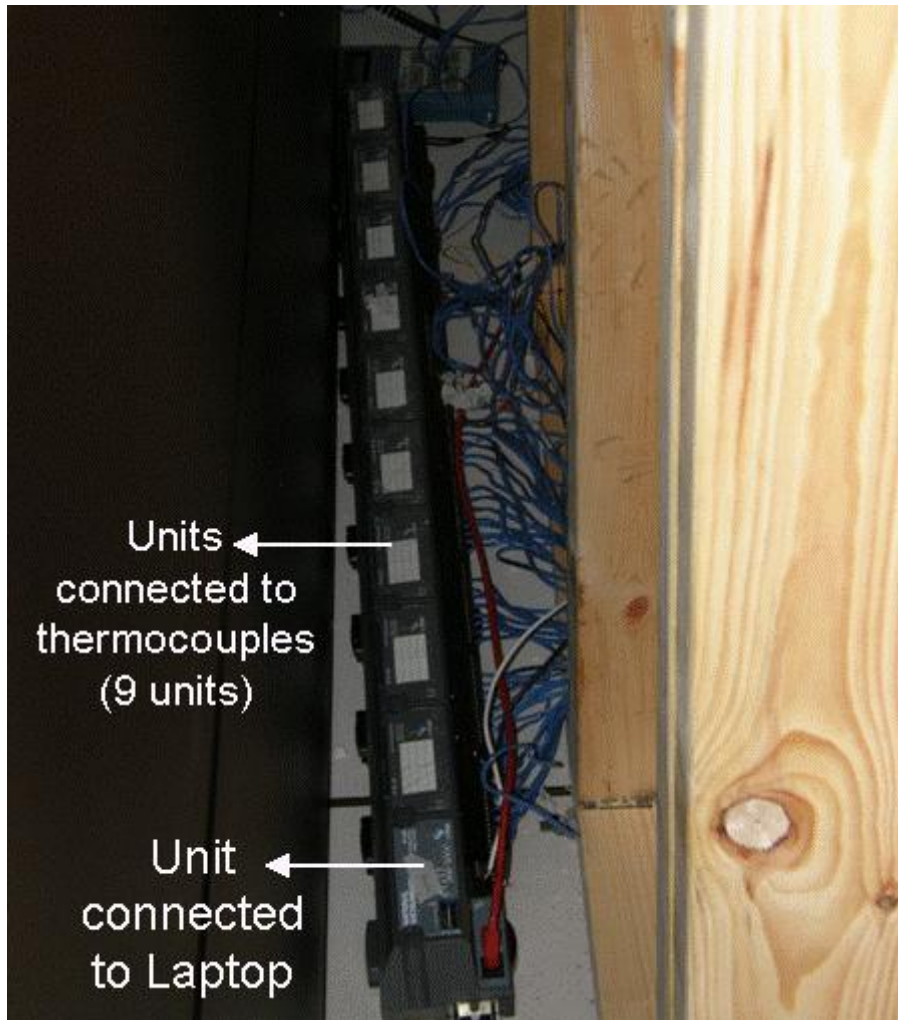


Figure 2.25. Thermocouple base units outside control volume

The placement of anemometer is shown in Fig. 2.23(b). Since this is a handheld anemometer, the handset of this device is kept over floor. The actual device affixed in the plenum, encounters direct flow coming out of the centrifugal fans of the CRAC unit.

## **2.8. Experimental Results and Discussions**

### **2.8.1. Conditions simulated**

The only parameter altered in the study was the power dissipation values of the server simulators. The fan is fixed at set point 10 of the 1-10 calibration. The actual value of the flow rate is not relevant to the study, as the only requirement is the server fan velocity to be constant for each simulation. The set point of the Liebert CRAC unit is set to 72°C and 45% relative humidity. Another requirement of this experiment is to keep  $V_{\text{CRAC}}$  constant. As the access to the Variable Frequency Drive (VFD) was not possible, the CRAC operation is set to a particular algorithm called “Intelligent Control”. The velocity is measured by the anemometer and the average is calculated. It is noticed that the velocity varies from 2.03 m/s to 3.30 m/s for all the experiments conducted. The average velocity over a time period of 30 min (duration of the experiment) varied between 2.35 – 2.42 m/s. Thus the average  $V_{\text{CRAC}}$ , while not constant at a particular value, had an error of  $\pm 3\%$ . This includes the variation due to operational conditions and inherent errors.

### 2.8.2. Repeatability

To check the repeatability of the experiments, the CRAC and the server fans are set to the above explained set points. The server is run with the heat loads as given in Table 2.3:

Table 2.4. Server heat load map for repeatability experiments

<b>Server Number</b>	<b>Heat Load (W)</b>
1	2000
2	2500
3	2000
4	2750

This experiment is run two times for 30 min. with different initial conditions, by running different experiments in between. For this heat load, the maximum variation in temperature for all the thermocouples used was found to be 0.41°C and the average variation was 0.25°C.

### 2.8.3. Setting up AILM – Calibration of the Data Center.

#### 1. *Running the “Base Case”:*

The base case, as explained before in AILM section, simulates the lowest load. For this, the servers are switched on to generate 1,000 W each of heat load. The experiment is run for 30 min. The plot shown in Fig. 2.26 also shows that within this time span, the control volume reaches a near steady state. The CRAC outlet

temperature oscillates initially due to the water chiller response. The temperatures at the server inlet are recorded.

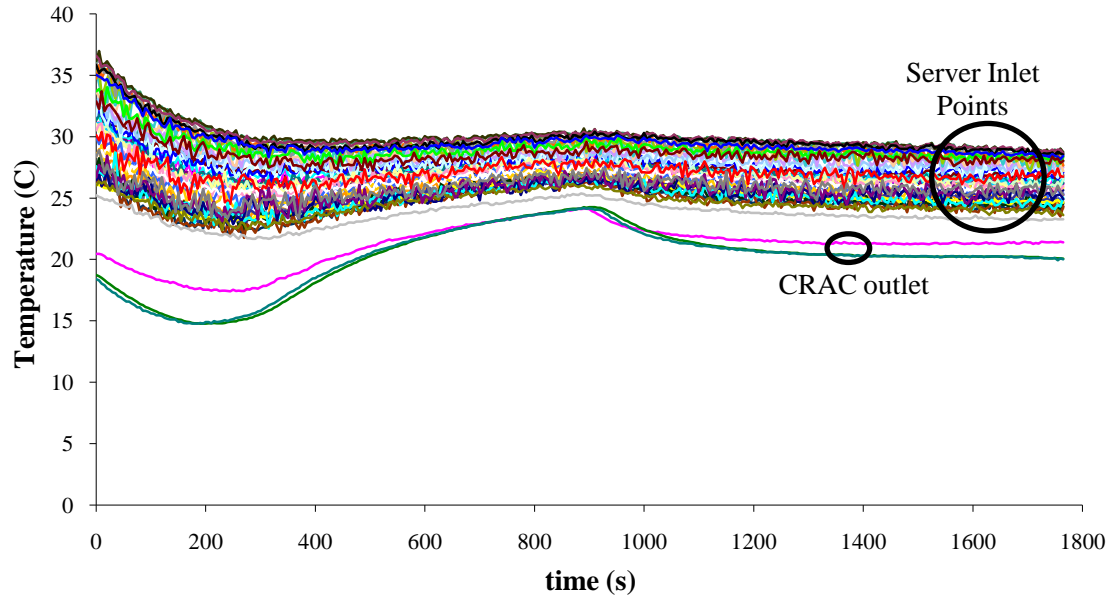


Figure 2.26. Temperature stability plot for experiments for a 30 min time interval

2. *Calibration of each server:*

Now an experiment is run with server 1 dissipating 2,000 W load, while the rest dissipating 1,000 W load. The temperature values are recorded. Sequentially, three other experiments are run with each of these servers dissipating 2,000 W of load while the others dissipating 1,000 W

3. *Determination of the “maximum temperature” points:* The AILM is based on the linearity of the energy equation. To exploit this feature, a location has to be chosen at the server inlet which would be the representative maximum temperature point of the server inlet and thus leading to a conservative estimation of maximum load. Based on the 5 experiments conducted before, thermocouples 10, 19, 28 and 32 are the

maximum temperature points for servers 1, 2, 3 and 4 respectively. It is interesting to note that these points are at the right-bottom of each server, except server 4. This is because, on the left side of the rack wires are present which prevent the recirculation from that side. For server 4, it is not at the bottom as the incoming cold air from the rack is not pushed enough from its path for the first server height.

4. *Making the AILM matrix:* Based on the procedure explained in the “Formulation” section, the AILM matrix is created. It is given in following table:

Table 2.5. The "A" matrix of AILM

		2000 W in server i.			
		server 1	server 2	server 3	server 4
$\Delta T$ server inlet	server 1	0.965	1.774	0.903	1.401
	server 2	1.058	1.183	0.84	0.342
	server 3	0.902	1.058	1.525	0.405
	server 4	0.779	0.872	0.872	0.28

#### 2.8.4. Case Study 5: $T_C = 32.2^\circ\text{C}$

**Experimental AILM (E-AILM):** The maximum temperature limit is set to  $32.2^\circ\text{C}$ . The optimization algorithm is run in MATLAB to obtain the following server heat load matrix:

Table 2.6. Server heat load map for AILM, Case Study 5

Server Number	Heat Load (W)
1	5750
2	0
3	2000
4	2750

Thus the maximum heat load this control volume can dissipate is 10.5 kW with server 1 operating at maximum heat load. The results can be intuitively compared with the AILM matrix drawn before. Change in the heat load of server 2 to 2000 W in Table 2.4 resulted in maximum change in other servers' inlet temperature. Thus shutting it down, other servers can be taken to higher capacities. This heat load distribution is tested and the steady state temperatures determined.

**Computational AILM (C-AILM):** With the heat loads of Table 2.5 and the AILM matrix of Table 2.4, the expected temperature values at the inlet of the servers can be calculated. This temperature distribution is labeled as C-AILM and will be used to compare between predicted and actual results.

**Experimental Uniform (E-Uniform):** To compare this with a standard load allocation case, the heat loads are uniformly distributed between the various servers. Since only discrete values are available for heat loads, the following is the case simulated:

Table 2.7. Server heat load map for "Uniform", Case study 5

Server Number	Heat Load (W)
1	2750
2	2750
3	2500
4	2500

Computational Uniform (C-Uniform): With the load distribution of Table 2.6 and the AILM matrix of Table 2.4, the predicted temperature distribution for the uniform case is determined.

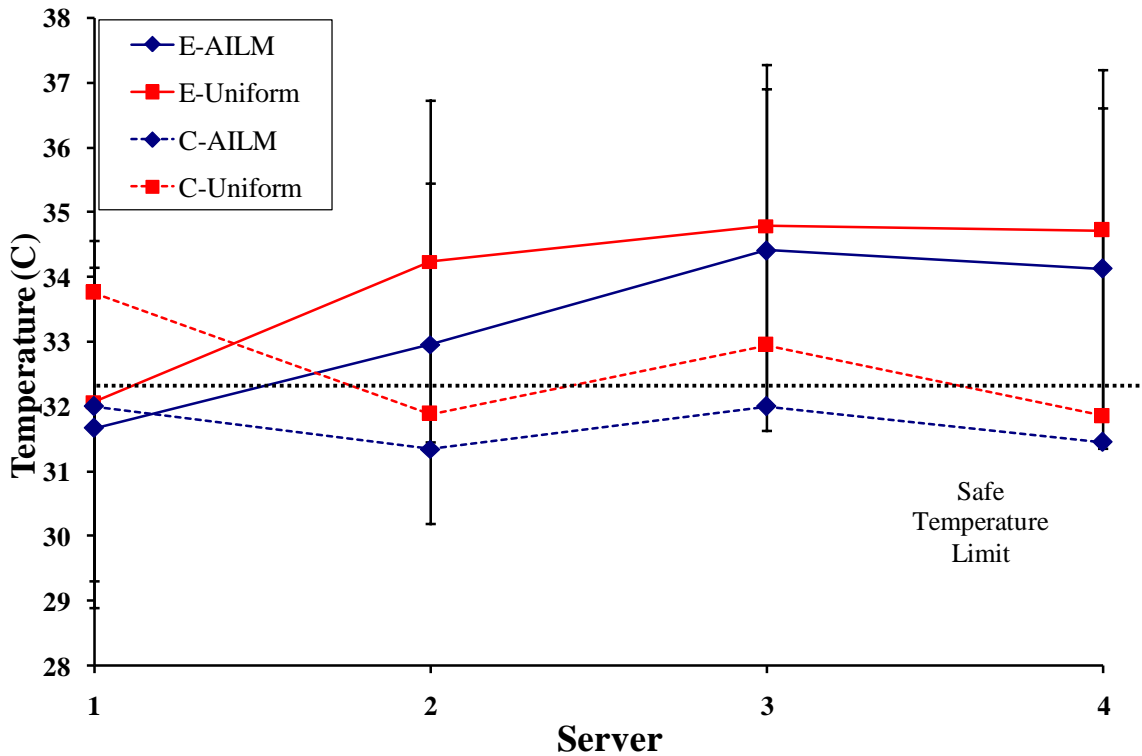


Figure 2.27. Comparison of temperature with server inlet for experimental and computational values for Case Study 5

Fig. 2.27 gives the comparison between the above cases. Following are the observations based on the chart:

1. For given net heat load, the AILM temperatures are lower than the Uniform temperatures for both the experimental and computational phases.
2. Both the E-AILM and the E-Uniform cases exceed the safe temperature limit for servers 2, 3 and 4.
3. The C-AILM temperatures are lower than the safe temperature limits. The C-Uniform temperatures are higher than the limit for 2 servers.
4. The error bars are  $+2.5\text{ }^{\circ}\text{C}$  and  $-2.8\text{ }^{\circ}\text{C}$ . The components of the common uncertainty, i.e.  $\pm 2.5\text{ }^{\circ}\text{C}$ , is based on the error due to variation in velocity of  $\pm 3\%$  and the thermocouple error of  $\pm 0.3\text{ }^{\circ}\text{C}$ . The model of the control volume presented in Fig. 2.22 is simulated in a CFD/HT environment. The variation of velocity leads to an average variation of  $\pm 2.44\text{ }^{\circ}\text{C}$ . Since, the other uncertainty related to thermocouples is independent of the former, a root mean square value is chosen as the net uncertainty [46]. An additional bias of  $-0.3\text{ }^{\circ}\text{C}$  exists since there are leakage issues in the setup. Based on the temperature data inside the room and the ambient temperatures, infiltration was calculated to be around 2.5% of the mass flow rate coming out of the CRAC units. This leads to an average bias of increase in the temperature by  $-0.3\text{ }^{\circ}\text{C}$ .

Thus, it is observed that though the AILM approach is giving better results than the uniform approach, it is still giving higher temperatures than allowed. The maximum exceed temperatures is by  $2.4^{\circ}\text{C}$ . Thus, another case is tried in which the maximum allowable temperature is reduced to  $28^{\circ}\text{C}$  and similar AILM and Uniform results are generated.

2.8.5. Case Study 6:  $T_C = 28^\circ\text{C}$

**E-AILM:** The maximum temperature limit is set to  $28^\circ\text{C}$ . The optimization algorithm is run again and the results are presented in Table 2.7.

Table 2.8. Server heat load map for AILM, Case Study 6

Server Number	Heat Load (W)
1	2750
2	0
3	1000
4	2500

Thus the maximum load the control volume can take is 6250 W. These conditions are run in the experiment.

**E-Uniform:** For net load of 6250 W and the given discrete setting allowed, following is the load vector used for experiments:

Table 2.9. Server heat load map for Uniform, Case Study 6

Server Number	Heat Load (W)
1	1750
2	1500
3	1500
4	1500

The results are presented in Fig. 2.28:

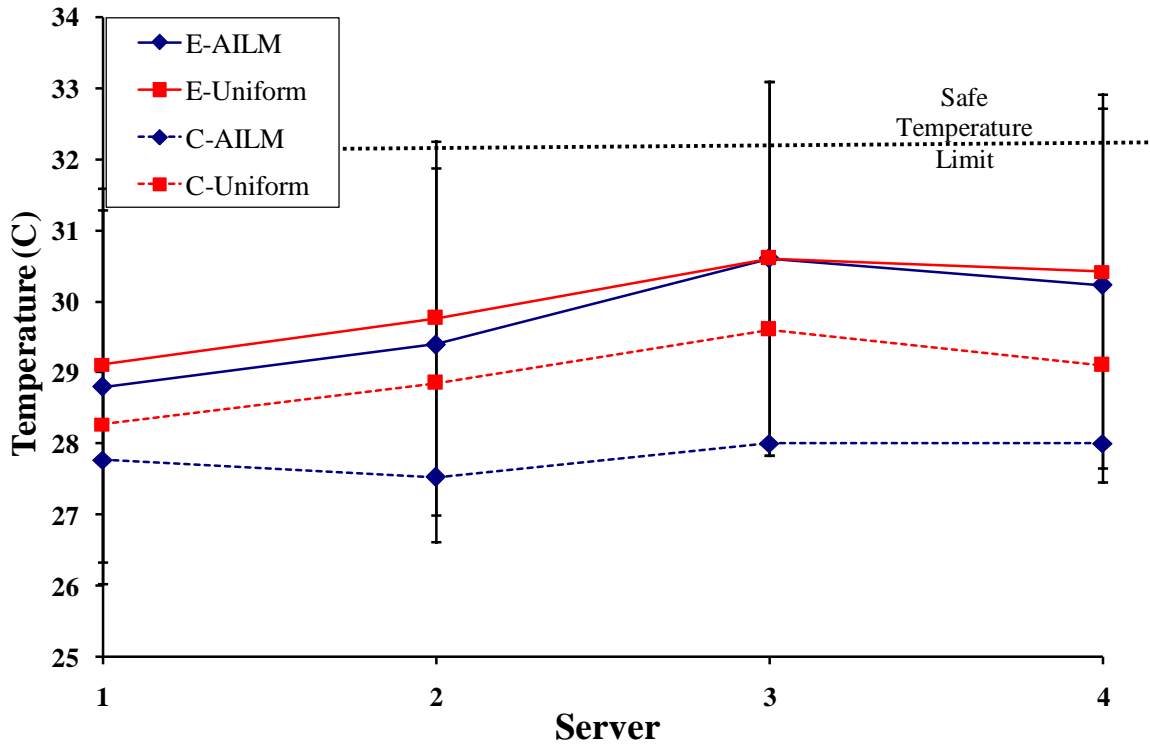


Figure 2.28. Comparison of temperature with server inlet for experimental and computational values for Case Study 6

Following are the observations:

1. The E-AILM temperatures are below the actual 32.2°C mark. Since the E-Uniform temperatures are also below that mark, the benefits obtained are not substantial.
2. The AILM temperatures are again lower than Uniform temperatures for both experimental and computational cases.

### **3. SCALABLE PODS BASED CABINET ARRANGEMENT AND AIR DELIVERY**

#### **3.1. Introduction**

The Scalable-Pod (S-Pod) arrangement [47] refers to the layout of a heat producing space in which the heat generating basic elements are arranged in groups of near circular forms - pods, such that the cooling agent is supplied to the internal space in between the heating elements and is exhausted outside the pod. This is subsequently collected and routed back to a cooling agent conditioning unit which re-supplies the cooling agent back in the pod.

A model application of this scheme is in a data center, i.e. the heat producing space. Here the racks filled with servers are the heat generating units and the cooling agent is generally air. The Computer Room Air Conditioning (CRAC) unit provides cooled air, which after passing through the under or over floor plenum, is released into the room through perforated tiles (HACA layout).

The HACA layout is prone to mixing of hot return air with the cold supply air and thus reducing the cooling potential of the cold air as well as forcing some rack inlets to be devoid of direct supply of cold air altogether as explained in [7]. Also, low ceiling heights, which are common in office space type data centers, can further enhance the mixing of hot and cold aisle air [6].

Thus to overcome the above mentioned issues, the CRAC units oversupply coolant to ensure safe conditions which leads to wastage of energy. As given in a study

by [6, 15], attempts have been made to place barriers to block mixing of hot return and cold supply air, which in turn makes traversing within the data center for server maintenance/replacement etc. more difficult. These attempts along with other derived layouts of HACA layout are given in Fig. 3.1. These include CRAC outside the room (Fig. 3.1(iii)), return air plenums (Fig. 3.1(iv), 3.1(vi)), overhead supply pipes in cold aisle (Fig. 3.1(v)), barriers on cold aisle side (Fig. 3.1(vii)), return air ducts (Fig. 3.1(viii)), supply side heat exchanger (Fig. 3.1(xii)), return side heat exchanger (Fig. 3.1(xi)) and overhead heat exchangers (Fig. 3.1(ix), 3.1(x)). Overhead plenums were introduced because they have higher static pressures, leading to more uniform flow distribution. But in this case, the under-floor plenum is still required for cable networking and data/fiber distribution. Baffles and dropped ceiling were discouraged because when cooling load increases, there is a possibility of servers starving due to the lower amount of airflow provided, than the exhaust flow rate of the servers. Also, fire and safety codes prevent the usage of such configurations.

In the S-Pod layout, as presented in Fig. 3.2, the perforated tiles are surrounded by racks on all four sides i.e. a pod structure. The air coming out of the tiles is pulled into the racks with the help of fans and is exhausted outside the pod. Then it is returned to the CRAC unit for supply. The arrangement of these pods in the data center can be in-line or staggered. Fig. 3.2 presents a staggered arrangement. It also presents a few possible schemes which can be employed for server access for maintenance.

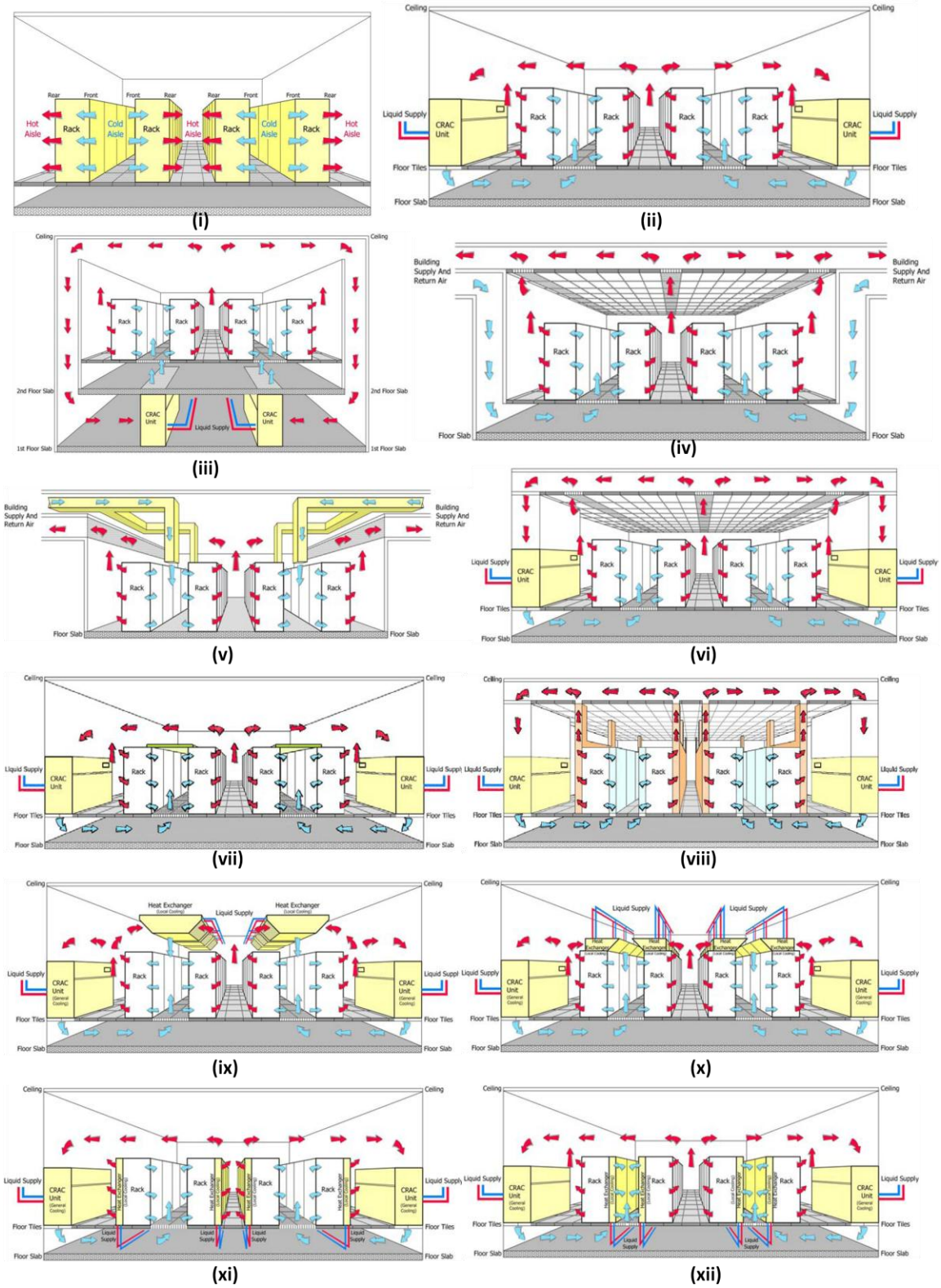


Figure 3.1 Various modifications to the Hot Aisle-Cold Aisle layout [6]

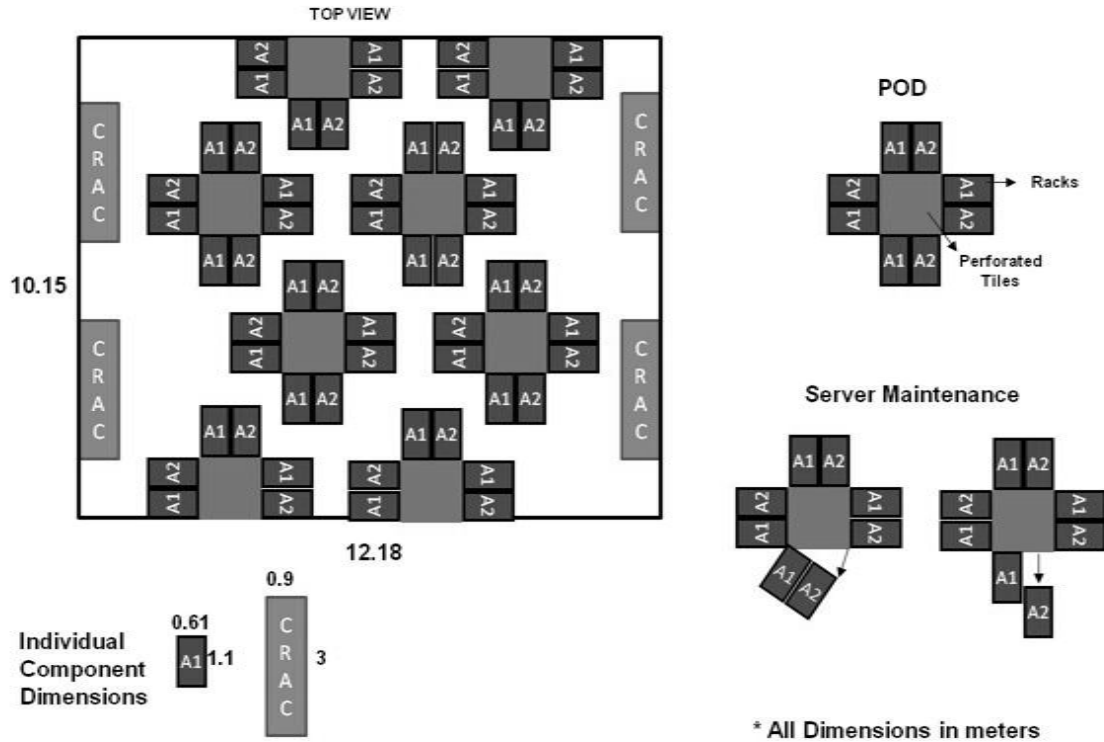


Figure 3.2. Scalable Pods (S-Pod) based layout with maintenance options

For the same floor space area, the HACA layout (Fig. 3.1) houses 44 racks while the S-Pod layout (Fig. 3.2) can accommodate 56 racks (27.3% more). The limitation in deciding the number of racks that can be housed is the minimum distance required between the CRAC and the nearest perforated tile, as below it a negative flow rate has been observed from the perforated tiles [48].

### 3.2. Computational Model

Fig. 3.3 and 3.5 shows the layouts for the two cases. The floor area has been kept the same (12.12 m × 10.2 m) for both. The height of the roof above the floor is 3 m and the plenum depth is 0.86 m. The whole facility consists of 44 racks for the HACA case and 56 racks for the S-Pod case. For the HACA case, considering the symmetric nature of

the layout, only a one-fourth model of the whole facility has been simulated, as shown in Fig. 3.3. Thus two of the faces in the reduced model are given a symmetry boundary condition.

Each rack has been modeled as a black box with a footprint of 0.61 m×1.1 m, height 2 m and a constant volumetric heat generation to simulate the heat dissipated by the chips in servers. Each rack is further compartmented into 6 sections equivalent to 6 servers to prevent flow of the coolant from one server mixing with another, inside the rack. CRAC units have dimensions of 3 m×0.9 m×2 m. The dimensions of racks and CRAC units are kept the same for both the layouts.

Gambit 2.4.6 has been used for constructing and meshing the model. For the S-Pod case, the mesh size chosen is 0.14 m. Tet/Hybrid elements of type TGrid have been used for the mesh. It has 1.2 million cells.

### 3.2.1. Boundary conditions

Fluent 6.3.26 code has been for the simulations and Fluent/Tecplot 360 have been used for post-processing. The racks inlets have been modeled as a porous jump with 0.35 m thickness and 20% open area [37], over which the pressure change is a combination of Darcy's Law and an additional inertial loss term and varies with velocity according to the equation:

$$\Delta p = - \left( \frac{\mu}{\alpha} v + C_2 \frac{1}{2} \rho v^2 \right) \Delta m \quad (3.1)$$

The perforated tiles are governed by the same boundary condition as the rack inlets but with a different  $C_2$  value. The values of the parameters in Eq. 3.1 and other boundary conditions are chosen according to the HACA layout model in Fig. 2.1.

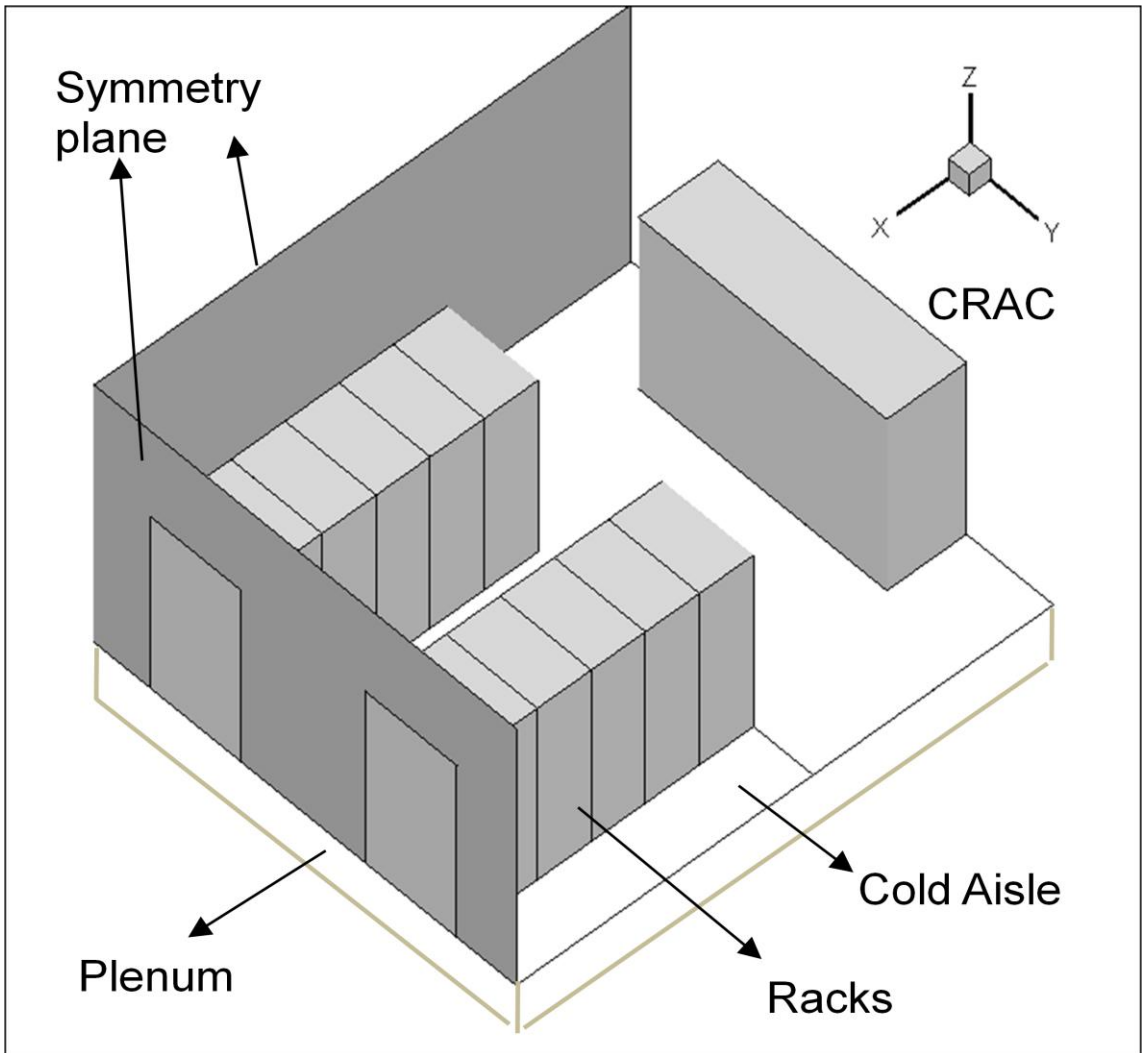


Figure 3.3. Symmetrical model simulated for the HACA layout

The face of the CRAC unit attached to the plenum is inlet to the facility and the face open to the room is outlet from the facility. The inlet conditions are 288.15 K and constant velocity (4 m/s and 7 m/s for the two studies). Since the simulations are run for steady state, the outlet temperature and velocity are calculated by applying energy balance between the inlet-outlet conditions and the heat dissipated by the racks. All the other surfaces are defined as walls.

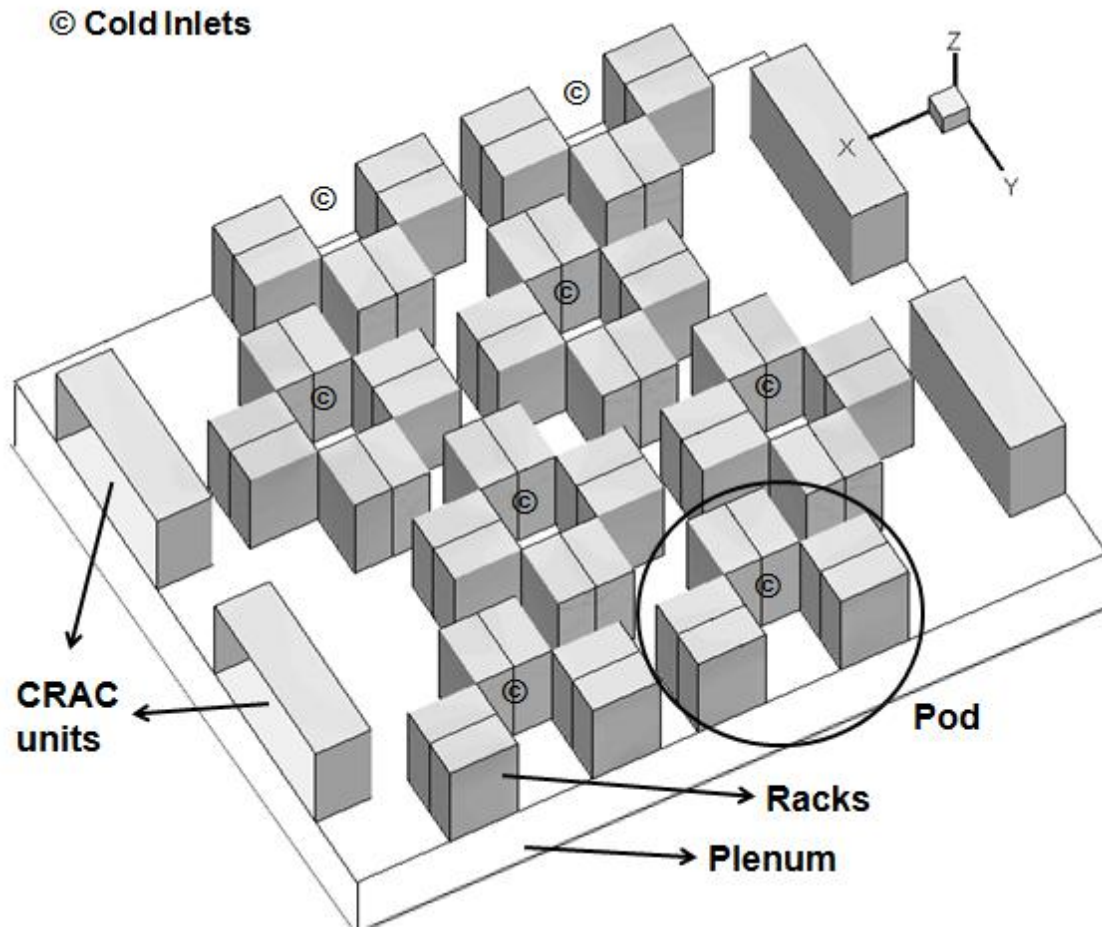


Figure 3.4. Symmetrical model simulated for the S-Pod layout

### 3.2.2. Solution and convergence

A first order upwind has been chosen for momentum discretization for initial set of iterations and is then switched to a second order upwind to achieve faster and accurate convergence. For all the others a second order upwind scheme has been used. Standard  $k-\epsilon$  turbulence model has been used to model viscous flow.

Density has been assumed to be constant because the effect of natural convection is negligible [36]. Accounting for natural convection, using the Boussinesq

approximation leads to a maximum error of less than 2% in the temperature difference between the fan outlets and CRAC supply.

### 3.3. Results and Discussions

#### 3.3.1. Case 1: $V_{CRAC} = 7$ m/s, $q'''$ (for each rack) = 16161.61 W/m<sup>3</sup>

Fig. 3.5 shows the rack inlet temperature contours for the two layouts for  $V_{CRAC}$  of 7 m/s. The volumetric heat generation rate for each rack is kept constant at 16161.61 W/m<sup>3</sup> (21.7 kW racks) for both the cases. For the S-Pod case, the maximum temperature at rack inlets is 302 K which is below the ASHRAE recommended limit of 305.15 K. On the other hand, for the HACA case the maximum rack inlet temperature is 321.94 K which is about 16 °C higher than the safe limit. In the HACA layout, the hot spots occur at the lower end of the rack inlets while the upper regions of the inlets are within the safe limits. This is because the servers at the bottom do not get enough flow due to high  $z$  momentum of flow coming from the perforated tiles. For the S-Pod case, the 3-column pods are well below the safe limits. In general, the 4-column pods have higher inlet temperatures than the 3-column Pods. This trend is observed because the mass flow rate received by both type of Pods is almost the same and varies between 11.569-11.581 kg/s (0.1% variation). Thus the same mass flow rate serves 8 racks in the 4-column pod, while it serves 6 racks in the 3-column pods. Maintaining the upper limit of allowed temperature for inlets, the maximum allowed HACA case uniform heat generation limit is 15 kW per rack, while for the S-Pod case it can be increased to 22 kW. Thus, the S-Pod case can take up to 53% higher heat load per rack for this velocity and since it has 27.3% more number of racks, the net heat load capacity of the data center with S-Pod layout increases by 95%.

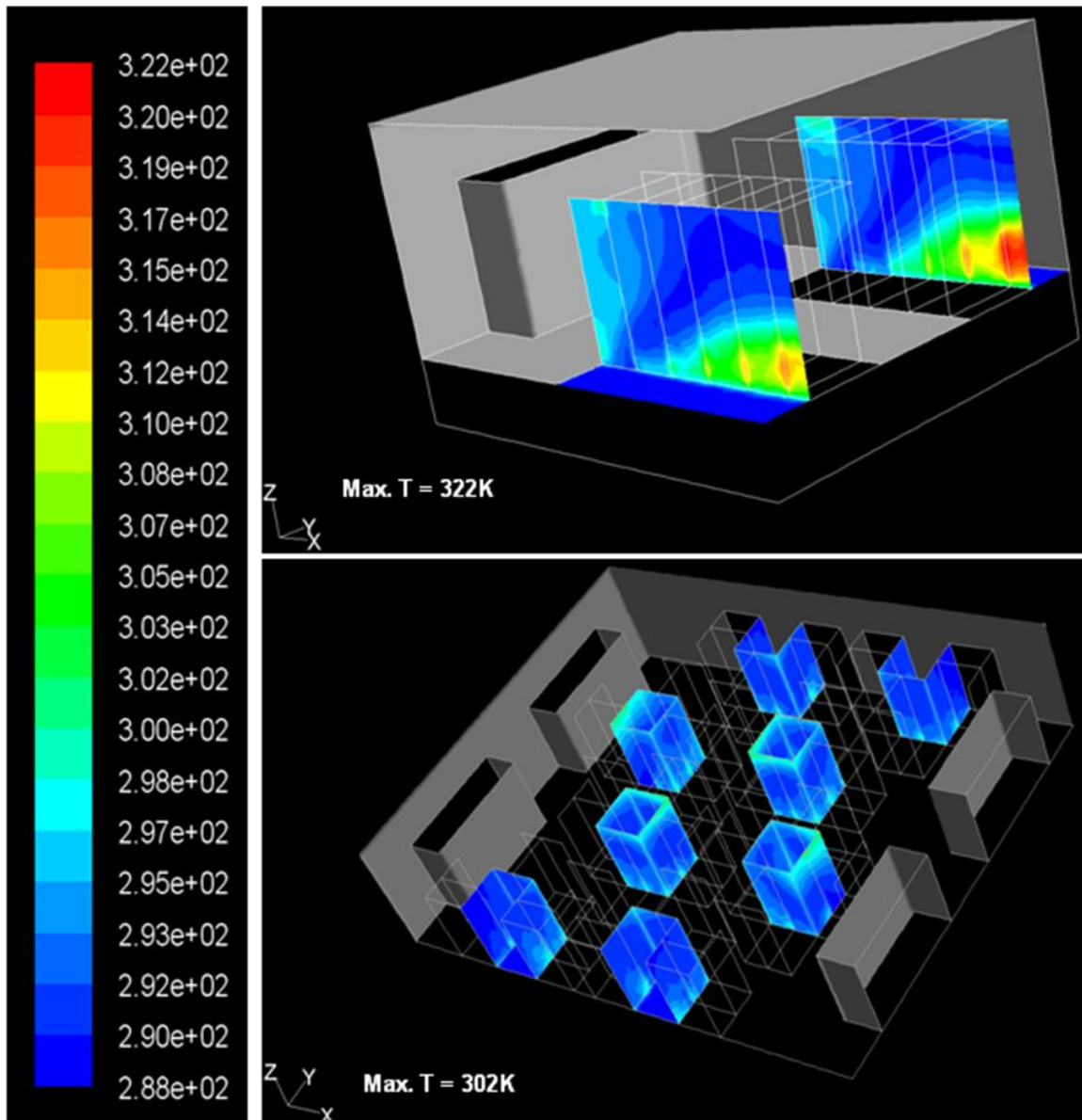


Figure 3.5. Rack inlet temperature contours for HACA and S-Pod layouts,  $V_{CRAC} = 7 \text{ m/s}$

3.3.2. Case 2:  $V_{CRAC} = 4 \text{ m/s}$ ,  $q'''$  (for each rack) =  $8888.89 \text{ W/m}^3$

Fig. 3.6 shows the results for  $V_{CRAC} = 4 \text{ m/s}$ . Here the rack heat generation rate is  $8888.89 \text{ W/m}^3$  (11.93 kW racks). In the HACA layout, it is noticed that the flow coming out of the perforated tiles lacks the momentum to reach the topmost levels of the rack. Thus, these regions get re-circulated air from the hot aisle. In the S-Pod case, the flow

rate remains almost same for all the Pods as in the previous case and varies between 6.610-6.61 kg/s (0.1%). The pods closer to the CRAC units have the highest inlet temperatures. This can be explained by the fact that the airflow after coming out of the CRAC outlets hits the floor and is forced towards these tiles, thus having a directional nature. This directional nature leads to non-uniform mass flow rate through the tiles, more specifically, higher for the racks away from the CRAC units, thus avoiding the racks closer to CRAC units to get sufficient mass flow rate of cool air at inlet. Thus, to fulfill the requirements of the mass flow rate, hot air present in the room is sucked from the top end of the pods. Recirculation is much more prominent for 4 m/s especially at the racks which are closest to the CRAC units.

For 4 m/s case, it is shown in Fig. 3.7. The streamline shows flow coming from the plenum through the perforated tiles inside the pod. It is then sucked in by one rack inlet. The flow coming out of this rack's fan is pulled from the top to satisfy its, as well as its neighboring racks' mass flow rate needs, as pointed out by the two circles. The maximum rack inlet air temperature for S-Pod case is 307.89 K. The maximum inlet air temperature for the HACA case has gone up to 311.66 K. Maintaining the upper limit of allowed temperature for inlets, the maximum allowed HACA case uniform heat generation limit is 8.6 kW per rack, while for the S-Pod case it can go up to 11 kW. Thus, S-Pod has 27.8% more heat dissipation capacity per rack and it has 27.3% more racks. Thus the net heat dissipation capacity for the data center in S-Pod case increases to around 63.6%.

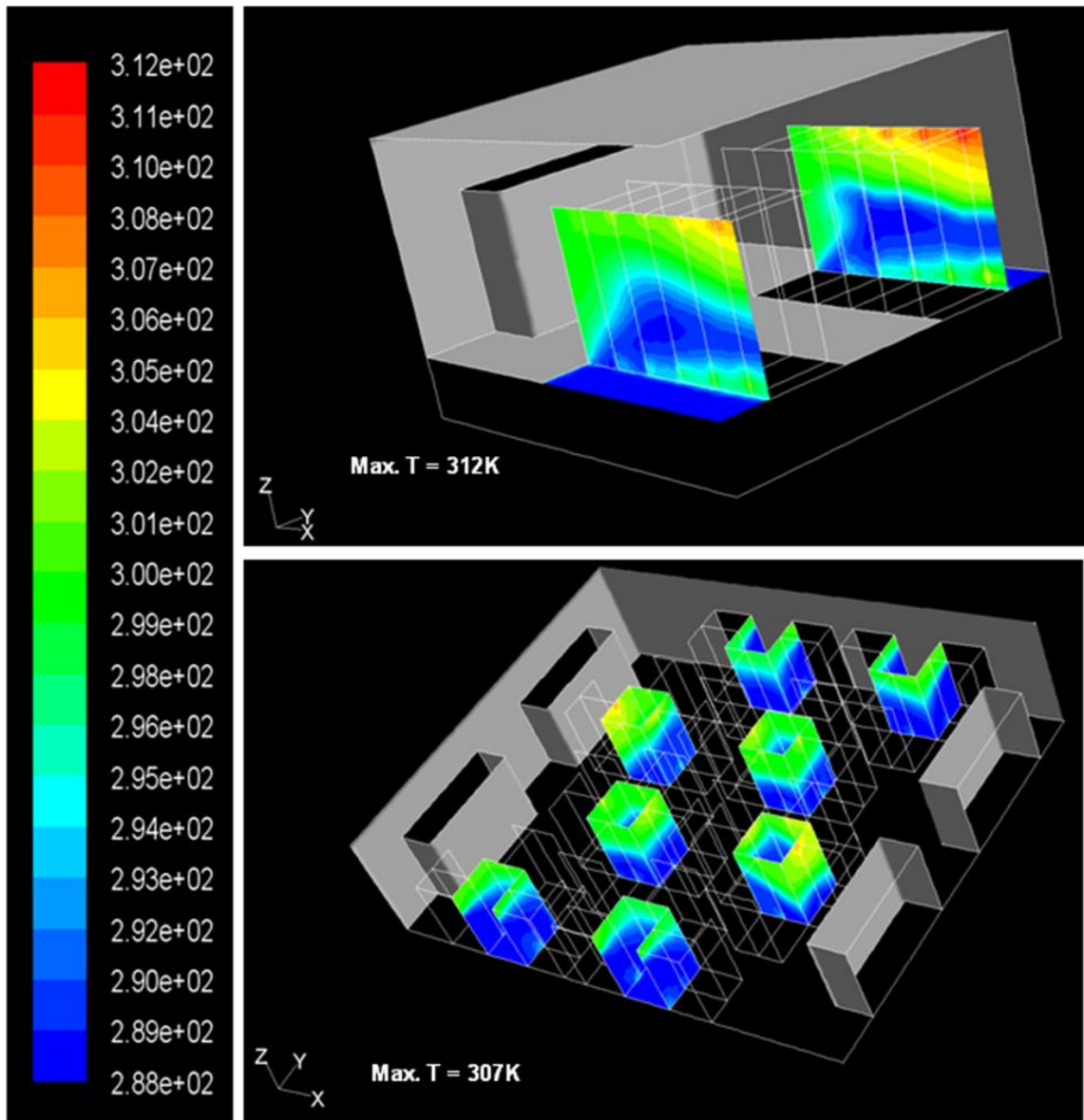


Figure 3.6. Rack inlet temperature contours for HACA and S-Pod layouts,  $V_{CRAC} = 4 \text{ m/s}$

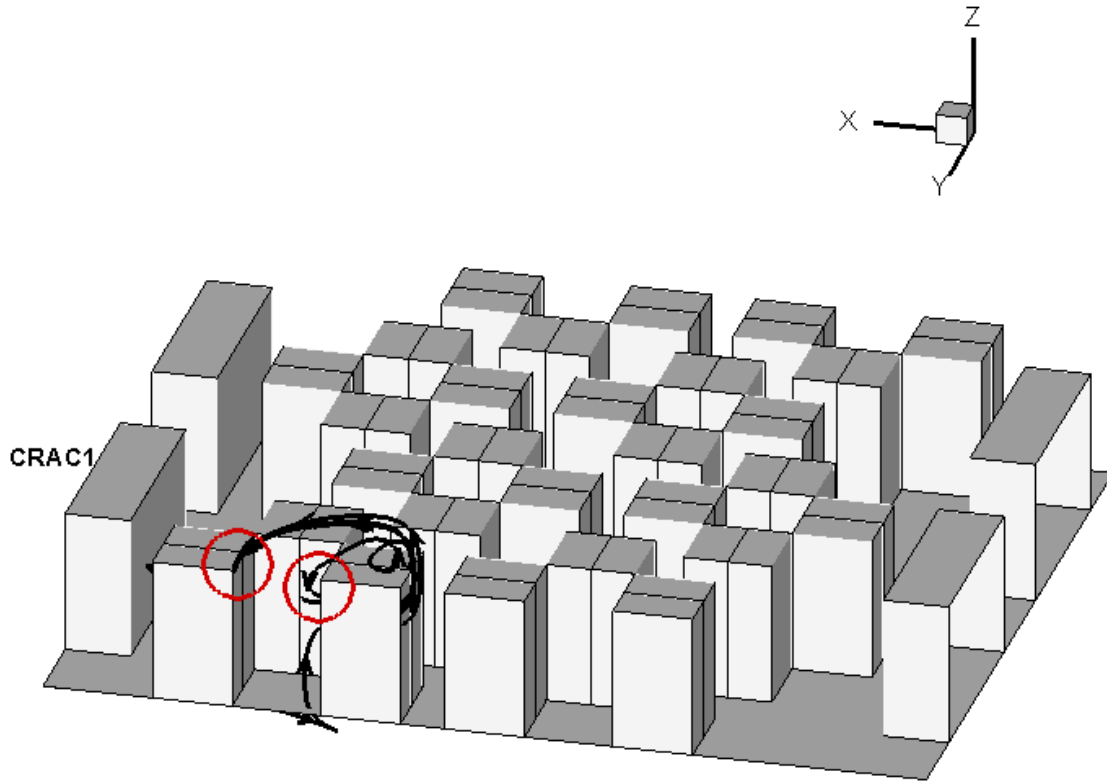


Figure 3.7. Recirculation patterns in S-Pod layout

In Fig. 3.8, the differences between the various maximum rack inlet temperatures and the CRAC outlet temperature, which is fixed at 288.15 K are displayed. This has been done for both HACA and S-Pod layout for  $V_{CRAC} = 4$  and 7 m/s. This temperature difference gives the loss of cooling potential for each rack, and is a measure of how much recirculation has adulterated the inlet flow to racks. The plot shows values for the 44 hottest racks in the S-Pod layout and the 44 racks of the HACA layout (the 12 racks simulated have been re-plotted based on symmetry). There are 2 important trends noticed:

1. Most of the HACA points are higher than the corresponding velocity hottest racks of the S-Pod case. This indicates that there is lower recirculation in the S-Pod case.
2. Most of the points corresponding to 4m/s are higher than those corresponding to 7m/s. This shows that recirculation is more prevalent in the 4m/s case. While this shows that the inlet conditions lead to reduced recirculation with increased CRAC velocity, it needs to be confirmed whether the same trend continues at even higher velocities.

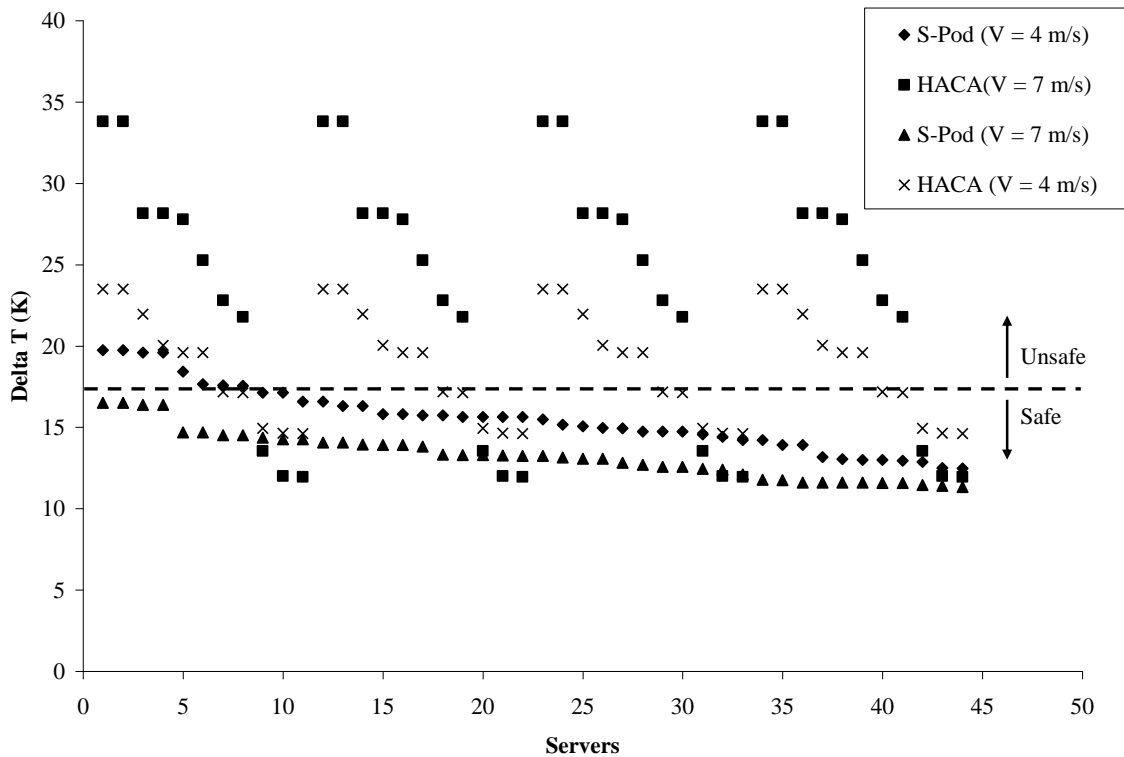


Figure 3.8. Temperature differential between rack inlet and CRAC outlet for various cases

### 3.3.3. Case 3: Optimal velocity for the S-POD case

Table 3.1: Mass Flow rates and other features for S-Pod Layout

<b>CRAC Velocity (m/s)</b>	<b>Net CRAC mass flow rate (kg/s)</b>	<b>Net mass flow rate at rack inlets (kg/s)</b>	<b>Temperature profile remarks</b>
4	52.9	86.9	Hot at top
7	92.6	86.11	Uniform
10	132	85	Hot at bottom

Table 3.1 shows the results for the simulations performed at a constant heat dissipation rate of  $16161.61 \text{ W/m}^3$  for three velocities. As per Table 3.1, the supply mass flow rate from the CRAC units for the complete facility with an S-Pod layout increases linearly with velocity. The net mass flow required by the racks on the other hand decreases very slowly with increasing velocity because of the changing pressure differences affecting the fan velocity. It also shows, that the 4 m/s case provides insufficient mass flow rate and lacks the momentum at the perforated tiles, thus leading to recirculation at top. The 10 m/s case provides 55% extra mass flow rate. However, most of this just leaves the pod from the top, called short circuiting of air, leading to wastage of cool air and thus energy. Also, because of the higher momentum of the flow coming out of the perforated tiles, the flow near the lower servers in the rack is highly directional in nature and is difficult to be sucked in. This leads to higher temperatures near the lower end. The 7 m/s case provides roughly the mass flow rate required by the racks and provides an optimal flow rate which reaches the top ends of the rack without abandoning the lower ends. Also, the average inlet temperature for 7 m/s case is lowest amongst the three velocities considered. This explains that there exists an optimal velocity for this layout. As the heat dissipation increases, the return air temperature at

CRAC inlets also increases. Since there is a cap on this temperature due to cooling limitations of the CRAC unit, one still needs to increase CRAC velocities to achieve higher heat load dissipations. While, 7 m/s case provides almost the same mass flow rate as required by the racks, and also yields the lowest maximum inlet temperature of the racks, it is not the optimal velocity in every case. To determine the optimal velocity, one has to increase from lowest velocity till an operating point which ensures maximum inlet temperature to be lower than 305 K.

#### 3.3.4. Case 4: Possible modifications in the S-POD layout.

##### *Case 4(a): Introduction of barriers*

It is noticed for lower velocity case (4 m/s) for both layouts, that since the momentum of the incoming air from the tiles is lower, it doesn't reach the top of the racks efficiently, leading to hot spots near the top of the racks. Thus there is mixing from the upper side of the racks for both HACA and S-Pod arrangements, as shown in Fig. 3.7. To avoid this in the S-Pod case, we propose that the cold air supply space above the raised floor inside the pod be closed from the top using a physical barrier (Fig. 3.9). Providing barriers at top will also prevent leakage of excess flow rate from the tiles to the hot aisle. In this case, it has to be ensured that the flow rate provided by each perforated tile is greater than or equal to the flow rate coming out of each pod to avoid creating negative pressures, ensuring a safe mass flow rate through the racks.

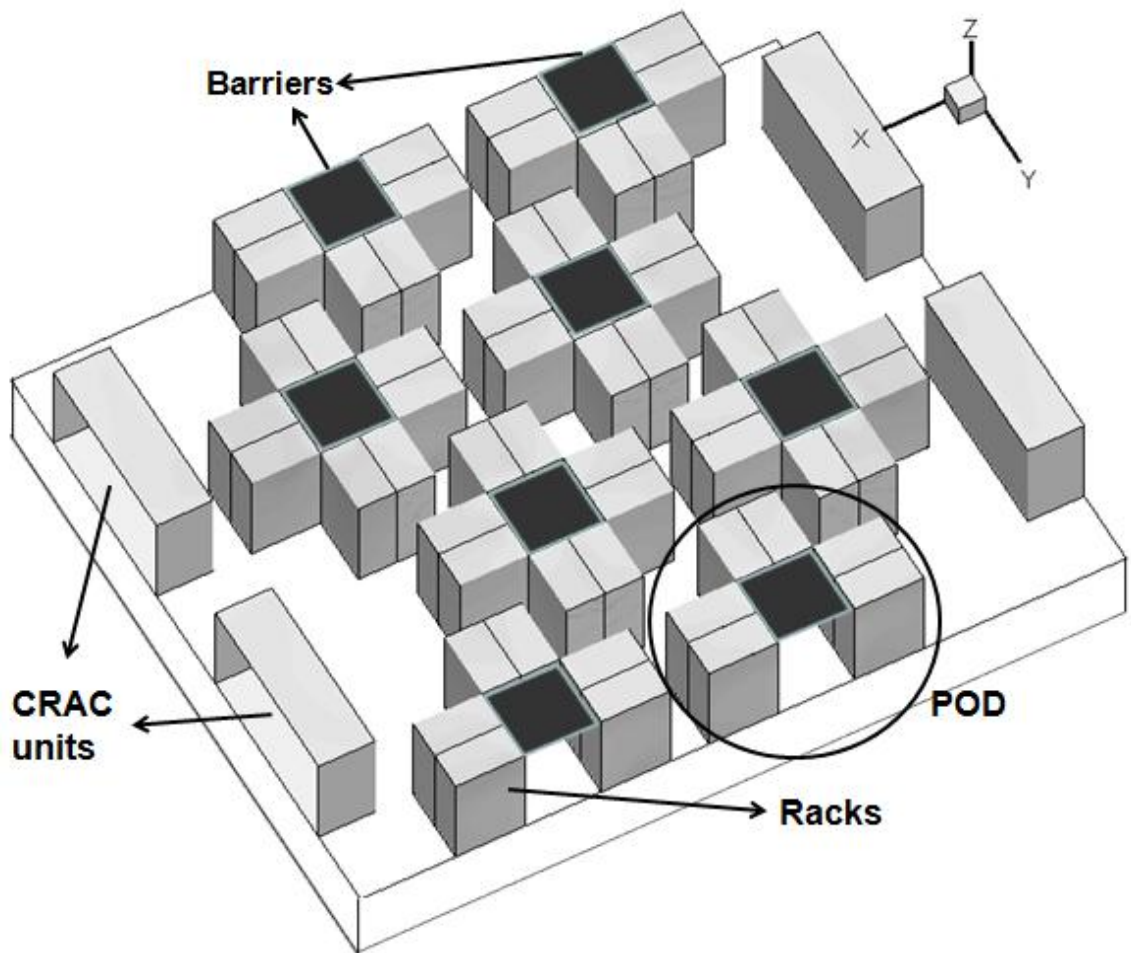


Figure 3.9. Barrier placement for S-Pod layout

*Case 4(b): Heterogeneous Data center*

There has been research conducted on Liquid Cooled Racks (LCR) [6]. While, their operating costs are lower than air cooled racks (ACR), the initial expenditure is usually higher. Thus, one could envision some of the ACRs to be replaced by these leading to a more heterogeneous environment. The biggest challenge that such legacy data centers will be facing is the optimal placement of the LCRs. One ad-hoc way of

going about it is by listing out the ACRs in decreasing order of their average inlet temperatures and then replacing the rack on the top of the list with every new LCR.

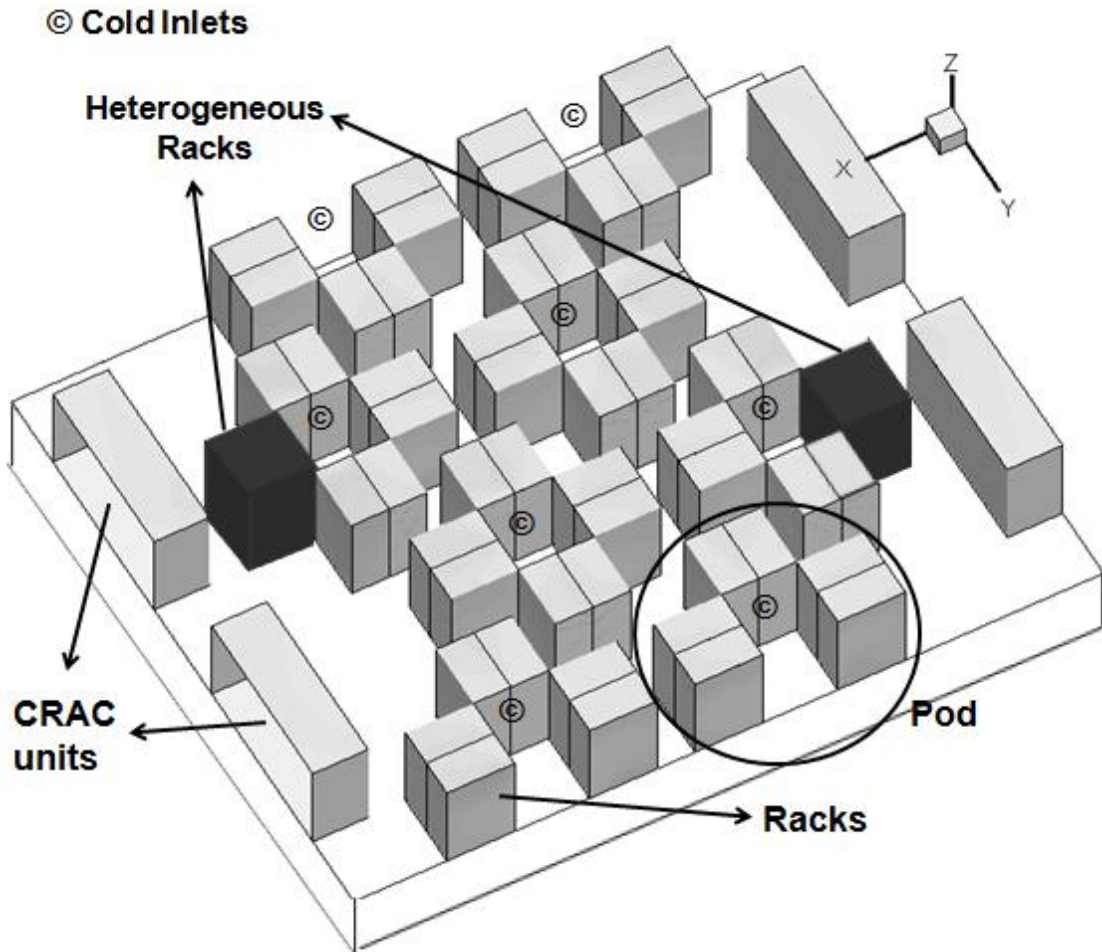


Figure 3.10. Location of Liquid Cooled racks in a primarily air-cooled datacenter

Consider the S-Pod layout. It currently has 56 ACRs. Also, assume that 8 new LCRs are brought in to replace the ACRs. For  $V_{CRAC} = 4$  m/s, had we made the above mentioned list and replaced the top 8, these would have been the 4 racks closest to the CRAC in the 2 pods closest to the CRAC as shown in Fig. 3.10 and the 4 racks adjacent to them. But a detailed CFD/HT analysis in Fig. 3.11 shows that the replacement of the

first 2 sets of racks reduces the temperature inside both of the pods and the replacement of the next set of racks would not have been ideal. Instead, it would have been wiser to replace 2 racks each in the middle pods, which will reduce the temperature of other racks in this pod.

Fig. 3.12 shows the maximum rack inlet temperatures for the heterogeneous and the baseline case. Clearly the benefit is larger in the 4 m/s case where there is about 1.4 °C reduction in the average rack inlet temperature as compared to the 7 m/s case where there is 0.61 °C change in the average rack inlet temperature. This maybe because there is higher airflow deficit and recirculation in the 4 m/s case, as compared to the 7 m/s case. So, using liquid cooling at the hottest racks not only increase the airflow available to the other racks, but also leads to reduction in the hot aisle temperatures which causes a considerable change in the rack inlet temperatures due to increased recirculation. As seen in Fig. 3.12, 8 racks above the threshold in the homogeneous case are now brought within safe limits.

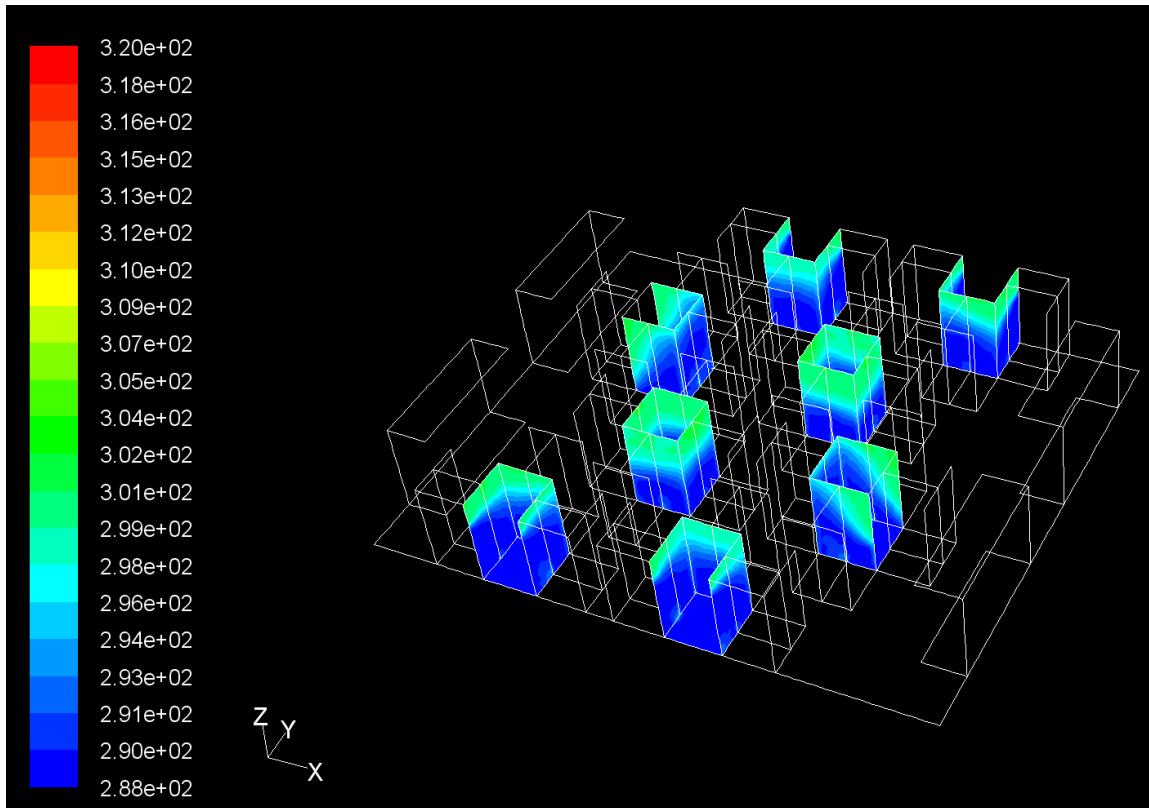


Figure 3.11. Temperature Contours and rack inlets for heterogeneous case

Based on the simulations, the following reasons are why S-Pod layout is better than the HACA layout:

1. The increase in the inlet temperatures of the racks closest to the CRAC unit in the HACA case for higher velocities such as 7 m/s (Fig. 4) indicates that side air recirculation exists. The inherent design of the S-Pod blocks the side air mixing if it is made leakage proof.
2. The number of racks per perforated tile (if same dimensions are maintained for both) is 2 for the S-Pod layout and 1 for HACA layout. Thus the pressure with which the flow comes out of the perforated tiles in S-Pod case will be higher, which will again avoid top side recirculation reaching the topmost servers in each rack.

3. The effective hot aisle dimensions reduce in S-Pod case, even further with a staggered arrangement. Thus more racks can be accommodated in the same space.
4. Higher heat dissipation is achieved in the S-Pod layout by increasing the effective inlet temperature from the room to the CRAC for a given velocity.

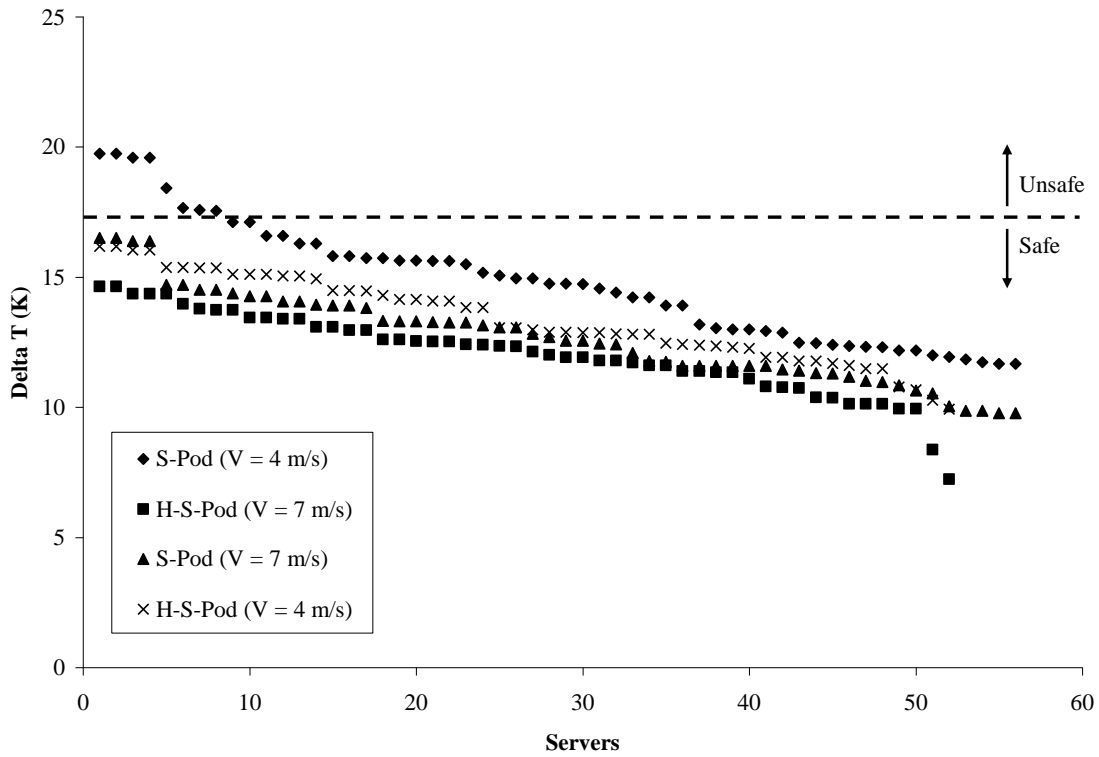


Figure 3.12. Temperature differential between rack inlet and CRAC outlet for S-Pod and Heterogeneous S-Pod cases

#### **4. Quantifying Inefficiency – Data Center Metrics**

For a Data Center Manager (DCM), minimization of the underlying costs for maintaining a data center without affecting the reliability standards and meeting the SLAs (Service Level Agreements) is the ultimate goal. The power consumption of server filled racks housed in such data centers has increased more than 200% in the past decade. While the industry average is still close to 3 kW, newer racks with blade centers can have up to 40 kW of power consumption. From a facilities standpoint, this directly translates to the increasing requirement of cooling capacity for the data centers. This problem can be dealt with a three step approach i.e. understanding the causality of inefficiencies, quantifying and logging their effects, and developing an intelligent and practical scheme to overcome it.

The understanding of inefficiencies was dealt with in Chapter 1 which talked about issues identified by government regulatory bodies, industry and academic institutions. Thus the scope of research is well understood. The development of the control schemes has been attempted in Chapter 2 and 3, where two long term and short term strategies for existing and new facilities have been laid out. While these schemes achieve their goals numerically and experimentally, parameters have to be determined which can benchmark the facilities in their current work setup and portray the improvements achieved by the control schemes. From a local standpoint, this can be observed by measuring the inlet temperatures at the rack inlets which ensures the safe working conditions, but one has to present the benefits on the entire data center level –

macro level, also. Thus logging and quantifying the different aspects of a data center on a time axis can help in resolving the above issues.

This chapter deals with parameters called “Data Center Metrics” which cover different facets of energy efficiency in a data center. Different metrics found in the literature are grouped under broad genres and summarized. The quantities to be measured, along with their locations, are presented via a power flow diagram listing out the various components. Also, the best devices/methods to achieve this data are identified. While, sensing and monitoring is one of the first steps for improving energy-efficiency, controlling various IT and facilities assets based on the collected values is the next step. This has been explained at a higher level in this chapter.

## **4.1. Metrics**

### **4.1.1. Issues in defining metrics**

Defining the constituent parameters of energy-efficiency metrics to uncover different inefficiencies is vital. Before delving into this, we should understand the differences between various commonly used terms which have been used interchangeably in literature but hold different meaning when used in a metric.

- *Energy-Efficiency and “Greenness”*: These terms have been used to refer to the same objective often in literature. It is implied that since in a data center, the energy consumption for a given floor area is much higher than any commercial building, the major source of environmental impact is the energy consumed. Thus greater energy efficiency means lower net energy consumption for the same footprint, leading to

- lower Greenhouse Gas (GHG) emission for the same facility, thus moving towards a “Greener Data Center”. From a life-cycle perspective though, for “greenness” of a data center, other issues have to be taken into account such as the building material, recycling potential for equipments, or pollution caused by the evaporation of water vapor in cooling towers (nutrient content of water increases leading to bio-films) etc.
- *Power Vs. Energy*: On a time axis, Power represents a profile, Energy is just an integrated scalar value. Monitoring of power is useful when the peak power consumption has to be determined. This is because, firstly, the equipment has to be designed to be at least functional, if not most efficient, at the peak power. Secondly, peak power consumption also defines the tariff brackets, according to which, utility meter bills will be decided. The final cost although, will be based on energy consumption. Also, energy values help in determining the greenness and sustainability metrics.
  - *Performance vs. productivity*: As mentioned in [49], productivity is an important parameter to be considered. Performance generally is on a per server basis. Thus with each component being optimized separately, the entire system is supposed to have high performance potential. Productivity deals with getting maximum throughput from the system. Thus, a combination of new and old systems can be used for distributing data, which are frequently read/written and dormant respectively for achieving higher productivity of the system. Higher performance results if all the devices used are energy efficient and can get maximum throughput on the whole.
  - *AC vs. DC Power Measurement*: While one issue is to measure the right type of power (AC or DC), the other is the knowledge of the type of power utilized at each

location, thus better informing the power measurement devices. For example, if power is measured at the plug where the servers are connected, the AC power is being measured. This does not consider the downstream losses because of AC to DC conversion for the server. There might be data centers though, which have a DC power distribution [6]. Also, the server manufacturers can better inform their customers about the relationship between the inlet and utilized power by standardizing losses.

- *Quantifying Losses:* While many metrics try to quantify losses related to say, IT or Cooling Facilities end, they have to be further broken down to finer levels. This also depends on where a particular measurement is taken. For example, if an external device is used to measure the power going from one unit to another, then power measured if assumed for the unit downstream also includes the line losses from the measuring device to the downstream unit. One needs to understand whether the source of higher inefficiency these is the equipment itself or the connecting lines.

#### 4.1.2. What is a Good Metric?

Based on the abovementioned conflicts and deviations in various measurements, following are the guidelines the author believes will help in determining the right energy metrics which would be effective and achievable:

1. The metric should not involve measurements which would require certain equipment to be taken down periodically. This affects the normal functionality of the data center and the benefit of the metric versus the loss due to the downtime is always questionable.

2. The metric should not lead to a “synthetic workload” which itself uses substantial amount of computing power of the data center.
3. A time wise trend is always a better tool to understand the variations as compared to a single point which has to be benchmarked periodically and compared with other “similar” data centers.
4. The metric should only cover a unique inefficiency. The motivation to use a metric is to understand how an aspect/feature is affecting the energy efficiency/greenness. Thus it going up or down should clearly indicate whether it is good or bad for a data center. Also, while an all-encompassing metric is good for one look conclusion, the granularity can be dealt by several hierarchies of metrics employed.
5. It should not be confused by the interchangeable use of the above mentioned terms.
6. They shouldn't be data center specific as then their global meaning would not be well understood. A certain class of data centers might have a metric though, specific to their work.
7. The metric should be measurable, independent of the hardware/manufacturer and should be developed looking from a future purview.
8. Lastly, some of the metrics employed should pave the way for future developments and thus decide the control aspect of assimilated data too.

There have been a lot of metrics reviewed from the literature in the coming section with their advantage, disadvantage and scope listed out. Also, to understand how these metrics can be calculated in one typical facility, Fig. 4.1 is referred each time. Fig. 4.1 gives a schematic diagram of the various devices and tools used in a data center, also giving power flow and in some cases, coolant flow diagram. This figure also gives

various power, temperature, flow and pressure meters installation locations which the author believes will be helpful in defining the metrics in the upcoming section. Fig. 4.1 is followed by a table referring to the specific metering device, the location of installation, its usefulness and what method or device should be used to monitor it.



Power Metering Device	Between		Usefulness / Application	Method / Device
	From	To		
1	Power Utility Supply	Switch Gear	Since, this gives the total power coming from the primary supply, it is a major parameter used to determine the efficiency of the Data Center	Usually these devices are already metered. While the electromechanical meters do not have this capability, the newer solid state meters can be connected through an Ethernet cable. The building management system (BMS) can also interface with these units
2	In house power Generator	Switch Gear	This power can also be used for efficiency calculations. Also, the % of energy coming from the generator per year can give an estimate of % of time, there was power failure and thus decide the importance of this unit for future facilities.	An electronic and network compatible power meter can be dedicated to this line.
3		Cooling Tower	For facilities side power breakdown	Main Circuit Breaker (MCB). An MCB can be purchased which gives direct Ethernet connection with each separate line. This can then directly interface with the BMS.
4		Chiller	To calculate metrics such as kW/ton	
5	Switch Gear/MCB	Chiller Pump	For facilities side power breakdown	
6		UPS	To calculate IT side of power inclusive of the losses from switch gear to UPS, at the UPS and further distribution lines.	
27	Switch	Lighting	To get power consumed by Lighting	

Table 4.1. List of Power metering devices, location and applications

Power Metering Device	Between		Usefulness / Application	Method / Device
	From	To		
	Gear/MCB	Sources/Panel		
7		CRAC	CRACs are now being placed under the UPS line owing to the reliability requirements of a Data Center. This is used to measure CRAC side power consumption. It is also used to calculate metrics such as cfm/kW	If the CRAC is connected to the UPS, there has to be a metering device to measure power. Otherwise, the connection can be monitored by the BMS. Also, if it is directly connected to the MCB, its power consumption can be taken from there.
8	UPS	Rack Power Distribution Unit (RPDU)	The power measurement happens at the RPDU. Thus, it gives the equivalent AC power at the plugs of the servers.	The RPDU should be network compatible and should at least give out the net power going out of it.
9		Metered PDU/Network	To calculate the net power consumed by the network	The PDU should be at least capable of being networkable and giving out net power going out of it. This will include losses in the power line from PDU to Network
10, 10'	RPDU	Servers/Storage	To calculate power consumed by each server/storage device. This value will be necessary for the AILM algorithm.	The RPDU should be outlet monitored. Also, during the selection, one should be careful about matching the specifications of the server and the RPDU like V, A and even cord type (C19-C20 vs. C13-C14).
11	RPDU	Server	This will give the DC power consumed by the server and thus	This has to be done through some standard software provided by the manufacturer or IPMI tools. This

Table 4.2. List of Power metering devices, location and applications (Contd.)

Power Metering Device	Between		Usefulness / Application	Method / Device
	From	To		
	outlet	Motherboard	eliminate the AC-DC conversion losses and the transmission losses from the power cord. Also, some racks employ cooling mechanisms like extra fans or heat exchangers. To get the TRUE IT load, one has to subtract all those values.	value is difficult to measure in practice.
12	RPDU outlet	Fans	This gives the cooling load inside a rack/server. This value can be used for further efficiency calculations.	This will be the (server power – RPDU outlet power). Since this might include errors due to distribution losses, the breakdown given by the software will be more accurate.

Table 4.3. List of Power metering devices, location and applications (Contd.)

Temperature /Humidity Sensor	Location	Usefulness / Application	Method / Device
13	Under floor or above ceiling Plenum/just below the floor tile	This can be used to get the inefficiencies introduced in the flow inside the plenum. Can also account for plenum leakage losses	Thermocouples (explained in Chapter 2) with wireless readouts are gaining popularity in the industry [ <a href="http://www.sensicast.com/uploadedFiles/WP-WSNs_for_Data_Centers.pdf">http://www.sensicast.com/uploadedFiles/WP-WSNs_for_Data_Centers.pdf</a> ]
14	Rack/Server Inlet	This temperature point is the primary measurement required for control algorithms like AILM. Even for monitoring, this measurement will inform the Data center managers of rack inlet temperatures which have to be lower than 32.2°C, as specified by ASHRAE. An array of these is required with at least one per server inlet. Some control schemes require this temperature. It is also helpful in calculating metrics like Return Heat Index (RHI), given later in the chapter and can help in quantifying the short-circuiting of cool air in the room.	Thermocouples (explained in Chapter 2) with wireless readouts; RPDUs equipped with temperature/humidity sensors.
15	Rack/Server Exhaust Temperature		
16, 17	CRAC Inlet Outlet	This is essential for calculating the total cooling provided by the CRAC unit. They are also helpful in calculating indexes like SHI and RHI, as discussed later.	Most new CRAC units have internal temperature sensors which govern their “intelligent control”. These could be accessed by an internal control box which can then be connected to the internet. Prioritizing user orders over the systems algorithm might vary from equipments.

Table 4.4. List of Temperature sensors, location and applications

Temperature /Humidity Sensor	Location	Usefulness / Application	Method / Device
19, 20	Chiller in and out pipes to the CRAC unit	This can be used to calculate metrics such as kW/ton. In some facilities, the chiller supplies cold water to the entire building and the data center. In such a case, instantaneous values of kW/ton of cooling at the chiller could be recorded. This when applied to the cooling achieved at the CRAC units with the specified time lag, gives the kW of the chiller dedicated just to the data center. This is clearly an approximate estimate assuming a linear relationship between power and cooling. This power value is essential for calculating the cooling facilities side load.	Again, this can be accessed by the internal control box of the chiller. Or otherwise, installation of thermocouples inside the pipe for an existing system is a tough task. So, a reasonable estimate can be the temperature of the surface of the non-insulated pipe coming/going from chiller kept in controlled environment.
21, 22	Chiller in and out pipes to the cooling tower	For checking the performance of the Cooling Tower	Either through chiller or the above explained approximate over-the-pipe temperature.

Table 4.5. List of Temperature sensors, location and applications (Contd.)

Velocity/Flow Rate Sensor	Where	Usefulness / Application	Type / Device
18	Chiller Pipe to CRAC	This can be used to calculate metrics such as kW/ton and the approximate power calculation explained above.	For new facilities, ideal would be an in-flow flow meter like a thermal mass or turbine type flow meter. For existing facilities, since the above involves intrusive methods, an ultrasonic flow meter could be used which can be fixed around the pipe.
23	Chiller to Cooling Tower	To check the performance of the Chiller	Some CRACs have Variable Frequency Drives (VFD). These can be accessed using an Ethernet cable and a user preferred control/monitoring scheme can be initiated.
24	CRAC inlet to the Plenum	To calculate the net cubic feet per minute (CFM) so that metrics like cfm/kW can be calculated. Also, this is used in control schemes like AILM and is also the chief parameter which has to be changed. For changing the flow however, the link should be to the CRAC unit and not the flow meter itself.	By Fan curves or by an anemometer and then calibrating it with the velocity of the fan can give a curve which can be used for these types of servers. Similar process has to be repeated for different kinds of servers.
28	Server inlet/outlet	To get an approximation of the net flow going out of the server. This measurement can also come from the velocity of the fan of the server. It is used to calculate metrics like RHI	

Table 4.6. List of velocity sensors/flow meters, location and applications

Pressure Sensor	Where	Usefulness / Location	Method/ Device
25, 26	In the plenum and the room	To calculate the pressure differences between the room and the plenum. Doing this at multiple locations will help one understand how the variations in pressure differences are as this is directly related to the flow rate coming out of the tile. Some Datacenter managers might also want to install under-floor fans based on this study at appropriate positions to facilitate higher flow rate	There is no specific specification since most transducers work in the ranges existing in a Data center.

Table 4.7. List of Pressure sensors, location and applications

#### 4.1.3. General Metrics

The following list gives the various metrics that have been used in the industry or will become increasingly common in near future are:

- 1. Average Watts per Square Foot (W/ft<sup>2</sup>) [6]:** This is a very broad metric. It calculates the ratio of the total power coming in to the total floor area of the data center. It was applicable initially when the actual distribution of load in the data center was not considered. There may be ambiguity in identifying the correct power going to the facility and the specific area it is being distributed over. Various utilities and support systems were included and excluded in some definitions. Moreover, with trends pointing towards higher power densities [6], the future watts per foot was tough to determine initially. From Fig. 4.1, the metric can be calculated by:

$$\frac{(1 + 2)}{\text{Total floor area of a data center}}$$

- 2. kW per rack [6]:** This metric considers the heat load per rack. The total heat load is then obtained by multiplying it with the number of racks existing in a data center. It is easier to extrapolate to a given facility than watts per square foot. Though there are some issues related to it. For example, since during the planning phase of a data center the equipment configurations are not decided, hence the total heat load will be tough to determine. This is essential information for selecting facilities equipment such as chillers, chiller pumps, CRAC units etc. Also, the power data is generally listed by the rack/server manufacturers is often more conservative than the actual power used. The above two metrics give a load map for a point in time and the trends

for future prediction cannot be generated. For each rack, power per rack will be given by (Fig. 4.1).

$$\sum_{\substack{\text{for all RPDUs } j \\ \text{for a rack}}} 8_j$$

- 3. MIPS/kW [9]:** This is the acronym for millions of instructions per second per kW. While MIPS is a metric used to characterize just computing, this ratio gives how much computing is achieved each second for every unit of power that goes into a data center. The use of total power and not just the IT power has faced some criticism. For every 1 kW of IT power needed for a Data center, around 0.5 kW is taken up by the cooling and support systems [5]. If only the IT power is used for division, then this metric fails to reflect the effects of cooling power wastage. Also, for systems which are just live but not doing any useful work, useful MIPS will be essentially 0 but there will be significant amount of power consumed. While, MIPS is a value to be obtained directly by connecting to the server, kW can be obtained by (Fig. 4.1):

$$\sum_{\substack{\text{for all} \\ \text{servers } i}} 8_i$$

#### 4.1.4. Component Specific Metrics

- 1. kW/ton [9]:** This metric is used to characterize the chiller efficiency. It is the power used by the chiller to provide 1 ton of cooling. It is an essential metric for data centers using chillers which also supply to other parts of the system such as office space etc.

The exact method to calculate chiller power using this metric for a particular system is given in a later section. It can be represented as (Fig. 4.1):

$$\frac{18 \times C \times |19 - 20|}{4}$$

Where, C is a constant.

- 2. CFM/kW [9]:** This is a metric which can be used to quantify the efficiency of air conditioning unit, like CRAC. It gives the volumetric flow rate of cool air released by the CRAC unit for every kW of power consumed by it. This is an incomplete metric for the CRAC as there are other important parameters not considered, including set point temperature and humidity as the power consumed will change with change in these values. It can be represented as (Fig. 4.1):

$$\frac{24}{7}$$

#### 4.1.5. Server-Level Metrics

- 1. Expected Maximum Load (EML) or Expected Maximum Heat (EMH) [9]:** The design of a new data center is dependent upon the kind of equipments which will be housed in it. The power dissipation data for servers is a requirement as the cooling facilities are designed accordingly. The current practice is to provide the “nameplate” data which is the maximum power consumption by the equipment incorporating a safety factor. Designing the facility with nameplate power data instead of the real data can lead to over-sizing by a factor of 4 or more [9]. Thus a new metric has to be initiated, which on the basis of previous monitoring of the equipment, gives the Expected Maximum Load (EML). This in turn will help determine the Expected

Maximum Heat (EMH) rejected by equipments. The sizing of data center cooling equipments based on this will lead to operating conditions near their highest efficiency points. The mathematical expression based on Fig. 4.1 is not possible in this case as this is the maximum value of load observed for a given time or based on similar data center values.

- 2. Embedded Watts per \$1000 of 1U Server Spending [50]:** There has been an increasing need for understanding how the energy costs can be related to actual power consumption of the servers, since one has to understand the tradeoff between investments in buying a new server, versus operating costs for maintaining the old servers. One of the ways to quantify it is by using this metric. The power consumption of servers per \$1000 invested initially has risen from 8 to 109 W between 2000 and 2006. By 2012, according to the current and most conservative estimates, it will be 1650 and 157 embedded Watts respectively [50]. The metric can be presented as (Fig. 4.1):

$$\frac{\sum_{\text{servers } i} \text{for all } P_i}{(\text{Total Server Spending}/1000)}$$

- 3. Site Infrastructure Power Overhead Multiplier (SI-POM) [49]:** SI-POM is the ratio between the data center power consumption at the utility meter to the AC power consumed by the IT equipment at the plug. It is the power ratio and not the energy ratio. Thus the value of SI-POM varies with time, depending on the workload of the data center. The losses captured in this metric include UPS losses, PDU losses, power/cooling equipment losses, lighting losses and other unidentified loads which add to the inefficiency. The Uptime Institute with participant companies has

suggested that improving operation can lead to higher impact in lower time period, compared to upgrading a data center. Also, this metric will not show improvement due to savings of energy by free cooling. This is because it gives the power ratio and thus the greatest use of it is to get the peak value of SI-POM which will help in sizing of the future equipment. The peak value then varies with each of the two parameters reaching their peaks. In a data center, it is intuitive that with peak load in IT hardware, the peak load in cooling equipments will exist, although there might be a time lag between the two events. Also, if free cooling is incorporated such that it affects power consumption throughout the year and thus components like UPS and PDUs can be undersized, then it will show a drop in SI-POM too. To display seasonal improvements, another metric can be defined which calculates the above ratio terms of energy. This can be referred to as the Site Infrastructure Energy Overhead Multiplier (SI-EOM). For the setup mentioned, it can be calculated by:

$$\frac{(1 \text{ or } 2 \text{ or } (1 + 2))}{8}$$

- 4. IT Hardware Power Overhead Multiplier (H-POM) [49]:** It is defined as the ratio of the AC power given out from the plug, to the DC power utilized by the actual IT equipment. This metric accounts for the power conversion loss from AC to DC, and the power loss in adjunct equipments such as servers fans. This can be used by the manufacturers to market their “higher efficiency” devices. Another fact linked to it is that H-POM will not vary a lot for servers whose 100% CPU utilization power is similar to power consumed in “comatose” state. H-POM can be defined as:

$$\frac{10}{11}$$

**5. DC Hardware Compute Load per unit of Computing Work Done (DC-CLCW)**

[49]: As H-POM addresses the power losses from the plug to the servers, this metric accounts for losses inside the IT equipment. Thus for various IT equipments like processors, storage devices and networking devices this metric defines the efficiency for processing, reading and writing data and routing data packets respectively. This has been identified to be difficult to measure at the data center operation level by the Uptime Institute. However, it gives the manufacturers a parameter to benchmark their processes.

**6. Deployed Hardware Utilization Ratio (DH-UR) [49]:** This metric is used to quantify the efficiency of using a server. It is the ratio of the number of servers running live applications (not in just a comatose stage with only basic applications), to the total number of servers deployed. For the storage devices, it is the ratio of the terabytes of storage accessed in the last 90 days to the total terabytes of storage for the data center. While the usage of this metric is questionable because the knowledge of servers not being used will lead to their shutting down by the IT staff, it still can give the industry averages and thus help Data Center Managers understand the level of flexibility and redundancy required. It can be calculated by:

$$\frac{\text{No. of servers running live applications}}{\text{Total number of servers}}$$

**7. Deployed hardware Utilization Efficiency (DH-UE) [49]:** This metric has been designed by the Uptime Institute to encourage virtualization in data centers. It is the ratio of the minimum number of servers to handle peak load to the total number of servers deployed. Please note, that even for fully virtualized case, this ratio will always be less than equal to 1 since the numerator of this metric can be a fraction.

Furthermore, the numerator should also include the overhead due to inclusion of the virtualization layer. This can be expressed as a fraction of the total workload capacity of a server. Again, with the virtualized setting, this ratio is calculated with the denominator now being the integer number of virtualized servers required. The idea thus is to increase the value of this metric. The motivation of doing so is that many servers use almost the same amount of power for their 100% utilization and comatose stages. Thus shutting down as many servers as possible will reduce the net power consumption. However, the performance and reliability of servers can be lower for higher rates of utilization, as we know well from our own Personal Computers. Also, the performance is limited by the RAM in the system and it has to be compatible with full CPU utilization.

$$\frac{\textit{Minimum number of servers required to handle total load}}{\textit{Total number of srevers}}$$

#### 4.1.6. IT Metrics

- 1. IT productivity per embedded watts (IT-PEW) [50]:** It considers the IT aspect of the data center accounting for architectural, reliability and operating conditions. It is used for equipment selection and system architecture decisions. There are two levels at which decisions have to be made and comparisons have to be performed. From a company's strategic perspective, to ensure certain reliability levels, Service Level Agreements (SLA) and response times, the upper management incorporates considerable redundancy in the IT design. This has to be monitored at facilities that are in similar businesses and thus benchmarking the redundancy allowed. While this

will give the management a parameter to compare how their performance is with respect to the industry, it can also be used from a marketing perspective leading to competition and innovative ways of achieving lower IT-PEW. At a lower level, since this can be at equipment level, this metric can be utilized to determine the vendor of choice for various equipments. Again transparency will be the key achievement, along with marketing motivated efforts for enhanced IT-PEW. IT productivity has to be parameterized with respect to each data center. Total power taken up by servers and storage is  $\delta$ .

- 2. Power Usage Effectiveness (PUE) [51]:** PUE is defined as the ratio of the total power coming into a data center (at the utility meter) to the power consumed by the IT equipment. The IT power consists of compute, storage, network and supplemental equipment, even the computer used to measure this value. The idea behind the development of this metric is to understand whether performance improvement is possible in a facility before moving to a new facility. This may lead to a short term strategy for operational improvement. It also can be benchmarked with other facilities, and a relative performance can be derived. Moreover, this can give a measure of improvements due to introduction of new design/retrofits. PUE can be calculated by (Fig. 4.1):

$$\frac{(1 \text{ or } 2 \text{ or } (1 + 2))}{(8 + 9)}$$

- 3. Data Center Efficiency (DCE) [51]:** This metric is just the reciprocal of PUE. Thus, it gives the efficiency of the power brought inside the data center in being utilized by the IT equipment. Thus a DCE value of 0.5 indicates that power required from the

power grid would be twice the power consumed by the IT equipments. It can be calculated by:

$$DCE = \frac{1}{PUE}$$

Calculation of the PUE and DCE values can be performed in an approximate manner, or by data manipulation of values taken at several locations. It is difficult to calculate these values for buildings where the power to the data center does not come through a utility meter. Also, increasingly new racks are being introduced which have some kind of refrigeration, blowing or cooling mechanism incorporated in them. For such cases, it is tough to determine the division of power between the IT and facilities. So, planning for instrumenting the data center has to be done in detail before aiming for generating the metrics. Thus to understand the grey areas, PUE has also been defined in the following manner:

$$PUE = \text{Cooling Load factor (CLF)} + \text{Power Load Factor (PLF)} + 1.0$$

Here, CLF is cumulative power consumed by CRACs, chillers, chiller pumps, cooling towers etc. normalized by IT load. PLF is power consumed by UPS, switch gear, Power Distribution units (PDUs) divided by the IT load. The 1.0 factor comes because of IT load normalized by itself.

#### 4.1.7. Facilities Metrics

- 1. Site Infrastructure Energy Efficiency Ratio (SI-EER) [50]:** This is a metric to identify the efficiency of the data center site infrastructure system, including the power and cooling support systems. It is the ratio between the net power going into the data center to the power consumed by the IT services only. This metric is very

similar to PUE (Power Usage Effectiveness), which is discussed later in this section. Thus, this accounts for the cooling system efficiencies, UPS and transmission losses and other inefficiencies specific to a data center layout and operations. The Uptime Institute has found that the average value of this in the industry currently is 2.5 [50]. This means that for every 2.5 kW which enters the data center, only 1 kW is utilized by the IT equipments. The Institute believes that in the best case scenario, it can be 1.6. The reduction of SI-EER to 2.0 is definitely achievable in near future. For the setup mentioned, it can be calculated by:

$$\frac{(1 \text{ or } 2 \text{ or } (1 + 2))}{(8 + 9)}$$

#### 4.1.8. Air-management metrics

1. **Supply Heat Index (SHI) [52]:** Inside the data center, to maximize the performance of any air management scenario, there are three objectives. Firstly, the infiltration of hot aisle air into cold aisle, thus increasing the temperature of the inlet air to a server has to be minimized. Secondly, the short circuiting of cold air without being used up by the servers has to be minimized. Thirdly, the decrease in temperature of the return air between the server exit location and CRAC inlet has to be minimized. This motivates the introduction of two new parameters which can quantify these inefficiencies. This can be characterized by taking the ratio of enthalpy rises between suitable locations. Since, our main objective is to take out heat from the servers, the net enthalpy rise at the server exhaust from the CRAC outlet to the plenum is the normalizing factor. Supply Heat Index (SHI) is the ratio of the enthalpy rise at the

inlet to the servers due to infiltration of hot air in cold aisle to the total enthalpy rise across the server as explained before.

$$SHI = \frac{\sum_{i - \text{all servers}} (14_i - 13)}{\sum_{i - \text{all servers}} (15_i - 13)}$$

2. **Return Heat Index (RHI) [52]:** RHI is the ratio of the heat extraction at the CRAC unit to the net enthalpy rise across the server with reference to the temperature at CRAC outlet to the Plenum. This will show the flow efficiency of the data center. Thus the four locations where the temperature has to be measured are CRAC inlet from the room, CRAC outlet to the plenum (if there is a plenum), server inlet temperatures and server outlet temperatures.

$$RHI = \frac{\sum_{j - \text{all CRACs}} 24_j \cdot (16_j - 17)}{\sum_{i - \text{all servers}} 28_i (15_i - 17)}$$

#### 4.1.9. Panoptic Metrics

1. **Data Center Energy Efficiency and Productivity Index (DC-EEP) [50]:** This metric is the multiplication of the IT-PEW and the SI-EER. It gives the IT productivity achieved for every unit power going into the site infrastructure systems. Clearly, a higher value is better. While monitoring this for a single facility might not be very beneficial, benchmarking this across data centers can help develop a set of best practices for higher DC-EEP.

$$DC-EEP = IT-PEW \times SI-EER$$

2. **Data Center Performance Efficiency (DCPE) [51]:** This has been proposed by the Green Grid for long term strategies. It is the ratio of the useful work done by the data

center and the total facilities power required to do so. In some ways, it is very similar to another metric by Uptime Institute, called the Data Center Energy Efficiency and Productivity Index (DC-EEP).

- 3. Free Cooling Metric (FCM) [49]:** With more cognizance in the data center world about using air-side/water-side economizers, free cooling is becoming more and more popular. Thus a metric is required to understand the energy savings by free cooling. This is defined as follows:

$$FCM = \frac{\text{thermal kW of free cooling used per year} \times \text{kW consumed at utility meter per kW delivered to the cooling system}}{\text{COP of Mechanical Plant} \times \text{kWh used by the /Cooling System facility in that year}}$$

This metric is the ratio of energy saved by free cooling to total energy consumed by the system. Thus, a higher value of this is more desirable.

- 4. Enabling Energy Savings Metric (EES) [49]:** This metric is used for understanding how putting a server to hibernate/sleep or power save mode can help achieve lower power consumption than leaving it in idle condition where it consumes power almost equal to a full CPU utilization state. Thus to calculate the energy savings over a period of time, the following formula is suggested to be used:

$$\int_{\text{all pieces of equipment } i} \left( \int_{\text{all possible hibernate time}} (idle\_power_i - hibernate\_power_i) \bullet (SI - POM) dt \right)$$

- 5. Data Center Productivity (DCP) [53]:** Data center productivity are a general class of metrics introduced by the Green Grid. This involves the ratio of the useful work

done by the data center to any particular resource used by the data center. The general class will be represented as DCxP where x is the resource against which the useful work has been calculated. X can, for example, be the average power of the data center or the peak power in a certain period.

- 6. Data center Energy Productivity (DCeP) [53]:** This is, as defined before, a class of the DCP which calculates the ratio of the useful work to the total energy consumption of the data center for producing this work. Thus, decreasing PUE will lead to lower energy consumption and thus improving DCeP. The definition of Useful Work uses a utility function which is based on the SLA's and response time. For ex: if the task is done within time, then it is one or else 0. Thus from a time window perspective, it is represented as:

$$Useful\ Work = \sum_{i=1}^M V_i \cdot U_i(t, T) \cdot T_i$$

Here  $V_i$  is a normalization function assigning value to tasks,  $U_i$  is the utility function,  $t$  is the time window of assessment and  $T$  is absolute time for task completion. If  $V_i$  and  $U_i$  are 1, then the metric will just give the tasks completed in the given time window and maximizing it will be the objective of the operator.

- 7. Corporate Average Data Efficiency (CADE) [40]:**

CADE is defined as:

$$CADE = Facilities\ Efficiency \times IT\ Assets\ Efficiency$$

Both of these efficiencies are products of utilization % and energy efficiency %. While utilization % for facilities is the IT load divided by the facilities capacity, for IT asset, it is the average CPU utilization. Energy efficiency for Facilities is the IT

load divided by the total energy consumption by the Data center. For IT asset, it is the future manufacturer specified energy efficiency metric. This is an attempt to combine the IT and Facilities metrics to get one single parameter for efficiency.

## 4.2. Control Parameters

Fig. 4.3 gives the control architecture/strategy of a data center. The workload entering in the data center is directed to the brain of the control strategy. This gets feedback from the various metering devices about power and energy data and parameters like temperature, velocity, humidity and pressure. The protocols specific to the facility are hardcoded here. Also issues like SLAs can be flexibly altered with the kind of job entering the system. Thus the control strategy can be employed on two ends of the system:

1. **Facilities end:** This is both for increasing energy efficiency by curbing direct wastage and by optimally using the available resources. Following should be controlled:
  - a. *CRAC unit:* There are three different parameters to be varied. One is the flow rate which is directly related to the compressor operation. This can be controlled by accessing through the network the Variable Frequency Drive (VFD) unit. Secondly the set point has to be varied. The set point consists of two values- Temperature and Humidity. The intelligent CRAC units have a way of sensing the flow rate and the input and exit temperature values and determining this for optimal conditions. Another way is to connect to the internal “brain” of the CRAC unit and setting the priorities according to the user.
  - b. *Flow rate and Temperature for Chillers:* This considers the trickling effect of the change in CRAC velocity. Some chillers have a feedback unit which

automatically changes the conditions. For others, the chiller has to be accessed through network or by interfacing to the higher level Building management System (BMS)

- c. *% Porosity of tiles*: This is a parameter which can change the net flow rate coming out of the tiles. While it doesn't save energy directly, it can be helped to channel the flow for better airflow management. For example, if a section of the Data center is not in operation for the current load state, the tiles can be completely closed, thus allowing other functional sections to achieve higher flow rates.

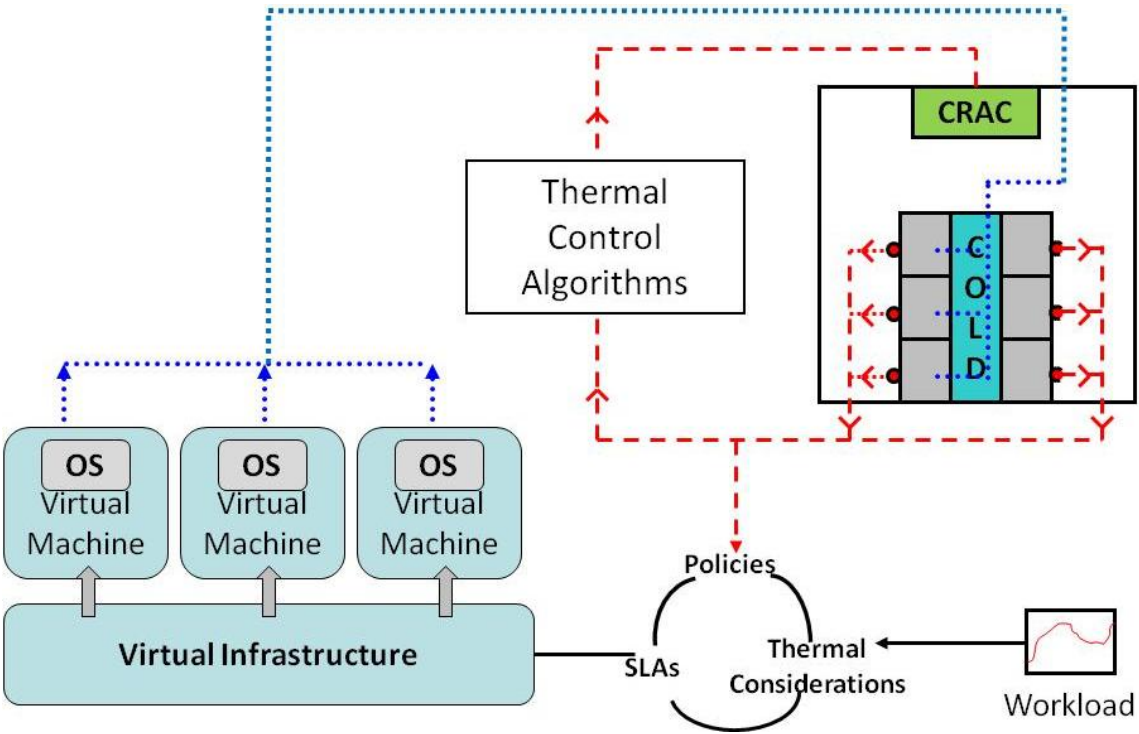


Figure 4.2. Control Scheme for a data center

2. **IT end:** This should consider the virtualization of the servers. As Fig. 4.3 shows, this would help direct load according to the outcome of the strategies set in the brain of the control system. The virtual layer while optimizing in the given subspace, should consider thermal bounds for each server.

Thus, as the industry is moving towards higher energy dissipation densities and increased prices of supplied power, the costs going in a data center have to be broken down and inefficiencies have to be identified. Further, sensing the right quantities with a comprehensive system to filter and analyze the data is required. Finally an all-control scheme has to be devised keeping the right trade-off between the increased costs of installing such controls and the savings on the energy end.

## 5. CONCLUDING REMARKS

### 5.1. Summary

The objective of the study was to lay down strategies for increasing energy-efficiency in data centers. Such facilities can be broadly classified into existing or legacy, and the ones still in design/conceptual stage. For the legacy facilities, it is very important to instrument the data center to achieve the granularity desired for mapping inefficiencies. To quantify such values, metrics have to be conceived which uncover various losses. These should be able to differentiate between, otherwise interchangeably used parameters. Twenty-five different metrics which have been used by industry are presented, and their advantages/disadvantages given. Instrumentation devices and locations are recommended to measure these metrics. This will help industry-wide benchmarking and comparison of processes to come up with a set of best practices.

The local approach for mitigation of losses is to have a feedback loop based on the monitored values, directing various devices to operate optimally. The requirements for such an algorithm are expressed based on model simulations of a simplistic scenario. The model derived here is called *Ambient Intelligence based Load Management (AILM)*. This is based on the concept of temperature linearity, when the flow field is the same. The formulation of the algorithm in the calibration and operational phases is given. A computational test bed is used to investigate the functionality, and the improvement obtained. Scenarios like a futuristic heterogeneous data center are also considered. An experimental test bed is setup, which evidenced the improved performance of AILM

versus a standard case. The temperature prediction solely based on computations does not completely match the experimental values and causes for such a deviation are mentioned.

Since the new facilities are still in design phase, it gives a lot of flexibility in trying out different options for layouts. The Hot Aisle-Cold Aisle layout suffers from recirculation and short-circuiting of cold air. Thus, variations to this have been suggested in the form of a new layout called Scalable Pods (S-Pods). Here racks are deployed as pods with cool air coming from the enclosed perforated tile. It is shown to achieve significant improvement over HACA. Its advantages/disadvantages, along with various design aspects have been studied through case studies.

Thus, this study attempts to provide a holistic approach towards increasing the energy efficiency within the datacenter.

## **5.2. Future Work**

These are the following three venues which the author believes should be explored:

1. *Large Scale Implementation of AILM*: The experimental implementation of AILM presented in this study, has been done for a single rack, filled with server simulators. In future, a larger scale experimental validation should be performed with larger number of racks and real servers to determine the variation due to changing fan velocity. Also, cases should be considered where more than 1 CRAC units are involved, as this will address the question of the minimum number of base recordings required for complex scenarios. Research should be done in which AILM calibrates itself during normal operations by taking  $n$  linearly independent load scenarios with time, and auto-updates with time based on flow variations in datacenter due to

additional equipments/barriers. The current study focuses on steady state values. Research should be conducted to derive the best approach for shifting from one velocity to another with changing loads, such that the temperature at rack inlets do not exceed critical levels.

2. *Scalable Pod Implementation:* The experimental study for S-Pods has not been undertaken here and is recommended. Combining S-Pod with AILM will also be an interesting study.
3. *Metric Evaluation and Design:* Based on the above two designs, new metrics should evolve, which can monitor the efficacy of the above mentioned algorithms. Also, instrumentation of a model lab is required to capture all the metrics mentioned in this study and eliminate the redundant and non-specific ones. This can be achieved by fine tuning the values of various working systems of a datacenter and observing the change in the corresponding metrics.

## **APPENDIX A: Computational fluid dynamics-heat transfer**

While analytical solutions are desirable for describing the fluid flow dynamics and related heat transfer, such solutions are often not available for complex geometries. Thus to determine the characteristics of flow and related parameters, one has to employ CFD/HT techniques.

CFD uses numerical techniques and pre set algorithms to solve temperature, velocity and pressure fields at every point. Since point is a more subjective term, the domain is discretized in small cells which resemble the “point” mentioned before. The equations for continuity, momentum and energy (described in next section) are solved for each of these cells. The equations are also discretized using a Finite Volume Method. An initial condition is set by the user for all of these cells. Then based on the set boundary conditions, the solver iteratively solves the above mentioned equations and converges on residuals specific to each equation. The minimum residual level for convergence is set by the user. The solution is said to be converged if the convergence levels are satisfied for all the residuals. The algorithm used in the current case for determining pressure and velocity in a coupled way is called SIMPLE. This algorithm involves under relaxation factors for each of the converging properties like momentum, pressure etc. and this can be altered by the user depending on the problem, though a set of default values are provided by FLUENT.

There are two parameters of interest while evaluating a CFD code, Time and Memory. While they are not completely independent of each other, the value of one is not sufficient to determine how good a model is. The underlying common variable for

both of these is the mesh size and total number of meshes. For a given geometry, lower mesh size will lead to larger number of meshes. This would require the code to store more number of values and thus the memory required is higher. Also, the number of cells for which the equations have to be solved will increase leading to higher computational time. Since this all indicates towards going for a higher mesh size, one has to understand the convergence characteristics. Higher the mesh size, more are the approximations for the “point”, and thus while getting a converged solution might be difficult for the given boundary conditions, also the error in the final computational value will increase. Thus one has to judiciously pick the mesh sizes based on tradeoff between computational size and solution accuracy.

### **5.3. Governing Equations of Fluid Flow and Heat Transfer**

There were three basic equation referred in the previous section, Mass Momentum and Energy. All of these equations can be derived from the general principles laid out in the Reynolds Transport Equation. The equations and some text has been taken from [54].

Mass Equation: This is also called the continuity equation. The convergence of mass claims that the rate of change of mass in a domain is zero as mass cannot be created or destroyed (this is for fluid flows for velocities below speed of light). Thus the mass entering a domain is equal to the sum of mass stored in the domain and the mass leaving the domain. It is represented mathematically by:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m$$

where,  $\rho$  is the fluid density,  $v$  is the velocity, and  $S_m$  is the mass source term

Momentum: The conservation of momentum principle states that sum of the time derivative and convective transport of the linear momentum of fluid is equal to the sum of the forces acting on the domain. The forces can be body forces which act on the bulk of the fluid like gravity or could be surface forces acting on the boundaries of the domain, like shear stresses. These equations are also referred to as the Navier-Stokes Equations.

Mathematically they are represented as:

$$\frac{\partial}{\partial t}(\rho \vec{v}) + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot (\bar{\tau}) + \rho \vec{g} + \vec{F}$$

where,  $p$  represents the static pressure and  $g$  is the acceleration due to gravity,  $F$  is a term that includes any other model-dependent body forces like force is due to pressure drops involved in a porous media model and  $\tau$  is the viscous stress tensor. The stress tensor is a complex term can be written as:

$$\bar{\tau} = \mu \left[ (\nabla \vec{v} + \nabla \vec{v}^T) - \frac{2}{3} (\nabla \cdot \vec{v}) I \right]$$

where,  $\mu$  is the molecular viscosity, and  $I$  is the unit tensor.

Energy: The energy equation states that any increase in the energy of a region is equivalent to the sum of work done to or by the region and also the heat transfer to or from that region. Mathematically, the energy equation may be written as follows:

$$\frac{\partial}{\partial t}(\rho E) + \nabla \cdot (\vec{v}(\rho E + p)) = \nabla \cdot \left[ k_{eff} \nabla T - \sum_j h_j \vec{J}_j + (\bar{\tau}_{eff} \cdot \vec{v}) \right] + S_h$$

Where,  $k_{\text{eff}}$  is defined as the effective thermal conductivity of the region, and it is the sum of the thermal conductivity of the fluid in the region ( $k$ ) and the turbulent thermal conductivity ( $k_t$ ), estimated by various turbulent models.  $J$  is the diffusion flux of the species  $j$ ,  $S_h$  is volumetric heat generation term (For example, in our case the heat dissipated by the chip). The variable  $E$  used in the L.H.S. is defined as:

$$E = h - \frac{p}{\rho} + \frac{v^2}{2}$$

Where,  $h$  is the sensible enthalpy.

## REFERENCES

1. LBNL, *High Performance Data Center - A Design Guideline Sourcebook*. 2006, Pacific Gas and Electric Company, pp 25.
2. EPA, *Report to Congress on Server and Data Center Energy Efficiency*. 2007, U.S. Environment Protection Agency, pp 47.
3. Patel, C.D., et al. *Thermal considerations in cooling large scale high compute density data centers*. 2002. San Diego, CA, United States: Institute of Electrical and Electronics Engineers Inc.
4. Friedrich, R. and C.D. Patel. *Towards Planetary Scale Computing - Technical Challenges for Next Generation Internet Computing in Thermal Challenges in Next Generation Systems: Proc. Thermes Conference*. 2002: Millpress, 2002.
5. Sullivan, R.F., *Alternating Cold and Hot Aisles Provides More Reliable Cooling for Server Farms*. 2002, The Uptime Institute.
6. ASHRAE Editors, *Datacom Equipments, Power Trends and Cooling Applications*, ASHRAE, Editor. 2005, pp 68.
7. Tschudi, W., *LBNL Data Center Energy-Efficiency Activities*. 2007.
8. DOE, *Annual Energy Outlook 2007, with Projections to 2030*. 2007a, US Department of Energy: Washington, pp 35.
9. Tschudi, W., et al., *High Performance Data Centers - A Research Roadmap*, LBNL, Editor. 2003, pp 48.

10. Mitchell-Jackson, J., *Energy Needs in an Internet Economy: A Closer Look at Data Centers*, LBNL, Editor. 2001.
11. Stein, J. *More efficient technology will ease the way for Future Data Centers*. in *ACEEE Summer Study on Energy Efficiency in Buildings*. 2002.
12. Thompson, C.S., *Integrated Data Center in the New Millennium*, in *Energy User News*. 2002.
13. Staff, T., *Heat Density Trends in Data Processing, Computer Systems and Telecommunications Equipment*. 2000, The Uptime Institute.
14. Sullivan, R.F., L. Strong, and K.G. Brill, *Reducing Bypass Airflow is essential for Eliminating Computer Room Hot Spots*. 2004, The Uptime Institute.
15. Blackburn, M., *Five Ways to Reduce Data Center Server Power Consumption*. 2008, The Green Grid.
16. Industry, *Guidelines for Energy-Efficient Data Centers*. 2007, The Green Grid.
17. Sharma, R.K., et al., *Balance of Power - Dynamic Thermal Management for Internet Data Centers*. Jan 2005, Hewlett-Packard White Paper.
18. Brooks, D. and M.Martonosi. *Dynamic Thermal management for High Performance Microprocessors*. in *Proceedings 7th International Symposium High-Performance Computer Architecture (HPCA 01)*. 2001: IEEE CS Press.
19. Skadron, K., T. Abdelzaher, and M.R. Stan. *Control- Theoretic Techniques and Thermal-RC Modeling for Accurate and Localized Dynamic Thermal Management*. in *Proceedings 8th Int'l Symp. High-Performance Computer Architecture (HPCA 02)*. 2002: IEEE CS Press.

20. Moore, J., *Automated Cost Aware Data Center Management*, in *Department of Computer Science*. 2006, Duke University, pp 87.
21. Moore, J., et al., *A Sense of Place: Toward a Location-aware Information Plane for Data Centers*. HP White Paper, 2004. HPL-2004-27.
22. Moore, J., et al., *Data Center Workload Monitoring, Analysis and Emulation*. HP White Paper, 2004. HP Labs, Palo Alto.
23. Moore, J., J.S. Chase, and P. Ranganathan. *Weatherman: Automated, online, and predictive thermal mapping and management for data centers*. 2006. Dublin, Ireland: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States.
24. Moore, J., et al., *Making scheduling "cool": temperature-aware workload placement in data centers*, in *Proceedings of the annual conference on USENIX Annual Technical Conference*. 2005, USENIX Association: Anaheim, CA.
25. Chase, J.S., et al. *Managing energy and server resources in hosting centers*. 2002. Banff, Alta., Canada: Association for Computing Machinery.
26. Pinheiro, E., et al. *Load Balancing and Unbalancing for Power and Performance in Cluster-Based Systems*. in *Proceedings of the Workshop on Compilers and Operating Systems for Low Power*. September 2001.
27. Elnozahy, M., M. Kistler, and R. Rajamony. *Energy-Efficient Server Clusters*. in *Proceedings of the Second Workshop on Power Aware Computing Systems*. Feb 2002.
28. Bilodeau, S., *Energy Management of a Chilled Water Plant using Thermal Storage*, in *10th International Conference on Thermal Energy Storage*. 2006.

29. Dietrich, J. and R. Schmidt, *The Green Data Center for education*. September 2007, IBM.
30. Group, I.S.a.T., *Keeping your Data Center Cool: There is Another Way*. Feb 2008, IBM.
31. Robert Francis Group , *The Rise of Power to Power: Dealing with the new Data Center Constraint*. 2006, White Paper (Robert Francis Group).
32. Scaramella, J. and M. Eastwood, *Solutions for the Datacenter's Thermal Challenges*. January 2007, IDC White Paper.
33. Patterson, M., et al., *Data center TCO; a comparison of high-density and low-density spaces*. Jan 2007, Intel White Paper.
34. Filani, D., et al., *Dynamic Data Center Power Management: Trends, Issues and Solutions*. Intel Technology Journal, Feb. 2008. **12**(01).
35. Microsoft, *Best Practices for Energy Efficiency in Microsoft Data Center Operations*. Microsoft (Fact Sheet), Feb. 2008.
36. Rambo, J., *Reduced order Modeling of Multiscale Turbulent Convection: Application to Data Center Thermal Management*, in *School of Mechanical Engineering*. 2006, Georgia Institute of Technology: Atlanta, pp 69.
37. Samadiani, E., Y. Joshi, and F. Mistree. *The thermal design of a next generation data center: A conceptual exposition*. 2007. Cairo, Egypt: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States.

38. Y.Joshi, E. Samadiani, and F. Mistree, *Reduced Modeling Based Robust Thermal Design of Energy Efficient Data Centers*, in *The Eighteenth International Symposium on Transport Phenomena*. Aug 2007: Daejeon, Korea.
39. Gupta, S., *Thermal Aware Task Placement in Data Centers*. 2008, IMPACT Laboratory, Arizona State University: Phoenix.
40. McKinsey, *Revolutionizing Data Center Efficiency*. 2008, The Uptime Institute, White paper presentation.
41. Ferziger, J.H. and M. Peric, *Computational Methods for Fluid Dynamics*. Vol. 3rd review edition. 2002: Springer.
42. Somani, A. and Y. Joshi, *Ambient Intelligence Based Load Management (AILM)*, USPTO, Editor. 2008: USA.
43. Ashford, R., *Mixed integer programming: A historical perspective with Xpress-MP*. *Annals of Operations Research*, 2007. **149**(1): p. 5-17.
44. Nelson, G., *Development of an Experimentally-Validated Compact Model of a Server Rack*, in *School of Mechanical Engineering*. Dec. 2007, Georgia Institute of Technology: Atlanta.
45. TSI, *Operation and Service Manual for Model 8350 VELOCICALC Air Velocity Meter*, TSI, Editor. Aug. 1988.
46. Taylor, J.R., *An Introduction to Error Analysis*. 1982: Oxford University Press. 270.
47. Somani, A. and Y. Joshi, *Scalable Pods Based Cabinet Arrangement and Air Delivery for Energy-Efficient Data Center*, USPTO, Editor. 2008.

48. Schmidt, R., K. Karki, and S. Patankar, *Raised-Floor Data Center: Perforated Tile Flow Rates for Various Tile Layouts*, in *Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, IITHERM*. 2004: Las Vegas, NV.
49. Stanley, J.R., K.G. Brill, and J. Koomey, *Four Metrics Define Data Center Greenness*. 2007, The Uptime Institute.
50. Brill, K.G., *Data Center Energy-Efficiency and Productivity*. 2007, The Uptime Institute.
51. Green Grid Committee, *Green Grid Metrics: Describing Datacenter Power Efficiency*. The Green Grid - Technical Committee White Paper, Feb. 2007.
52. Sharma, R.K. and C.E. Bash. *Dimensionless Parameters for Energy-Efficient Data Center Design*. in *IMAPS-ATW*. Oct 26th 2002. Palo Alto: HP Labs.
53. Anderson, D., et al., *Framework for Data Center Energy Productivity*. 2008, The Green Grid White Paper.
54. Panton, R.L., *Incompressible Flow, Second Edition*. 1996, New York: John Wiley & Sons Inc.