ABSTRACT

Title of Document: RISK ASSESSMENT AND MITIGATION OF TELECOM EQUIPMENT UNDER FREE AIR COOLING CONDITIONS

Jun Dai, Doctor of Philosophy (Ph.D.), 2012

Directed By: Chair Professor and Director, Michael G. Pecht,

Department of Mechanical Engineering

In recent years, about 40% of the total energy is devoted to the cooling infrastructures in data centers. One way to save energy is free air cooling (FAC), which utilizes the outside air as the primary cooling medium, instead of air conditioning, to reduce the energy consumption to cool the data centers. Despite the energy saving, the implementation of free air cooling will change the operating environment, which may adversely affect the performance and reliability of telecom equipment.

This dissertation reviews the challenges and risks posed by free air cooling. The increased temperature, uncontrolled humidity, and possible contamination may cause some failure mechanisms, e.g., conductive anodic filament (CAF) and corrosion, to

be more active. If the local temperatures of some hot spots go beyond their recommended operating conditions (RoC), the performances of the equipment may be affected.

In this dissertation, a methodology is proposed to identify the impact of free air cooling on telecom equipment performance. It uses the performance variations under traditional air condition (A/C) to create a baseline, and compares the performance variation under variable temperature and humidity representing FAC with the baseline. This method can help data centers determine an appropriate operating environment based on the service requirements, when FAC is implemented. In addition, a statistics-based approach is also developed to identify the appropriate metric for the performance variations comparison. It is the first study focusing on the impact of FAC on the telecom equipment performance.

This dissertation also proposes a multi-stage (design, test, and operation) approach to mitigate the reliability risks of telecom equipment under free air cooling conditions. Specifically, a prognostics-based approach is proposed to mitigate the reliability risks at operation stage, and a case study is presented to show the implementation process. This approach needn't interrupt data center services and doesn't consume additional useful life of telecom equipment. It allows the implementation of FAC in data centers which were not originally designed for this cooling method.

RISK ASSESSMENT AND MITIGATION OF TELECOM EQUIPMENT UNDER FREE AIR COOLING CONDITIONS

By

Jun Dai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Michael G. Pecht, Chair
Professor Michael M. Ohadi
Professor Peter Sandborn
Associated. Professor Gary A. Pertmer
Associated. Professor F. Patrick McCluskey
Senior Research Scientist Dr. Diganta Das

Dedication

This dissertation is dedicated to my family and my parents, who always support me in my life.

Acknowledgement

I would express my deep and sincere thanks to my advisor, Prof. Michael Pecht, for his support and guidance all the time during my graduate study in CACLE. I thank Prof. Gary A. Pertmer, Prof. Sandborn, Prof. Michael M. Ohadi, Prof. F. Patrick McCluskey, and Dr. Diganta Das for serving on my dissertation committee and providing me valuable advice.

I want to express my special thanks to Prof. Michael M. Ohadi and Dr. Diganta Das. They provided a lot of great suggestions on my dissertation work. Their support and guidance are very important for this dissertation.

I am thankful to the CALCE staff, including Dr. Azarian, Dr. Osterman, Sony Mathew, Anshul Shrivastava, Bhanu Sood, Joan Lee, David Eisner and Mark Zimmerman, for their kindness and support.

I thank the current and former students the discussions and help provided in my research, including Jie Gu, Haiyu Qi, Lei Nie, Shunfeng Cheng, Weiqiang Wang, Xiaofei He, Hyunseok Oh, Yuxun Zhou, Yong Wang, Hongbo Bi, Anto Peter, Yan Ning, Wei He, Fei Chai, Minzhi Dong, Yunhan Huang, Preeti Chauhan, and Jing Tian.

Finally, I would like to thank my parents, my wife (Lei Wang) and my baby (Allen Dai) for their constant encouragement and support.

Table of Content

List of Figures

List of Tables

Chapter 1. Introduction

The telecommunication industry is concerned about the energy costs of its operating infrastructure, i.e., data centers. Data centers include the buildings, facilities, and rooms that contain enterprise servers, communication equipment, cooling equipment, and power equipment, and they generally provide the computing backbone of the telecommunication infrastructure [1].

1.1 Energy Consumption in Data centers

Approximately 1.5% to 3.0% of the energy produced in industrialized countries is consumed by data centers. A single data center can house hundreds or thousands of servers, storage devices, and network devices, and continued growth in the number of servers is expected as industries expand data center capabilities. Data centers are forty times more energy-intensive than conventional office buildings [2].

In 2007, the Environmental Protection Agency (EPA) published its report to the Congress on Server Data Center Energy Efficiency [3] to evaluate the energy consumption of government and commercial data centers in the US. This EPA report projects the near-term growth of energy consumption based on current trends and evaluates potential savings related to improving energy efficiency.

U.S data centers roughly doubled their energy consumption between 2001 and 2006, using 61 TWh[1] in 2006, at a costing of USD 4.5 billion [3]. The electricity

1 Tera is equal to 10^{12}

consumed by data centers is equivalent to the amount consumed by 5.8 million average U.S. households and is similar to the volume used by the entire U.S. transportation manufacturing industry [3].

In the European Union (EU), the electricity consumed by data centers is also expected to contribute significantly to total electricity consumption in the near future. The electricity consumed by data centers in Western Europe was about 56 TWh in 2007, and is projected to increase to 104 TWh per year by 2020 [1]. Furthermore, the data centers of online giants such as Google, Amazon and eBay can consume more energy than manufacturing companies. For example, the Google data center in Oregon is estimated to consume 103MW (Megawatts)[2] of energy in 2011 [4].

Data centers generally exhibit high power intensities, with increased density of servers. A survey of power consumption in more than twenty data centers in 2006 found that a data center's IT equipment, including servers, storage devices, telecom equipment and other associated equipment, can use from about 10 to almost 100 watt/sq.ft[3], over forty times more than is used by a conventional office building [3]. The energy consumption of a single rack of servers can reach up to 20 - 25kW, which is equivalent to the peak electricity demand of about fifteen typical California homes [3]. Moreover, about 50 percent of the energy consumed by data centers is devoted to the power and cooling infrastructure that supports electronic equipment [3][5], as shown in Figure 1.

---

2 MW is Megawatts, 10^6watts
3 watt/sq.ft is watts per square foot

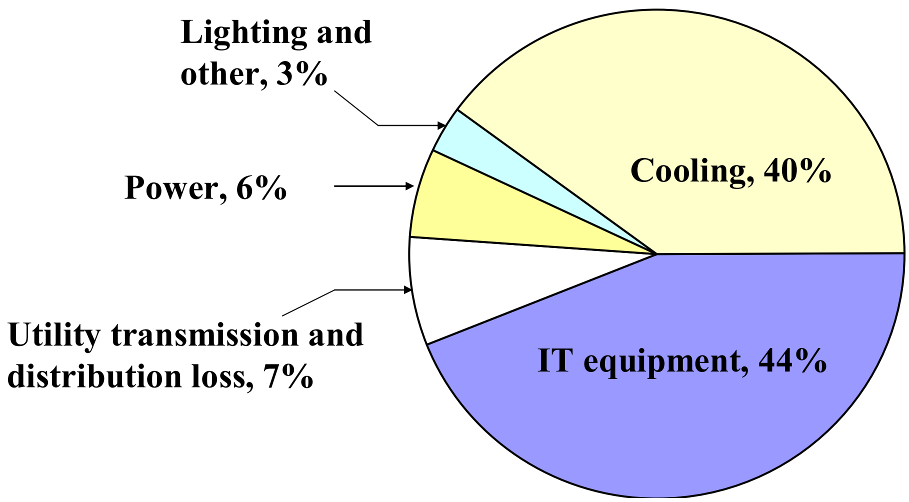


Figure 1: About 50% of the Energy Consumed by Data Centers is Used for Power and Cooling [5].

The public commitments of several leading companies on energy reduction are listed in Table 1. Besides the consideration for cost, energy reduction also can help increase the reputation of companies being socially responsible.

Table 1: Company Public Commitments to Energy Reduction [6]

Company	Commitment
Intel	Reduce normalized energy use in operations by 4% pa of 2002 level by 2010, and by annual 5% of the 2007 level by 2012.
Hewlett-Packard	Reduce energy consumption of desktop and notebook PC families by 25% (per unit) of 2005 level by 2010.
Nokia Siemens Networks	Reduce energy use of office facility by 6% of the 2007 level by 2012.
France Telecom	Reduce energy consumption by 15% below 2006 level to 2020.
Nokia	Reduce energy consumption of office facility to 6% of the 2006 level, by 2012.

1.2 Environmental Concerns

The telecom industry accounts for nearly 2% of global greenhouse gas emissions [6]. The "Smart 2020 Report" [6], published in 2008 by the Global e-Sustainability Initiative, a partnership of technology firms and industry associations, and the Climate Group, a non-profit environmental club, predict that the rapidly growing amount of carbon dioxide generated by information and telecommunication technologies could make them among the biggest greenhouse gas emitters by 2020.

In 2007, the gas emission of telecom industry was 830 Mt (megatons, 10^6 ton) CO_2e (CO_2 equivalent), about 2% of the global emissions from human activity in the world. Even if efficient technology is developed to reduce the energy consumption, the gas emission by telecom industry will increase at an annual rate of 6% until 2020, which will reach 1.43 Gt (gigatons, 10^9 tons) of CO_2. About one quarter of the gas emission comes from the telecom equipment materials and manufacturing processes, and the rest is generated during their operation.

Generally, there are three subsectors in the telecom industry: PC and peripherals (workstations, laptops, desktops and peripherals such as monitors and printers), data centers and Telecoms devices (fixed line, mobile phones, chargers, internet protocol TV (IPTV) boxes and home broadband routers). PCs are already widely used in developed countries, however, they are experiencing an explosive growth in some developing countries. The emerging middle classes in China and India are purchasing PCs in a manner similar to how their counterparts in developed countries did before, and as a result, this will substantially increase the gas emission due to the large population. The carbon footprint of PCs and monitors is expected to be 643 $MtCO_2e$ in 2020, with an annual growth of 5% based on 200 $MtCO_2e$ in 2002 (the peripherals gas emission will be about 172 $MtCO_2e$ in 2020). Another main contributing factor would be the growth in the number and size of data centers. The "information age" requires a great amount of servers to store data and make them instantly available upon request. The carbon footprint of data centers is projected to reach 259 $MtCO_2$ emissions by 2020, compared with 76 Mt CO_2 emissions in 2002. The telecom device gas emission is also expected to increase to 349 $MtCO_2$ in 2020. The growths of the three subsectors in telecom industry are shown in Figure 2.

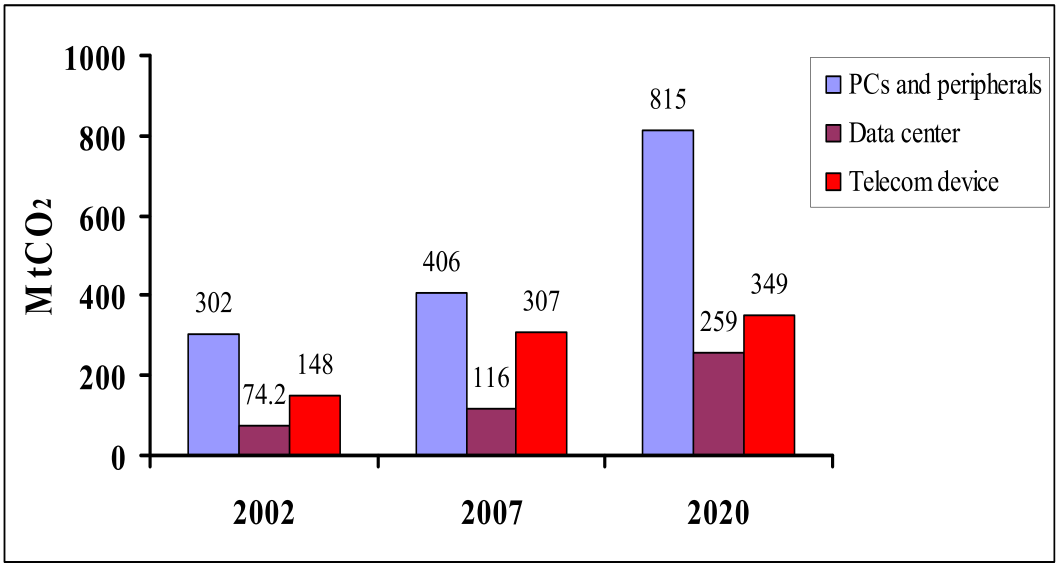


Figure 2: the Gas Emission Growth of Global Telecom Industry Subsectors [6]

The public environmental commitments of several prominent companies are listed in Table 2 [6]. These commitments show how corporations are seeking to improve their environmental impact.

Table 2: Public Environmental Commitments of Companies [6]

Company	Commitment
British Telecommunications	Reduce the worldwide CO_2 emissions per unit of BT's contribution to GDP by 80% from 1996 levels by 2020, and reduce UK CO_2 emissions in absolute terms by 80% below 1996 levels by Dec 2016.
Microsoft	Every two years through 2012, halve the difference between annual average data centre PUE[4] and the ideal PUE (1.0).
Sun	Reduce US CO_2 emissions 20% from 2002 levels by 2012.
Alcatel–Lucent	Reach a 10% reduction in total CO_2 emissions from facilities from the 2007 baseline by the end of 2010.
Dell	Reduce operational carbon intensity[5] by 15% from 2007 levels by 2012.

1.3 Cooling Methods to Improve Energy Efficiency

The pressures of energy costs and environmental concerns are forcing the telecom industry to find ways to improve energy efficiency in data centers' electrical equipment in general. Several cooling methods already are in use: air conditioning with new power-management technologies, liquid cooling, tower free cooling and free air cooling. This chapter introduces and compares the various applicable cooling methods to assist in the appropriate selection under different environment situations.

[4] PUE is defined as the ratio of the total power drawn by a data center facility to the power used by the IT equipment in that facility. It is a key metric of energy efficiency advocated by Green Grid, a global consortium dedicated to advancing energy efficiency in data centers and business computing ecosystems.

[5] Carbon intensity is the total carbon dioxide emissions from the consumption of energy per dollar of gross domestic product (GDP).

1.3.1 Air-Condition with New Power Management Technologies

Air conditioning is currently the dominant cooling method in data centers, with room temperature usually set at a fixed temperature. However, new power measurement and management technologies have been developed to monitor, manage, and improve the energy efficiency of air conditioning unit. One example is IBM's Measurement and Management Technologies (MMT) [7].

MMT is a tool set that helps visualize and understand the thermal profile of an existing data center and its power and cooling systems. MMT provides a detailed assessment of the heat distribution throughout the center by creating a three-dimensional chart that pinpoints power and cooling inefficiencies. MMT also provides a real-time solution for monitoring and managing the cooling and energy efficiency of the center. After a measurement survey, sensors are installed and coupled to a software system encoded with the survey results to provide ongoing reporting and analysis of the room conditions. This facilitates setting optimal energy and cooling system levels and reduces over-provisioning and over-cooling. This in turn can allow an increase of data center densities and delay capital expenditures on power and cooling improvements. The key benefits include reducing data center costs by improving cooling efficiency and the availability and reliability of new and existing IT systems [7].

Under collaboration with IBM, a five-month test of MMT was implemented in Toyota Motor Sales and Southern California Edison, one of the largest electric utilities in U.S.[8] The test was executed in Toyota's 20,000-square-foot Torrance data center. MMT showed Toyota where to make quick, inexpensive changes to air flow and temperature set points in the room. It allowed Toyota to safely shut down two computer room air conditioners, resulting in significant energy and cost savings. The test results showed that air conditioning was reduced by thirty percent, including an overall reduction in hot spots and a cooler consistent ambient temperature throughout the center. Southern California Edison, Toyota Motor Sales' energy provider, quantified the company's energy savings as a demand reduction of more than ten percent, and recommended MMT for broad application by its clients for energy and demand reductions [8].

1.3.2 Liquid Cooling

Liquid cooling is a traditional cooling method with an extensive history in data centers. The liquid can be water, Freon or another coolant. Some variants of liquid cooling, such as heat pipes or vapor chambers, are increasingly used in data centers, but they are considered as air cooling methods, since it is server airflow to remove the heat instead of liquid [9].

Liquid usually has large heat transfer coefficient and can carry great heat during the heat exchange process, and then liquid cooling is an appropriate solution for high-density components, such as CPU. However, the implementation of liquid cooling increases the overall cost of data centers significantly, and as well as the complexity of the operational use models of IT equipment. Some companies suggest that the use of liquid cooling is limited to special high-density equipment/components [9].

There are several implementation methods of liquid cooling to the rack. One involves a liquid-cooled door, which is usually located on the back of the rack and cools the air flowing from the rack to (or near) ambient room temperature, and then remove the heat. Another implementation method is a closed-liquid rack that is sealed; which uses a heat exchange to remove the heat from the airflow fully contained within the rack. The heat exchange is connected with a liquid cooling system and transfer the heat to liquid. This design is thermally and airflow neutral to the room, and usually also quiet. However, when the closed-liquid system fails (although it is a rare event), the rack would need to be opened manually to prevent overheating [9].

Another implementation method of the related rack-cooling strategy are in-row liquid coolers and overhead liquid coolers, which are similar as the liquid-cooled door. These two methods remove the heat very near the heat sources and the room-level air-cooling system are not stressed by heat load, although there is local room airflow. The advantage is that they are rack-independent, and not limited to some specific server or rack manufacturers. However, a disadvantage of both the methods is that they occupy a lot of racks or overhead real estate. The trade-off need be considered in their implementation [9].

Another liquid cooling technology is offered by Iceotope [10], a manufacturer of liquid cooling equipment. In this method, the server motherboard is completely immersed in an individually sealed bath of an inert liquid coolant. The generated heat is removed by the coolant from the sensitive electronics to a heat exchanger, which is formed by the wall of the bath. The coolant is continuously re-circulated and cooled in the bath. Icetope claims this cooling technology can dramatically reduce data center energy costs if IT managers can get comfortable with the idea of liquids in their centers.

A great potential risks is the leaking of the liquid pipes close to the IT equipment, and it usually can cause the system failures even if the liquid may evaporate, which should be taken into consideration when the cooling method is decided. One possible solution is using a leak detection system with the liquid cooling implementation although it needs additional cost [9].

1.3.3 Free Air Cooling

Free air cooling bypasses the traditional air conditioner by pulling in cold air from the outside, cycling it through the server racks, and flushing it back outside. This is usually implemented by an airside economizer, which brigs outside air into the building and distributes it using a series of dampers and fans.

Intel conducted a ten-month test to evaluate the impact of using only outside air via air-side economization to cool a high-density data center in New Mexico in October 2007 [11]. The center had nine hundred heavily utilized production servers. In this test, the system provided one hundred percent air exchange, a temperature variation in the supply air from 18^{o}C to more than 32^{o}C, with no humidity control (4% RH to more than 90% RH) and minimal air filtration. The result showed that about USD 2.87 million (a 67% savings of total energy) was saved by the new cooling method. Internet giants, such as Google and Microsoft, have used free air cooling in their data centers. Google operates such a data center in Belgium and the room temperature can be maintained above 27°C [12], which can allow the application of free air cooling during most of a year. Microsoft operates a data center with free air cooling in Dublin, Ireland, in which the room temperature can reach 35°C [13]. Vodafone runs its base stations at a standard temperature of 35ºC now, rather than the previous norm of 25-30ºC, in order to save energy for cooling. It claims that the higher temperature did not reduce the equipment's reliability or life expectancy. Vodafone plans to replace air-conditioning with free air cooling in the majority of its base stations between 2008 and 2011 as part of a plan to reduce its carbon footprint by fifty percent between 2006 and 2020 [14].

1.3.4 Tower Free Cooling

Beside free air cooling with airside economizer systems, there is another cooling method tower free cooling (or simply free cooling) for data centers. This cooling method is usually implemented by waterside economizers, which conjunct with a cooling tower evaporative cooler or a dry cooler remove the heat from the rooms. A waterside economizer system has cooling coils to cool the room air and then carry the heat to the heat exchanger, which is connected with a cooling tower outside the room with cool water. This process is shown in Figure 3.

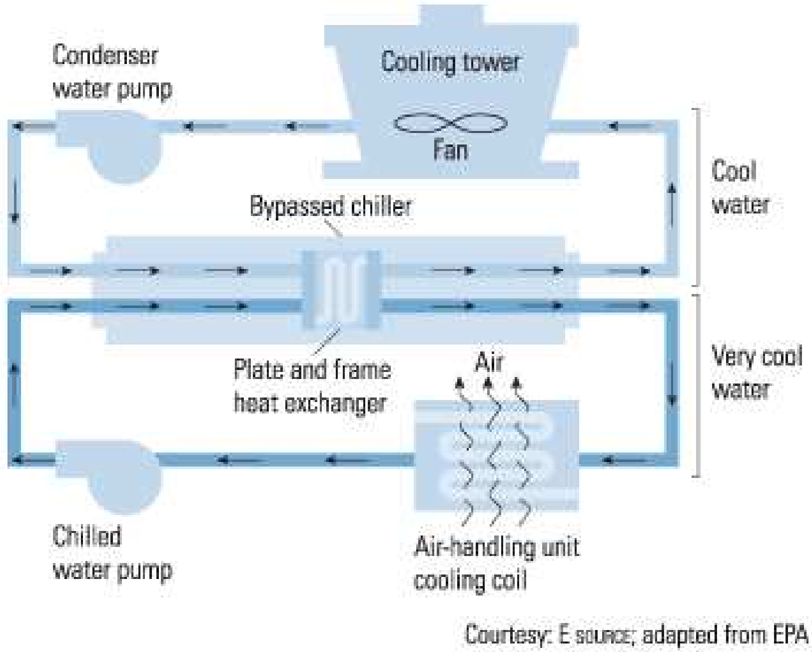


Figure 3: Typical Waterside Economizer [17]

Airside economizers are typically used for more hours than waterside economizers since free air cooling is used in mild conditions, whereas tower free cooling can only be used in cold conditions. Airside economizers are also easier to design, operate, and maintain than waterside economizers, which are complicated and include cooling towers and heat exchangers. However, there are still some reasons to choose tower free cooling over free air cooling. For example, waterside economizers (WSE) do not require large outside air and relief air ducts and associated louvers. This makes them a prime candidate for, say, a large historic building where the building envelope cannot be altered. Tower free cooling can be also used where it may not be practical to create large floor openings in some facilities to accommodate the outside air and relief ducts.

Wells Fargo's bank introduced tower free cooling for its data center in Minneapolis, Minnesota, in 2005, and achieved significant energy savings. The added investment was $1 million, due to the implementation of tower free cooling, about one percent of the total construction costs. The waterside economizer was used when the outside air temperature dropped to about 2^{o}C, and can be operated about four months a year. The energy savings amounted to $150,000 in 2006 and up to $450,000 a year subsequently as the bank continued to expand operations [15].

1.3.5 Comparison

An appropriate cooling method must be selected to maximize energy efficiency. To assist with this decision, this subchapter compares these methods and identifies their advantages and disadvantages.

Energy Efficiency

The energy efficiency of air conditioning with the new power management technologies is only moderate, since the efficiency is better than with traditional A/C, but it doesn't utilize the cool air outside the room which means more energy saving. IBM claims liquid cooling is very efficient for high-power density subsystems (e.g., CPUs) due to the high heat transfer coefficients [16], but Intel doubts its efficiency for entire data centers with many low-density units [9]. Intel [11], Google [12], and Microsoft [13] consider that the energy reduction with free air cooling is highly efficient, and Google and Microsoft have operated new data centers with free air cooling. For the Fargo bank, tower free cooling has also proved very energy-efficient.

Flexibility of Equipment Location

Air conditioning with new power management technologies is very flexible and the equipment can be put almost anywhere in the room. However, liquid cooling is not flexible, since the equipment can only be installed near access to the liquid pipes. Free air cooling and tower free cooling have high flexibility, much like air conditioning with new power management technologies.

Retrofit Cost

The cost of retrofitting air conditioning with new power management technologies is moderate. Generally, the retrofit costs for liquid cooling are higher than for other cooling methods, because the pipes for liquid recirculation must be installed, and sometimes re-installed. For example, when Icetope is installed, motherboards need be removed from the servers and then completely immersed in a sealed bath of coolant, which results in high costs in existing data centers with traditional A/C. Retrofitting free air cooling and tower free cooling entails moderate costs, since airside economizers, waterside economizers and associated pumping equipment and filtration equipment are needed.

Weather Dependence

Air conditioning and liquid cooling with new power management technologies are not weather-dependent, but both free air cooling and tower free cooling have high weather dependence. Mild weather conditions can maximize the operating hours of airside economizers and cold weather conditions can maximize the operating hours of waterside economizers. This comparison is summarized in Table 3.

Table 3: Comparison of Cooling Methods

	A/C	Liquid Cooling	Free Air Cooling	Tower Free Cooling
Energy efficiency	Medium	High for high power density subsystem but medium for whole data centers	High	High
Flexibility	High	Low	High	High
Retrofit cost	Medium	High	Medium	Medium
Weather dependence	Low	Low	High (mild)	High (cold)

1.4 Summary:

The telecom industry has become more concerned with its energy consumption and the associated environmental effects. Since about half of the total energy consumption in the telecom industry is devoted to power and cooling equipment in data centers, there is a great opportunity to modify cooling methods to improve the energy efficiency of telecom industry. This can have the benefit of not only meeting environmental requirements, but can also help lower operating cost.

Cooling equipment provides great opportunities to improve the energy efficiency by modifying the cooling methods. Free air cooling is one of the most promising cooling methods for data centers and has been adopted by some leading companies, although it has a high dependence on the local weather. In order to maximize the energy efficiency, data centers may implement several cooling methods in various seasons in a year.

References:

[1]. European Commission, “Code of Conduct on Data Centres Energy Efficiency—Version 1.0”, Oct. 2008.

[2]. M. Hodes, et al., “Energy and Power Conversion: A Telecommunications Hardware Vendor’s Prospective,” Power Electronics Industry Group (PEIG) Technology Tutorial and CEO Forum, Dec. 6, 2007, Cork, Ireland.

[3]. U.S. Environmental Protection Agency Energy Star Program, “Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431,” Aug. 2, 2007.

[4]. J. Cradden, “The Greening of the IT Sector,” Technology Ireland, Dec. 2009.

[5]. P. Johnson, and T. Marker, “Data Center Energy Efficiency Product Profile”, Pitt & Sherry, Report to Equipment Energy Efficiency Committee (E3) of The Australian Government Department of the Environment, Water, Heritage and the Arts (DEWHA), Apr., 2009.

[6]. The Climate Group on Behalf of the Global eSustainability Initiative (GeSI) “SMART 2020: Enabling the Low Carbon Economy in the Information Age,” Brussels, Belgium, 2008.

[7]. International Business Machines (IBM), “IBM Measurement & Management Technologies (MMT) Data Center Thermal Analysis,” IBM Systems Lab Services and Training Solutions Brief, 2010.

[8]. International Business Machines (IBM), “IBM Collaborates with Toyota Motor Sales, U.S.A., Inc. and Southern California Edison to Create Green Data Center”, Press Release, Oct. 13, 2009.

[9]. M. K. Patterson and D. Fenwick, “The State of Data Center Cooling,” Intel Corporation white paper, Mar, 2008.

[10]. R. Miller, “Iceotope: A New Take on Liquid Cooling”, Data Center Knowledge, Nov, 2009.

[11]. D. Atwood , and J. G. Miner, “Reducing Data Center Cost with an Air Economizer”, Intel Information Technology White Paper, August 2008,

[12]. R. Miller, “Google’s Chiller-less Data Center,” Data Center Knowledge, July 2009.

[13]. R. Miller, “Microsoft’s Chiller-less Data Center,” Data Center Knowledge, September 2009.

[14]. The Economist, “How Green Is Your Network?” Dec. 4, 2008.

[15]. R. L. Mitchell, “Case Study: Wells Fargo's Free Data Center Cooling System”, Computer World, Nov.5, 2007.

[16]. S. O'Donnell, “IBM claim that water cooled servers are the future of IT at scale”, the Hot Aisle, Jun. 3, 2009.

[17]. U.S. EPA, “Heat and Cooling”, Energy Star Program, Jan, 2008.

Chapter 2. Free Air Cooling

Free air cooling (chiller-less cooling) is one of the best options for significantly cutting cooling costs in data centers, and is increasingly being used to improve energy efficiency. Free air cooling is usually implemented using air economizers. With an economizer, when the outside air is cooler than the return air, the hot return air is exhausted and replaced with cooler, filtered outside air, essentially "opening the windows" for free cooling. This chapter introduces the implementation of free air cooling.

2.1 Introduction of Airside Economizer

There are slight differences between airside economizers designed by different companies[1][2][3]. This chapter introduces a typical air economizer in detail and describes its method of operation. An airside economizer system regulates the use of outside air for cooling the equipment in a data center, as shown in Figure 4. It takes advantage of the time in the year when the ambient conditions are favorable and provides a free air cooling cycle to meet the cooling requirements. Generally, it consists of sensors, ducts, dampers and containers that supply the appropriate volume and temperature of air to satisfy cooling demands. Before airside economizers are used, a temperature range for the supply air temperature must be set; this can be controlled by the airside economizer [1]. The outside air is brought into the containers and then distributed to cool the equipment in data centers via a series of dampers and fans. The supply air cools the equipment, transfers heat, and then returns to the containers of the air economizers. Instead of being recirculated and cooled, the exhaust air is simply directed outside. The sensors measure the outside and inside air conditions. If the outside air is particularly cold and beyond the set temperature range of the supply air, the economizer will mix the inlet and exhaust air, ensuring that the temperature of the supply air is within the set range. But if the outside air is far outside the set range, the airside economizers cannot work and air conditioning is needed. Thus, the set temperature range of the supply air determines the operating hours of airside economizers in a year.

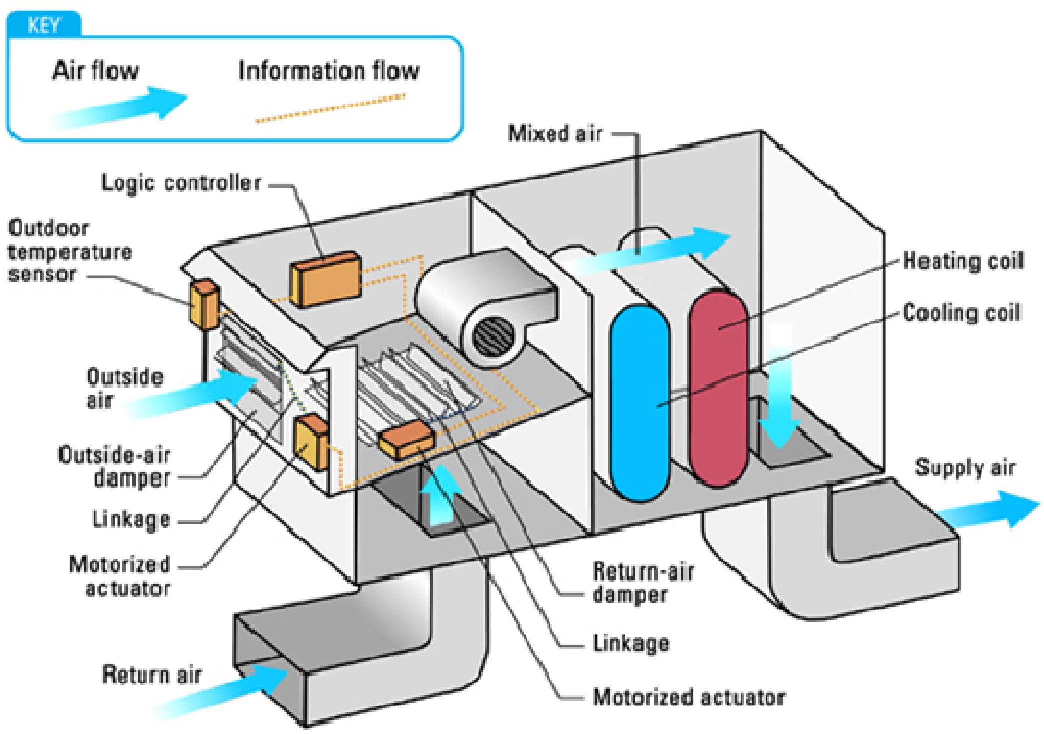


Figure 4: Schematic of Airside Economizer with Airflow Path [12]

The dampers used for closing and opening the economizer windows, improved filters, fans, actuators, logic modules, and sensors add to the cost of an airside economizer. Though the costs of the logic controller and sensors are for the most part independent of the economizer size, the cost of all other components depends on the size of the data center and the amount of outside air it brings in. The implementation of free air cooling also incurs the cost of installation. An example of the costs for an airside economizer is shown in Table 4.

Table 4: Example: Costs of Airside Economizer [8]

Components	**Cost**
Dampers	USD 2000
Honeywell Actuator ML8185	USD 500
Honeywell W7215 Logic Module	USD 500
Honeywell C7400A1004 Enthalpy sensors	USD 272
Honeywell C7835A Discharge Sensors	USD 300
Air Inlet fan	USD 400
Total	USD 3972

2.2 Operating Environment Setting

The operating environment setting is the key factor in the implementation of free air cooling. The operating environment affects not only the reliability of telecom equipment, but also the operating hours per year of the airside economizers. An appropriate operating environment setting should consider a survey of the climate, equipment specifications, standards, and identified hotspots.

Weather Survey

Surveying the weather is necessary for determining the operating environment. Obviously, the weather varies greatly throughout the year from place to place.

Generally, once the operating environment has been set, the local weather conditions determine the operating hours per year of airside economizers. When the ambient air conditions exceed the range of the set operating environment, the heating or cooling coils of the economizers adjust the air to meet the set temperature. However, some alternatives have been implemented. For example, one Google data center in Belgium doesn't use chillers or heating coils at all [5]. When the weather gets hot, it redirects the workload, turning off equipment as needed and shifting the computing load to other data centers. The weather information for some cities can be found from Climatetemp.info (http://www.climatetemp.info/), and the examples of four different cities (Seattle, New York, Beijing, Shanghai) are shown in Table 5~6. Notice that New York and Beijing have lower temperatures than the others, and that Seattle and Shanghai have higher humidity levels than the other two.

Table 5: New York Weather Average [13]

Month	Temperature (^oC)			Humidity (RH)
	Avg Min	Avg Max	Avg	
Jan	-3	4	0.5	61
Feb	-2	4	1	58
Mar	1	9	5	56
Apr	6	15	11	55
May	12	21	17	56
Jun	17	26	22	58
Jul	20	28	24	57
Aug	19	27	23	60
Sep	16	24	20	60
Oct	10	18	14	59
Nov	4	12	8	60
Dec	-1	6	2.5	60

Table 6: Beijing Weather Average [13]

Month	Temperature (^{o}C)			Humidity (RH)
	Avg Min	Avg Max	Avg	
Jan	-10	1	-4.5	50
Feb	-1	8	3.5	50
Mar	2	12	7	48
Apr	7	21	14	46
May	13	27	20	49
Jun	18	31	25	56
Jul	21	31	26	72
Aug	20	30	25	74
Sep	14	26	20	67
Oct	6	20	13	59
Nov	-2	10	4	56
Dec	-8	3	-2.5	51

Standards Requirements

Standards are an important reference for determining the operating environment setting. Generally, the recommended temperature range in the standards is 18~27^{o}C, and the allowable temperature is 5~40^{o}C. The allowable humidity range is 5%~80% (The details will be introduced in chapter 4). The purpose of the recommended envelope is to give guidance to data center operators on maintaining high reliability and energy efficiency. Telecom equipment manufacturers test their equipment within the allowable range to verify that it will function within those boundaries. Typically, manufacturers will perform a number of tests prior to announcing a product to be sure it meets all the functionality requirements within the range. This is not a certification of reliability,) but the recommended range is a statement on reliability. The IT manufacturers recommend that data centers maintain their environments within the recommended envelope for extended periods of operation. Exceeding the recommended limits for short periods of time should not cause a problem, but running near the allowable limits for months could increase reliability issues [6].

Equipment Specifications

The specifications for the equipment in data centers are important references for setting the operating environment ranges. The specifications can usually be found in the datasheets for individual equipment items. For example, the operating temperature of Cisco 3600 Series routers is specified as 0~40^{o}C, and the humidity range is 5~95% [7]. Note that the operating environment requirements in datasheets refer to the local operating environments, rather than the inlet temperature of supply air. In addition, it may be that the specifications are higher than the standard limits. When this happens, comprehensive consideration is needed in order to make the determination. Generally, standard operating environment limits are assumed to be the foundational specifications by both equipment manufacturers and customers. However, the manufacturer and customer may agree on additional specifications that may allow the equipment to be operated at higher specifications than normal.

Identification of Hotspots

Hotspots are among the most dangerous points in data centers. When the operating environment is being set, it is important to consider the effect of free air cooling on hotspots. The operating temperatures at those places must be anticipated, and be reflected in the datasheet specifications. The temperatures at hotspots should be continually monitored during operation.

2.3 Operation of Airside Economizers

Figure 5 shows the logic used by the controller for deciding the opening and closing of the airside economizer dampers. Heat and enthalpy sensors are located outside as well as inside the data center. The control logic looks at the differential of the sensible heat of the return air and the outside air, as shown in Figure 5 [8].

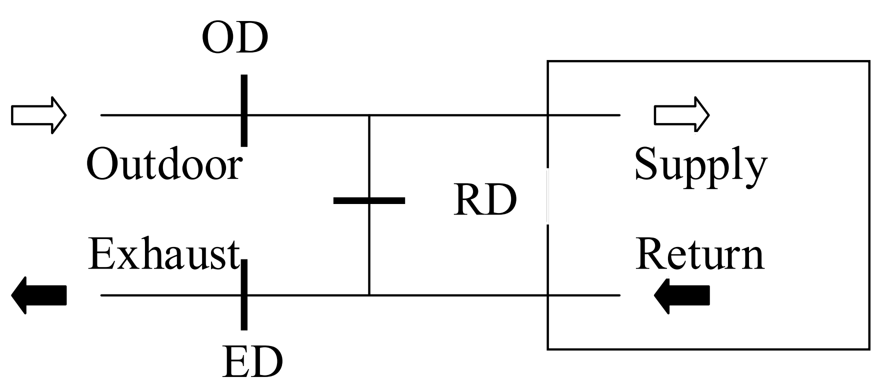


Figure 5: Control logic of the airside economizer for damper actuation [8]

Figure 6, 7, and 8 show when the outside air can be used. For example, when the outside air temperature (OAT) is higher than the return air temperature (RAT), no outside air is brought into the data center space, as shown in Figure 6. In this case, the outside air damper (OD) is completely closed, resulting in no economization. The cooling coils inside airside economizers are used to cool the return air [8].

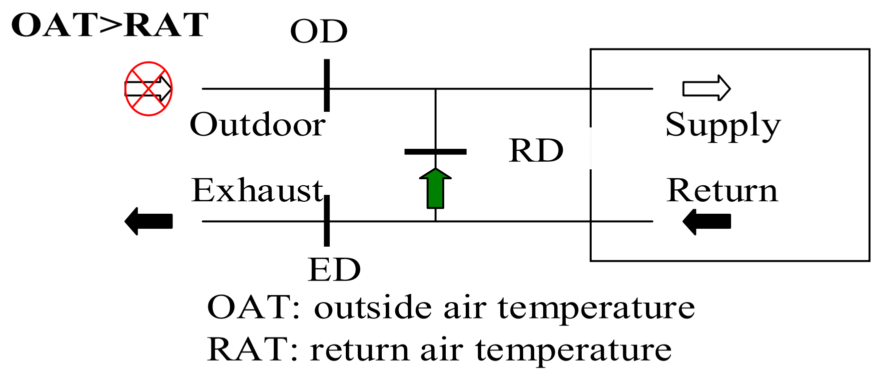


Figure 6: Case for which no outside air cooling is available from the economizer [8]

Figure 7 illustrates when the outside air temperature (OAT) is sufficiently lower than the return air temperature (RAT). In this case, the return air damper (RD) is completely closed, allowing no return air to mix with the outside air. This results in partial economization and additional cooling should be provided with the chillers.

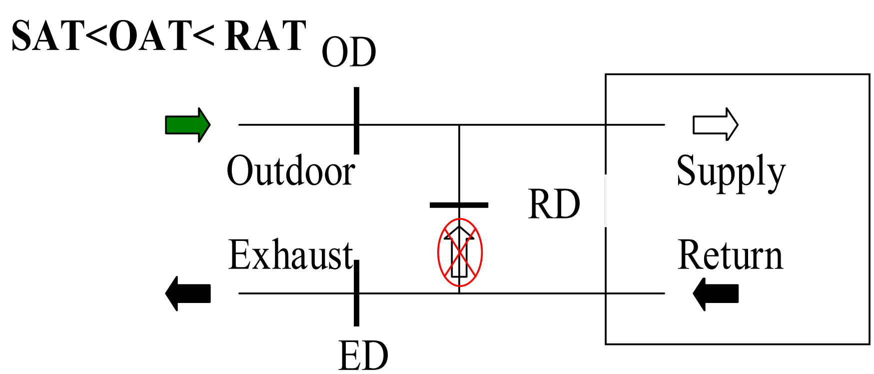


Figure 7: Case for which partial outside air cooling is available from the economizer [8]

Figure 8 shows the case when the outside air temperature (OAT) is lower than the supply air temperature (SAT) requirement, thus outside air is brought into the data center space. In this case the return air is mixed with the outside air so that the temperature falls within the set operating environment range. If the outside air temperature is extremely cold and the air mixed with return air is still lower than the

SAT requirement, the heat coils inside the airside economizers are used to warm the mixed air [8].

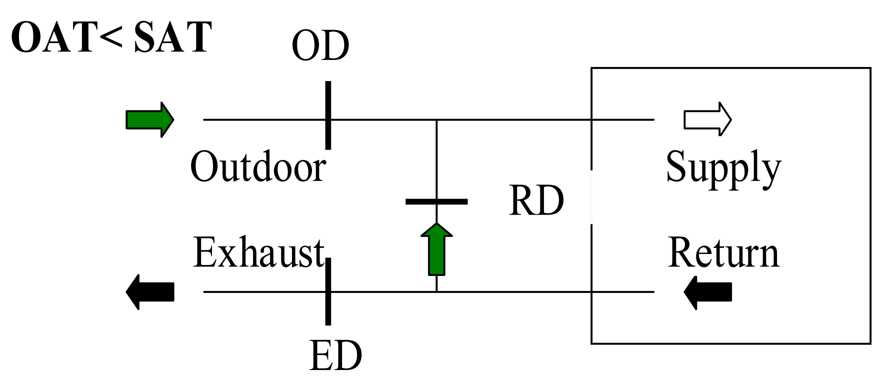


Figure 8: Case for which 100% cooling is available from the economizer [8]

2.4 Benefits of Free Air Cooling

The energy saving of free air cooling depends on the local climate of data centers and the design characteristics. In 2008, the Department of Civil and Environmental Engineering in University of California Berkeley and Lawrence Berkeley National Laboratory Berkeley published a report to estimate the energy saving in several climate zones of California [10]. In order to identify quantitatively the energy saving, the report compared free air cooling (with airside economizers cooling) with tower free cooling (with waterside economizers cooling) and baseline (with traditional air conditioning unit cooling) based on energy models and simulations. This subchapter introduces the energy saving benefits of free air cooling in California as an example.

2.4.1 Data Center Design Scenarios

In this paper, the baseline considers the traditional "computer room air conditioning" (CRAC) units. In this case, CRAC units are placed on the server room floor. The air is cooled by entering the top of CRAC and passing over the cooling coils, and then is discharged into the underfloor plenum. The cold air in the underfloor plenum goes through the perforations in the floor tiles located in front of the server racks, and pass across the server racks to remove their heat with the help of the servers. The exhausted air exits the backside of the server racks and becomes warm, which will rise to the intake of the CRAC unit. In the baseline scenario, the air circulation is usually internal to the data center. A rooftop air handling unit (AHU) provides a small amount of air to positively pressurize the room and supply outside air for occupants. Refrigerant in water-cooled chiller plant use heat exchangers to cool water from the CRAC units of the data center. The chiller refrigerant uses heat exchangers to transfer the waste heat to the condenser water piped from the cooling towers, in which the warm water can be cooled by the outside air. This baseline design is widely used in most mid-size and large-size data centers [10].

In the water-side economizer (WSE) scenario, it includes a CRAC unit similar to that of the baseline scenario, except that additional heat exchangers are installed between the chilled water supplied to the CRAC units and the condenser water in the cooling towers. If the local climate is cold enough, the chiller plant can be not used, and the condenser water in the cooling towers can is cold enough to cool directly the chilled water supplied to the CRAC. The CRAC units and chiller plant are the same as those in the baseline scenario. The energy saving is, therefore, through replacement of compressor-driven chilling with fan-driven evaporation cooling [10]. The schematic is shown in Figure 9.

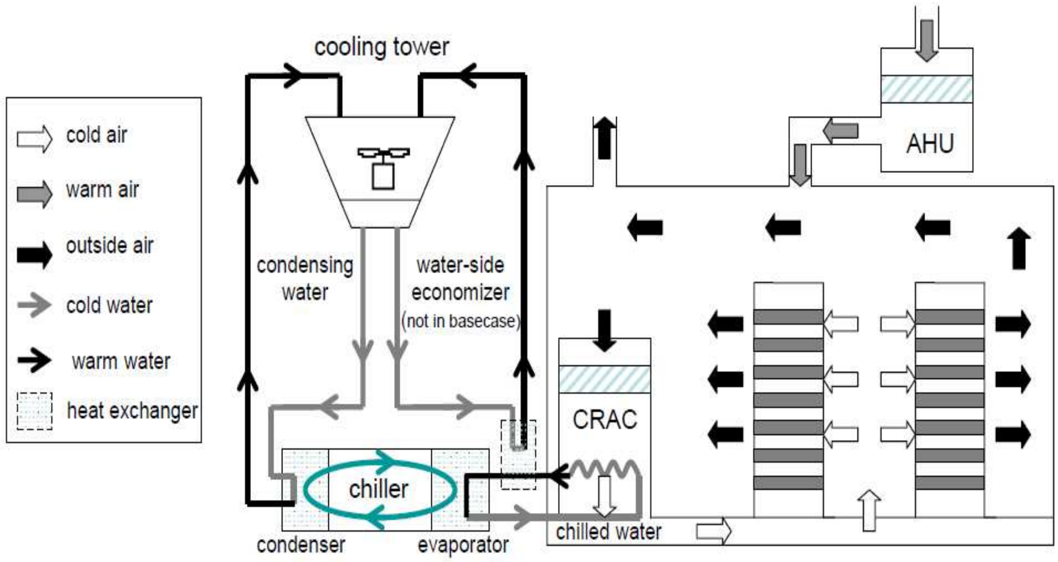


Figure 9: Schematic of waterside economizer scenario [10]

In the air-side economizer (ASE) scenario, there are some differences of air delivery from the traditional CRAC units in typical data centers. AHUs are usually placed on the rooftop which is outside of the data center room, and some ducts are used to deliver air to and from the server racks. The ducted system design can prevent the cold air and warm air from unintentionally mixing in the data center, but it has greater air resistance other that in a traditional CRAC unit. When the outside air temperature is inside the set range, the AHU directly supply the outside air into the data center, pass over the servers, and move the return air with heat removal to the outside of the room. In this process of 100% outside air cooling, fans consume more energy than that in the baseline case. However, the economizer design saves the energy for operating the chiller, chilled water pumps, and the cooling tower fans. If the outside air temperature is higher than the set range, the chiller needs be operated like the baseline case [10].

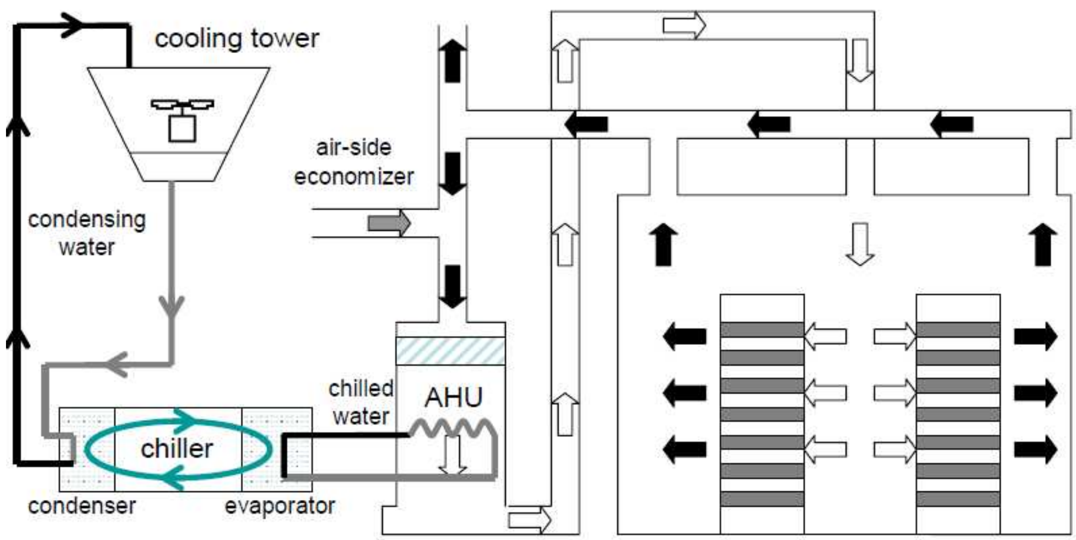


Figure 10: Schematic of airside economizer scenario [10]

2.4.2 Energy Modeling Protocol

The model calculations assume that each design is implemented in (2800 m^2) data center with a size of a 30,000 ft^2. The heat density of the data center is assumed to about 80 W/ft^2 (0.86 kW/m2, 2.4 MW total), which is currently considered to be of low-range to mid-range. Table 7 shows the data center basic properties in the all three scenarios. Total Energy demand is calculated as the sum of energy consumption by servers, chiller use, fan operation, transformer and uninterruptible power supply (UPS) losses, and building lighting [10].

Table 7: Data center characteristics common to all design scenarios [10]

Data Center Parameters	
Floor Area	30000 ft^2
UPS Waste Heat	326 kW
Data Center Lights	30 kW
Total Rack Load	2000 kW
Total Internal Load	2,356 kW
Average Internal Load Density	79 W/ft^2
Minimum Ventilation	4,500 ft^3/min
Supply Air Temperature	13^oC
Return Air Temperature	22^oC
Chiller Capacity	1750 kW
Number of Chillers	3

The chiller system includes coolant compressor, chilled water pumps, condensing water pumps, humidification pumps, and cooling-tower fans. Energy consumption of servers, UPS, and lighting are considered as constant in the three design scenarios. The humidity is conventionally restricted by ASHRAE 2005 (40%~55%) in the baseline case and WSE scenarios, and is typically not restricted in the ASE scenario. Air-side economizer scenario also has a different fan parameters listed in Table 8.

Table 8: Data center fan properties [10]

Fan System Parameters	**Baseline and WSE**			**ASE**	
	MUAH	Exhaust	CRACs	Supply	Relief
Total Air Flow (cfm)	4500	4500	49500	437758	437758
Fan Motor Size, Nominal (hp)	7.5	3	10	30	50
Number of Fans	1	1	30	10	5
Fan Efficiency	53.3%	44.0%	55.6%	63.8%	67.5%
Fan Drive Efficiency	95%	95%	95%	95%	95%
Fan Motor Efficiency	89.6%	86.2%	90.1%	92.5%	93.2%
VFD Efficiency	n/a	n/a	n/a	98%	98%
Total Static Pressure Drop	3.5	1	1.6	2	1

2.4.3 Result

This model considers five cities in California (Sacramento, San Francisco, San Jose, Fresno, Los Angeles) and assumes that a data center is located in each city. The annual energy consumption of each data center is calculated based on the three design scenarios, and the ratio of total data center energy to server energy consumption is also calculated as in Table 9. In the baseline scenario, the performance ratio of building energy consumption over server energy consumption is 1.55, which is the same for all of the five data centers, since the operation under this design is practically independent of outdoor weather conditions. The performance ratios of the ASE scenario show air-side economizers can reduce the energy consumption compared with the baseline case. The WSE scenario can save energy compared with the baseline, but the saving is less than that of the ASE scenario. In a data center with a large amount of energy consumption, a small change in the performance ratio represents a significant saving. For example, the performance ratio change from 1.55 to 1.44 in San Jose data center can save energy of about 1.9 million kWh/y, equivalent to about $130,000/y (assuming $0.07/kWh) [10].

Table 9: Ratio of total building energy to computer server energy (PUE) [10]

	San Jose	San Francisco	Sacramento	Fresno	Los Angeles
Baseline	1.55	1.55	1.55	1.55	1.55
ASE	1.44	1.42	1.44	1.46	1.46
WSE	1.53	1.54	1.53	1.53	1.54

The energy consumption of the five data centers under the three design scenario is shown in Figure 11. The result shows that the ASE scenario savings in San Francisco provides the greatest energy saving while that in Fresno provides the least energy savings. Under the WSE scenario, the data center in Sacramento obtains the greatest benefits while those in Los Angeles and San Francisco gain the minimal energy savings. The San Francisco WSE scenario would be expected to gain significant savings due to the cool climate, but the chiller part-load inefficiencies reduced the saving. The San Francisco air contains a relatively higher moisture content which increases the latent cooling load in the model and often reaches the capacity limit of the first chiller plant, and a second chiller needs be activated. Since the cooling load is equally shared by the two chillers and the cooling is transferred from the first chiller to the second chiller, both chillers have cooling loads slightly above half their capacity limits, which results in inefficiencies of the chillers. So the data center with the WSE scenario in San Francisco requires modeling the hour-by-hour load of the chiller instead of the peak load, and then operates the appropriate number of the chillers to maintain the chillers near their most efficient operating point at any moment [10].

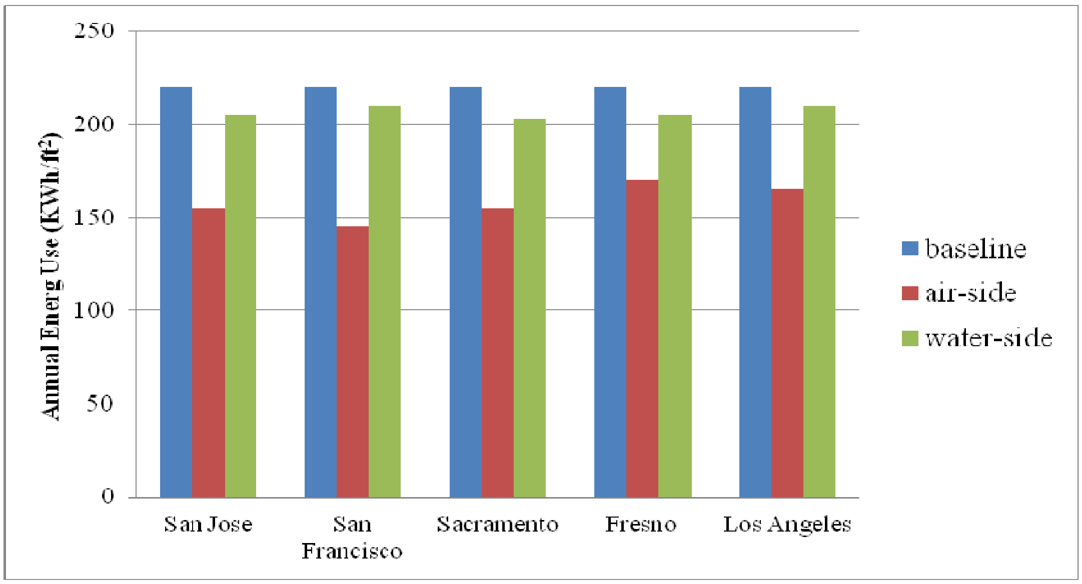
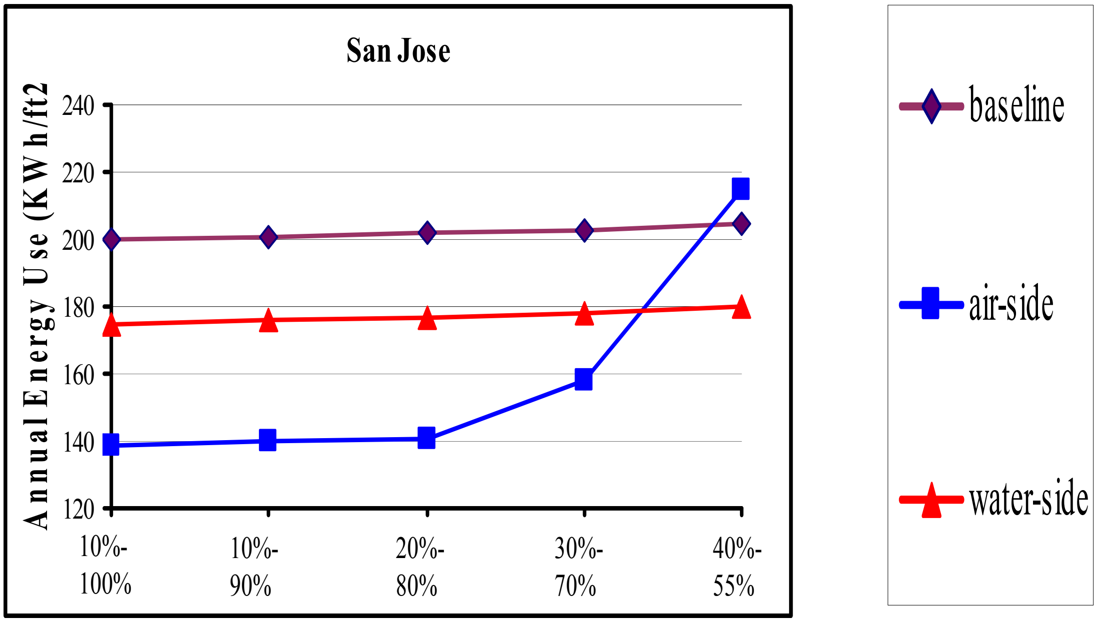
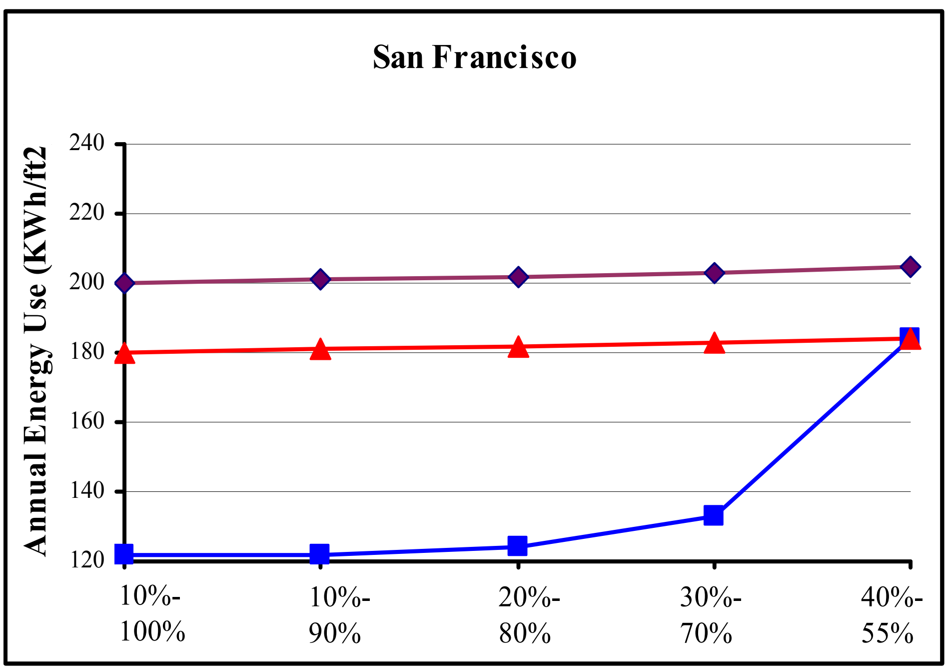


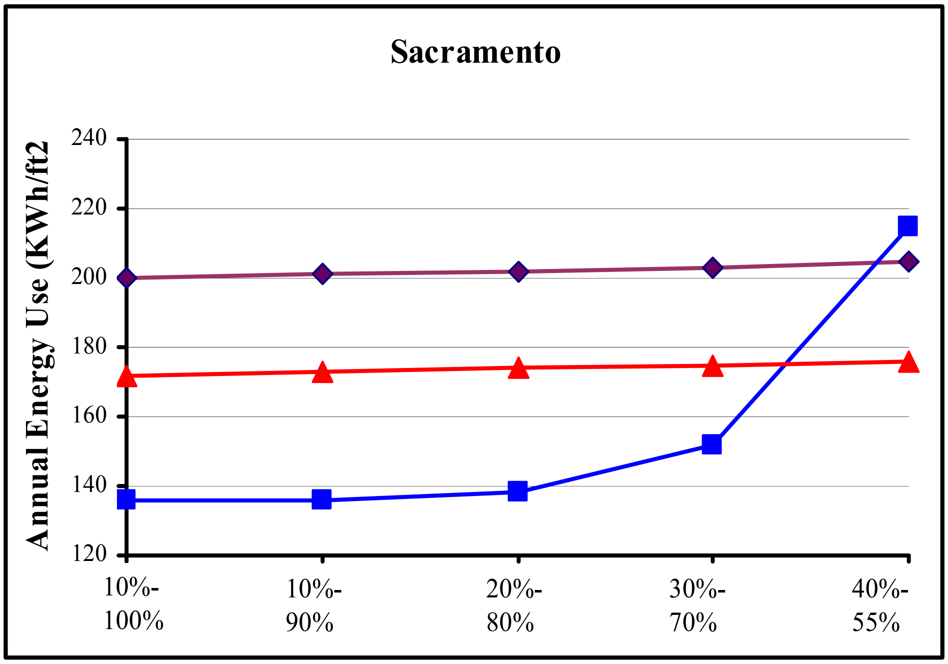
Figure 11: Energy consumption under economizer scenarios [10]

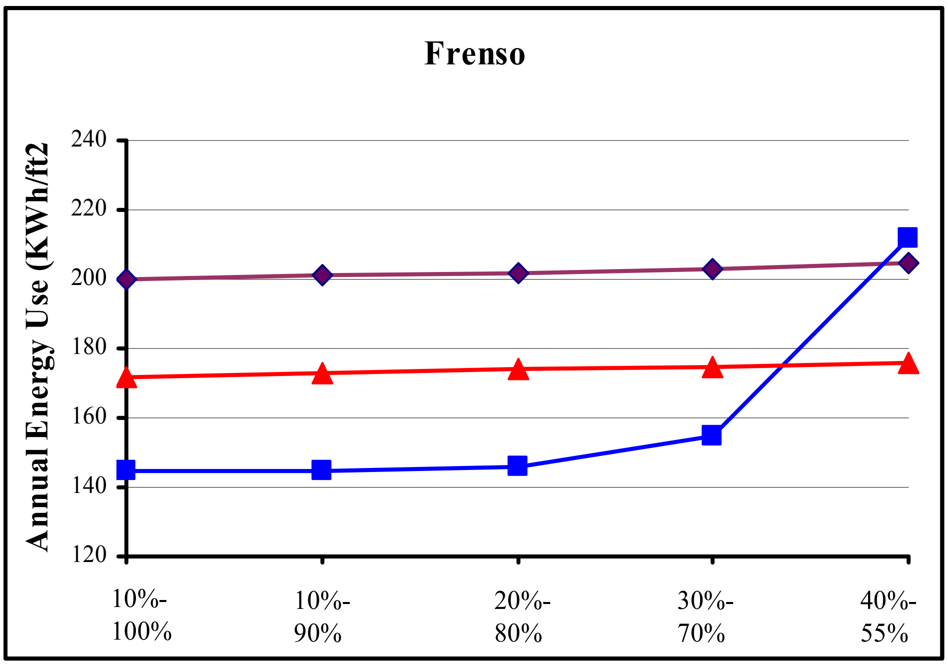
Figure 12: shows the annual energy consumption with the different humidity restrictions in the five data centers. Among the three design scenarios, the baseline and WSE scenario are generally independent with the humidity restrictions, however, the ASE energy consumption will increase sharply and it may exceed even those in the other scenarios when the relative humidity (RH) restriction ranged is narrowed. The humidity range for data centers recommended by ASHRAE thermal guidelines 2004 version is 40%-55% (The humidity ranges in 2008 and 2011 version are represented by both relative humidity and dew point, and the detail will be introduced in chapter 4). In order to gain the maximum energy saving, the humidity level in the

ASE scenario may go far beyond the recommended range. It may cause some reliability risks to the equipment, which will be discussed in the following chapter.






Sacramento
Annual Energy Use (KWh/ft2
240
220
200
180
160
140
120
10%-
100%
10%-
90%
20%-
80%
30%-
70%
40%-
55%


Frenso
Annual Energy Use (KWh/ft2
240
220
200
180
160
140
120
10%-
100%
10%-
90%
20%-
80%
30%-
70%
40%-
55%

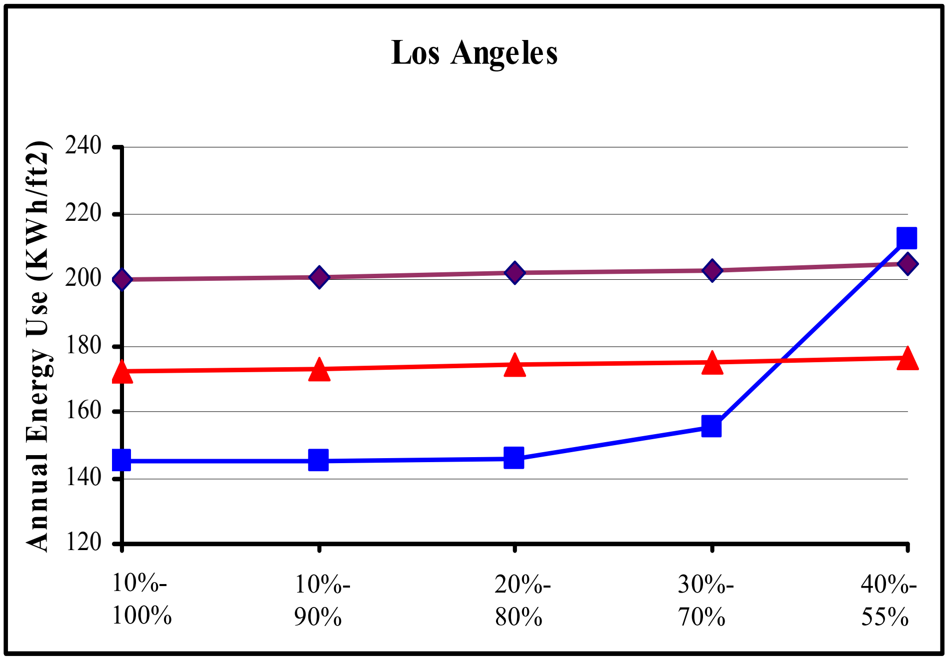


Figure 12: Chiller and fan energy resulting from humidity restrictions [10]

2.5 Implementations of Free Air Cooling

Free air cooling is being increasingly accepted and implemented in companies, shown in Table 10. It has been widely used in data centers located from U.S.A, Europe, Asia to Australia. Due to the climate diversity of the data center locations and designs, the annual days of free air cooling are different with various energy saving benefits. This subchapter introduces the implementations of free air cooling in two example cases of commercial company.

Table 10: Implementation of free air cooling in companies [9]

Company	Location	Description
Facebook	Oregon, U.S.A	The Facebook's first company-built facility, 147,000 sq ft, with Power Usage Effectiveness (PUE) of 1.15.
Microsoft	Chicago, U.S.A	One of the largest data centers, 700,000 sq ft.
Citigroup	Frankfurt, Germany	230,000 sq ft, 65% of a year with free air cooling.
Digital Realty Trust	California, U.S.A	More than 65% of a year with free air cooling, annual 3.5M KWH energy saving ($250,000), with a PUE of 1.31.
VMWARE	Washington, U.S.A	The mechanical system uses hot air/cold air physical separation to extend the operation hours of air-side economizers
Microsoft	Dublin, Ireland	303,000 sq ft, Microsoft's fist Mega data center in Europe.
Internet Initiative Japan	Japan	It is expected to reduce the cost of cloud service by 40%, reducing annual CO_2 output by about 4,000 tonnes.
Equinix Sydney	Sydney, Australia	The $32M data center is one of the first centers with free air cooling in Sydney.
Advanced Data Centers	California, U.S.A	237,000 sq ft, use of air-side economizers and recycled grey water as a redundant water supply.
Google	Brussels, Belgium	Operated above 80^{o}F, with temperatures above the acceptable range only about seven days per year on average.
Weta Digital	Wellington, New Zealand	10,000 sq ft, running full time and often at full capacity, with no air-conditioning.
IBM Cloud	N Carolina, U.S.A	More than 100,000 sq ft, with $362M, annual use of free air cooling for half year.
Fujitsu	Perth,	Span about 8,000 sq ft, and potentially decreasing

	Australia	the cooling load by up to 50%.
HP Wynyard	Newcastle, United Kingdom	300,000 sq ft, Data center Leaders' award 2008.
Verne Global	Keflavik, Iceland	100% free cooling utilizing the low ambient temperature.

Case Example: The Implementation of Intel

In a typical Intel data center, more than 60% of the energy consumption is spent on the power and cooling equipment. The design of increasingly complex semiconductors requires the support of computing capacity in data centers, which results in the rapid growth of energy consumption. Intel implemented free air cooling to minimize the energy consumption for a given computing capacity in data centers.

As a proof of concept (PoC) test, the free air cooling was implemented in one of Intel data centers in New Mexico with a high performance and high density standard for ten months. These blade servers in the data center are highly utilized to deliver great computing capacity, however they also generate a lot of heat. With air conditioning unit, the supply air to cool the severs is 20^{o}C. After the air passes across the servers, its temperature increases by 32^{o}C and reaches 52^{o}C. If Intel wants to re-circulate the air, it needs to be cooled by 32^{o}C, which will consume substantial energy to do it with air conditioning units [1].

The implementation method of free air cooling is shown in Figure 13. In order to avoid the equipment downtime due to the severe operating environment, free air cooling was implemented in a trailer originally designed to provide temporary additional computing capacity. The trailer was 1,000-square-foot (SF) and divided into two approximately 500 SF compartments. One compartment with 448 high utilized blade servers was cooled by airside economizers, which were modified from low-cost, warehouse-grade direct expansion (DX) air conditioning equipment. The air economizers expel the hot and exhausted air from the servers to outdoors and supply the cold outside air to cool the servers. The other compartment with another 448 blade servers was cooled by the traditional air conditioning units in order to identify the impact of free air cooling on reliability. The sensors installed in the two compartments were used to monitor the temperature and humidity [1].

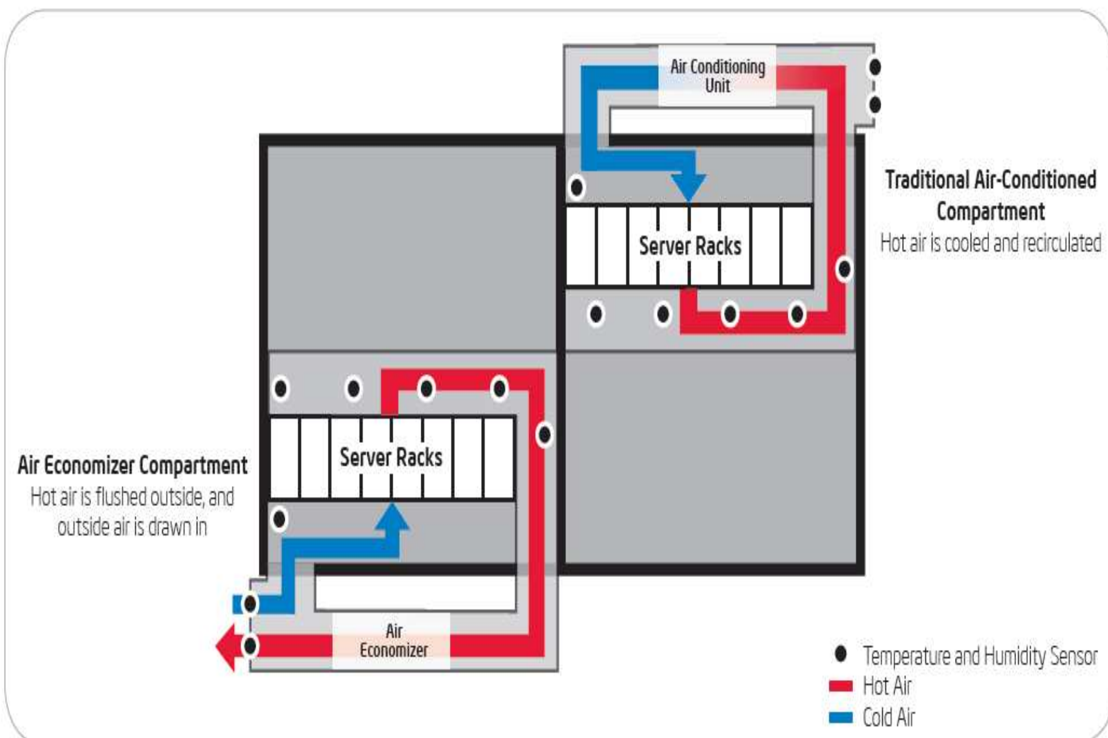


Figure 13: Implementation of free air cooling in Intel [1]

In order to maximize the energy saving, the outside supply air temperature was set to have a wide range of 18°C to 32°C, since the servers can work under the temperature as high as 37°C according to the manufacturer ratings. This temperature range of the supply air was maintained by the air conditioning units inside the airside economizers. When the outside air exceeded 32°C, the air conditioning units would start to cool the supply air to 32°C. If the temperature of the supply air was below 18°C, the hot return air from the servers would be mixed with the supply air into the set range. There are no controls on the humidity in this case. For the contamination, minimal filtering was applied to remove only large particles in the supply air but permit fine dust into the room. During the PoC test, the servers were used to run large workload to maintain a very high utilization rate of about 90% [1].

The PoC test started in October 2007 and ended in August 2008. The average temperature at the data center location is shown in Table 11. The servers with free air cooling were subjected to wide operating condition variations. Due to the slow response of the low cost air conditioning units inside the airside economizers, the temperature of the supply air exceeded slightly beyond the set range. The records showed that the supply air temperature varied from 17.7°C to 33.3°C. The relative humidity varied from 4% to more than 90% with rapid changes at times. The compartment and the servers with free air cooling were covered with dust.

Table 11: Average temperatures at Intel free air cooling location [1]

Month	Ave high (^oC)	Ave low (^oC)
Jan	9	-5
Feb	13	-2
Mar	17	1
Apr	21	5
May	23	9
Jun	32	15
Jul	33	18
Aug	31	17
Sep	28	13
Oct	22	7
Nov	14	0
Dec	9	-5

If both the compartments were cooled by air conditioning units, the total energy consumption in the trailer was about 500 kilowatts (KW). When the economizer was used, the cooling load of the DX air conditioning units was reduced from 111.78 KW to 28.6 KW in the economizer compartment, which saved the energy consumption as high as 74%. It was estimated that 67% energy consumption could be saved with 91% use of airside economizers, which potentially could reduce the annual energy cost by up to USD 2.87 million in a 10-megawatt (MW) data center. The failure rate in the compartment with DX air conditioning units cooling was 2.45%, and the failure rate in the economizer compartment was 4.46%, which almost doubled due to the dust and wider ranges of temperature and humidity [1].

Case Example: The Implementation of Dell

Free air cooling was also implemented in one of Dell's data center with 50000 ft^2 in Austin, TX. This implementation realized $179k (about 15% reduction in overall energy cost in the data center) by free air cooling in the first four months of 2010, although the climate in Austin is not ideal for the implementation of this cooling method. This power consumption is shown in Figure 14.

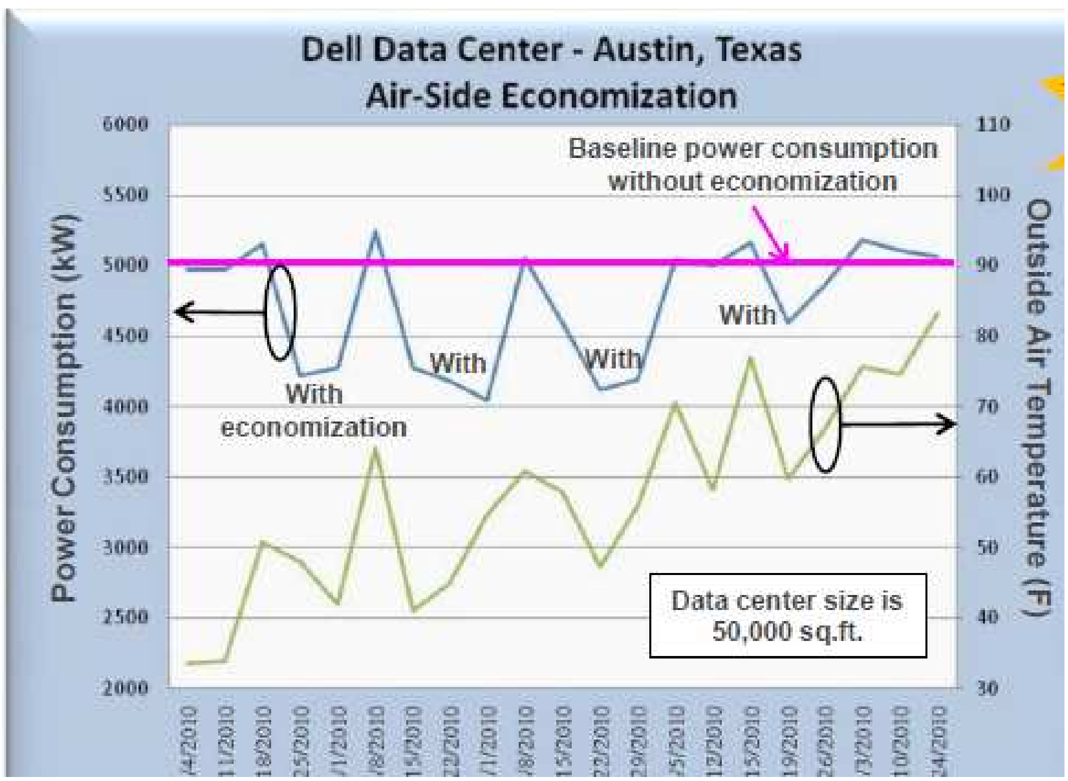


Figure 14: Energy consumption of Dell data center with free air cooling [11]

2.6 Summary

Free air cooling is increasingly accepted in data centers and is usually implemented with airside economizers. During the implementations, the data center operators need set operating environment ranges of the supply air. The energy saving benefits from free air cooling depends on the set operating environment and the local climate in data centers. The potential hotspots should be considered and removed by local cooling or other methods during the implementation. Researches show that the humidity restrictions have great impacts on the energy saving. In order to maximize the energy saving of free air cooling, the humidity of the supply air is usually uncontrolled under free air cooling conditions.

References:

[1]. Intel Information Technology, "Reducing Data Center Cost with an Air Economizer,", IT@Intel Brief; Computer Manufacturing; Energy Efficiency; August, 2008.

[2]. D. Pickut, "Free Cooling: Economizers in Data Centers," Equinix, Inc., Interop presentation, March 2008.

[3]. V. Sorell, "OA Economizers for Data Centers," ASHRAE Journal, December 2007.

[4]. ASHRAE TC 9.9 2004, "Thermal Guidelines for Data Processing Environments," 2004.

[5]. R. Miller, "Google's Chiller-less Data Center," Data Center Knowledge, July 2009.

[6]. ASHRAE, "2008 ASHRAE Environmental Guidelines for Datacom Equipment," Atlanta, 2008

[7]. Cisco, "Cisco 3600 Series — Modular, High-Density Access Routers," March 2002.

[8]. Kumar, "Use of Airside Economizer for Data Center Thermal Management," Master's thesis at Georgia Institute of Technology, 2008.

[9]. Datacenterdynamics, "Free Cooling Guide", Apr. 30, 2010.

[10]. A. Shehabi1, S. Ganguly, K. Traber, H. Price, A. Horvath1, W. W. Nazaroff, A. J. Gadgil, "Energy Implications of Economizer Use in California Data Centers", ACEEE Conference Proceedings, Monterey, CA, Sep., 2008.

[11]. T. Homorodi, J. Fitch, “Fresh Air Cooling Res earch”, Dell Techcenter, July 2010.

[12]. Energy Star, “Air-side Economizer”, access on Jun., 2012. http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_economizer_airside

[13]. Climatetemp.info, access on Jun., 2012 http://www.climatetemp.info/

Chapter 3. Risks of Free Air Cooling

Application of free air cooling modifies the operating environment in data centers. Normally, most data centers need to refer to the operation conditions required in the industry standards such as “ASHRAE Thermal Guidelines for Data Center and Other Data Processing Environments”. But under free air cooling, operating environments usually go beyond those in current data centers and standards. This modification may result in some reliability issues. This chapter reviews the required operation condition in data centers and summarizes the risks arising from free air cooling, and discusses their impact on the reliability of telecom equipment in data centers.

3.1 Operating Environment in Standards

Operating environment settings directly affect cooling energy efficiency. It is estimated that data centers can save four to five percent of energy costs for every one degree increase in server inlet temperature by the traditional A/C cooling method. In most data centers, the operating environment is maintained at fixed air inlet temperatures and narrow humidity ranges. Some standards have expanded their requirements for operating environment to improve energy efficiency.

There are several standards recommending operating conditions for telecom equipment. The American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. (ASHRAE) published its “Thermal Guidelines for Data Centers and Other Data Processing Environments” in 2004 [1], which provide allowable and recommended operating condition limits for data centers, including temperature and humidity. The recommended condition is "to give guidance to data center operators on maintaining high reliability and also operating their data centers in the most energy efficient manner", and the allowable condition is "where the IT manufacturers test their equipment in order to verify that the equipment will function within those environmental boundaries" [1]. In 2008, ASHRAE expanded the limits to save energy [2]. The revision has increased both the temperature and moisture ranges recommended for data center equipment, as shown in Table 12 (The moisture range is expressed in term of dew points, since ASHRAE thinks equipment failure is not directly related to relative humidity but is strongly related to dew points.). ASHRAE specified four classes in 2008 version, and two of which (class 1 and 2) are applied into data centers. The detailed information of these classes will be introduced in the following.

Table 12: Operating environmental limits of class 1 and 2 per ASHRAE [2]

	Recommended Limits		Allowable limits
	2004 version	2008 version	
Low temperature	20°C	18°C	15°C
High temperature	25°C	27°C	32°C
Low moisture	40%RH	5.5°C DP[6]	20%RH
High moisture	55%RH	60% and 15°C DP	80%RH

ASHRAE revised the thermal guidelines in 2011 and the new version have more data center classes to "accommodate different applications and priorities of IT equipment operation" [3]. The specification of the classes are [3]:

Class A1: Typically a data center with tightly controlled environmental parameters (dew point, temperature, and relative humidity) and mission critical operations; types of products typically designed for this environment are enterprise servers and storage products.

Class A2: Typically an information technology space or office or lab environment with some control of environmental parameters (dew point, temperature, and relative humidity); types of products typically designed for this environment are volume servers, storage products, personal computers, and workstations.

Class A3/A4: Typically an information technology space or office or lab environment with some control of environmental parameters (dew point, temperature,

6 Dew point is the temperature at which the air can no longer hold all of its water vapor, and some of the water vapor must condense into liquid water.

and relative humidity); types of products typically designed for this environment are volume servers, storage products, personal computers, and workstations.

Class B: Typically an office, home, or transportable environment with minimal control of environmental parameters (temperature only); types of products typically designed for this environment are personal computers, workstations, laptops, and printers.

Class C: Typically a point-of-sale or light industrial or factory environment with weather protection, sufficient winter heating and ventilation; types of products typically designed for this environment are point-of-sale equipment, ruggedized controllers, or computers and Personal Digital Assistants (PDA[7])s.

These classes are also shown in Table 13. The environmental specifications of the classes in 2011 version are shown in Table 14. The The purpose of the recommended envelope is to give guidance to data center operators on maintaining high reliability and also operating their data centers in the most energy efficient manner. The allowable envelope is where IT manufacturers test their equipment in order to verify that the equipment will function within those environmental boundaries.

[7] PDA is a handheld computer for managing contacts, appointments and tasks.

Table 13: 2011 and 2008 ASHRAE Thermal Guideline classes [3]

2011	2008	Application	IT equipment	Environmental Control
A1	1	Data Center	Enterprise servers, storage product	Tight Control
A2	2		Volume servers, storage product, personal computers, workstations	Some control
A3	N/A			
A4	N/A			
B	3	Office, home, transportable environnent, etc.	Personal computers, workstations, laptops, printers.	Minimal control
C	4	Point-of-sale, factory, industry, etc.	Point-of-sale equipment, ruggedized controllers, or computers and PDAs.	No control

Table 14: Environmental limits 2011 ASHRAE Thermal Guideline classes [3]

Class	Product Operation			Product Power Off		
	Temp.	Humidity range	Max dew point	Temp.	Humidity range	Max dew point
Recommended						
A1 to A4	18 to 27	5.5°C DP to 60% and 15°C DP				
Allowable						
A1	15°C to 32°C	20% to 80%RH	17	5°C to 45°C	8% to 80% RH	27°C
A2	10°C to 35°C	20% to 80%RH	21	5°C to 45°C	8% to 80% RH	27°C
A3	5°C to 40°C	-12˚C DP & 8% RH to 85% RH	24	5°C to 45°C	8% to 85% RH	27°C
A4	5°C to 45°C	-12˚C DP & 8% RH to 90% RH	24	5°C to 45°C	8% to 90% RH	27°C
B	5°C to 35°C	8% to 80% RH	28	5°C to 45°C	8% to 80% RH	29°C
C	5°C to 40°C	8% to 80% RH	28	5°C to 45°C	8% to 80% RH	29°C

The recommended and allowable operating environment limits for the data center classes are also expressed by ASHRAE Psychrometric Chart, shown in Figure 15 [3].

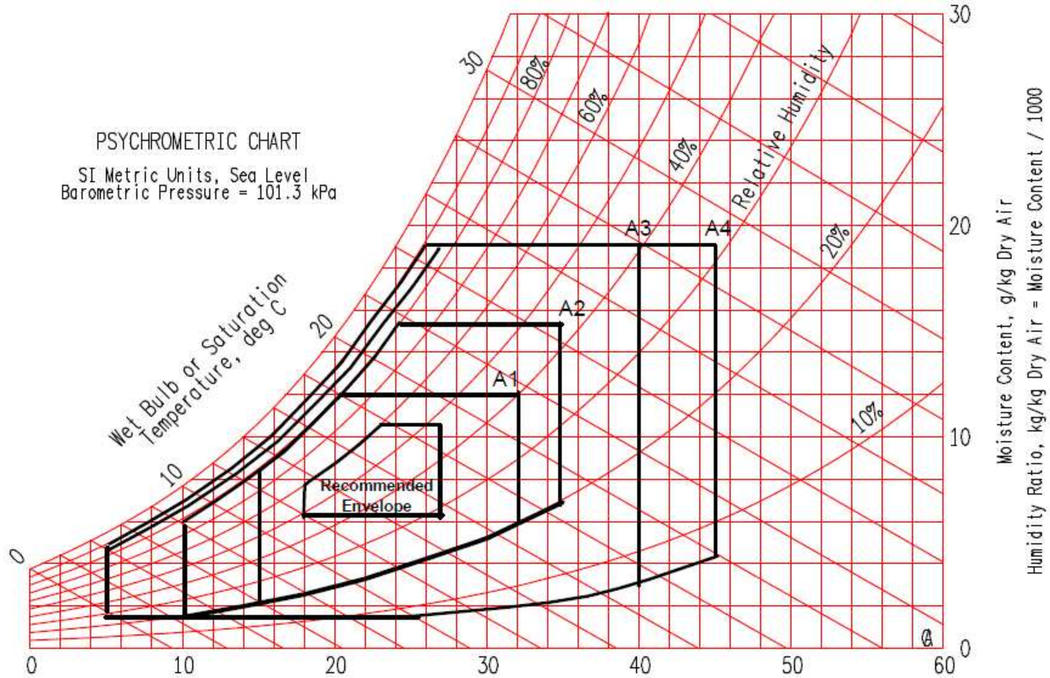


Figure 15: ASHRAE 2011 Psychrometric Chart [3]

The Telcordia Generic Requirements GR-63-CORE [4] and GR-3028-CORE [5] also provide recommended operating environments for telecom equipment. The recommended temperature range is the same as that in the ASHRAE 2008 version (another reason for expanding the ASHRAE recommended temperature, since Telcordia GR-63-CORE is widely accepted and used in the industry). The humidity ranges are slightly different from those in the ASHRAE guidelines, as shown in Table 15.

Table 15: Operating environmental limits in Telcordia Generic Requirements [4][5]

	Recommended Limits	Allowable limits
Low temperature	18°C	5°C
High temperature	27°C	40°C
Low relative humidity	5%RH	5%RH
High relative humidity	55%RH	85%RH

ETSI 300 019 [6] was published by the European Telecommunications Standards Institute (ETSI) in 1994, and provides environmental requirements for telecom equipment under different conditions. It defines eight environmental conditions: storage, transportation, stationary use at weather-protected locations, stationary use at non-weather-protected locations, ground vehicle installations, ship environments, portable and non-stationary use, and stationary use at underground locations. There may be several classes within each condition. As an example, stationary use at weather-protected locations includes six classes:

Temperature-controlled locations (class 3.1): a permanently temperature-controlled, enclosed location with usually uncontrolled humidity, a combination of classes 3K3/3Z2/3Z4/3B1/3C2(3C1)/3S2/3M1 in IEC standard 60721-3-3.

- Partly temperature-controlled locations (class 3.2): an enclosed location having neither temperature nor humidity control.

- Non-temperature-controlled locations (class 3.3): a weather-protected location with neither temperature nor humidity control.

- Sites with a heat-trap (class 3.4): a weather-protected location with neither temperature nor humidity control, affected by direct solar radiation and heat trap conditions.

- Sheltered locations (class 3.5): a shelter where direct solar radiation and heat-trap conditions do not exist.

- Telecommunication control room locations (class 3.6): a permanently temperature-controlled, enclosed location, usually without controlled humidity, a combination of classes 3K2/3Z2/3Z4/3B1/3C2(3C1)/3S2/3M1 in IEC standard 60721-3-3.

The allowable environments for the classes of stationary use at weather-protected locations are shown in Table 16.

Table 16: Allowable environments for stationary use at weather protected locations [6]

Environment classes	**Unit**	**Classes**						
		3.1		3.2	3.3	3.4	3.5	3.6
		N[8]	E[9]					
Low temperature	^oC	5	-5	-5	-25	-40	-40	15
High temperature	^oC	40	45	45	55	70	40	30
Low relative humidity	%	5	5	5	10	10	10	10
High relative humidity	%	85	90	95	100	100	100	75

The purpose of the recommended operating range is to give guidance to data center operators for maintaining high reliability [2]. When the equipment is used

[8] N: normal
[9] E: exceptional

beyond the recommended operating conditions, it may be less reliable. Currently, most data centers operate at temperatures between 20^{o}C and 25^{o}C. A survey of fourteen of Sun's (now Oracle) data centers in 2007 showed that eight had the inlet temperature set at 20^{o}C, five at 22^{o}C, and one at 23^{o}C [27].

3.2 Failure Causes of Free Air Cooling

Generally, the operating temperature in a free-air cooled data center is subjected to increased temperature variations, which may affect the lifetime of equipment and result in some reliability concerns. The humidity during free air cooling is usually uncontrolled in order to save the energy required by humidification and dehumidification, but this, too, may cause some failure mechanisms (such as electrostatic discharge (ESD) and conductive anodic filament (CAF) formation) to be more active. In addition, the contamination in free air cooling is a potential failure cause.

3.2.1 Increased Temperature and Temperature Cycling

When free air cooling is used in data centers, a major concern is that the operating temperatures may rise and exacerbate existing hotspots. Increases in operating temperatures, particularly in power infrastructures and cooling systems, may affect the performance of communication equipment. It is not clear whether the component junction temperatures increase linearly with the ambient temperature under a free air cooling environment. The thermal resistance may also change in the new environment. So it is possible that the local temperatures of the existing hotspots may increase rapidly and damage the equipment.

Changes in temperature can impact the electrical parameters of components and systems. As a result of these variations, particularly at hot spots, there is a risk of intermittent out-of-specification behavior. Free air cooling elevates the operating temperature. The temperature-dependence of equipment reliability is assessed by the effect of temperature stresses on dominant failure mechanisms. Failure mechanisms depend heavily on operational stresses and usually are more active in elevated temperature. Some failure mechanisms are sensitive to temperature and some equipment components will weaken under increased operating temperatures. Consequently, some components and systems may not be reliable for their expected lifetimes if operated under free air cooling.

Some equipment is particularly sensitive to temperature increases. For example, the batteries in data centers and base stations may react to operating temperature changes. The recommended temperature for batteries is usually 25^{o}C, and the allowable temperature range is 15^{o}C to 30^{o}C. A decrease in temperature may cause a drop in battery capacity; a one-degree decrease generally results in a one percent drop in battery capacity. A temperature increase may accelerate the corrosion of bipolar plates in batteries, with more water consumed, then affect the lifetimes of batteries. Generally, the lifetimes of batteries are maximized around 25^{o}C, and it is estimated that the expected life may drop by fifty percent when the operating temperature increases fifty percent [7].

Increases in operating temperatures also lead to increased failure rates of some components. For example, the failure rate for a typical switched mode power supply

(SMPS) doubles with every 10 to 15^{o}C temperature rise above 25^{o}C [8]. Increased operating temperatures also result in increased conduction and switching losses in switching transistors. With a temperature rise, the mobility of charge carriers also is reduced. In switching transistors such as power MOSFETs, reduced mobility leads to increased device resistance, and hence increased conduction loss [9]. Reverse bias leakage currents increase exponentially with increases in temperature, and this leads to greater power loss [10].

High operating temperatures lead to an increase in switching turn-off time for power transistors, which in turn results in increased switching loss. Another result of higher operating temperatures is degradation in the gate oxide of switching transistors, which may result in time-dependent dielectric breakdown. High temperatures can lead to activation of the parasitic bipolar junction transistor in a power MOSFET and the parasitic thyristor in an IGBT, leading to destruction of the device due to latch-up. In aluminum electrolytic capacitors, increased operating temperature leads to the evaporation of the electrolyte. This will cause a reduction in the capacitance, an increase in the equivalent series resistance (ESR), and increased power dissipation. High operating temperatures cause structural overstress within the Schottky diode, causing cracks that can propagate into the Schottky contact region leading to catastrophic failure.

The free air cooling application can increase the temperature variation and result in additional temperature cycles for the equipment. For example, during a proof-of-concept study performed by Intel, the average diurnal temperature variation ranged from 13°C to 17°C [11]. For a piece of equipment with a lifetime of five years, this would result in an additional 1825 temperature cycles, which may well accelerate the wear-out of the equipment [11].

Free air cooling may also accelerate wear-out in cooling equipment like fans. When the operating temperature is increased, the cooling algorithm may increase the fan speed to offset the temperature increase. This can affect the lifetime and reliability of the fans.

3.2.2 Uncontrolled Humidity

Data centers require continuous air conditioning to address the high internal heat loads generated by equipment and to maintain indoor temperatures below the maximum recommendation for operating computers. Air economizer cycles, which bring in large amounts of outside air to cool internal loads when weather conditions are favorable, could save substantial cooling energy. However, there is reluctance from many data center owners to take advantage of this common cooling technique due to concerns about airborne pollutants and the potential loss of humidity control. In practice, the humidity is usually uncontrolled to maximize the energy efficiency (e.g., at Intel). This uncontrolled humidity may cause reliability risks. Typical humidity levels in data centers based on ASHRAE Guidelines are between 40% and 60%RH. This range provides effective protection against a number of corrosive failure mechanisms, such as electrochemical migration and conductive anodic filament (CAF) formation. The introduction of free air from the outside could potentially cause significant swings in data center humidity. In the Intel case, the humidity varied from 4%RH to more than 90%RH. Both overly high and overly low humidity can activate failure mechanisms. CAF can be caused by high humidity, and electrostatic discharge (ESD) is more common in low humidity environments. These failure mechanisms can result in equipment failure [12].

3.2.3 Contamination

Contamination is another potential risk with the application of free air cooling. In some cases (e.g., the Intel test [11]), there is no control of dust and gas. There has been a recent increase in the rate of hardware failures in data centers high in sulfur-bearing gases, especially in centers located near industrial operations and other sources of pollution.

The effects of airborne contamination on datacenter equipment fall into three main categories: chemical effects, mechanical effects, and electrical effects. Two common chemical failure modes are copper creep corrosion on circuit boards and the corrosion of the silver metallization in miniature surface-mounted components. Mechanical effects include heat sink fouling, optical signal interference, increased friction, and so on. Electrical effects include changes in circuit impedance, arcing, and the like. It should be noted that the reduction of circuit board feature sizes and the miniaturization of components, necessary to improve hardware performance, also makes the hardware more prone to attack by contamination in the data center environment. Manufacturers constantly struggle to maintain the reliability of their hardware with ever-shrinking feature sizes, without taking the added costly measure of hardening all their IT equipment, most of which is not installed in corrosive environments [14].

Extensive failures of telecommunications circuit boards can be caused by the combined effect of the deposition of hygroscopic dust and elevated relative humidity [15]. The presence of water-soluble materials in the dust leads to surface resistance degradation at high relative humidity [16]. Hygroscopic dust is always present in the atmospheric air but is particularly abundant in industrialized urban environments [17]. To prevent premature failure, the level of dust deposited on devices needs to be limited. More importantly, the physical (particle size) and chemical (PH value, solubility) properties of the dust should be characterized. For example, small amounts of magnesium chloride ($MgCl_2$) in the dust can cause ion migration failure under low relative humidity conditions because of its low deliquescent relative humidity.

Sulfur-bearing gases, such as sulfur dioxide (SO_2) and hydrogen sulfide (H_2S), can cause corrosion of silver metallization and board finishes by the formation of silver sulfide. There is an important synergistic effect between gases. For example, hydrogen sulfide alone is not very corrosive to silver, but the combination of hydrogen sulfide and nitrous oxide is highly so [18]. Similarly, neither sulfur dioxide nor nitrous oxide alone is corrosive to copper, but together they attack copper at a very fast rate [19]. The anion in the gas dissolved in the water is normally more corrosive than that of a water-soluble salt. For example, the corrosive effects of gaseous chlorine and hydrogen chloride in the presence of moisture tend to be stronger than those of chloride salt anions, due to the acidic character of the former species [20].

All metal components of telecom equipment are affected by corrosion reactions in a contaminated environment. External connectors, sockets, and wiring of electronic devices are problematic sections in telecom equipment. Pore corrosion is a concern for noble-plated contacts, and fretting corrosion can occur between two solid surfaces in contact. High density printed circuit boards with smaller feature size and spacing are vulnerable to ionic migration and creep corrosion. Fine particles can also lead to electrical shorts caused by corrosion [20]. If water layers are formed on critical surfaces and interfaces, they can result in resistance degradation, leading to soft and/or hard equipment failures [25]. The purpose of the case and frame structure is to protect the electronics device from external environmental conditions, but the outside of the case and frame structure may suffer from excessive atmospheric corrosion. The polymer material can swell due to moisture absorption.

Most data centers are well designed and are in areas with relatively clean environments, and most contaminants are benign. Data centers generally should not experience hardware failures due to particulate or gaseous contamination. However, some data centers, particularly these located in highly populated municipal areas, may have harmful environments arising from the ingress of outdoor particulates and/or gases in free air cooling. This could be a reliability concern for equipment in the center.

3.3 Reliability Risks of Free Air Cooling: Failure Mechanisms

Due to the failure causes implicit in the implementation of free air cooling, some potential failure mechanisms become more active under the new operating environment. Three main mechanisms (CAF, ESD and corrosion) are introduced in the following subchapters.

3.3.1 Conductive Anodic Filament Formation (CAF)

CAF occurs in substrates and printed circuit boards (PCBs) when a copper conductive filament forms in the laminate dielectric material between two adjacent conductors or in plated-through vias under an electrical bias [26]. CAF can be a significant and potentially dangerous source of electrical failures in IC packaging substrates, PCBs and the overall system (package, module) that they are part of. The increased board density that has being driven by the chip scale packaging (CSP) revolution of the early 90s, along with the increased input-output density on the chips, has forced the PCB industry to decrease via wall-to-wall distances and feature sizes. The trend in the electronics industry to place as many components as possible on a minimum of PCB real estate has increased the reliability requirements for bare PWBs and is raising possible reliability issues caused by CAF formation within the multilayer structure.

The action of CAF is shown in Figure 16. The hydrolysis reaction at the anode creates hydronium ions, while the one at the cathode creates hydroxide ions. The acidic hydronium ions accumulate at the anode, while the basic hydroxide ions accumulate at the cathode. If there is a pH drop at the anode then the copper corrosion products become soluble. These soluble products will try to travel from the anode to the cathode, due to the PH gradient, through any weak opening in the laminate. When these conductive filaments reach the cathode, CAF is formed and the insulation resistance between the cathode and anode drops significantly. Eventually an electrical short is created.

For the conductive salts to migrate from one end of the conductor or hole wall to the next, an easily accessible pathway, such as an area of poor adhesion of glass to resin or high ionic impurities in the resin, is needed. The pathway between these two conductors becomes an electrochemical cell with moisture from the laminate.

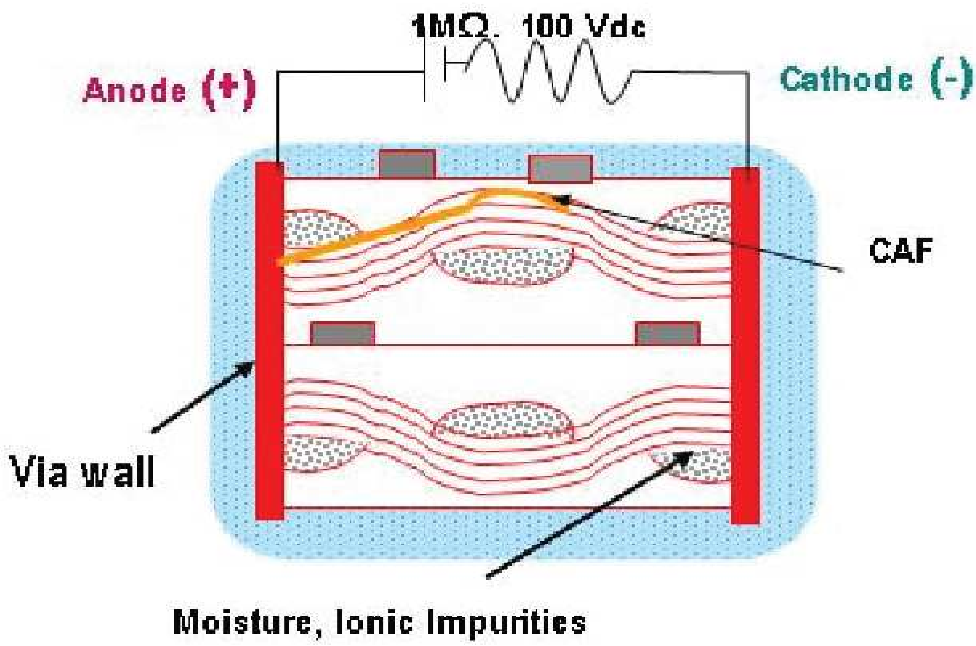


Figure 16: Cross-sectional view of CAF pathways [26].

A combination of bias voltage (voltage applied during the test) and high relative humidity can cause a CAF failure. The electrical failure is caused when a filament grows from a copper anode to a copper cathode. It is postulated that the CAF failure proceeds in two stages. The first stage involves the degradation of the resin-glass interface, followed by an electrochemical migration process, which allows the filament growth. This first step is believed to be reversible; the material's insulation resistance returns after baking and drying. The second step, actual CAF growth, is believed to be irreversible. The mean time to failure is a function of voltage bias, relative humidity, hole-to-hole and line-to-line spacings, temperature, and the resin system.

The reaction process of CAF occurs due to moisture absorption. Most laminate materials absorb moisture through surface absorption and diffusion into the interior, especially when exposed to high temperature and humidity environments, which accelerate the absorption and can result in quicker degradation and path formation. The different moisture absorption rates of resin and glass fiber can also lead to interface stress. Resin swells due to the absorption of moisture, which can lead to debonding at the resin/glass fiber interface [13].

3.3.2 ESD

If the humidity is too low, data centers can experience electrostatic discharge (ESD), akin to giving someone a shock after shuffling along a carpeted floor in stocking feet [25]. That sort of event can shut down electronic equipment and possibly damage it. Generally, ESD can be caused by humidity, temperature, pressure, airborne particles, and air recirculation, but the most significant environmental factor in ESD control is the relative humidity (RH).

In very dry areas, humidification is desirable because it makes antistatic materials with "sweat layers" function better, and reduces (but does not eliminate) triboelectric charging for all materials. A high relative humidity--over 30% RH--reduces the resistance of most dielectrics, resulting in an increase in return current (which opposes a charge buildup). When an object is undergoing tribocharging in a high humidity environment, the object will reach an equilibrium point where the tribocharging current equals the return current. For objects that undergo charging to a high potential, the primary impact of humidity is to encourage or discourage corona, and affect the rise time of the discharge current.

Normally, the moisture content in the air tends to lower the surface resistance of floors, carpets, table mats, and so on by letting wet particles create a vaguely conductive (less than 10^{-9} Ohms/square) film over an otherwise insulating surface. If the relative humidity decreases, this favorable phenomenon disappears. The air itself, being dry, becomes a part of the electrostatic build-up mechanism every time there is an air flow (wind, air conditioning, blower) passing over an insulated surface.

3.3.3 Corrosion

Corrosion is another failure mechanism affected by humidity. Corrosion occurs in the metalization in the presence of moisture and contaminants. Atmospheric corrosion is a ubiquitous phenomenon that affects materials of all sorts exposed to the open environment. Telecom components are also affected by other failure mechanisms under free air cooling conditions based on their specific structures, materials and applications. In this chapter, life models will be presented to provide predictive times to failure for corrosion-related failure mechanisms. Acceleration testing needs to be carefully designed and implemented with respect to the specific failure mechanism. Corrosion control and knowledge of various protection methods have become a major subfield of design in order to ensure product reliability and competitiveness in a demanding market.

3.4 Performance Risks of Free Air Cooling

A traditional data center usually use multiple A/C units to “fine tune” data center temperature in addition to the use of air flow, however, in a free air cooled data center, the temperature across the data center is controlled largely by air flow. Thus, additional designs and air flow simulations, as well as measurement and control during the operation must be considered for the proper implementation of free air cooling. When a data center accepts free air cooling, more air flow and temperature optimizations are required to eliminate unwanted hot spots compared with those in a traditional data center. This kind of optimizations may even have to be performed on selected rack configurations, which helps make sure that all the equipment within the rack are working within the required operating condition envelope. The temperature and air flow optimization also help identify the “weak link” hardware, i.e., that has a lower thermal margin than the other pieces of equipment and then limits the ability of the data center to function at its target maximum temperature.

For data centers that are already in operation, the “weak link” hardware are limitation for the free air cooling implementation since major design modifications are no longer feasible. One solution is to modify the rack configuration or provide more air flow to cool the equipment back into the required operating conditions. However, in some cases, the only option is to replace the "weak link" hardware with equipment that has a wider thermal margin. This can be very costly and time consuming and could interrupt the services of data centers. In these cases, data center operators will need to find the operating condition limits of the data centers based on that of each piece of equipment.

Changes in temperature and humidity levels can impact the electrical parameters and life of components and systems. As a result, the equipment may intermittently fail to meet specifications, particularly if those are located at hot spots. Possible loss of performance will arise if it is necessary to apply uprating methods, such as stress balancing for components or equipment, which trade off temperature for performance (e.g., reduced operating frequency or speed with increases in operating temperature). This approach is common in computer systems (e.g., reduced access speed for memories or lower frequency of operation for microprocessors) and can be applied to telecommunication systems and data centers, but any decreased performance must be assessed against customer expectations regarding service quality and availability.

With free air cooling, some system-level operational performance changes are expected, due to the dependence of the electrical parameters of the system components (parts and materials) on temperature and humidity. In general, good design practices that account for component performance based on their temperature dependence (e.g., worst case analysis), will ensure acceptable operation, as long as the temperature and humidity levels are within those stated by the recommended operating conditions of the datasheets. However, if the free air cooling results in the operation of components at environmental conditions outside the datasheet recommended operating conditions (either on the low or high side), then additional analysis will be necessary to quantify the associated electrical operating parameters' characteristics and assess their impact on the system-level performance.

3.5 Summary

This chapter introduces the possible risks for telecom equipment in data centers when free air cooling is implemented. The changes in temperature, humidity and contamination may make the failure mechanisms more active compared with those under traditional A/C conditions, and then affect the reliability of the telecom equipment. The effects may become significant when the data centers are located in the areas where the air is not clean enough. These risks need be analyzed when free air cooling is considered by data center staff.

References:

[1]. ASHRAE Technical Committee (TC) 9.9 2004, "Thermal Guidelines for Data Processing Environments", 2004.

[2]. ASHRAE Technical Committee (TC) 9.9, "2008 ASHRAE Environmental Guidelines for Datacom Equipment," Atlanta, 2008.

[3]. ASHRAE Technical Committee (TC) 9.9, "2011 Thermal Guidelines for Data Processing Environments – Expanded Data Center Classes and Usage Guidance," Atlanta, 2011.

[4]. Bell Communications Research Inc., Generic Requirements GR-63-CORE, "Network Equipment-Building System (NEBS) Requirements: Physical Protection," Piscataway, NJ, March 2006.

[5]. Bell Communications Research Inc., Generic Requirements GR-3028-CORE, "Thermal Management in Telecommunications Central Offices," Piscataway, NJ, Dec. 2001.

[6]. ETS 300 019, "Equipment Engineering (EE); Environmental Conditions and Environmental Tests for Telecommunications Equipment," European Testing Standards Institute, Sophia Antipolis, France, 2003.

[7]. Yu Xiaobing, Song Qi, " Research of Environmental Temperature Increase in Base Stations," Telecom Engineering Technica and Standardization, December 2008, Chinese version.

[8]. K. Billings Pressman and T. Morey, Switching Power Supply Design, McGraw Hill, New York, NY, 2009.

[9]. J. Barak Haran, D. David, N. Refaeli, B. Fischer, K. Voss, G. Du and M. Heiss, “Mapping of Single Event Burnout in Power MOSFETs,” IEEE Transactions on Nuclear Science, 2007, pp. 2488-2494.

[10]. P. McCluskey, R. Grzybowski and T. Podlesak, High Temperature Electronics, CRC Press, Boca Raton, FL, 1997.

[11]. Intel Information Technology, “Reducing Data Center Cost with an Air Economize,r, IT@Intel Brief; Computer Manufacturing; Energy Efficiency; August 2008.

[12]. Konstantine Karavakis and Silvio Bertling, “Conductive Anodic Filament (CAF): The Threat to Miniaturization of the Electronics Industry,” Circuitree Magazine, Dec., 1, 2004.

[13]. K.C. Yung Winco, “Conductive Anodic Filament: Mechanisms and Affecting Factors,” HKPCA Journal, Issue 21, 2006, pp.1-6.

[14]. ASHRAE Technical Committee (TC) 9.9, “Gaseous and Particulate Contamination Guidelines for Data Centers,” white paper, 2009.

[15]. W.H. Burnett, F.S. Sandroff, and S.M. D’Edigo, “Circuit Failure Due to Fine Mode Particulate Air Pollution,” Proceedings of the 18th International Symposium for Testing & Failure Analysis, Los Angeles, California, USA, 1992, pp. 329-333.

[16]. M. Tencer, “Conductive Aqueous Layer Formation at the Gel–Substrate Interface in Equilibrium with 100% RH Environment.” IEEE Transactions on Component Packaging Technology, No. 23, 2000, pp. 693–9.

[17]. (a) J.D. Sinclair, L.A. Psota-Kelty, C.J. Weschler, and H.C. Shields , Measurement and Modeling of Airborne Concentrations and Indoor Surface Accumulation Rates of Ionic Substances at Neenah, Wisconsin," Atmospheric Environment, No, 24, 1990, pp. 627–38; (b) J.D. Sinclair, L.A. Psota-Kelty, C.J. Weschler , "Indoor/Outdoor Ratios and Indoor Surface Accumulation of Ionic Substances at Newark, New Jersey," Atmospheric Environment, Volume 26, Issue 5, April 1992, pp. 871-882..

[18]. L. Volpe and P.J. Peterson, "Atmospheric sulfidization of silver in tubular corrosion reactor," Corrosion Science, Vol. 29, No. 10, 1989, pp. 1179-1196.

[19]. L. Johansson, "Laboratory Study of the Influence of NO_2 and Combination of NO_2 and SO_2 on the Atmospheric Corrosion of Different Metals," Electrochemical Society Extended Abstracts, vol. 85-2, 1985, pp. 221-222.

[20]. J.D. Sinclair, "Corrosion of Electronics," Journal of the Electrochemical Society, March 1988, pp. 89C - 95C.

[21]. C.J. Weschler and H.C. Shield, "The Impact of Ventilation and Indoor Air Quality on Electronic Equipment," ASHRAE Transactions, Part I: Symposia, 1991, pp. 455-463.

[22]. C.J. Weschler, S.P. Kelty, and I.E. Lingovsky, "The Effect of Building Fan Operation on Indoor-Out-Door Dust Relationships," Journal of Air Pollution, Control Association, Vol 33, 1983, pp. 624-629.

[23]. W.W. Nazaroff and G.R. Cass, "Mathematical Modeling of Chemically Reactive Pollutants in Indoor Air," Environmental Science Technology, Vol 20, No. 9,1986, pp. 924-934.

[24]. M. Tencer, “Deposition of Aerosol (“Hygroscopic Dust”) on Electronics – Mechanism and Risk,” Microelectronics Reliability, Vol. 48, Iss. 4, April 2008, pp 584-593.

[25]. R. Moss, "Exploding the Humidity Half-Truth and Other Dangerous Myths," EOS/ESD Technology Magazine, April 1987, p. 10.

[26]. K. Karavakis and S. Bertling, “Conductive Anodic Filament (CAF): The Threat to Miniaturization of the Electronics Industry,” Circuitree Magazine, Dec., 1, 2004.

[27]. R. Miller, “Data Center Cooling Set Points Debated”, Data Center Knowledge, Sep, 2007.

Chapter 4. Assessment of Free Air Cooling on the Performance of Telecom Equipment

Free air cooling usually changes the operating conditions in data centers compared with those in traditional A/C. The equipment and data center design rules, equipment test and acceptance conditions, and overall operating cost estimates for data center equipment are impacted by operating conditions. This chapter assesses the potential performance risks associated with the implementation of free air cooling.

4.1 Assessment Method

The network architecture in a data center consists of a set of routers and switches, and the main function is to channel incoming data from one or more input ports to a specific output port, which will take the data toward its intended destination. The main performances of network [1]:

- Availability: whether the network is working.
- Response time: The time that it takes a packet to travel between two points on the network.
- Network utilization: it represents the percent of time that the network is in use over a given period.

- Network throughput: The throughput of a network represents the amount of network bandwidth available for a network application at any given moment, across the network links.

- Network bandwidth capacity: the total amount of bandwidth available between two network endpoints.

Among the five elements, network utilization and network bandwidth capacity depend on the workload and the network bandwidth design, and are not related with free air cooling implementation. The availability just shows whether the network is working and it is not sufficient to identify the completed impacts of free air cooling impacts on telecom equipment. In addition, it can be also indicated by the throughput. The responses time can be derived from the throughput. So, this dissertation focuses on the throughput of telecom equipment to represent its performance. In order to identify the impact of free air cooling on telecom equipment performance, we conducted a performance monitoring and recording experiment on a typical switch under simulated free air cooling condition, and traditional air condition (A/C), respectively. Some metrics and baselines are required to compare and evaluate the performance variations of the switches under the different operating conditions. Since most data centers currently use A/C units as the cooling method, the performance of telecom equipment under A/C is used as the baseline for measuring the impact of free air cooling. In this case, the number/percentage of throughput off the baseline can be used as the possible metrics:

– 1% off baseline

– 2% off baseline

– 5% off baseline

This chapter uses the experimental data to compare equipment performance under free air cooling to equipment performance under traditional A/C, based on the. The performance variations is analyzed to identify the impact of free air cooling on the performance of telecom equipment

4.2 Setup of Experiment

The network equipment selected in this case was a Zonet zfs 3015P switch. It is widely used in offices and small enterprises, and its primary function is to send packets to their intended destinations. Its rated operating conditions are 0-40^{o}C and 10-90% RH, which are typical ranges for routers and switches. A photo of the switch is show in Figure 17.

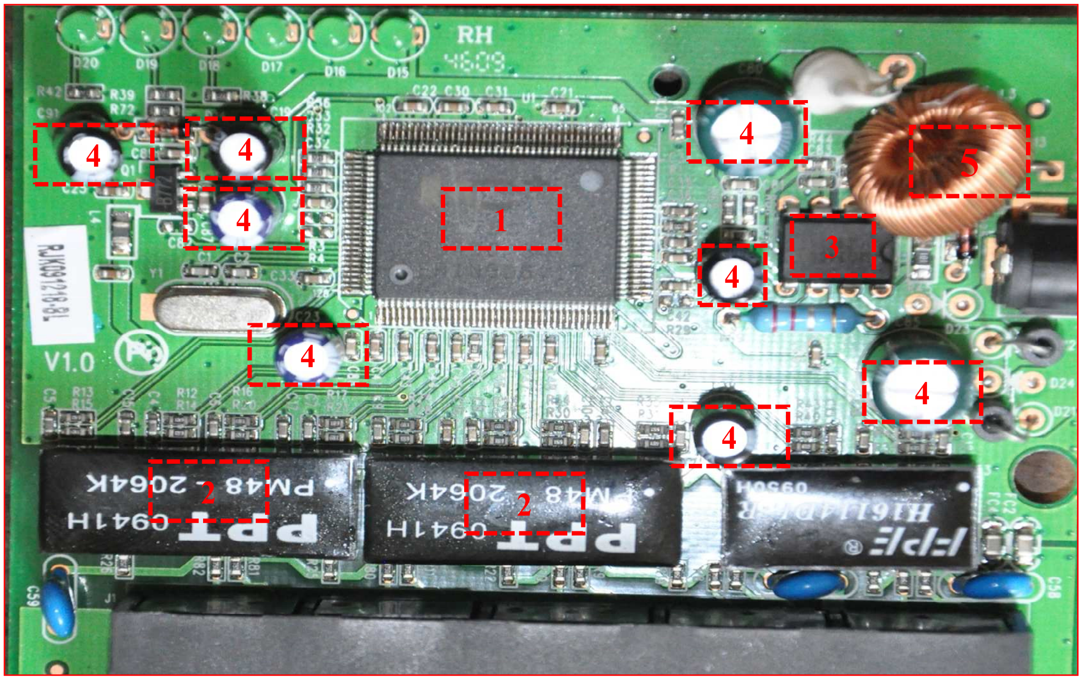


Figure 17: Zonet zfs 3015P switch
(1) IC+ IP 178C chip; (2) Magnetic components;
(3) MC 34063A; (4) Capacitor; (5) Coil

In the experiment, data packets were sent from one computer to another computer throughout the switch, which was put inside an environmental chamber shown in Figure 18. To monitor the performance of the switches in the experiment, we used NetIQ Chariot, a network testing software package. This software sent large files with sizes up to 10^8 bytes through the switch continuously and then calculates its performance parameters. In this experiment, the files are generated by the testing tool itself, rather than by the hard drive of computer; and so it is not dependent on the limitations of hard drive access speeds.

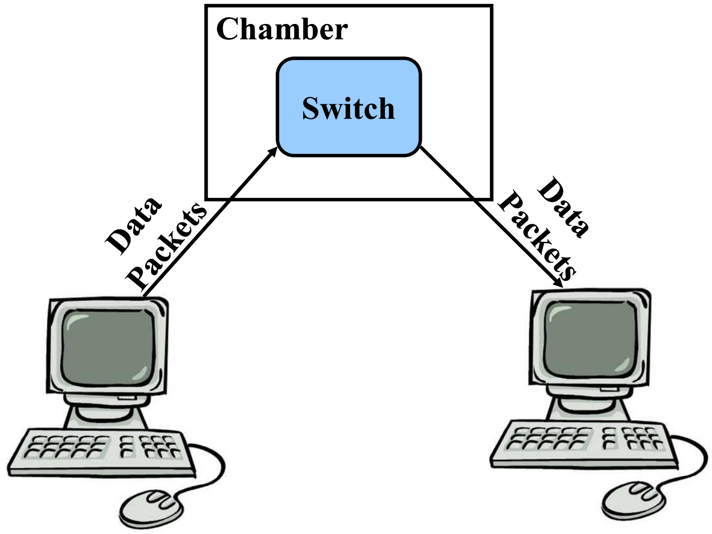


Figure 18: Overview of switch experiment system

There were two test conditions in this experiment: variable temperature and humidity representing FAC and A/C. The operating conditions (temperature/humidity) of free air cooling are usually decided by data center operators based on the local climates, industry standard requirements, equipment specifications, hot spot identification, and other analyses. In this case, the variable temperature and humidity representing FAC was set to be between 10^{o}C and 50^{o}C, and 15% to 85% RH, with the following considerations:

- The ASHRAE allowable ranges covering A1 and A2[10] data center class are 10^{o}C to 35^{o}C, 20% to 80%RH [3].

[10] In the ASHRAE data center classification, typically, the equipment selected for data centers is designed to meet either Class A1 or A2 requirements [3]

– The European Union (EU) is considering legislation for 2012 that would require data center hardware to be temperature excursion tolerant up to 45°C [4].

– The selected ranges give 5°C degree and 5%RH margin to the current temperature, and also provide information for the possible future inlet air temperature increase.

The profile shown in Figure 19 was used to mimic this condition. Based on the survey of data centers shown in chapter 3.1, the A/C in the experiment was set to be 20°C and 50%RH. Three switches were exposed to each test condition.

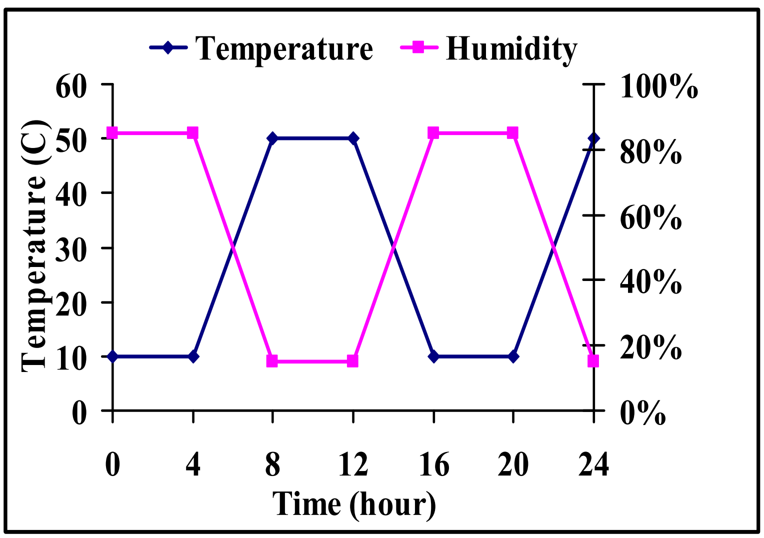


Figure 19: Variable temperature and humidity representing FAC profile

4.3 Test Result Analysis

The experiment was conducted for four months. Three critical performance parameters of the switch were monitored in the experiment: throughput, response time, and transaction rates. The three parameters are related and can be derived from each other, so only throughput was considered in this case. The monitored switch throughputs under A/C and free air cooling condition are shown in Figure 20 to Figure 25, respectively. Based on the monitored samples, the switch throughputs under free air cooling conditions have larger variations than those under A/C.

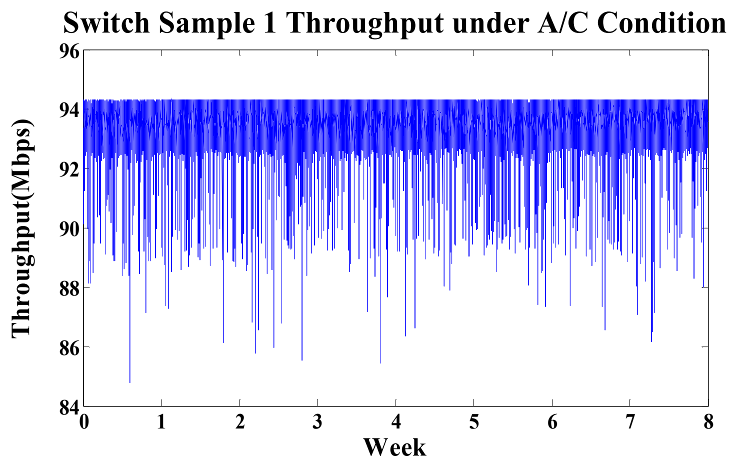


Figure 20: Monitored throughput of switch sample 1 under A/C condition

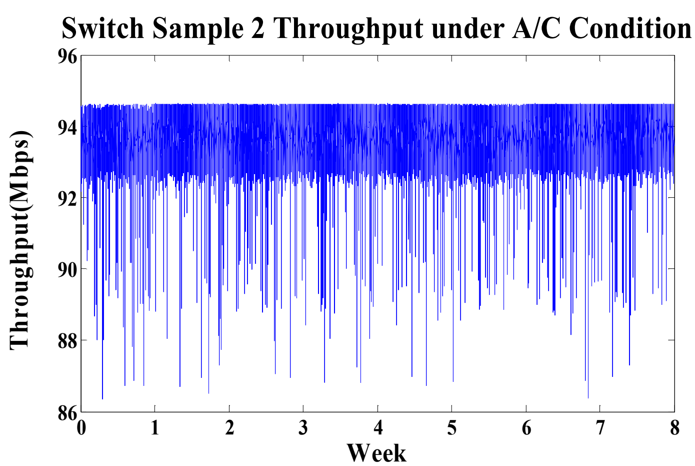


Figure 21: Monitored throughput of switch sample 2 under A/C condition

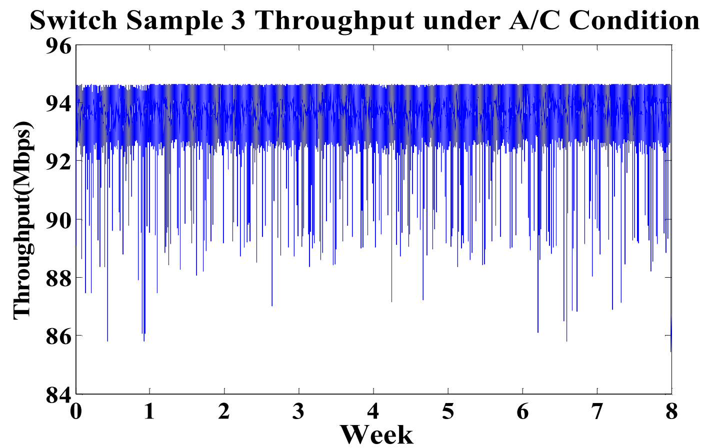


Figure 22: Monitored throughput of switch sample 3 under A/C condition

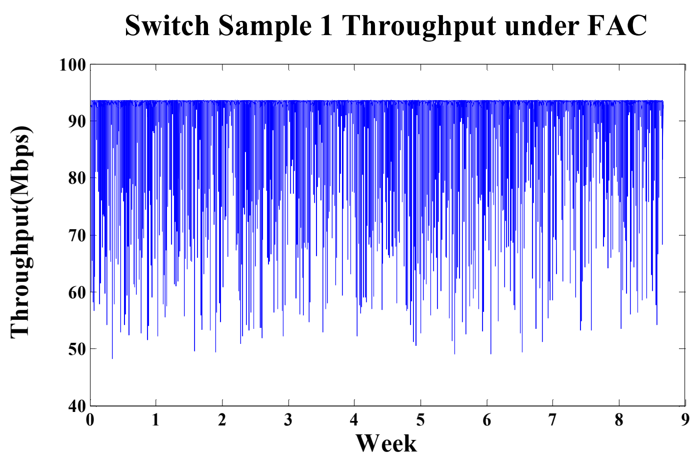


Figure 23: Monitored throughput of switch sample 1 under variable temperature and humidity representing FAC

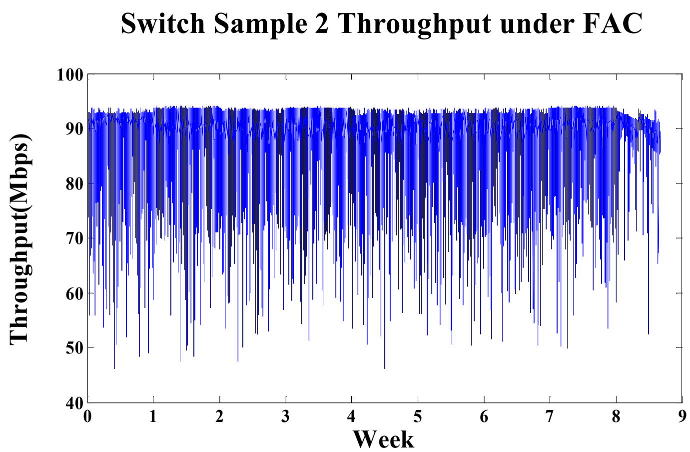


Figure 24: Monitored throughput of switch sample 2 under variable temperature and humidity representing FAC

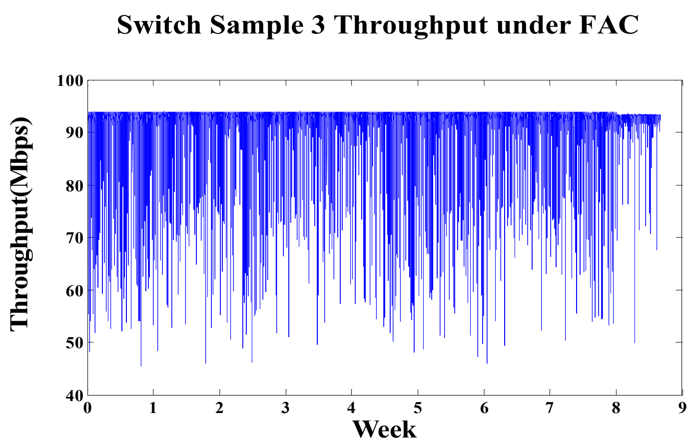


Figure 25: Monitored throughput of switch sample 3 under variable temperature and humidity representing FAC

In order to evaluate the throughput variations of the switches, throughput baselines were created to compare them under the two conditions. In this case, the baseline of every switch was considered as its average throughput of the first 10000 data packets (about one day). The throughput baselines of the six switches are shown in Table 17. Generally, the throughput baselines under the free air cooling condition were a little lower than those under the A/C, since there were more variations decreasing the throughput ability.

Table 17: Throughput baselines of switches

Sample		Baseline (Mbps)	1% off	2% off	5% off	10% off	20% off
A/C	Switch 1	93.579	92.643	91.708	88.900	84.221	74.863
	Switch 2	93.736	92.798	91.861	89.049	84.362	74.989
	Switch 3	93.734	92.797	91.859	89.047	84.361	74.987
Free Air Cooling	Switch 1	92.969	92.040	91.110	88.321	83.672	74.376
	Switch 2	91.322	90.409	89.495	86.756	82.190	73.058
	Switch 3	92.783	91.855	90.927	88.144	83.505	74.226

Some metrics are required to compare and evaluate the performance variations of the switches under the two conditions. The maximum throughput variation of every sample under the two conditions is shown in Table 18.

Table 18: Maximum samples throughput variations

	Maximum throughput variation off the baseline
Sample 1 under A/C condition	9.4%
Sample 2 under A/C condition	7.9%
Sample 3 under A/C condition	8.9%
Sample 1 under variable temperature and humidity	48.1%
Sample 2 under variable temperature and humidity	49.7%
Sample 3 under variable temperature and humidity	51.0%

As described in chapter 4.1, we consider the number/percentage of throughput off the baseline as the possible metrics: 1% off baseline, 2% off baseline, 5% off baseline. The throughput variation comparisons of the three samples under A/C condition with the three metrics are shown in Figure 26-28.

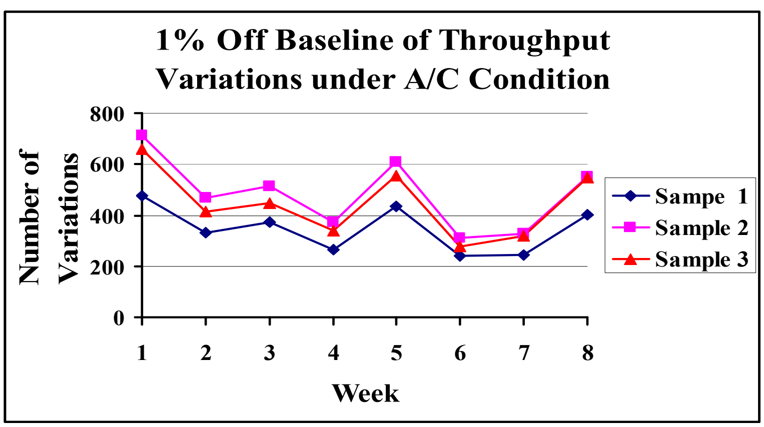


Figure 26: Throughput Variations Comparison under A/C Condition: 1% Off Baseline

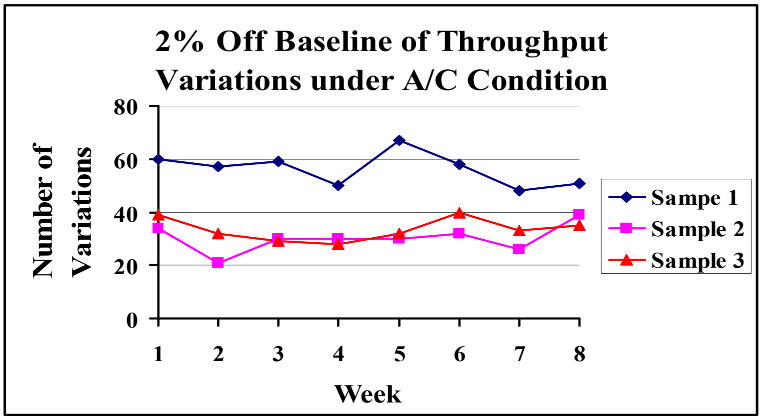


Figure 27: Throughput Variations Comparison under A/C Condition: 2% Off Baseline

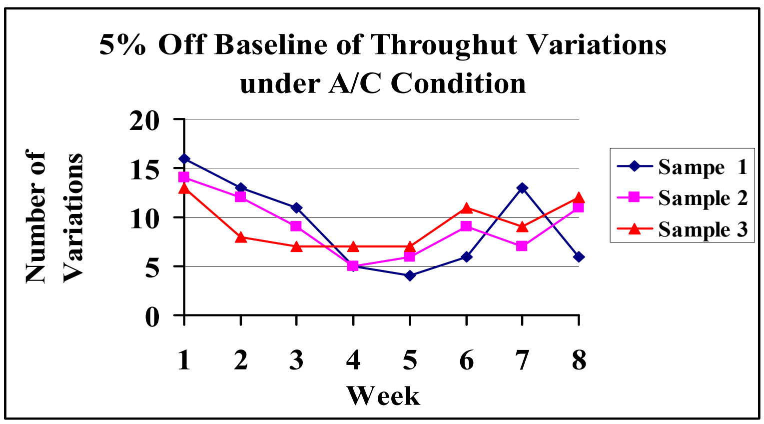


Figure 28: Throughput Variations Comparison under A/C Condition: 5% Off Baseline

The comparisons between the three sample throughput variations under the free air cooling condition and the throughput variations (average of the three samples) under the A/C condition with different metrics are shown in Figure 29-Figure 31[11]. Generally, The switch throughput variations frequencies under variable temperature and humidity representing FAC are significantly increased compared with those under A/C, but the increase rates are various with different samples. There are some jumps in the throughput variation frequency under variable temperature and humidity

[11] FAC week 9 data is extrapolated based on four days test data.

representing FAC condition, which can be considered as the temporary throughput degradations:

– Sample 2: jumps in week 2 with 1% and 2% off baseline, and jump in week 2 and 3 with 5% off baseline.

– Sample 3: jump in week 8 with 2% off baseline.

– Sample 1: is the most stable one and no significant jump is observed.

In addition, there are no 10% off baseline, and 20% off baseline of throughput variations observed under A/C condition, but they occur under variable temperature and humidity representing FAC conditions.

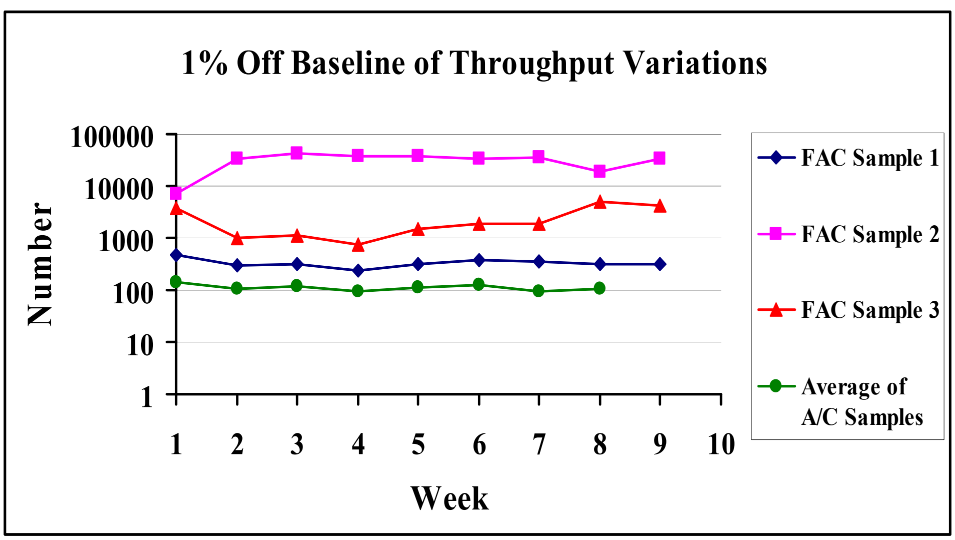


Figure 29: Performance variations of 1% off baseline

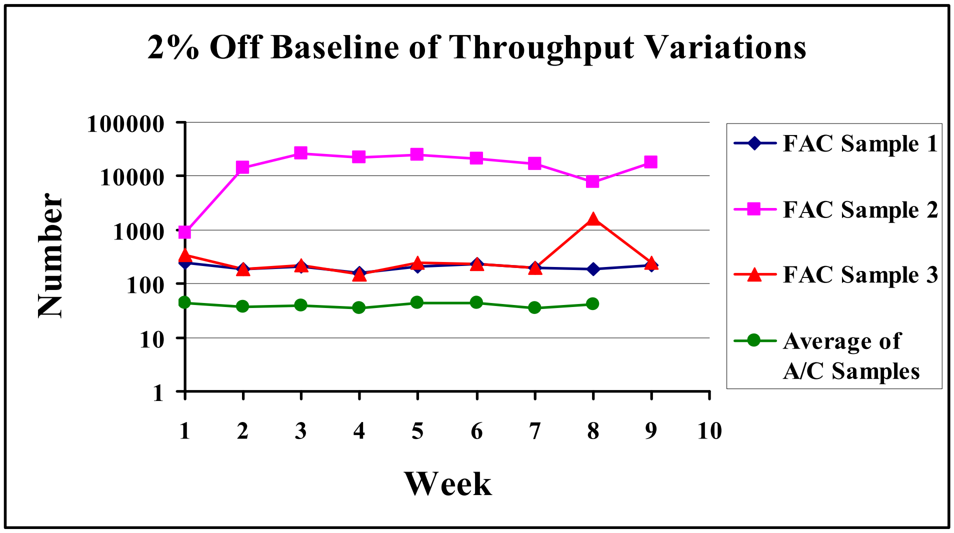


Figure 30: Performance variations of 2% off baseline

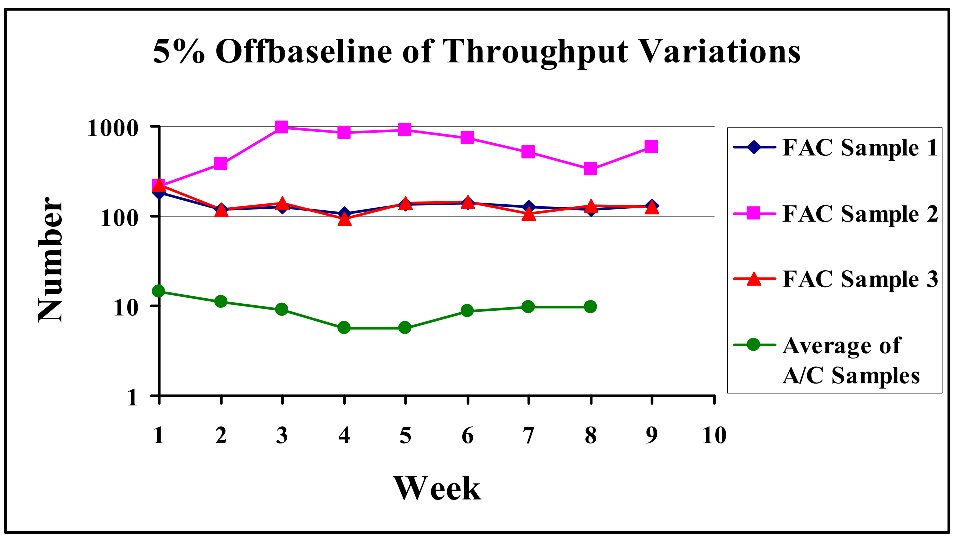


Figure 31: Performance variations of 5% off baseline

The increases over the eight weeks of the average throughput variation frequencies between the two conditions are shown in Table 19. Under simulated free air cooling conditions, the increases of throughput variation frequencies compared with those under air conditioning are various depending on the different metrics. The table shows the variation frequencies of 1% off baseline, 2% off baseline, and 5% off baseline.

Table 19: Increases of throughput variation frequencies

	A/C	Free Air Cooling	Increase
1% off baseline	424	11170	2634%
2% off baseline	40	5806	14515%
5% off baseline	9	146	1622%

4.4 Selection of Appropriate Metric

An appropriate metric needs to be selected from the above-referenced three metrics to compare the performance (throughput) variations between the two conditions. The performances under the A/C condition are used to identify the appropriate metric, since they are the baselines for the comparisons to identify the impact of free air cooling on the performance variations of telecom equipment. Table 20 shows the total throughput variation frequency under the A/C condition. Since there are no throughput variations beyond 10% off baseline and 20% off baseline, the table shows the variation frequencies of 1% off baseline, 2% off baseline, and 5% off baseline.

Table 20: Total throughput variation frequency under the A/C condition

	Switch 1			Switch 2			Switch 3		
Week	1%	2%	5%	1%	2%	5%	1%	2%	5%
1	476	60	16	713	34	14	660	39	13
2	709	117	29	1182	55	26	1076	71	21
3	1083	176	40	1697	85	35	1525	100	28
4	1347	226	45	2069	115	40	1866	128	35
5	1781	293	52	2677	145	46	2420	160	42
6	2020	351	58	2987	177	55	2698	200	53
7	2264	399	71	3313	203	62	3017	233	62
8	2668	450	77	3865	242	73	3564	268	74

Since the lifetime of a switch is usually expected to be more than three years, we assume that there are no significant degradations of the three switch samples in the two month test, and the performance variation of any of the three switches is expected to be more or less constant over the one week period. Therefore, the total performance variation frequency of each sample is expected to increase linearly with each week, and an appropriate metric is expected to represent this linear increase.

For every metric, the expected performance variation frequency under the above assumption can be calculated by least square linear regression, which minimizes the sum of squared errors between the observed values and the expected values [4]. Figure 32 shows an example calculation of the expected total performance variation frequency of 1% off baseline (sample 1) under the A/C condition by least square linear regression[12], which is estimated as:

$$y = 337x \tag{5.1}$$

Where y is the expected total performance variation and x is the number of weeks. For example, when x=1, y=337, which means that there are 337 expected variations of 1% off baseline (sample 1) occurring under the A/C condition every week.

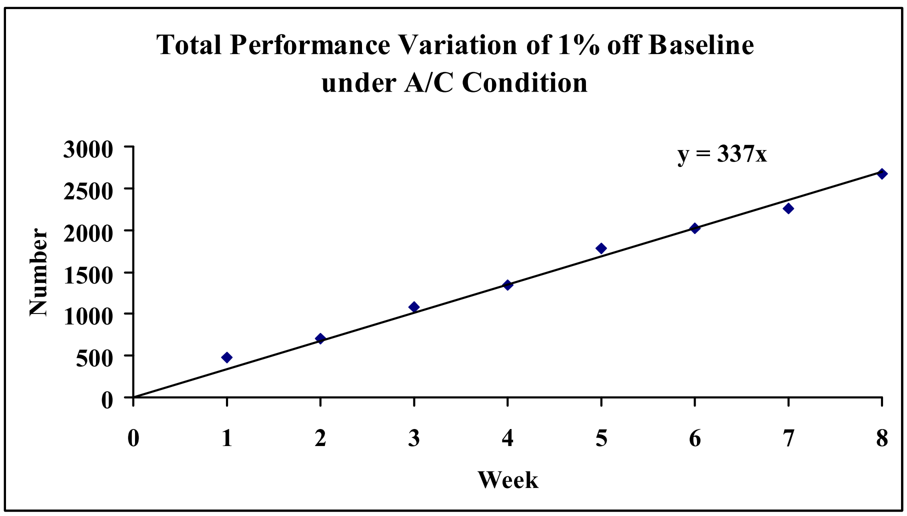


Figure 32: Calculation of expected performance variation frequency of 1% off baseline under the A/C condition

The goodness of linear curve fit represents the effectiveness of the metric to demonstrate the linear increase of the performance variation over time. The Residual Sum of Squares (RSS), the sum of every squared deviation between the observed

[12] There is a boundary condition: there is no performance variation before the test starts, that is y=0 when x=0.

(actual) value and the expected value, is widely used to evaluate the goodness of the linear curve fit, and it is expressed as [5]:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5.2}$$

Where y_i is the observed value and $\hat{y}_i$ is the expected value.

In this case, we propose a new feature normalizing RSS to evaluate the goodness of the linear curve fit and the effectiveness of the metric, which is defined as:

$$\alpha = \frac{\sqrt{RSS}}{\beta} \tag{5.3}$$

Where α is the new feature and β is the expected weekly performance variation frequency (that is the expected performance variation frequency in the first week). For example, the RSS for 1% off baseline (sample 1) under the A/C condition is calculated as 44760, shown in Table 21, and then α is calculated as:

$$\alpha = \frac{\sqrt{44760}}{337} = 0.63 \tag{5.4}$$

Table 21: RSS calculation for 1% off baseline (sample 1) under the A/C condition

Week	Observed value	Expected Value	Error square
1	476	337	19321
2	709	674	1225
3	1083	1011	5184
4	1347	1348	1
5	1781	1685	9216
6	2020	2022	4
7	2264	2359	9025
8	2668	2696	784
RSS			44760

Similar to the above process, α is summarized in Table 22 for all the three metrics and samples. The low α represents the high goodness of the linear curve fit, and as a result, effectively represents the linear increase of performance variation over time. Therefore, an appropriate metric should have a low α with the three samples. According to Table 22, 2% off baseline has the lowest α, so it can be considered an appropriate metric to identify the impact of free air cooling on the performance variations of telecom equipment in this case.

Table 22: α calculation under the A/C condition

Switch 1			Switch 2			Switch 3		
1%	2%	5%	1%	2%	5%	1%	2%	5%
0.63	0.25	1.52	0.91	0.36	1.24	0.86	0.33	0.72

4.5 Conclusions

This chapter identifies the impacts of free air cooling on the performance of telecom equipment. When free air cooling is implemented in data centers, the operating conditions may increase beyond the recommended operating conditions (RoC) of telecom equipment to maximize the energy saving. Furthermore, the hot spot equipment is more likely to run beyond its RoC than equipment cooled by A/C units. The performance variations may be one key concern of free air cooling implementation.

The performance variations vary using different metrics, and thus an appropriate metric is critical to identify the impact of free air cooling on the performance of telecom equipment. Since the performance variations exist even under the A/C, and 2% off baseline was the most stable metric representing the assumption of constant throughput variations every week under A/C in our case, the 2% off baseline was selected as the metric for this comparison of performance variations.

The frequency of the performance variations increases under free air cooling compared with the variations under A/C. The performance variations under free air cooling are also larger than those under the A/C. This kind of large performance variation decreases the quality of data center services and may be unacceptable to the customers of data centers. The future research may include identifying the impact of temperature and humidity on the performance of telecom equipment, respectively, which can help data center operators set proper operating condition ranges without affecting the equipment performance significantly during the implementation of free air cooling.

References:

[1]. R.Blum, "Network Performance Open Source Toolkit: Using Netperf, tcptrace, NISTnet, and SSFNet", Wiley, ISBN: 978-0-471-43301-9, July 2003.

[2]. ASHRAE Technical Committee (TC) 9.9, "2011 Thermal Guidelines for Data Processing Environments – Expanded Data Center Classes and Usage Guidance," Atlanta, 2011.

[3]. Homorodi, T., Fitch, J., "Fresh Air Cooling Research", Dell Techcenter, July 2010.

[4]. J.Y. Audibert, and O. Catoni, "Robust Linear Least Squares Regression," The Annals of Statistics, Vol. 39, No. 5, 2011, pages 2766–2794.

[5]. A. Cottrell, "Regression Analysis: Basic Concepts," the course of Econometric Theory and Methods, Department of Economics, Wake Forest University, 2011.

Chapter 5. A Multi-Stage Risk Mitigation Approach for Free Air Cooling

This dissertation develops a multi-stage risks mitigation approach for the free air cooling, which considers three stages of the equipment life cycle (shown in Figure 33): design, test, and operation. Design is the process of originating a plan for creating a product, structure, system, or part. Test is the process of using machines, tools, equipment, and experiments to assess the capability of a product to meet its requirements. Operation means that the equipment is already in place and is being used by the end users. The risk mitigation process starts with the identification of operating condition ranges with free air cooling for data centers. Identification of an operating condition range is fundamentally necessary for the assessment.

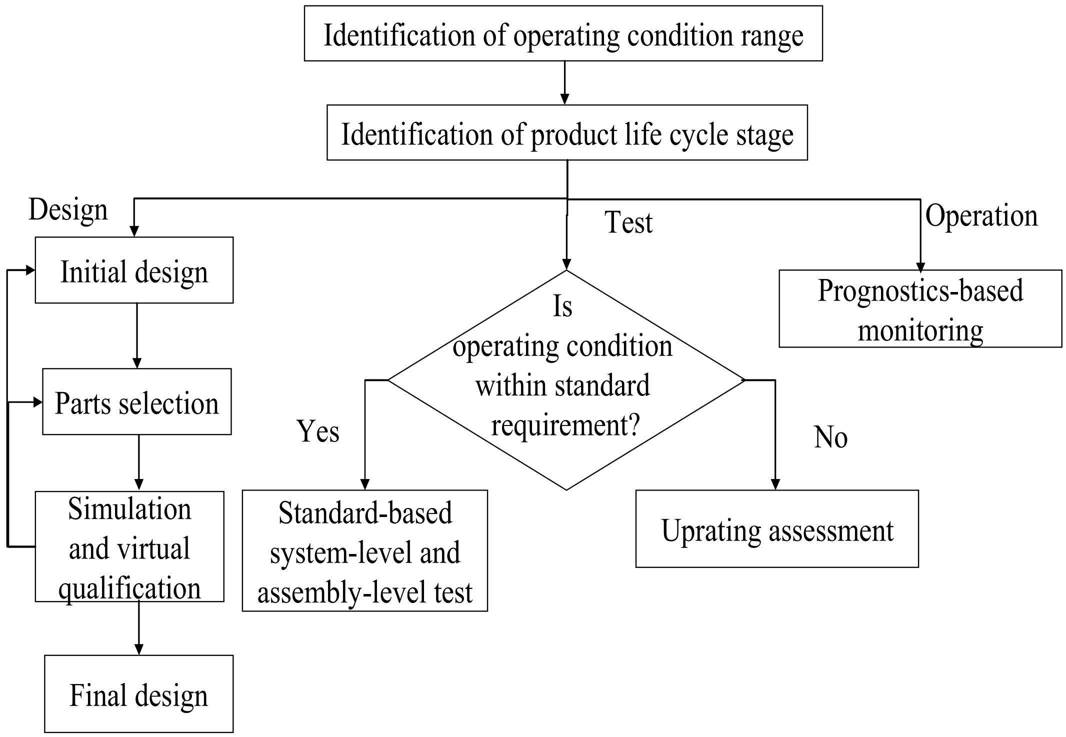


Figure 33: Schematic of the multi-stage risk mitigation process.

The next step in the assessment process is the identification of the product's life cycle stage. In the next three sections, the evaluation process is shown for the three identified product lifecycle stages. In each subsection, the available information and hardware are discussed first, and then the process is described within those constraints.

5.1 Design Stage

During the design stage, the functional requirements of the product are defined. However, the hardware is not yet finalized. Even though the product does not exist physically, the material and performance information regarding the potential parts can be used to assess the performance and reliability of the product. Prior experience with the use of similar parts and designs can also be used to augment existing information for part assessment. When the equipment is assessed at this stage, an iterative process is followed to finalize the design and bill of material.

In the design stage, the evaluation process includes the initial design, part selection, simulation and virtual qualification, and final design. The first step in the design stage is to design a product based on an understanding of the information about the new environment and the expected functional and reliability requirements. This paper deals with environmental conditions with wider ranges of temperature and humidity. The local operating temperatures of the parts in the system will need to be estimated for the new environmental range. Another issue that affects a part's temperature is the cooling algorithm of its local cooling equipment (e.g., fan). When the fan speed is dynamic and varies based on the part's temperature changes, the part temperature might not increase linearly with the ambient temperature changes, since the fan speed will increase [1]. Other information, such as data on part performance in similar operating conditions, can also be useful to estimate the part temperature.

Part selection is based on the part data sheets and local operating conditions. The data sheets contain information such as part type and category, electrical ratings, thermal ratings, mechanical ratings, electrical and thermal characteristics, and functional descriptions. In this case, two types of ratings are important for the selection: absolute maximum ratings (AMR) and recommended operating conditions (RoC). The IEC defines AMR as the "limiting values of operating and environmental conditions applicable to any electronic device of a specific type as defined by its published data, which should not be exceeded under the worst possible conditions"

[2]. Recommended operating conditions are the ratings on a part within which the electrical specifications are guaranteed [2]. The local operating conditions can be estimated based on historical data under similar operational conditions or calculated using computational fluid dynamics (CFD) and other simulation tools and data from similar products in the same technology family. Using the content of the datasheets and the local operating conditions, the parts can be initially selected.

In the third step in this phase, the initial design with the initially selected parts is evaluated and improved by a virtual qualification process. Performance simulation and virtual qualification are used to evaluate the functional performance and the reliability of the product, respectively. During the performance evaluation, the system design is assessed to determine whether it can meet or exceed the expected functional requirements under the life cycle conditions. An example of a performance simulation tool for semiconductors is the Simulation Program with Integrated Circuit Emphasis (SPICE) [3]. This tool and its many commercial and academic variations are used to evaluate the performance of semiconductor parts. The virtual qualification process uses physics of failure (PoF) models of the critical failure mechanisms [4]. The failure mechanisms can be identified and their impact on the parts can be estimated through this process. To begin, stress analysis is performed to determine the local operating conditions of the parts. These stress analysis results are then used as input for the failure models. For example, some versions of SPICE, such as HSPICE and PSPICE, have functionalities that can help evaluate electronic parts for their susceptibility to several electrically-driven failure mechanisms, such as time dependent dielectric breakdown (TDDB) and negative bias temperature instability (NBTI). When used in this manner, SPICE serves as a virtual qualification tool.

Based on the results of virtual qualification, the design is improved and parts are re-selected, if necessary. Then, the improved design with the new parts is re-evaluated by virtual qualification. This process is repeated until the results show that the design meets the expected requirements under the new environment.

This phase closes with the creation and release of the final design to the product manufacturing stage. From this point on, additional testing will continue to assess the manufactured product. This process is described in the next section.

5.2 Test Stage

Typically, the manufacturing process of a product has been completed by the time the test stage is reached. If a product's design is modified during its manufacturing process, the assessment should go back to the design stage and the evaluation should be re-started. Based on the identification of operating condition ranges under free air cooling conditions, it can be determined whether the operating conditions are within the standards' requirements. If so, the equipment will be evaluated by the test methods provided by the standard. Otherwise, the equipment will be evaluated by an uprating method.

5.2.1 Standards-based Assessment

This paper applies a widely used standard, Telcordia Generic Requirements GR-63-CORE [5], for system-level and assembly-level assessment for free air cooling conditions. The Telcordia GR-63-CORE provides a test method for equipment in a network equipment building system. Its operating condition requirements are shown in Table 15. If the ambient temperature and relative humidity associated with free air cooling conditions are within the required ranges, the tests in Telcordia Generic Requirements GR-63-CORE are valid for the equipment.

The operating temperature and humidity test in this standard can be used to assess the risks of free air cooling at the system level and the assembly level. The equipment operates during the test, which lasts about one week. The failure criteria are based on the ability of the equipment to operate throughout the test period. If a product can operate properly during the test, it can be considered satisfactorily functional. This test is performed for qualification, but cannot be used to predict reliability over the expected lifetime. In the test, the controlled conditions are temperature and relative humidity. The temperature range is from -5^{o}C to 50^{o}C, and the relative humidity range is from less than 15% to 90%. The temperature and humidity profiles are shown in Figure 34 (the test is operated for only one cycle with the temperature and humidity profiles in Figure 34).

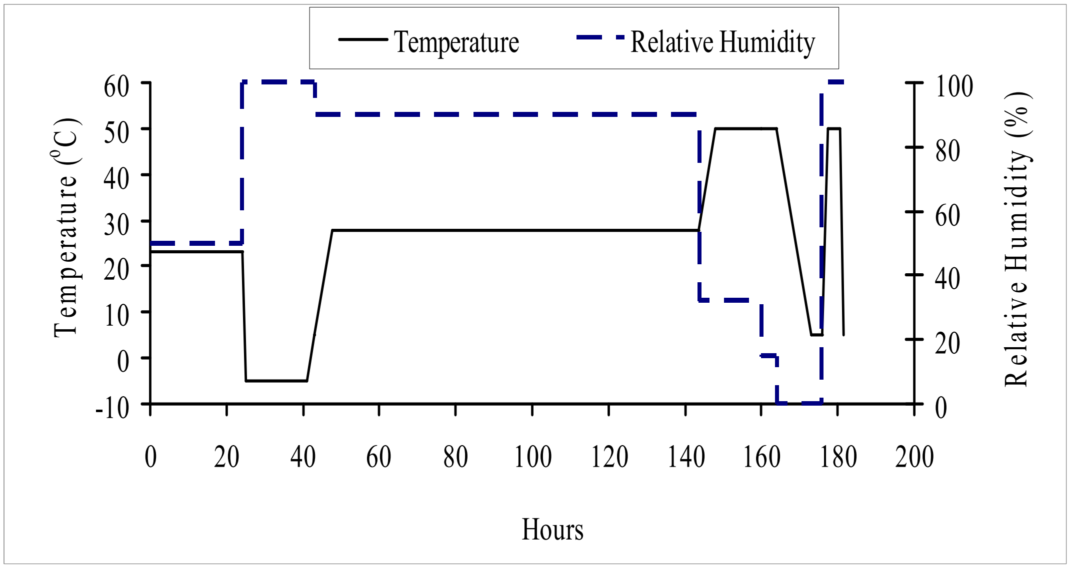


Figure 34: Operating temperature and humidity test[13] [5].

5.2.2 Uprating Assessment

If the operating conditions with free air cooling are outside the standards' requirements, the standards-based method is no longer valid. In this case, some parts located at hot spots may be operated beyond their AMR or RoC. The parts with small thermal margins are also at risk when operating beyond their specific ranges under free air cooling conditions. A practical alternative way to evaluate the risks of free air cooling conditions is through uprating assessment of parts at risk.

IEC Standard 62240 [6] provides uprating tests to ensure that products can meet the functionality requirements of applications outside manufacturer-specified temperature ranges. The first step in part-level testing is to identify whether the operating temperature for the part exceeds the manufacturer-specified temperature range. The operating temperature of the parts can be obtained from the system-level

[13] There are some special humidity profiles in the test that are ranges of RH instead of exact RH values, such as any RH and any RH below 15%. In this figure, 100% represents any RH and 0% represents any RH below 15%.

and subsystem-level testing results and additional analysis. If the operating temperature increases beyond the manufacturer-specified ranges, the entire process in Figure 35 should be performed. If the increased operating temperature is still within the specified range, only the part functionality test over the application temperature range is required to ensure that the part can operate as expected within the higher operating temperature range.

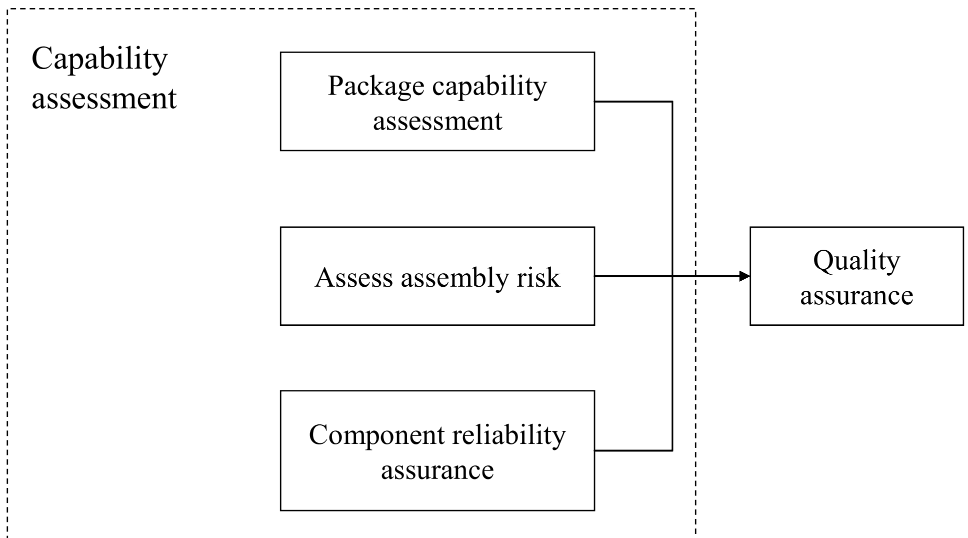


Figure 35: Testing flowchart in IEC Standard 62240 [6].

The process in Figure 35 starts with a capability assessment, which consists of three steps: package capability assessment, subsystem risk assessment, and part reliability assurance. The package capability assessment ensures that the package and internal construction can withstand a higher temperature and will not cause any material properties to change based on analyzing the part's qualification test data and other applicable data. A subsystem risk assessment estimates the ability of a device to perform under a higher temperature. Re-characterization is recommended because it uses processes similar to those used by part manufacturers for original part characterization. Part reliability assurance qualifies a part based on the application requirements and performance requirements over the intended range of operating conditions.

The flowchart of parameter re-characterization is shown in Figure 36. The re-characterization process consists of three main parts: (1) choosing the statistical features of the parameters and calculating the sample size *N*; (2) executing the test and calculating the margins *M*; and (3) conducting the necessary assembly functional test. If the calculated margin is large enough and the functional test is passed, the component is considered to have passed the uprating test.

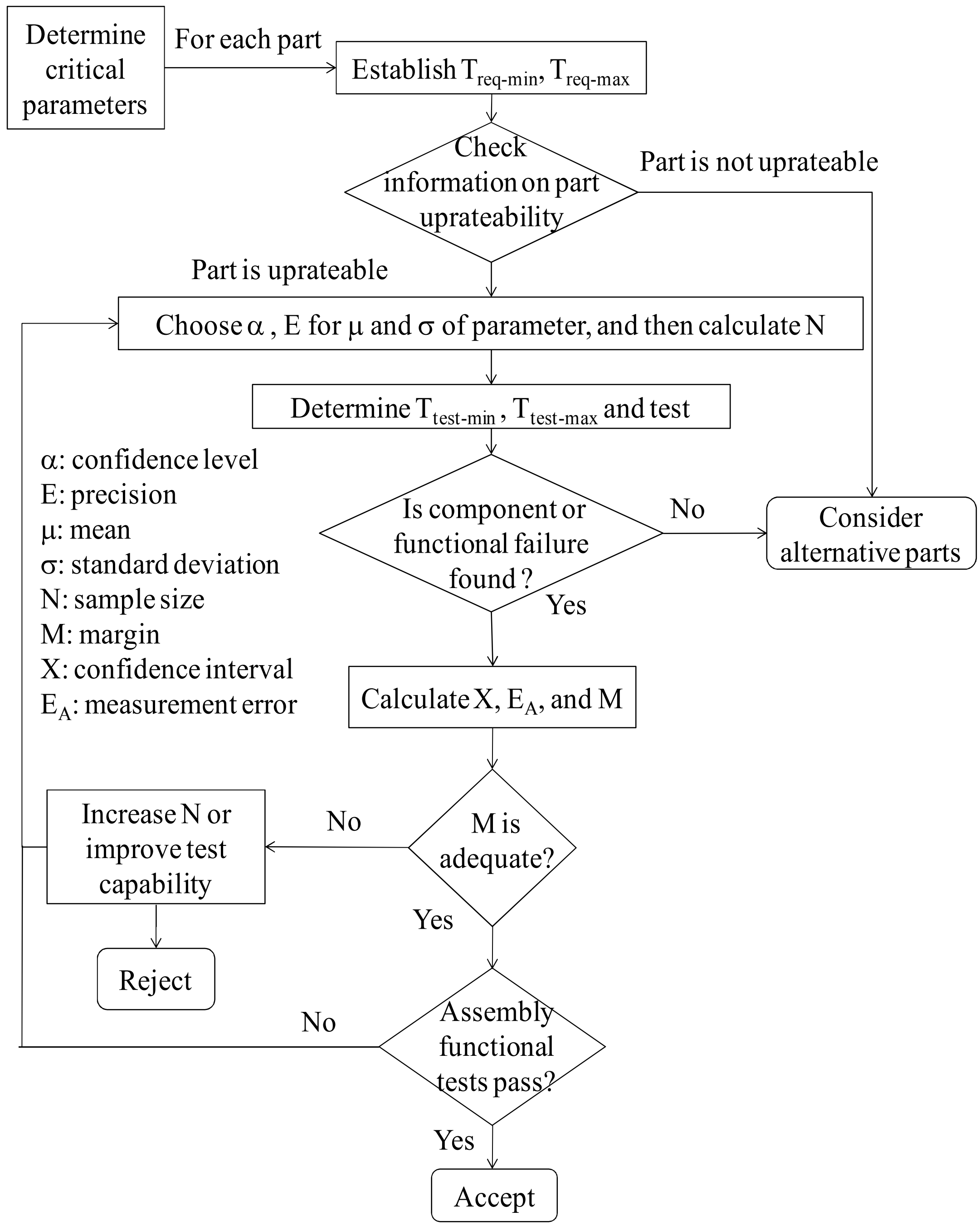


Figure 36: Flowchart of parameter re-characterization process [6].

Quality assurance secures the ongoing quality of successfully uprated parts by monitoring the part process change notices obtained from the manufacturers. In this assessment, parameter re-characterization testing can be used to assess incoming parts, and change monitoring can be used to give warning of a part change that would affect its ability to operate under an increased operating temperature.

It should be noted that uprating is a very expensive process, and the qualification of an uprated part has to be redone if anything about the part manufacturing process is changed. With good thermal design at the system level and careful component selection, uprating can usually be avoided.

5.3 Operation Stage

For data centers already in operation, it is usually not practical to re-qualify equipment and evaluate the risks when free air cooling is considered. Firstly, it is not practical to take equipment out of service for testing, especially in mission-critical data centers with 24/7 service requirements. Secondly, even if the test were attempted, the tested equipment would lose some of its useful life in the test process. This would offset the energy savings of free air cooling. Thirdly, it seems impossible to gather an appropriate sample size of equipment already in operation for the test. In addition, accelerating the life-cycle conditions for an entire data center is impractical, prohibitively expensive, and unlikely to provide useful information on the reliability of a system. We propose using prognostics and health monitoring as a retrofitting technique to assess and mitigate the risks and enable the implementation of free air cooling in data centers that were not initially designed for this cooling method.

5.3.1 Introduction of PHM

Prognostics and health management (PHM) uses in situ system monitoring and data analysis to identify the onset of abnormal behavior that may lead to either intermittent out-of-specification performance or permanent equipment failure. This novel method need not interrupt the service of data centers for the purpose of reliability assessment. PHM permits the assessment of the reliability of a product (or system) during operation [7]. Generally, PHM can be implemented using the physics-of-failure (PoF) approach, the data-driven approach, or a combination of both (the fusion approach) [8].

The physics-of-failure method uses knowledge of a product's life-cycle loading and failure mechanisms to perform reliability design and assessment [7][9][10][11][12]. The data-driven approach uses mathematical analysis of current and historical data to provide signals of abnormal behavior and estimate remaining useful life (RUL) [7]. The fusion model combines PoF and data-driven models for prognostics [8], overcoming some drawbacks of using either approach alone. The flowchart of the fusion model is shown in Figure 37.

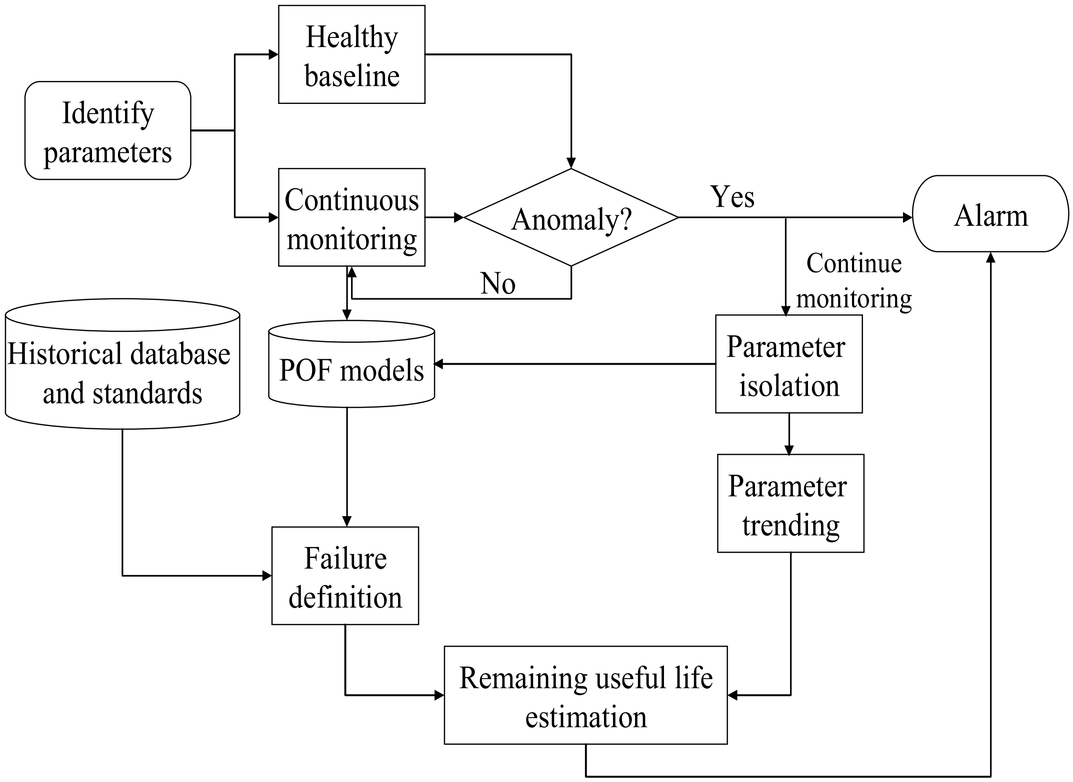


Figure 37: Flowchart of the fusion approach [8]

Each of the PHM approaches can potentially be implemented to assess the reliability risks under free air cooling in data centers. PHM can be applied to a collection of equipment such as servers, storage, and routers. It takes advantage of some unique features of data centers under free air conditions

Because the air side economizers work within set points of temperature to supply untreated air, the input and outlet air temperatures are key indicators of cooling system effectiveness. Typical data centers also use thermal simulation results to identify hot spots and air flow patterns. Noise is also generated in data centers, such as the movement of internal air due to fans and blowers and the rotation of hard disk drives. There are also functional specifications for the data centers in terms of access times, band width, and related performance features. The overall power consumption of the systems is a further operational consideration. These factors can all be monitored and recorded as inputs to a PHM system.

Operating data center equipment can also be monitored for function and performance. Functional parameter for storage and communications systems are defined and monitored at both the system and the component levels. This ability to monitor environmental and operational data allows the use of data-driven PHM at the system level with only a limited need for additional sensing, monitoring, storage, and transmission tools.

Critical subsystems must be identified through functional analysis. Based on in situ load monitoring of the parameters of critical subsystems (e.g., voltage, current, resistance, and thermal resistance), PoF can calculate the accumulated damage to these parts caused by known critical failure mechanisms, and predict the RUL of the subsystem.

The data-driven method detects system anomalies based on system monitoring that covers performance (e.g., uptime, downtime, and quality of service) and other system parameters (e.g., voltage, current, resistance, temperature, humidity, vibration, and acoustic signal). The data-driven method identifies failure precursor parameters based on system performance and the collected data. When there are pre-defined failure criteria for the system, the trending of the failure precursor parameters is used to predict the RUL. Note that the main purpose of PHM is to predict wear-out failures, rather than intermittent failures. However, PHM can help identify impending intermittent failures—that is, the loss of some performance characteristic for a limited period, with subsequent recovery of the function. This type of failure is usually not repeatable but can be disruptive. PHM can identify the possibility of intermittent failures by monitoring and analyzing the failure precursor parameter trending through the data-driven approach. The collected data is also useful for failure analysis.

5.3.2 A Prognostics-based Approach for Risk Mitigation in Free Air Cooling

A prognostics-based approach is implemented to assess and mitigate risks arising from free air cooling, as shown in Figure 38. The first step is identifying the operating condition range set by operators when free air cooling is implemented. Based on the identified operating condition range, a failure modes, mechanisms, and effects analysis (FMMEA) is conducted to identify the weakest subsystems/components because they are the ones most likely to fail first in a system. FMMEA is a methodology used to identify critical failure mechanisms, which is defined as "the process by which a specific combination of physical, electrical, chemical, and mechanical stresses induce failure" [13]. FMMEA uses a life cycle profile to identify active stresses and select the potential failure mechanisms. The failure mechanisms need to be prioritized based on knowledge of load type, level, and frequency, combined with the failure sites, severity, and likelihood of occurrence [13]. The failure sites associated with the critical failure mechanisms are identified as the weakest subsystems/components in a system, and the parameters of the system and its weakest subsystems/components (e.g., voltage, current, resistance, temperature, and impedance) will be selected to be monitored with consideration of measurement applicability. Based on the monitored data, the three PHM approaches

(PoF, data-driven, and fusion) can be used to detect anomalies and perform risk mitigation.

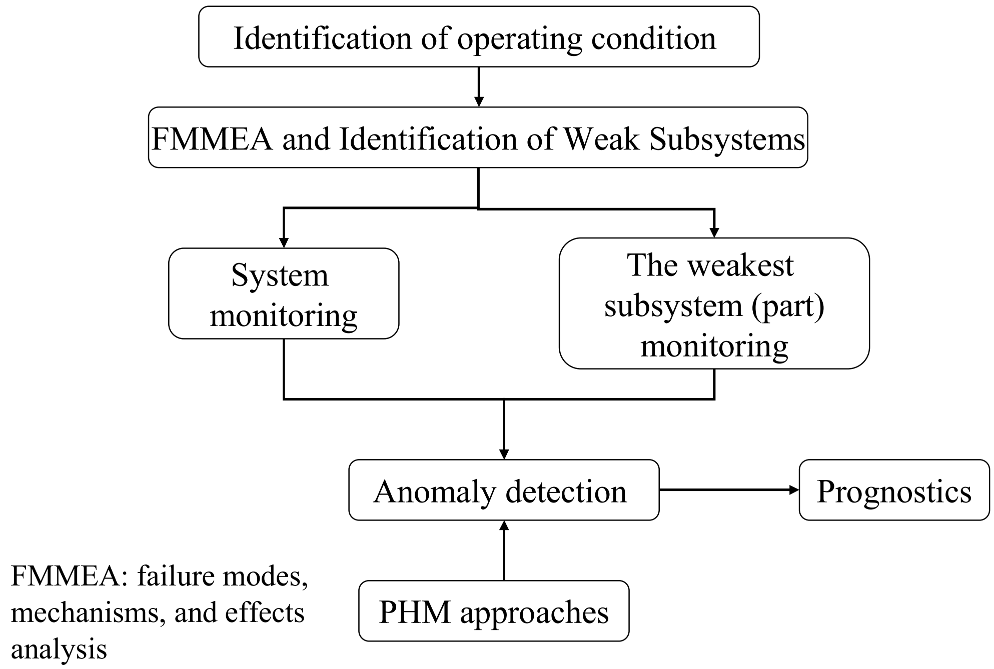
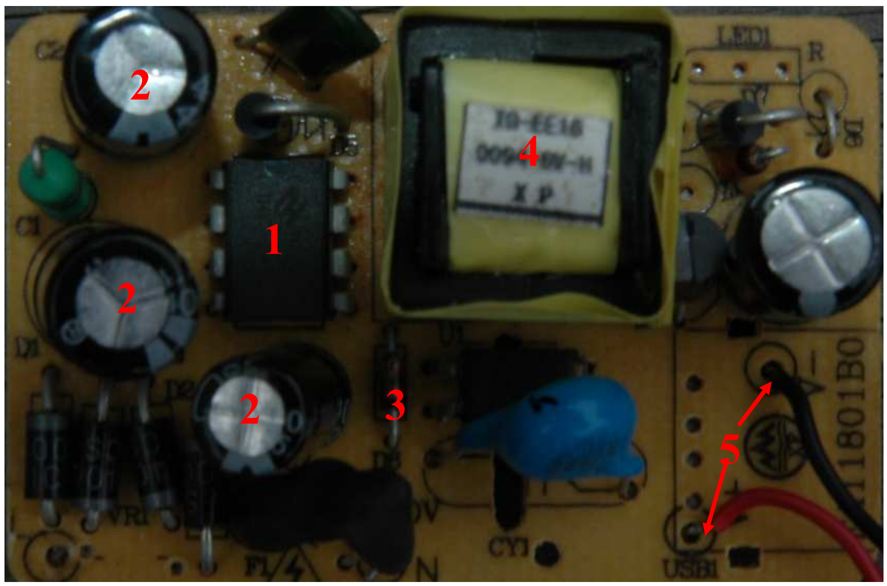


Figure 38: A prognostics-based approach to mitigate the risks at operation stage.

5.3.3 Case Study of Network Equipment in a Data Center

In addition to servers, the network equipment (e.g., routers and switches) are also key IT equipment in data centers. The network equipment selected for this study was the power adapter of a Zonet ZFS 3015P switch, which is widely used in offices and small enterprises. In this case study, we implemented a data-driven method to detect anomalies in the power adapter to provide early warning of failure, since a power adapter is necessary for the operation of telecom equipment. Also, a power adapter is an analog circuit, which makes the observation of parameter shifts and implementation of PHM to detect anomalies straightforward. A block diagram of a power adapter is shown in Figure 39. For the power adapter, the performance parameter is the output voltage, which has a rated output of 9 V. It is considered to have failed when the output voltage drops more than 10% from the rated value (i.e., 8.1 V).



#1—THX 202H IC; #2—aluminum electrolytic capacitor; #3—resistor; #4—power transformer; #5—output voltage supply.

Figure 39: Power adapter of Zonet ZFS 3015p switches.

5.3.3.1 Identification of Operating Condition

The operating conditions are set by data centers and usually determined by the amount of energy savings expected from the implementation of free air cooling. In this case, we assumed that the operating conditions were 0–50^{o}C and 5–95% RH in order to maximize energy savings[14] (the rated operating conditions of the power adapter are 0–40^{o}C and 10–90% RH). We used conditions of 95^{o}C and 70% RH in the experiment to increase the rate of degradation. The power adapter was placed inside an environmental chamber and was in operation for the duration of the experiment. An Agilent 34970A data acquisition monitor was used to monitor and record the parameter trends of the power adapter.

5.3.3.2 FMMEA and Identification of Weak Subsystems

The power adapter in this case is a kind of switched-mode power supply (SMPS). According to the FMMEA results of SMPS [15], the critical failure mechanisms are aging of the electrolyte, wire melt due to current overload, thermal fatigue, contact mitigation, time dependent dielectric breakdown, and solder joint fatigue. The components with a high risk due to critical failure mechanisms are the aluminum electrolytic capacitor, the diode, the power metal-oxide-semiconductor field-effect transistor (MOSFET), the transformer, and the integrated circuit (IC).

5.3.3.3 System and Weak Subsystem Monitoring

With consideration of measurement applicability, four parameters of capacitors and integrated circuits were monitored in this experiment: the voltages of the three capacitors (shown as #2 in Figure 39), and the output frequency of the THX 202H IC (shown as #1 in Figure 39). In addition, the output voltage across the power adapter was also monitored for the power adapter performance trends (shown as #5 in Figure 39). The first step was to monitor the parameter shifts that indicate the degradation

[14] A thermal analysis of the local component temperature based on the operating condition is not considered in this paper, since it is not a requirement for the implementation of the data-driven method used to detect anomalies and provide early warning of failure.

trend of the power adapter. In this case, there were three samples of switches in the experiment, and five parameters of every power adapter sample were monitored: the IC frequency, three capacitor voltage, and the output voltage (the supply voltage to the switch)[15]. The monitored results are shown in Figure 40 to Figure 48. It is observed that the output voltage dropped from 9.3V to about 3V, which means that the power adapters could not supply the rated voltage to the switches for their normal functions, and the power adapters failed in other words. When the failures occurred, the IC frequency also had dramatic drops from about 150KHZ to about 10KHZ, which are slight differences among the samples. The voltages of capacitor 1 have generally gradual degradation processes in the experiment, and the voltages of capacitor 2 and 3 kept general constant even when the power adapters failed.

The monitored parameters in these figures are explained as follows:

- Frequency: IC chip frequency
- V1: voltage of capacitor 1
- V2: voltage of capacitor 2
- V3: voltage of capacitor 3
- Vout: output voltage of power adapter

[15] In this case, the switches connected with the three power adapter samples were not required to transfer the data packets due to the safety concerns in the high temperature condition (95^{o}C). That was, the switches had no workload. The parameter trends in this condition may be different from those that the switches have full workload to transfer the data packets.

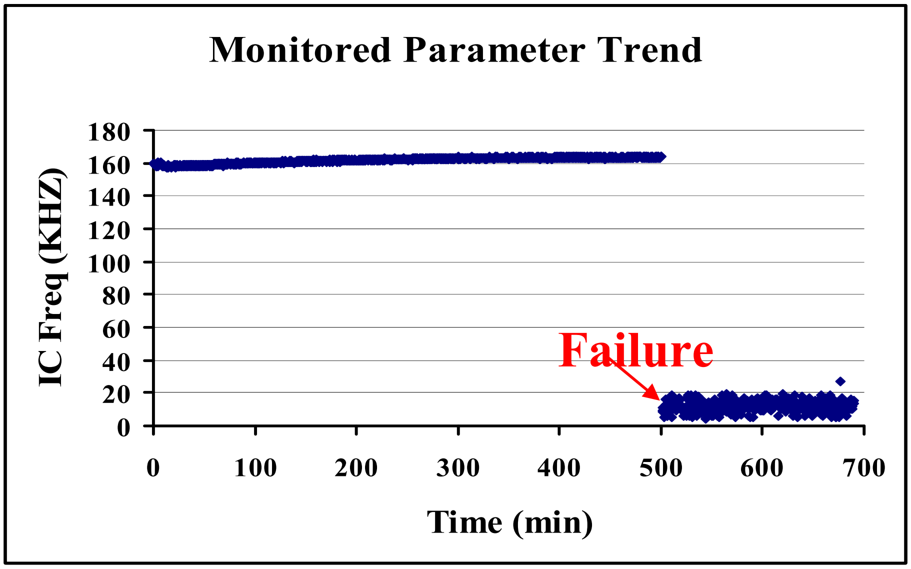


Figure 40: The IC frequency of power adapter 1.

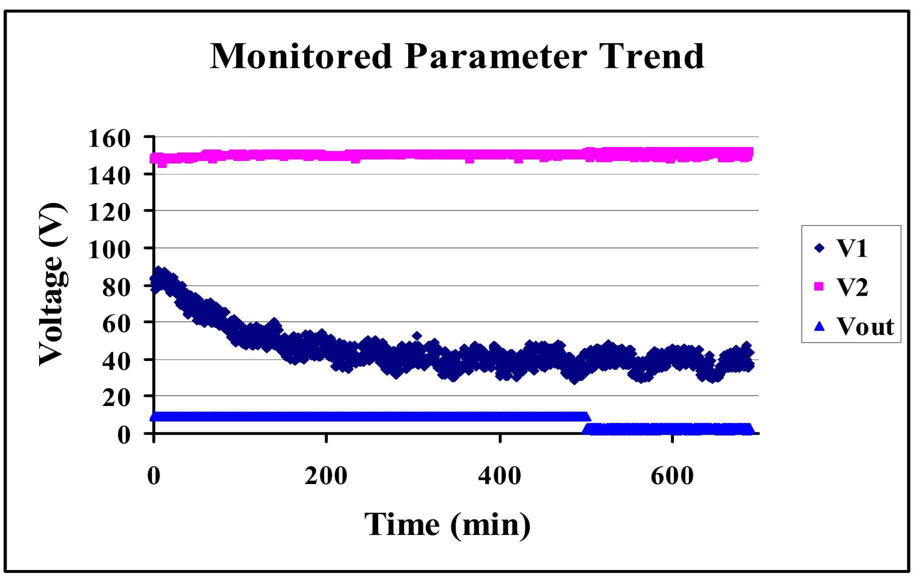


Figure 41: The voltages across capacitor 1 and 2 of power adapter 1

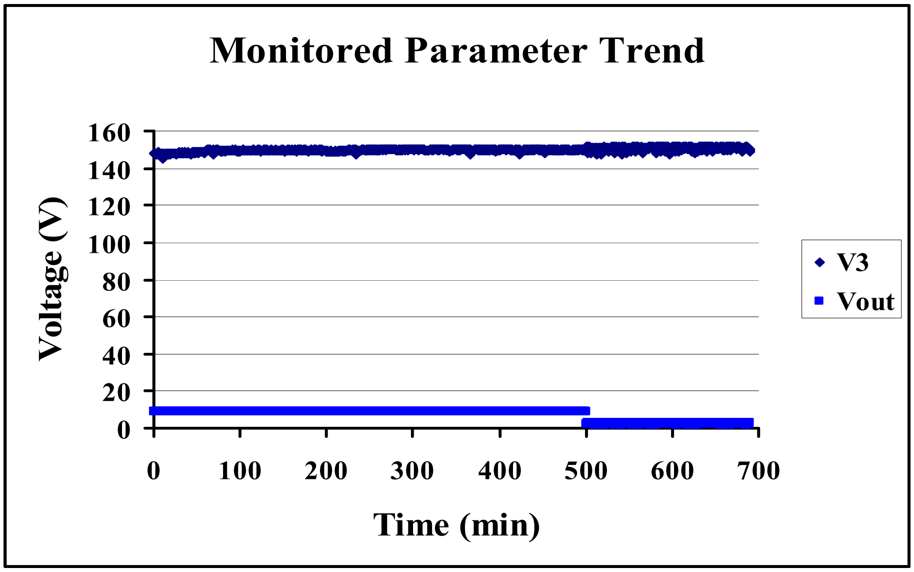


Figure 42: Output voltage and voltage across capacitor 3 of power adapter 1

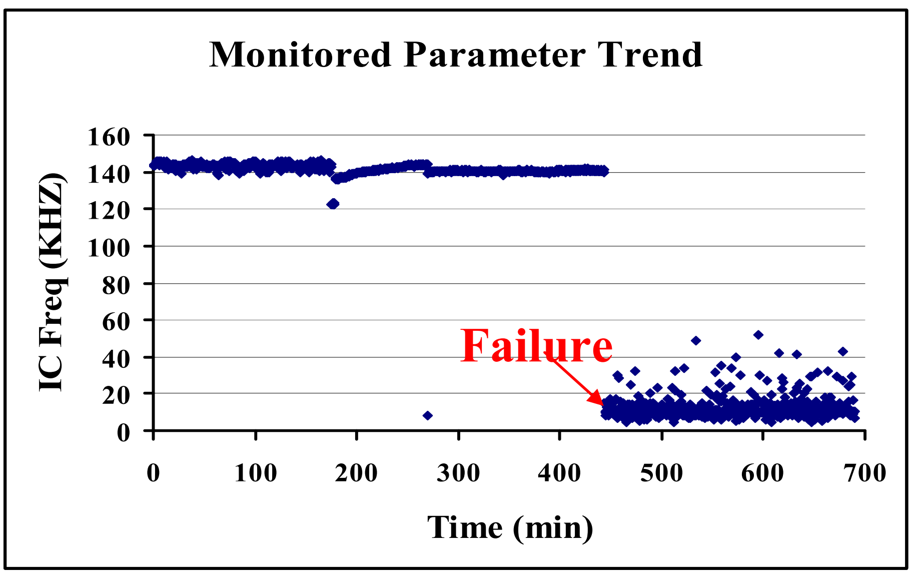


Figure 43: The IC frequency of power adapter 2

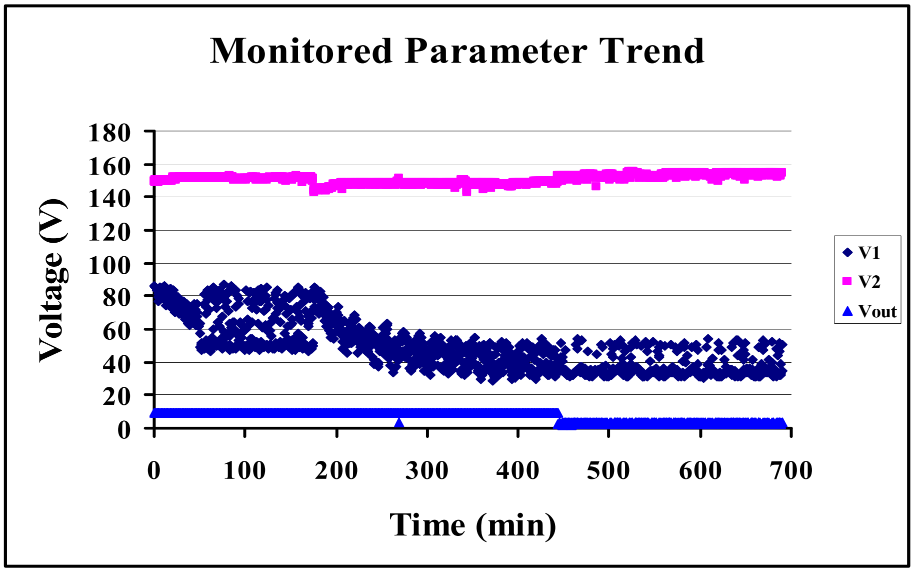


Figure 44: The voltages across capacitor 1 and 2 of power adapter 2

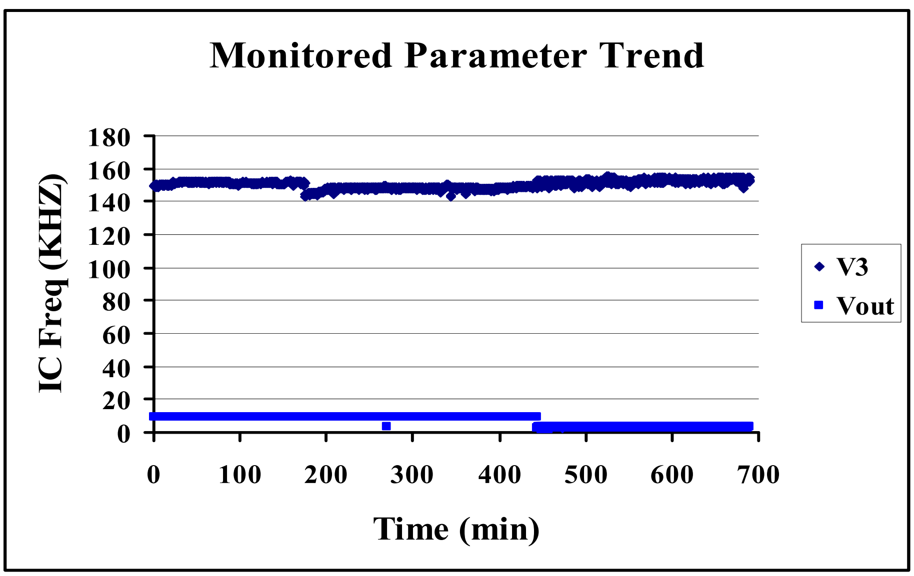


Figure 45: Output voltage and voltage across capacitor 3 of power adapter 2

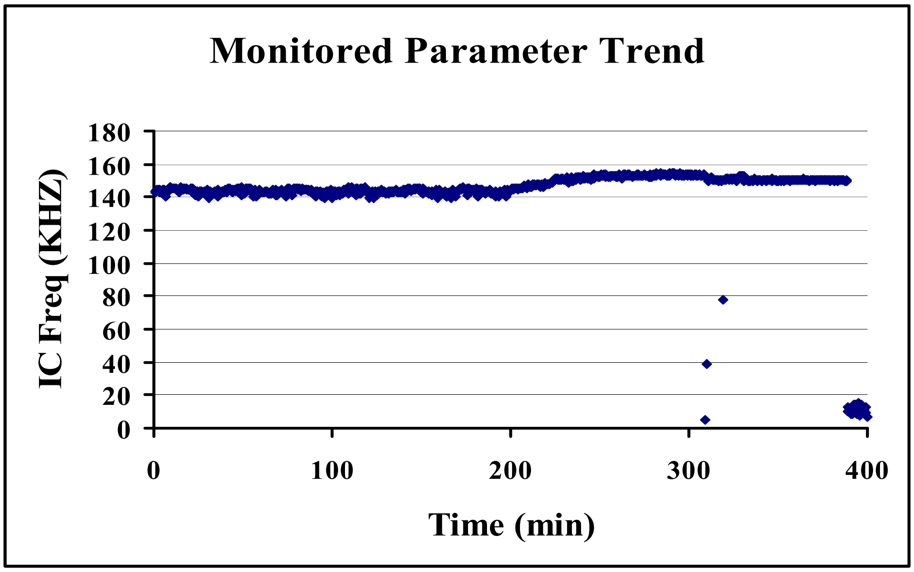


Figure 46: The IC frequency of power adapter 3

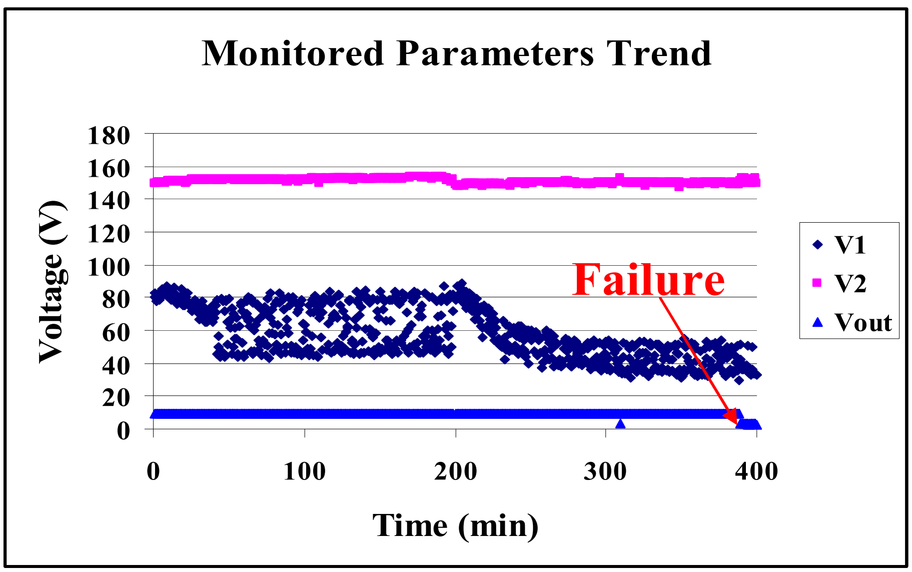


Figure 47: The voltages across capacitor 1 and 2 of power adapter 3

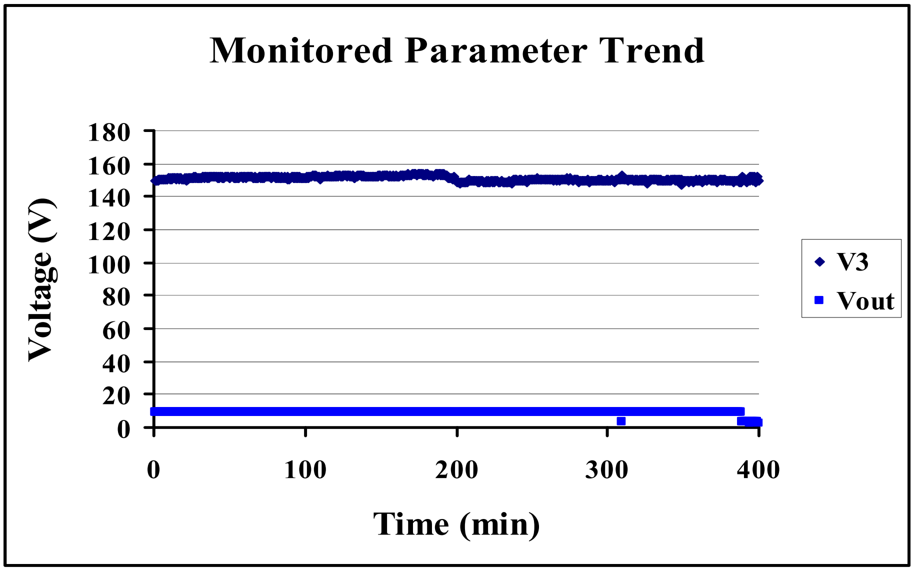


Figure 48: Output voltage and voltage across capacitor 3 of power adapter 3

The detailed comparisons between the health baselines and the final values are shown in Table 23. It is noted that the health baseline of every monitored parameter is the average value of its first 20 data (10 min) in the experiment, which are considered as the health data. The final value of the every monitored parameter is the average value of its first 20 data (10 min) after the power adapter failed. It is observed that the IC frequency and the voltage of capacitor 1 have big shifts during the failure process in the experiment.

Table 23: Summary of the monitored parameters shifts

	Failure time	Parameter	Baseline	Final value	Shift
Sample 1	501 minutes	V1 (V)	82.1	38.5	53.7%
		V2(V)	148.0	150.5	1.4%
		V3(V)	147.9	150.3	1.4%
		Vout (V)	9.34	2.74	70.7%
		Frequency (kHZ)	159.0	10.4	93.7%
Sample 2	444 minutes	V1 (V)	82.9	37.8	54.2%
		V2(V)	149.8	152.1	1.3%
		V3(V)	149.8	151.6	1.3%
		Vout (V)	9.35	2.90	69.0%
		Frequency (kHZ)	144.9	11.4	92.4%
Sample 3	389 minutes	V1 (V)	82.4	38.6	53.7%
		V2(V)	150.3	151.3	0.7%
		V3(V)	150.3	150.6	0.7%
		Vout (V)	9.35	3.28	65.8%
		Frequency (kHZ)	143.8	11.0	92.4%

5.3.3.4 Anomaly Detection

Back propagation neural network is adaptive statistical models based on an analogy with the structure of the brain. The general idea is to generate the output of a neuron based on a function of the weighted sum of the inputs plus a bias, as shown in Figure 49. The main benefits of back propagation neural network to be implemented to detect anomaly in this case include: (1) applicability for nonlinear statistical modeling, (2) no need for distribution assumptions, (3) no need for degradation models (4) supervise training with mapping input to desired output. These benefits exactly fit the data features in this case.

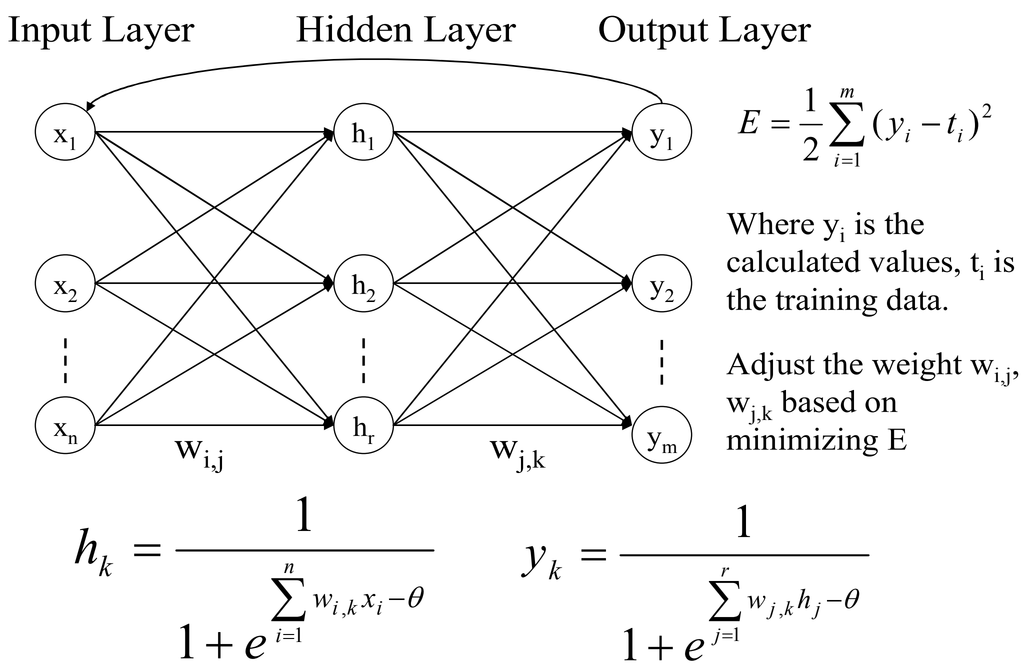


Figure 49: Back propagation neural network

Back propagation neural network can take supervised training, which supplies the neural network with inputs and the desired outputs, and the weights are modified to reduce the difference between the actual and desired outputs. The learning procedure is: (1). randomly assign weights (between 0-1); (2). present inputs from training data, propagate to outputs; (3). compute outputs Y, adjust weights according to the delta rule, back propagating the errors, and the weights will be nudged closer so that the network learns to give the desired output; (4). Repeat; stop when no errors, or enough epochs completed [16][17][18].

The implementation process is shown in Figure 50. The process starts with the pre-process of the experiment data. The experiment data is normalized as:

$$A_{nrom} = \frac{A - A_{mean}}{A_{std}} \tag{5.1}$$

Where A is V1, V2, V3, f_{Freq}, A is the test data, A_{mean} is the mean of the test data, A_{std} is the standard deviation of the test data.

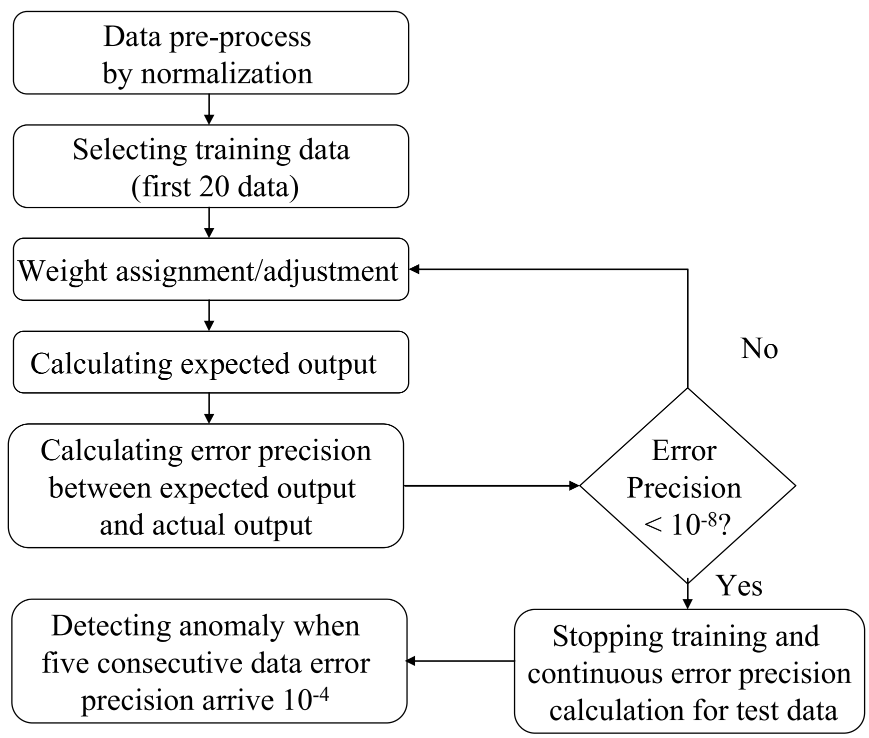


Figure 50: Neural network based anomaly detection

The first 20 data is considered as the health data and selected as the training data to train the neural network. The training purpose is to adjust the weights of the input parameters (IC frequency, V1, V2, and V3 in this case), which ensures that the expected output (Vout in this case) calculated by nerual network is close enough (error within pre-set precision) to the actual output under equipment health condition. The weights are adjusted by minimizing the error between the expected Vout and the actual Vout in this case. The preset error precision is 10^{-8} in this case, that is, the weight adjustment will stop when the error precision between the expected Vout and the actual Vout is equal or below 10^{-8}.

The error precision between the expected Vout and the actual Vout is selected to detect the anomaly. The expected Vout is calculated based on the relation between the input parameters (IC frequency, V1, V2, and V3) and the Vout, which was determined by the health data in the training phase. The error precision increase represents that the relation between the input parameters and the Vout determined in the training phase is not valid any more, that is, an anomaly occurs.

There are no clear and detailed criteria for the anomaly detection. In this case, an anomaly is considered to be detected when the five consecutive data error precision arrive or go beyond 10^{-4}, based on the error precision of 10^{-8} in the training phase.

5.3.3.5 Prognostics

When an anomaly is detected, the 30 minutes data before the anomaly detection point (including the anomaly detection point) are selected to predict the failure. This case uses the exponential model to predict the failure:

$$y = ae^{bx} + ce^{dx} \tag{5.2}$$

Where y is the error precision, and x is the time, and the least square curve fitting is used to determine the model parameter a, b, c, and d.

There are no clear and detailed criteria for failure. In this case, an failure is considered to be predicted when the five consecutive data error precision arrive or go beyond 10^{-3}, based on the error precision of 10^{-8} in the training phase. The process is shown in Figure 51.

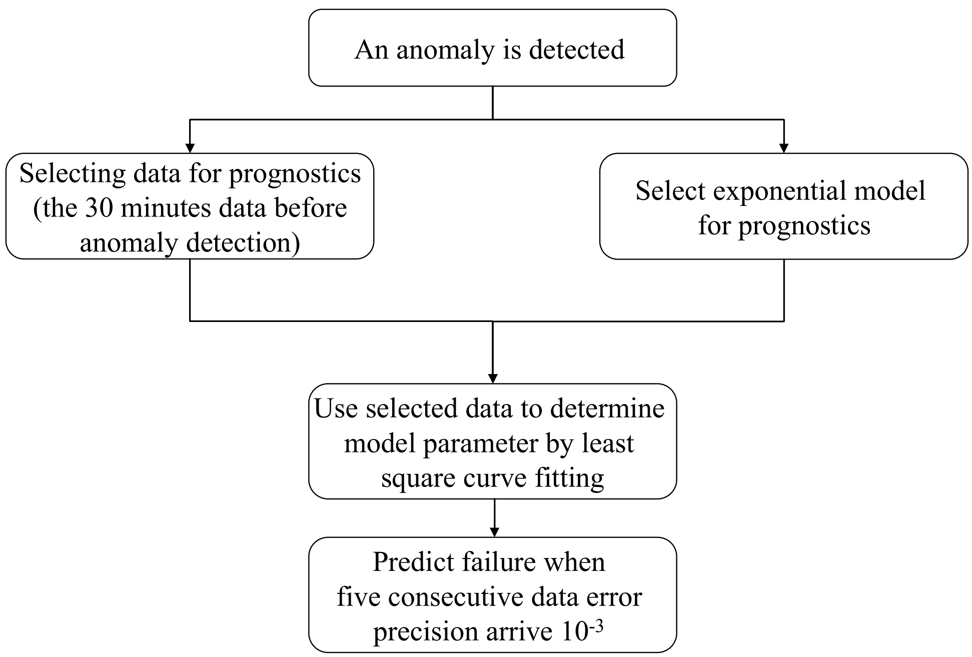


Figure 51: Prognostics by exponential model

5.3.3.6 Anomaly Detection and Prognostics Results

Since V2 and V3 are almost constant (no shift to indicate the power adapter degradation) in the experiment, we primarily consider use V1 and Frequency for anomaly detection and prediction, respectively. We also use all the parameters to detect the anomaly and predict the failures. The anomaly detections and failure prediction with only the IC frequency, v1, and all the parameters of the three power adapter samples are shown in Figure 52~60. In this case, the error precision of data for training (the first 20 data) is equal or below 10^{-8}. As mentioned in chapter 5.3.3.5 and 5.3.3.6, it is considered as anomaly when the error precision of 5 consecutive data is below 10^{-4}. When the anomaly is detected, the failure time is predicted, and it is considered as failure when the error precision of 5 consecutive data is below 10^{-3} according to the error precision in the training phase.

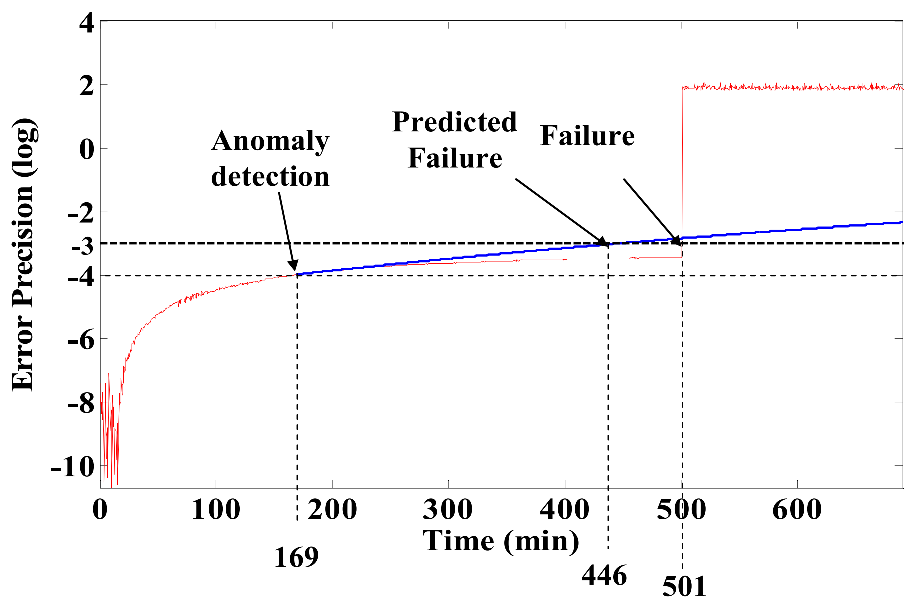


Figure 52: Anomaly detection and prediction with IC frequency: power adapter 1

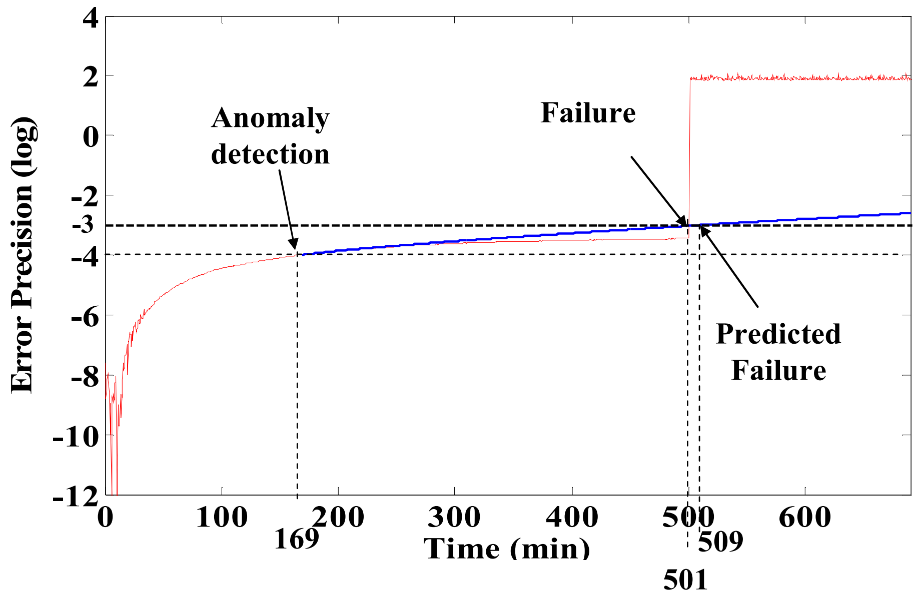


Figure 53: Anomaly detection and prediction with V1: power adapter 1

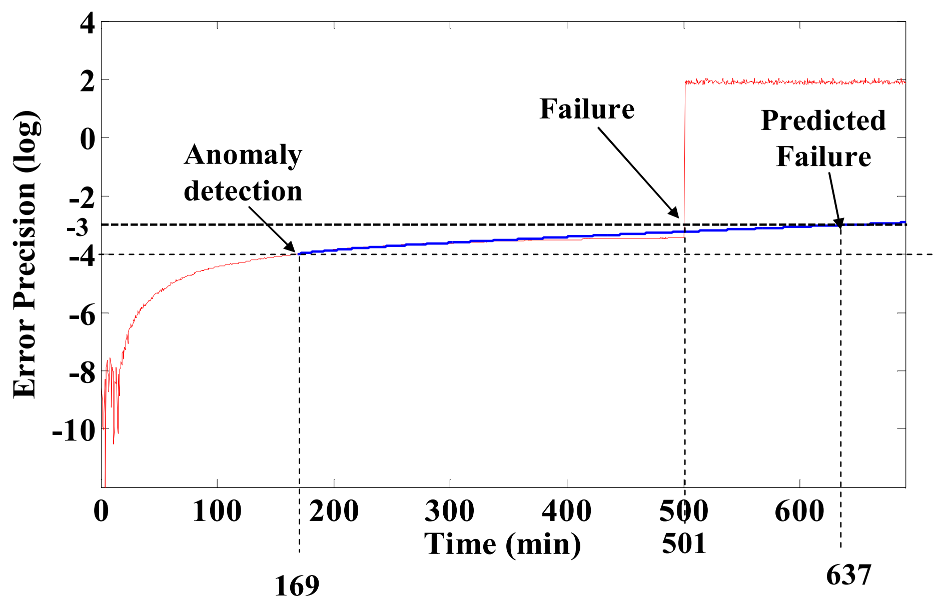


Figure 54: Anomaly detection and prediction with all the parameters: power adapter 1

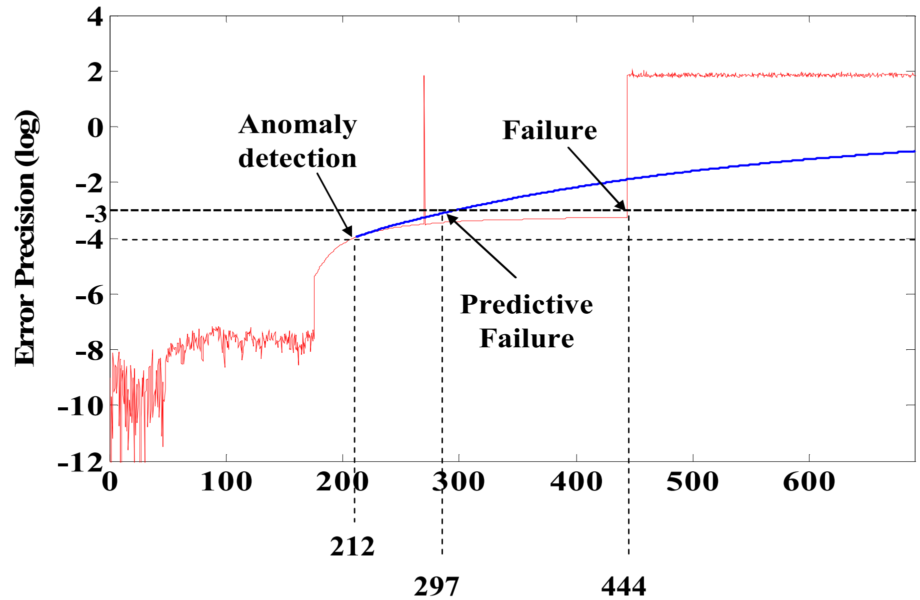


Figure 55: Anomaly detection and prediction with IC frequency: power adapter 2

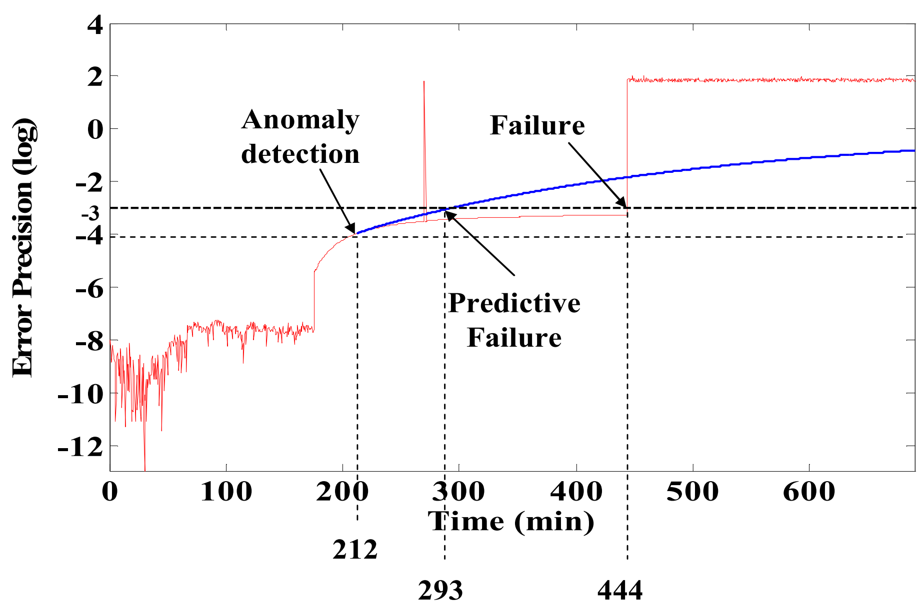


Figure 56: Anomaly detection and prediction with V1: power adapter 2

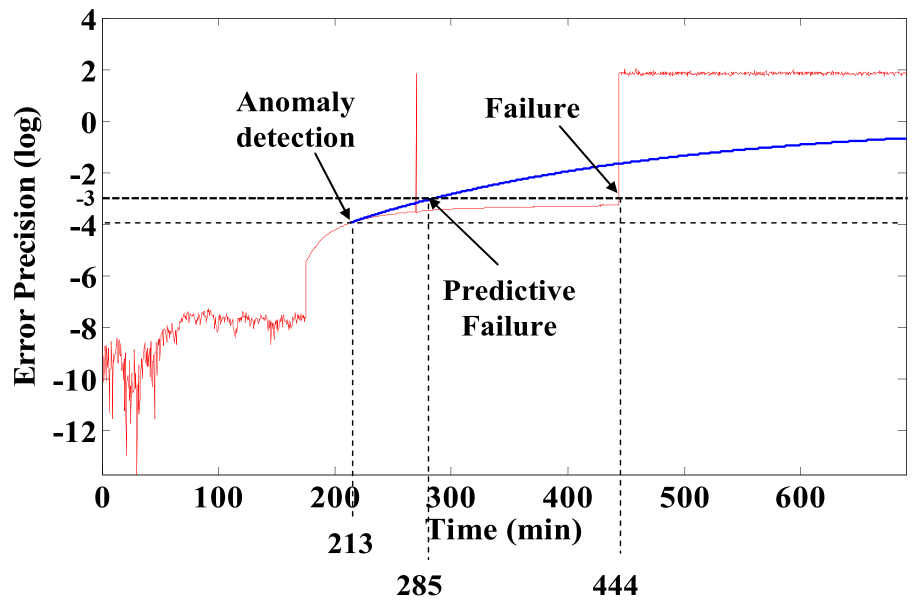


Figure 57: Anomaly detection and prediction with all the parameters: power adapter 2

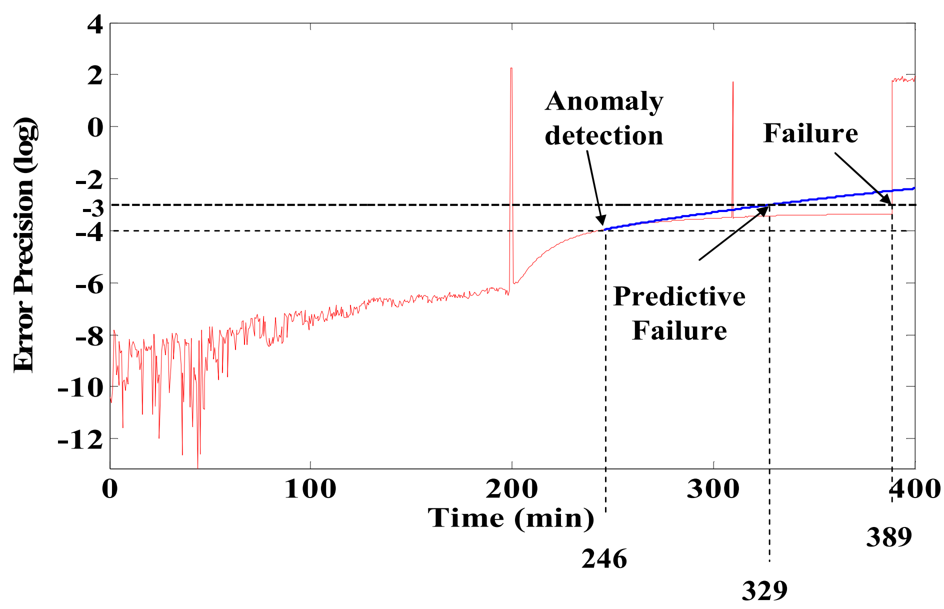


Figure 58: Anomaly detection and prediction with IC frequency: power adapter 3

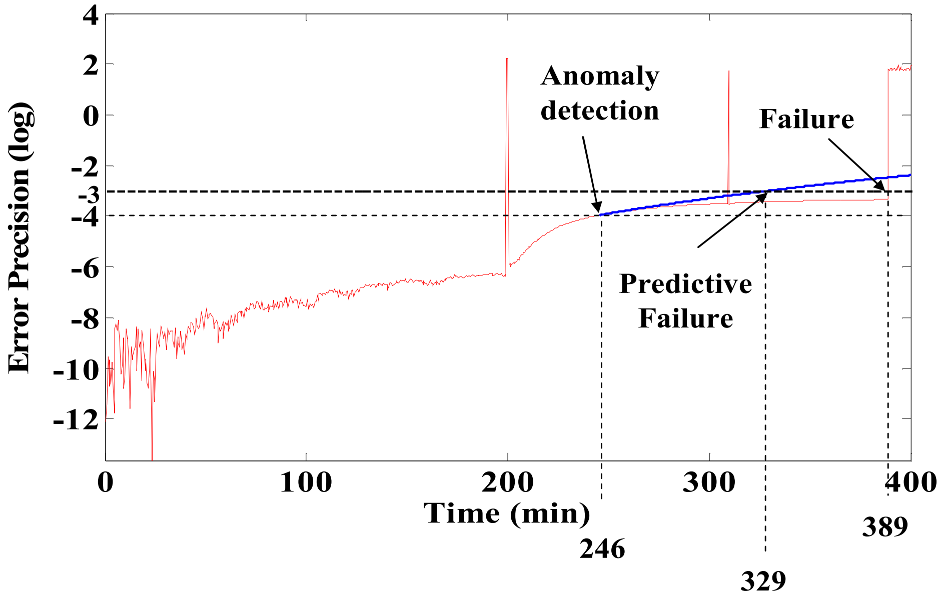


Figure 59: Anomaly detection and prediction with V1: power adapter 3

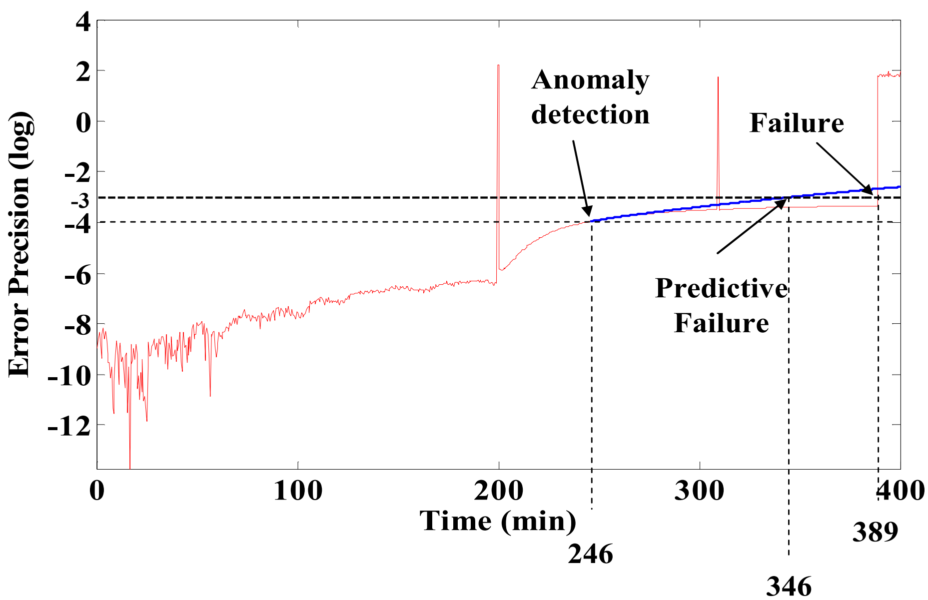


Figure 60: Anomaly detection and prediction with all the parameters: power adapter 3

The results of anomaly detection and failure prediction are summarized in Table 24. It is observed that the anomaly detection times with different parameters

(Frequency, V1, and all the parameters) are almost the same. The predicted failure times with different parameters are different, and no parameter has higher prediction accuracy than the others with all the three samples

Table 24: Anomaly detection and prediction: v1&freq. vs all the parameters

	Used Parameters	Failure time (min)	Anomaly detection time (min)	Predicted failure time (min)
Sample 1	Frequency	501	169	446
	V1		169	509
	All parameters		169	637
Sample 2	Frequency	444	212	297
	V1		212	293
	All parameters		213	285
Sample 3	Frequency	389	246	329
	V1		246	329
	All parameters		246	346

5.4 Conclusions

This dissertation develops a multi-stage risk mitigation approach for free air cooling. Specifically, prognostics and health management is implemented to identify and mitigate the risks to data center equipment arising from free air cooling at operation stage. The case study presented in this dissertation showed that the monitoring of equipment parameters can provide early failure warnings in the form of an alarm. When the remaining useful life is estimated using a suitable algorithm, data centers can schedule maintenance or replacement of equipment to avoid unplanned downtime. This method also helps identify whether the equipment failure is intermittent or permanent without having to interrupt data center service.

The PHM approach can be also used to other operating conditions (e.g., free air cooling conditions). When the power adapter fails, the output voltage (Vout) will has a shift from rated 9V, and some other parameters are expected to also have shifts corresponding to the Vout shift. The parameter shifts can be utilized for anomaly detection and prognostics based on the data-driven approach. For other operating conditions, the neural network needs be re-trained using the health data under the new conditions, since the health may be slightly different under different operating conditions, and the weight adjustment is very sensitive to the slight differences.

The PHM approaches can be also implemented on the system level, which starts with the identification of the weakest components. The following components may be the potentially weakest components which are the most likely to fail first in the system: (1).the components at risks with data center environment considerations (e.g., airflow designs) which may suffer from faster degradations than others under free air cooling conditions. Furthermore, the components on hot spots need get additional attentions; (2).the components with the local temperature beyond its manufactures' specified temperature under free air cooling conditions; and (3).the components with the highest hazard rates according to the history data under similar operating conditions. When the weakest components are identified, the previous approach can be used to detect anomaly and predict failures.

With the requirement of economizers in the 2010 version of ASHRAE standard 90.1, free air cooling will be increasingly implemented in existing and new data centers. The prognostics-based method can allow the implementation of free air cooling in data centers which were not originally designed for this cooling method. Furthermore, it also allows data centers to perform predictive maintenance (condition-based maintenance) instead of preventive maintenance (routine or time-based maintenance) by providing early failure warnings, which can reduce both the costs of equipment replacement and service down time. It is especially useful to mission-critical data centers.

When the next generation of data center equipment is designed for free air cooling, the primary concern should be the local operating conditions of the electrical parts, since this will be the deciding factor in data center reliability and availability. Toward that goal, PHM can help by gathering valuable life-cycle data. The knowledge of the life cycle and the operating parameters gained by using PHM in data centers will also help risk mitigation at telecommunications base stations that operate unattended at remote locations and where variations in the outside environment are even greater. Implementing PHM at such installations is critical for improving global data and voice communication networks and delivering the benefits of technological advances to the far reaches of the globe.

References:

[1] ASHRAE, "2008 ASHRAE Environmental Guidelines for Datacom Equipment," Atlanta, GA, 2008.

[2] IEC, "IEC Standard 60134 – Rating Systems for Electronic Tubes and Valves and Analogous Semiconductor Devices," 1961. (Last reviewed on July, 1994 by IEC Technical Committee 39 on Semiconductors).

[3] Kundert, K. S., "The Designer's Guide to SPICE and SPECTRE," Kluwer Academic Publishers, Boston, MA, 1998.

[4] JEDEC, "JEDEC91A – Method for Developing Acceleration Models for Electronic Part Failure Mechanisms," Arlington, VA, Aug. 2003.

[5] Bell Communications Research Inc., Generic Requirements GR-63-CORE, "Network Equipment-Building System (NEBS) Requirements: Physical Protection," Piscataway, NJ, Mar. 2006.

[6] International Electro-technical Commission, "IEC Standard 62240, Process Management for Avionics—Use of Semiconductor Devices outside Manufacturers' Specified Temperature Range," Switzerland, 2005.

[7] M. Pecht, "Prognostics and Health Management of Electronics," Wiley-Interscience, New York, NY, 2008.

[8] R. Jaai, and M. Pecht, "Fusion Prognostics," Proceedings of Sixth DSTO International Conference on Health & Usage Monitoring, Melbourne, Australia, Mar. 2009.

[9] P. Lall, M. Pecht, and M. J. Cushing, "A Physics-of-Failure (PoF) Approach to Addressing Device Reliability in Accelerated Testing," 5th European Symposium

on Reliability of Electron Devices, Failure Physics and Analysis, Glasgow, Scotland, Oct. 1994.

[10] J. Gu, and M. Pecht, “Prognostics-based Product Qualification,” 2009 IEEE Aerospace Conference, Big Sky, Montana, Mar. 2009.

[11] J. Gu, and M. Pecht, “Physics-of-failure-based Prognostics for Electronic Products”, Transactions of the Institute of Measurement and Control Vol. 31, No. 3/4, pages 309-322, 2009.

[12] J. Gu, and M. Pecht, “Prognostics Implementation of Electronics under Vibration Loading”, Microelectronics Reliability, Vol. 47, Issue 12, pages 1849-1856, Dec. 2007.

[13] D. Wang, Q. Miao, and R. Kang, “Robust Health Evaluation of Gearbox Subject to Tooth Failure with Wavelet Decomposition,” Journal of Sound and Vibration, Vol. 324, Issue.3-5, pages 1141-1157, Jul. 2009.

[14] W. Wang, M. Azarian, and M. Pecht, “Qualification for Product Development,” 2008 International Conference on Electronic Packaging Technology & High Density Packaging, Shanghai, China, Jul., 2008,

[15] S. Mathew, M. Alam, and M. Pecht, “Identification of Failure Mechanisms to Enhance Prognostic Outcomes,” MFPT: The Applied Systems Health Management Conference 2011, May 10-12, 2011, Virginia Beach, Virginia.

[16] Mirchandani, G., Cao, W., “On Hidden Nodes for Neural Nets”, IEEE Transactions on Circuit and Systems, Vol.36, No. 5, pp.661-664; 1989.

[17] J.Y. Audibert, and O. Catoni, “Robust Linear Least Squares Regression,” The Annals of Statistics, Vol. 39, No. 5, 2011, pages 2766–2794.

[18] A. Cottrell, "Regression Analysis: Basic Concepts," the course of Econometric Theory and Methods, Department of Economics, Wake Forest University, 2011.

Chapter 6. Contributions and Future Work

This dissertation develops an approach to determine the impacts of variable temperature and humidity ranges representing free air cooling on the performance of telecom equipment. This is the first research about the impact of free air cooling on the performance of telecom equipment. In this research, the frequencies and ranges of the throughput variations are focused as the main attributes of performance. Several metrics (1%, 2%, 5% off baseline) are used to determine the impacts of free air cooling. It is found that both the frequency and the ranges of the performance variations are significantly increased under variable temperature and humidity ranges representing free air cooling compared with those under traditional A/C condition. This method determines the impacts of corresponding variable temperature and humidity representative of that encountered in some FAC systems on the throughput of telecom equipment, thus assist data center manufacturers determine an appropriate operating environment when free air cooling is considered.

Another main contribution is the development of a multi-stage reliability risk mitigation approach for free air cooling. Specifically, a prognostics-based approach is developed to mitigate the risks at operation stage. This approach can mitigate the reliability risks by detecting anomaly and providing early warnings of failures. Data centers needn't be interrupted for the assessment purpose, and no additional useful life of telecom equipment are consumed by the assessment. It allows that free air cooling is used in data centers which were not originally designed for this cooling method. Furthermore, it allows data centers to perform predictive maintenance (condition-based maintenance) instead of preventive maintenance (routine or time-based maintenance).

Future work may include the investigation on additional humidity and temperature combinations (consistent with ASHRAE expanded envelops), and the identification of the effect of contamination on the reliability of telecom equipment in data center, when free air cooling is implemented. It also can include developing the physics-based degradation models of the weakest components to predict the remaining useful life of equipment.