Energy Efficiency in Data Centres and the Barriers to Further Improvements: An Interdisciplinary Investigation

Gemma Ann Brady

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy as part of the

Integrated PhD/MSc in Low Carbon Technologies

The University of Leeds

School of Chemical and Process Engineering

School of Mechanical Engineering

January 2016

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Section 3.5 to 3.7 and Chapter 6 are based on work published in:

Brady, G. A., Kapur, N., Summers, J. L., & Thompson, H. M., 2013. A Case Study and Critical Assessment in Calculating Power Usage Effectiveness for a Data Centre, *Energy Conversion and Management*, **76**, pp. 155-161, DOI:10.1016/j.enconman.2013.07.035

The candidate wrote the literature review and carried out the described case study, the candidate also developed the discussion and conclusions for the paper along with the appropriate presentation of the results. The candidate also responded to reviewer comments. The work was prepared under the guidance of all co-authors.

Acknowledgements

Sincere thanks and gratitude first go to my supervisors, Professor Nik Kapur, Dr Jon Summers and Professor Harvey Thompson, thank you for your supervision and guidance during this research. Thanks to Yaser Al-Anni and Adam Thompson in the Energy Building lab for their valuable help, especially when it came to understanding the inner workings of a server. Thanks also go to Professor Ian Bitterlin for reviewing and offering helpful comments on the work within Chapter 6.

Whilst they were not involved in this research, the following people all made great contributions towards my understanding of the data centre and the industry. First of all, Malcolm Howe and Matthew Winter for their help and guidance during a work placement during my first year of PhD study. Between them and their colleagues at Cundall, they introduced me to the industry and played a key role in me gaining interest in this sector. Next I would like to thank Michael Walker and his colleagues at Aimes, I am grateful for the time they spent with me and for helping me to understand the inner workings of the data centre. I would also like to say a thank you to all the willing participants within this study, I am grateful for the time that you gave up to help me with this research. To my new colleagues and friends at Arup in Sheffield and Newcastle for providing such a supportive environment over the past year. It has been inspiring to be surrounded by people with such passion and enthusiasm for engineering design, giving me the much needed energy to complete my academic studies.

The four years I spent in the Low Carbon Technologies DTC have been among my most memorable, the credit for this must go to my friends and colleagues within the DTC. I have thoroughly enjoyed working and laughing alongside you all, from the lunchtime debates, birthday cakes, culinary delights and Christmas dinners to adventures in China, the lake district and nights out in Leeds. To the first generation of DTC Drs, Pip, Shem, Hannah, Tom, Gilli and Sam and my own cohort Pip, Jayne, Zarashpe, Ramzi and David, I extend the warmest of thanks and gratitude for your company and friendship. I would like to wish you all the best in your future endeavours. To Jo Robinson for the sewing sessions and for introducing me to the delights of everything woolly, also to Brown Owl, Laughing Owl, Tawny Owl and the Brownies of 1st Meanwood, for providing a fun and creative outlet on a Monday night.

I would also like to extend a huge thank you to my family, particularly my mum and dad who have supported me through many years of university study. Finally, I would like to send special thanks to Lee who with his love and sense of humour has kept me sane throughout the process and helped me to keep my cool whilst under pressure.

Abstract

Creation, storage and sharing of data throughout the world is rapidly increasing alongside rising demands for access to the internet, communications and digital services, leading to increasing levels of energy consumption in data centres. Steps have already been taken towards lower energy consumption, however there is still some way to go. To gain a better understanding of what barriers there are to further energy saving, a cross-section of industry representatives were interviewed. Generally, it was found that efforts are being made to reduce energy consumption, albeit to varying degrees.

Those interviewed face various problems when attempting to improve their energy consumption including financial difficulties, lack of communication, tenant/landlord type relationships and physical restrictions. The findings show that the data centre industry would benefit from better access to information such as which technologies or management methods to invest in and how other facilities have reduced energy, along with a greater knowledge of the problem of energy consumption.

Metrics commonly used in the industry are not necessarily helping facilities to reach higher levels of energy efficiency, and are not suited to their purpose. A case study was conducted to critically assess the Power Utilisation Effectiveness (PUE) metric, the most commonly used metric, through using open source information.

The work highlights the fact that whilst the metric is valuable to the industry in terms of creating awareness and competition between companies regarding energy use, it does not give a complete representation of energy efficiency. Crucially the metric also does not consider the energy use of the server, which forms the functional component of the data centre. By taking a closer look at the fans within a server and by focussing on this hidden parameter within the PUE measurement, experimental work in this thesis has also considered one technological way in which a data centre may save energy. Barriers such as those found in the interviews may however restrict such potential energy saving interventions.

Overall, this thesis has provided evidence of barriers that may be preventing further energy savings in data centres and provided recommendations for improvement. The industry would benefit from a change in the way that metrics are employed to assess energy efficiency, and new tools to encourage better choices of which technologies and methodologies to employ. The PUE metric is useful to assess supporting infrastructure energy use during design and operation. However when assessing overall impacts of IT energy use, businesses need more indicators such as life cycle carbon emissions to be integrated into the overall energy assessment.

Table of Contents

Acknowledgements ...... **ii**

Abstract ...... **iii**

Table of Contents ...... **iv**

Figures ...... **ix**

Tables ...... **xiii**

Nomenclature ...... **xiv**

1 Introduction **1**

1.1 Data Centres and Carbon Emissions 1

1.2 Climate Change and the Problem of Increasing Emissions 3

1.3 How the ICT Sector Can Help to Reduce Carbon Emissions 4

1.4 Research Aims and Objectives 6

1.5 Thesis Outline 8

2 Background to Data Centres **10**

2.1 Introduction 10

2.2 Development of the Modern Data Centre 10

2.3 The Data Centre as a System 12

2.4 Worldwide Growth in Energy Consumption by Data Centres 14

2.4.1 Climate Change Policy and Data Centres 18

2.5 Maintaining the Data Centre Environment 20

2.5.1 IT Equipment Environmental Requirements 21

2.5.2 Heat Removal 26

2.5.3 Air Distribution 33

2.6 A Closer Look at Cooling the Server 35

2.6.1 Pressure Loss Across the Server 37

2.6.2 Air Distribution within the Server 38

2.7 Performance and Power Consumption of Server Fans 40

2.7.1 Measuring Fan Performance and Efficiency 43

2.7.2 Producing a Fan Curve 44

2.7.3 In Situ versus Manufacturers Stated Performance 46

2.7.4 Improving Fan Performance and Energy Efficiency 48

2.7.5 The Effect of Increased Inlet Temperatures 49

2.8 Power and Resiliency in the Data Centre 51

2.9 Conclusions and Research Themes 52

3 Energy Efficiency in Data Centres 54

3.1 Introduction 54

3.2 Defining Energy Efficiency 54

3.2.1 Energy Efficiency Potential 55

3.3 Exergetic Efficiency 56

3.4 Energy Efficiency in Data Centres 57

3.5 Measuring and Reporting Energy Efficiency 58

3.5.1 Energy Consumption Metrics 59

3.5.2 Thermal Management Metrics 63

3.5.3 Server and IT Efficiency Metrics 66

3.5.4 Exergy Analysis and Second Law Metrics 67

3.5.5 Carbon and Water Usage Effectiveness 69

3.6 Other Methods of Assessing Data Centres 70

3.6.1 Life Cycle Analysis and Embodied Carbon 71

3.7 Energy Consumption Case Studies 73

3.8 Increasing Energy Efficiency in the Data Centre 75

3.8.1 Technological Measures 76

3.8.2 Behavioural Measures 86

3.8.3 Policy-related Measures 87

3.9 Uptake of Energy Efficiency Measures 90

3.10 The Energy Efficiency Gap 92

3.11 Barriers to Energy Efficiency 94

3.11.1 Understanding Barriers Can Enable Change 94

3.11.2 Research Opportunities 94

3.11.3 Taxonomies and the Main Barriers to Energy Efficiency 95

3.11.4 Barriers Specific to Data Centres 98

3.12 Conclusions and Research Themes 101

3.12.1 Measuring and Presenting Energy Use and Efficiency 101

3.12.2 Barriers to Energy Efficiency in the Data Centre Industry 103

3.12.3 Summary 104

4 Qualitative Research Methodology 105

4.1 Introduction 105

4.2 Qualitative Research Aims 106

4.2.1 Ethical Considerations and Participant Recruitment 107

4.3 Research Methodology and the Application of Grounded Theory 108
4.3.1 Theoretical Sampling and Sample Saturation 109
4.3.2 Constant Comparative Analysis and Coding 110
4.3.3 Process Model 112
4.4 Development of the Interview Structure 114
4.4.1 Interview Aims 114
4.4.2 Avoiding Bias and Leading Questions 114
4.4.3 Interview Structure 115
4.4.4 Alterations during Data Collection 116
4.5 Transcribing and Analysing the Interview Data 118
4.6 Summary 119
5 Industry Views on Energy Efficiency in Data Centres 120
5.1 Demographics of the Participants 121
5.2 Qualitative Themes from the Interviews 125
5.3 Detailed Analysis of the Qualitative Data 126
5.3.1 Attitudes towards Energy Efficiency 127
5.3.2 Measuring and Monitoring Energy Use 133
5.3.3 Efforts Made Towards Reducing Energy Consumption 135
5.3.4 Incentives and Reasons for Reducing Energy Consumption 141
5.3.5 Difficulties and Barriers 143
5.3.6 Overcoming Difficulties 148
5.3.7 Use of Policies and Guidelines 150
5.3.8 Use of and Receptivity to Liquid Cooling Technology 153
5.3.9 Future of the Industry 156
5.4 Summary 159
5.4.1 Barriers to Saving Energy in the Data Centre 160
5.4.2 Summary of Other Findings from the Interviews 162
5.4.3 Research Themes 165
6 A Case Study and Critical Assessment of the Power Usage Effectiveness Metric 166
6.1 Introduction 166
6.2 Methodology 166
6.3 The Energy Consumption of the Servers 167
6.4 Energy Consumption of the Air Handling and Cooling Systems 169
6.4.1 Air Distribution – The Power Required to Operate the Fans 170
6.4.2 The Energy Consumption of the Evaporative Cooling System 171

6.5 Miscellaneous and Other Loads on the Electrical System 178
6.5.1 Lighting 178
6.5.2 Electrical Losses 179
6.6 Energy Analysis Results for the Prineville Data Centre 179
6.7 Sensitivity Analysis 180
6.8 Critical Assessment of the PUE Metric 184
6.8.1 Benefits 184
6.8.2 Limitations 184
6.9 Limitations of the case study 186
6.10 Summary and Discussion 186
7 Investigation of Server Air Flow and Fan Energy Use 189
7.1 Introduction 189
7.2 Experimental Methodology 190
7.3 System Description 191
7.4 Experimental Setup 193
7.4.1 Air Flow Test Bench 193
7.4.2 Server Fan Setup 194
7.4.3 Stress Linux 195
7.5 Server Characterisation 196
7.6 Fan Air Flow Rate Calculation 198
7.7 Results and Discussion 199
7.7.1 Server System Curve and Loss Coefficient 199
7.7.2 Fan Curves 204
7.7.3 Server Internal Fan Efficiency and Power Use 207
7.7.4 External Fan Efficiency and Power Use 208
7.7.5 Operating the Server without Using Internal Fans 210
7.8 Summary of Results and Energy Analysis 215
7.8.1 Effect on the PUE measurement 218
8 Conclusions and Suggestions for Further Research 220
8.1 Overview of Research 220
8.2 What barriers are there to further energy savings? 220
8.3 Is the PUE metric a useful tool for driving further energy savings? 222
8.4 Is there energy that can be saved in a standard server? 224
8.5 Recommendations and Future Research Endeavours 225
8.6 Final Summary 231

References **234**
Appendix **252**
Appendix A: Interview Script 252
Appendix B: Invitation Letter 257
Appendix C: Example Interview Transcript 260

Figures

Figure 1.1 CO_2-eq emissions in 2007 for subsectors (including data centres) within ICT sector (Malmodin et al., 2010) 2

Figure 1.2 Key technologies and mitigation techniques from the BLUE Map scenario (IEA, 2010) 4

Figure 1.3 Research themes 6

Figure 2.1 ENIAC (left) and a modern data centre (right) 11

Figure 2.2 A typical data centre layout (sourced from (Cho and Kim, 2011) based upon image from http://www.oriensoft.com/Infrastructure/) 12

Figure 2.3 Total electricity consumption of servers in the USA and the world for the years 2000 and 2005 [10] 17

Figure 2.4 Relationship between temperature and component failure rate for microprocessors, DRAM (Dynamic Random Access Memory) and PAL (Programmable Array Logic) (Remsburg, 2000) 22

Figure 2.5 Heat density trends in ICT equipment (Uptime Institute, 2000) 23

Figure 2.6 Environmental classes within the ASHRAE 2011 Thermal Guidelines for Data Processing Environments (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011) 25

Figure 2.7 Illustration of a CRAC based cooling system (Kennedy, 2009) 28

Figure 2.8 Illustration of a CRAH based cooling system (Kennedy, 2009) 28

Figure 2.9 Schematic showing the layout of a typical water cooled chiller system for a data centre (Beitelmal and Patel, 2006) 29

Figure 2.10 Number of free cooling hours available within Europe based on the ASHRAE envelope from 2009 30

Figure 2.11 Illustration of direct and indirect evaporative cooling (ESource) 31

Figure 2.12 Liquid submersion cooling from Green Revolution Cooling (Green Revolution Cooling) 33

Figure 2.13 Twelve alternate methods of delivering supply and removing return air in a data centre (Rasmussen, 2011) 35

Figure 2.14 An image showing the inside of a typical server, with the internal fans visible and the CPUs located underneath heat sinks (Ibrahim et al., 2012) 37

Figure 2.15 PIV measurement showing the flow through the system with the fan mounted at the inlet (Grimes and Davies, 2002) 39

Figure 2.16 Heat transfer coefficient for a fan speed of 9000 rpm and (a) a fan to plate distance of 15mm, (b) a fan to plate distance of 5mm. contour level: 5 $W/m^2 K$ (Stafford et al., 2010) 42

Figure 2.17 Fan peak total efficiency (Holahan and Elison, 2007) 43

Figure 2.18 Axial fan curve with a typical system resistance curve (Fukue et al., 2011b) 44

Figure 2.19 Flow through an orifice 44

Figure 2.20 Flow bench for measuring a fan or system curve for a Device Under Test (DUT) (Recktenwald, 2006) ... 46

Figure 2.21 Comparison between manufacturers, measured and predicted fan curves for an AHU fan (Liu and Liu, 2012) ... 47

Figure 2.22 Fan efficiency at different power levels for both baseline (fan performance without acoustic perturbations) and controlled cases (fan performance with acoustic perturbations), with a dimensionless fan flow rate on the x axis and the fan static efficiency on the y axis. (Greenblatt et al., 2012) ... 49

Figure 2.23 Power use percentage increase versus temperature for a conventional server and a blade server (Muroya et al., 2010) ... 50

Figure 2.24 Normalised power over a temperature range for a case model data centre (Muroya et al., 2010) ... 50

Figure 2.25 Polynomial curves for energy use over an increasing inlet temperature (Bean and Moss, [no date]) ... 51

Figure 3.1 Energy losses from generation to use (dos Santos et al., 2013) ... 54

Figure 3.2 Proportions of different materials that make up a physical server (Meza et al., 2010) ... 72

Figure 3.3 Carbon emissions from the IT equipment, facilities and the data centre building ... 73

Figure 3.4 Possible carbon footprint reductions for a typical data centre in the USA, with IT device management, efficiency best practice and low carbon energy source options (Masanet et al., 2013) ... 81

Figure 3.5 Power consumption of equipment within a typical data centre as a percentage of IT load (Auvil, 2013) ... 82

Figure 3.6 Power consumption of IT equipment for a given workload percentage (Islam et al., 2015) ... 84

Figure 3.7 Power consumption of servers under different resourcing algorithms (Islam et al., 2015) ... 84

Figure 3.8 Energy efficient technology adoption in data centres with either over 5,000 servers or over 1,000 servers, adapted from (Stansberry, 2013) ... 91

Figure 3.9 Percentage of data centres that report performance or cost metrics to high level management, adapted from (Stansberry, 2013) ... 91

Figure 3.10 The extended energy efficiency gap (Backlund et al., 2012) ... 93

Figure 4.1 Coding categories according to Hahn and how they apply to this research (Hahn, [no date]) ... 111

Figure 4.2 Coding paradigm for grounded theory as illustrated by Strauss (Strauss, 1987) ... 112

Figure 4.3 Research process incorporating grounded theory ... 113

Figure 4.4 Coding break down of questions, often with crossover of codes between questions ... 118

Figure 4.5 Image of NVivo coding showing ability to compare statements between interviews ... 119

Figure 5.1 Specific measures under the sub- theme of *'energy efficiency measures'* ... 137

Figure 5.2 Illustration of the discussion about the EU CoC ... 151

Figure 6.1 PUE measurement and analysis methodology using open source information ... 167

Figure 6.2 Drawing of an Open Compute server (Ogrey et al., 2013) ... 168

Figure 6.3 Diagram of air flow in the Prineville data centre (Park, 2011) ... 169

Figure 6.4 Average temperature and relative humidity for Redmond according to the TMY data set ... 172

Figure 6.5 Psychrometric chart with weather data for Redmond, Oregon and the cooling system modes A to G for the data centre as per the Open Compute information. Also shown is the ASHRAE recommended envelope ... 173

Figure 6.6 Number of hours per month and percentage of the year that each system mode (A-G) is in operation for, based upon TMY weather data ... 174

Figure 6.7 Psychrometric chart illustrating how each data point was treated during the analysis ... 175

Figure 6.8 Water consumption by the cooling system over a full year ... 176

Figure 6.9 Percentage changes in PUE as a result of altering different parameters within the metric: (i) humidification by ±20%, (ii) η_E by ±1%, (iii) IT load by -40% and +20% ... 182

Figure 6.10 Changes in PUE with increasing server load ... 183

Figure 7.1 Visual representation of the methodology used for the experimental work within this chapter, showing the steps taken between the main three stages of the research ... 190

Figure 7.2 Isometric diagram of the Sun Fire V20z server (Sun Microsystems, 2008) ... 191

Figure 7.3 Air flow bench, pictured with manometers for measuring differential and static air pressure (note, the right hand air flow settling screen was not in place for this photo) ... 194

Figure 7.4 Pololu Simple Motor Controller 18v15, with a maximum voltage of 30 V and maximum continuous current of 15 A (Pololu Corporation, 2001) ... 195

Figure 7.5 CPU temperature over time with server under stress ... 196

Figure 7.6 Layout of probe locations on the server casing relative to the internal fans, not to scale. ... 197

Figure 7.7 Pressure difference (Pa) between a certain point and the server exit for four cases; (a) fan at 10,463 RPM, (b) fan at 14,033 RPM, (c) fans unplugged, and (d) fans removed. Cases (c) and (d) were operated by drawing air through the server using an external fan. ... 198

Figure 7.8 System curves for SunFire v20z server, repeated tests with internal fans (a) present and (b) not present ... 200

Figure 7.9 System curves with and without internal fans present for the SunFire V20z Server 201

Figure 7.10 System curve from measured data (fans present in server) and calculated values from system loss coefficient....203

Figure 7.11 System curve from measured data (fans not present in server) and calculated values from system loss coefficient....203

Figure 7.12 Fan curves comparison between orifice plate sizes, both using an end plate with 15 mm diameter perforations 206

Figure 7.13 Comparison between measured and manufacturers fan curves....207

Figure 7.14 Fan curve for a single fan and six fans in parallel with the server system curve. The server system curve is measured here with the fans present and in operation. 208

Figure 7.15 Fan curve and efficiency for the large external fan, with system curves for 1, 2, 4 and 6 servers. The system curve is plotted for when the internal server fans are not present....209

Figure 7.16 Fan curve and efficiency for the large external fan, operating at full rated and 36 W power 209

Figure 7.17 (a) Server operating with internal fans present and in operation, (b) server operating with internal fans present but disconnected, and (c) server operating with internal fans disconnected and not present....211

Figure 7.18 Comparison between CPU temperatures for different tests 213

Figure 7.19 Total power use under different tests 213

Figure 7.20 Change in temperature across the server under different tests....214

Figure 7.21 Variations in internal server temperatures across each stress test for either server A and B when in situ together, or one server in isolation 215

Figure 8.1 Feedback loop from the findings of chapters 6 and 7 into further industry surveys....226

Tables

Table 3.1 Example data centre power use and metric results (Patterson, 2012) 62

Table 3.2 Comparison of the energy savings possible through using different cooling systems* 80

Table 5.1 Participant role and company description 122

Table 5.2 Details of data centres owned by Colo and OO participant companies 124

Table 5.3 Main themes from the interview data............126

Table 5.4 Sub-themes under the main theme of *'priority of energy'* 128

Table 5.5 Sub-themes under the main theme of *'customer interest'* 132

Table 5.6 Sub-themes under the main theme of *'monitoring and metrics'* 134

Table 5.7 Sub-themes under the main theme of *'energy efficiency measures'*............136

Table 5.8 Sub-themes under the main theme of *'incentives'*............142

Table 5.9 Sub-themes under the main theme of *'difficulties and overcoming problems'* 144

Table 5.10 Sub-themes under the main theme of *'policy'*............150

Table 5.11 Sub-themes under the main theme of *'future of the industry'* 157

Table 6.1 Description of each mode of operation for the Prineville data centre's cooling and air distribution system 170

Table 6.2 Maximum water flow and total water volume required for each mode of operation where evaporative cooling is used............176

Table 6.3 Power use of a Dell Inc. PowerEdge R820 (Intel Xeon E5-4650L, 2.60 GHz) server (SPEC, 2012). 183

Table 7.1 Architecture of the Sun Fire V20z server, information from the manufacturers datasheet (Sun Microsystems) 192

Table 7.2 System loss coefficient data 202

Table 7.3 Results from 1 mm perforated gate and fan power at 25% 205

Table 7.4 Results from 1 mm perforated gate and fan power at 100%............205

Table 7.5 Theoretical energy savings if one external fan were used to provide air distribution over two servers, with fans either still present or removed 217

Table 7.6 PUE measurements based upon data gathered in Chapter 6 and energy data from experimental work in Chapter 7 218

Nomenclature

Abbreviations

AHU	Air Handling Unit
ASHRAE	American Society for Heating, Refrigeration and Air Conditioning Engineers
BREEAM	Built Environment Environmental Assessment Method
CCA	Climate Change Agreement
CCL	Climate Change Levy
CEEDA	Certified Energy Efficient Data Centre Award
CEF	Carbon Emissions Factor
CFD	Computational Fluid Dynamics
CI	Capture Index
COP	Coefficient of Performance
CPU	Central Processing Unit
CRAC	Computer Room Air Conditioner
CRAH	Computer Room Air Handler
CRC	Climate Reduction Commitment
CUE	Carbon Usage Effectiveness
DCIE	Data Centre Infrastructure Efficiency
DCIM	Data Centre Infrastructure Management
DUT	Device Under Test
EETD	Environmental Energy Technologies Division
EPA	Environmental Protection Agency
ERE	Energy Reuse Effectiveness
EU ETS	European Union Emissions Trading Scheme
EU CoC	EU Code of Conduct on Data Centre Energy Efficiency
FVER	Fixed to Variable Energy Ratio
GHG	Greenhouse Gas
HPC	High Performance Computing
HVAC	Heating Ventilation and Air Conditioning
ICT	Information and Communications Technology
IEA	International Energy Agency
IP	Internet Protocol
IT	Information Technology
ITUE	IT Usage Effectiveness

LBNL	Lawrence Berkeley National Laboratory
LCA	Life Cycle Analysis
LEED	Leadership in Energy and Environmental Design
OECD	Organization for Economic Co-operation and Development
PDU	Power Distribution Unit
PIV	Particle Image Velocimetry
PoE	Power over Ethernet
PUE	Power Usage Effectiveness
PWM	Pulse Width Modulation
QoS	Quality of Service
RCI	Rack Cooling Index
RHI	Return Heat Index
RTI	Return Temperature Index
SERT	Server Efficiency Rating Tool
SHI	Supply Heat Index
SLA	Service Level Agreement
SPEC	Standard Performance Evaluation Corporation
TMY	Typical Meteorological Year
TUE	Total Usage Effectiveness
UNFCCC	United Nations Convention on Climate Change
UPS	Uninterruptible Power Supply
WUE	Water Usage Effectiveness

Symbols

A	Area	m^2
c	Rotor chord length	m
C_p	Specific heat capacity	J/kg.K
D	Diameter	m
d_t	Orifice diameter	m
g	Acceleration due to gravity	m/s^2
h	Head	m
I	Current	A
K, C	Loss coefficients	-
m	Mass flow rate	kg/s

P	Pressure	Pa
P_{dm}	Mean power	W
P_s	Saturation vapour pressure	Pa
Q	Heat energy	W
R	Rotor radius	m
Re	Reynolds number	-
T	Temperature	ºC
t	Time	s
V	Voltage	V
$\dot{V}$	Volumetric flow rate	m^3/s
W	Power	W
z	Elevation	m

Greek Symbols

ρ	Density	kg/m^3
ε	Expansion factor	-
η	Efficiency	-
θ	Moisture content	kg_v/kg_{air}
v	Velocity	m/s
φ	Relative humidity	-
ω	Rotational speed	rpm

Subscripts

air	Air	*man*	Pressure measured by manometer
amb	Ambient	*o*	Orifice
c	Cooling	*out*	Out
d	Distribution	*pd*	Power Distribution
db	Dry bulb	*T*	Total
dp	Dew point	*t*	Throat
f	Fan	*v*	Vapour
in	In	*wb*	Wet bulb
IT	IT equipment		
L	Loss		

1 Introduction

The past few decades has seen digital technology advance and grow to envelop most of today's society becoming a vital part of businesses, education, health care, government, communication networks, and many other services. Digital technology has enabled faster and more efficient business transactions, improved communications between people across the world, and increased access to knowledge and information. However this expansion in the digital world comes at a price. The electricity consumption of the Information and Communications Technology (ICT) sector has increased dramatically over the past two decades as the number of computers across the world has multiplied. With a rapidly rising amount of data being created in the world (90% of the world's data was created in only two years (IBM, [no date])), this is causing a rapid growth in energy consumption. 27% of electricity consumption in the ICT sector comes from data centres, hence it is important to consider these facilities in the effort to reduce worldwide energy use.

One of the key problems facing the world today is increasing levels of greenhouse gas emissions (GHG) and the concentration of these in the atmosphere, both of which have been linked to climate change (IPCC, 2007c). In order to address this problem, targets have been set to ensure that concentrations of GHG emissions in the atmosphere do not reach a critical level. The UK for example has a legally binding target of reducing its GHG emissions to 80% of 1990 levels by 2050 (Department of Energy and Climate Change, 2008). A key pathway to achieving lower emissions is to reduce the consumption of electricity, a subset of this pathway is to increase energy efficiency. All sectors (industry, transport, manufacturing etc.) need to be considered in terms of the amount of electricity that is consumed and the efficiency with which it is used. The ICT sector is no exception. Data centres are the second largest consumer of energy in the ICT sector (see Figure 1.1), and a reduction in their electricity consumption will help with the task of reducing GHGs.

1.1 Data Centres and Carbon Emissions

Computers form the backbone of our everyday tasks, from buying goods, browsing the internet and using social media platforms at home, to performing financial transactions at the bank, sending emails and storing information at work. They also facilitate communication between friends, families and colleagues throughout the world via technology such as emails and social networking. However, the sharing and storage of information relies on servers located in data centres and these consume a vast amount of energy in the form of electricity. Also, it is not just the ICT equipment that consumes electricity, but the multitude of support infrastructure components housed in the data centre such as power distribution equipment and cooling plant. The problem with consuming vast amounts of energy is that the generation of electricity produces GHGs, mainly in the form of carbon emissions (unless of course the electricity comes from renewable sources).

In the past few years there have been several estimates of the quantity of carbon dioxide (CO_2) emissions that are associated with the ICT sector and data centres, with variable credibility and depth of analysis. Each separate study has been conducted with varying boundary conditions and inclusions and so it is difficult to compare them directly. One widely quoted estimate from Gartner is that the ICT industry emits around 2% of worldwide CO_2 emissions (860 $MtCO_2$ in 2007). Of this, data centres make up 23%, or 0.5% of the total worldwide (Gartner, 2007). Malmodin et al provide another estimation of the ICT sector's contribution towards global GHG emissions, with their study showing that in 2007 the sector accounted for 1.3% of global GHG emissions (620 $MtCO_2$-eq). In this study it was found that the operation of data centres emits 170 $MtCO_2$-eq which is 0.5% of the world total (Malmodin et al., 2010) (see Figure 1.1). Both of these estimations are of the same order of magnitude however there were slight differences in the assumptions made during each analysis, such as the definition of ICT equipment and how much of the life cycle carbon emissions are included. There are also differences depending on whether the study uses carbon or carbon-*equivalent* emissions, the latter of which includes the quantity of all GHGs not just CO_2.

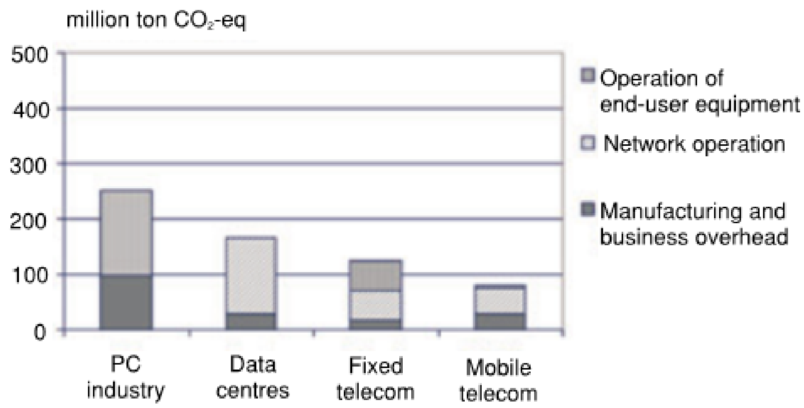


Figure 1.1 CO_2-eq emissions in 2007 for subsectors (including data centres) within ICT sector (Malmodin et al., 2010)

In terms of future growth in carbon emissions associated with the ICT sector, an article from McKinsey shows that by 2020 worldwide emissions could rise to 3% or 1.54 $GtCO_2e$

which is actually double the total amount of emissions produced by the UK in 2008 (Boccaletti et al., 2008). As the need for data centres is still increasing following the rising demand for ICT services (and is likely to continue to increase), energy efficiency is becoming a high priority if the level of associated carbon emissions is to be kept down. Being able to measure and quantify the energy consumption and efficiency of data centres is an important tool in aiding reductions in energy use.

1.2 Climate Change and the Problem of Increasing Emissions

One of the key issues in today's society is global warming and the effort to reduce carbon emissions and increase energy efficiency. Over the past two centuries, industrialization of developed countries has led to significant environmental challenges. Demand for energy is constantly on the increase across the world, as developed countries consume more energy and developing countries become more industrialized. It is predicted that there could be an increase of over two- thirds[1] in electricity generation demand by 2035 (International Energy Agency, 2013). The majority of this increase will come from Non-OECD[2] countries.

Increasing levels of demand for electricity is a problem as the majority of electricity today is generated using fossil fuels which emit greenhouse gases (GHG) as part of the generation process. Carbon dioxide and other GHGs within the atmosphere have increased steadily over the last century alongside the increase in worldwide demand for electricity (although the increasing levels are not solely due to electricity demand). Between 1970 and 2004, carbon emissions[3] rose by 80% (Pachauri, 2009). Over the 200 years prior to 1970 the concentration of carbon dioxide in the atmosphere increased by 50 ppm. Unfortunately, it only took another 30 years for that same increase to be experienced after 1970 (IPCC, 2007a). These increasing levels of carbon dioxide concentrations (along with other GHGs) are contributing towards a change in the global climate. If nothing is done to curb these increasing levels of emissions, then it is possible that by the end of the 21st century, the global average temperature will increase between 1.1°C and 6.4°C (Pachauri, 2009). Aside from increasing global temperatures, there is

---

[1] Figure based upon the New Policies Scenario

[2] Countries that are not part of the Organisation of Economic Co-operation and Development, for example China, India and the Middle East.

[3] For the purposes of this thesis, '*carbon emissions*' refers to emissions of all greenhouse gases.

also likely to be increasing occurrences of extreme weather events, sea level rises and even food shortages. In order to lower these emissions, measures need to be taken to reduce energy consumption along with improvements in the generation of energy.

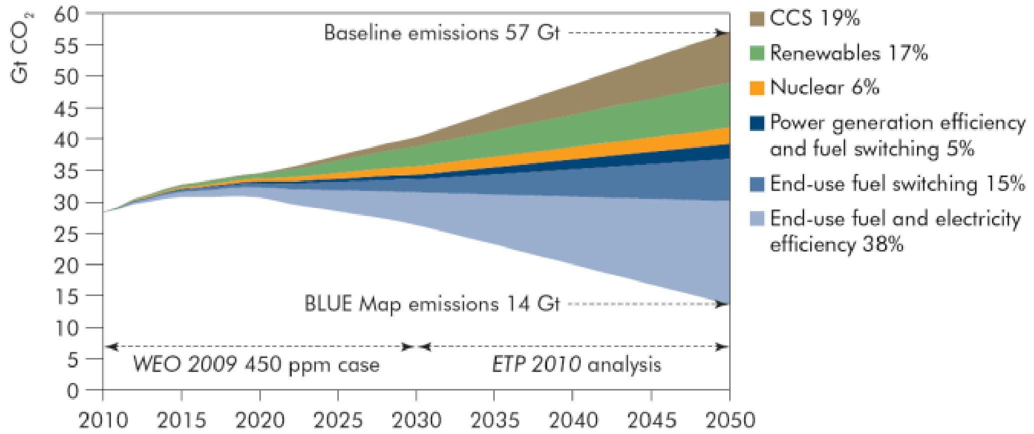


Figure 1.2 Key technologies and mitigation techniques from the BLUE Map scenario (IEA, 2010)

Figure 1.2 was produced by the International Energy Agency (IEA) and shows the key technologies and mitigation strategies that need to be used in order to maintain a safe level of carbon emissions by 2050. Increasing energy and fuel efficiency in all areas and sectors is a key part of meeting the targets set by world governments. Energy and fuel use efficiency is highly relevant to data centres along with efforts to use renewable energy sources which may become more accessible in the future.

This is the main driver behind the work in this thesis. If targets such as the one set by the IEA to increase energy efficiency in all sectors are to be met, then data centres and the ICT sector should be a key consideration.

1.3 How the ICT Sector Can Help to Reduce Carbon Emissions

When discussing the amount of energy consumption by data centres and the ICT sector as a whole, it is easy to forget the ability of this sector to reduce the consumption and emissions of others. This is an important consideration that should not be forgotten, data centres can actually *enable reduced emissions and energy consumption* elsewhere. The UK Government recognised this by publishing a POST (Parliamentary Office of Science and Technology) Note entitled 'ICT and CO_2 Emissions'. Whilst the document discusses the emissions from the ICT sector, it also recognises the contributions towards reducing emissions. These contributions include;

- smart technologies applied to city planning, appliances, buildings, industry and the electrical grid;
- dematerialisation of services (using ICT to replace physical interactions and products); and
- intelligent transportation systems, amongst many others (Parliamentary Office of Science and Technology, 2008).

Another report on this subject explores the role that technology (defined as ICT and consumer electronics) has in tackling climate change (Intellect, 2008). Conference calls, online shopping, internet communication and information access are all services made available through the use of ICT, and they all help to reduce carbon emissions. One particular good example is in the area of logistics. One postal company based in the USA, United Parcel Service Inc., have used ICT to reduce their fleet's fuel consumption and in turn reduce carbon emissions. By utilising routing technology, they are able to avoid taking left turns and hence save 9.8×10^7 minutes of idling time in one year[4] (Wagner, 2012). Intellect also note that the technology revolution has led to de-materialisation, whereby a product or process has minimal physical properties (Intellect, 2008). One example (of many) of this is emails being used instead of printed letters.

Opportunities for ICT to provide virtual equivalents of physical services and goods in order to increase energy efficiency was examined by the Climate Group in a 2008 report (The Climate Group, 2008). They point out that whilst making services more virtual is an important part played by ICT in reducing the individual's carbon footprint, the enabling of energy efficiency is more important (The Climate Group, 2008). Smart motor systems, smart logistics, smart buildings and smart grids are all cited as technologies enabling energy efficiency that are made possible by ICT (The Climate Group, 2008).

So whilst it is important to address the amount of carbon emissions accountable to the ICT sector and data centres, it is also important to recognise the benefits that they provide. Without this energy intensive sector, the emissions from other sectors would be higher. It is predicted that in 2020, ICT could be providing 7.9 $GtCO_2e$ of savings which would be equivalent to 15% of worldwide emissions (The Climate Group, 2008). Therefore, besides the benefits of ICT services, the sector also *delivers emissions reductions* by supporting other sectors to increase their energy efficiency. Having said this, despite all the energy and carbon benefits that ICT has enabled, the effort is fighting

[4] In the USA vehicles are permitted to turn right at certain junctions even if the traffic lights are red. This means that if left hand turns can be reduced, fuel can be saved by reducing vehicle idling time.

against rapid increasing levels of ICT usage which is leading to further energy consumption. Therefore it is important that whilst recognising the benefits, there is also a strong focus on reducing energy use in the data centre industry.

1.4 Research Aims and Objectives

Following on from the discussion above, it is clear that the problem of climate change and increasing energy use should be a driver for data centres to focus on their energy consumption and efficiency. As the literature review in chapters 2 and 3 will introduce, there are plenty of options available to the industry for reducing energy consumption. However as new facilities are approaching ever higher levels of efficiency (according to the most commonly used metric), older facilities are falling behind.

The industry faces the task of driving energy consumption down further whilst also expanding to cope with an increasing demand for services. This research asks what problems the industry may be facing along this pathway to lower energy consumption and whether the available tools are suitable for achieving this. The main aim of this thesis is to therefore;

> *Find out what barriers there are to the implementation of energy saving technologies and methodologies within data centres, and find opportunities to overcome them.*

This is addressed both from the perspective of investigating industry opinions through a series of semi-structured interviews, but also through technological assessment in two areas. Firstly through examining the most commonly used energy metric and secondly through looking at a new technological way to reduce energy loads, as illustrated in Figure 1.3.

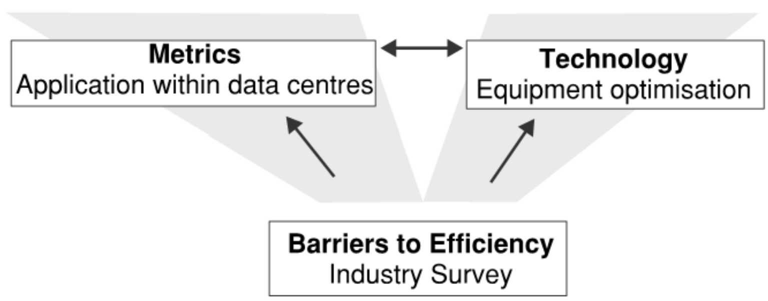


Figure 1.3 Research themes

Following a review of the existing literature, this research aim is addressed through setting the following objectives;

a) *Investigate industry opinions of, attitudes towards and efforts made relating to energy use in the data centre. Find out if there any existing barriers that may be preventing increased savings.*

Through informal discussion with industry members as part of early work for this research, it became clear that the data centre industry does not necessarily have the tools to adopt necessary energy saving measures. Therefore, through a series of recorded interviews, the opinions of various professionals within the data centre industry have been collated. With a focus on the measures that they may have taken to address energy use in the data centre, an understanding of industry efforts to reduce the use of energy will be gained. This objective examines perceived barriers to further savings in order to gauge what problems exist in the industry.

b) *Investigate the metrics and methodologies used to assess energy consumption and efficiency within data centres.*

Through the literature review, it can be noted that current metrics used within the industry are not necessarily fit for purpose (this is also reiterated by responses to the interviews as part of the first objective). A lack of suitable metrics is a barrier to the additional energy savings that could be made if energy use was be represented more effectively. Therefore metrics and methodologies that are currently used to assess energy use within data centres need to be reviewed in terms of their suitability and usefulness to the industry. One metric that has become an unofficial standard within the industry is also recognised as not being entirely suited to its purpose. The second objective addresses this issue by taking a closer look at the metric in question and discusses the issues. Part of this section of work involves a sensitivity analysis on this particular metric looking at which measured parameters have the most effect on the outcome.

c) *Conduct an investigation into the energy use of server fans as a technological means of improving energy efficiency*

From discussion in the literature review and interview participant's responses, it is clear that the IT (Information Technology) equipment within data centres has received less attention when it comes to energy use than the infrastructure has. As part of the IT equipment, internal server fans support the critical components by distributing air through the unit, but they consume energy in doing so.

The server fans respond, through the use of an automated internal control strategy, to the physical environment provided by the data centre. Maintaining the correct operating conditions, they increase or decrease in speed in response to changes in the temperature and pressure of the external environment. Although a component of the server, and therefore considered in the denominator of the PUE, these fans act as part of the mechanical infrastructure of the data centre and are hence part of the energy use overhead. The energy consumption of these fans is not currently represented in metrics such as the PUE. Server fans also offer an area where energy use can be reduced in the data centre.

The final aim of this research will be to conduct an energy analysis on the internal server fan and see if energy can be saved by using an external fan instead. Finding out whether it would be possible to switch to an external fan whilst still be delivering air to the correct areas within the server without it overheating is also a key aim. There has been a considerable amount of focus on the mechanical infrastructure in terms of energy use, rather than the IT equipment. Something that is not helped by the most commonly used metric which does not consider efficiency within the servers. The effect of this kind of energy efficiency measure will also be fed into the aforementioned metric to see how this kind of reduction in energy use affects the metrics output.

1.5 Thesis Outline

The objectives of this thesis are addressed through using interdisciplinary research techniques incorporating a blend of quantitative and qualitative investigations along with experimental analysis.

Chapters 2 and 3 present an in depth review of the existing literature, providing a background for the research. The qualitative methodology employed for initial investigative work in this thesis is described in Chapter 4, with a discussion of the results being detailed in Chapter 5. Considering the barriers to energy efficiency that exist, Chapter 5 examines the difficulties that industry representatives face. The work in this chapter studies the responses of interview participants regarding their views on the priority of energy saving, what metrics and measurement methodologies they use, which improvements they may have made and where they think the industry is headed in the future. An emerging theme of the interviews shows evidence of one choice the industry is making that could be improved, the use of metrics for assessing energy efficiency. The metric most commonly used within the industry is critically examined in Chapter 6, with the aim of further examining the differences of opinion raised in the interviews.

Another emerging theme from the interviews is that there is a lack of focus on the energy consumption of the servers, therefore Chapter 7 investigates a new technological approach through which it may be possible to reduce the energy use of this equipment. Finally, Chapter 8 discusses overall conclusions of the research and implications of the findings. Recommendations and suggestions for further research are also offered. An example interview transcript and participant interview invitation letter follows within the appendix.

2 Background to Data Centres

2.1 Introduction

Digital files and knowledge on the internet, information deposits for government, businesses and research, financial processes, telecommunications and many other IT based activities all rely on a collection of servers and their support equipment situated in data centres. In the digital society that we now live in, the world relies heavily upon the information and data which is processed through these facilities. However the expansion of communications and knowledge sharing has come at the cost of increased use of electricity within the ICT sector. Although as explained in 1.3, some of this increased use of electricity by servers and data centres has shifted from other sectors as dematerialisation due to the use of digital services and has helped to reduce energy use elsewhere.

This chapter provides an introduction to data centres and a background to the research in this thesis. The concept of a data centre is presented first and the different technologies for providing a stable temperature controlled environment to these facilities are outlined. Worldwide growth in the electricity consumption of data centres will then be discussed and the relevance of taking it into account when thinking about reducing carbon emissions. A closer look at the air flow through computer servers and the fans that enable the flow of air through them then follows. A brief reflection on the effect that climate change and emissions reduction policies have on the data centre industry, along with the regulations and policies that exist to enable reduced electricity consumption by data centres completes this chapter.

First of all, a broad context will be given to the modern data centre followed by the issues surrounding increasing carbon emissions and links to global climate change.

2.2 Development of the Modern Data Centre

When computers systems were first developed and became more powerful, they were very large pieces of equipment and required storage in dedicated rooms fit for the purpose. The equipment in these rooms needed careful organization and generated a heat load that needed rejecting to the external environment. Large computers such as the ENIAC machine required a considerable amount of cooling in order to keep them operational (Figure 2.1). ENIAC had 70,000 resistors, producing heat along with the vacuum tubes and other equipment, heat which needed to be removed in order to lengthen resistor life (Eckert, 1946). As computers became smaller and denser through the 1980s, demand for these machines grew and sectors within the economy became more dependent upon the ability to process and store data.

In the 1990s came the age of the client server where servers where positioned in a dedicated room in offices which soon became known as the data centre. With the rapid expansion of the internet, companies no longer had enough space to keep their servers within their offices, and dedicated facilities were constructed. There are now numerous examples of these dedicated facilities which house computer servers across the world, providing the ICT backbone to companies, businesses and services such as the internet.

Figure 2.1 ENIAC (left) and a modern data centre (right)

Today, data centres are large facilities which provide a suitable working environment for hundreds or even thousands of servers. Hardware within these facilities includes servers, data storage equipment, communications, routers and switches. All of this is supported by cooling systems, uninterruptible power supply (UPS) units, power distribution units (PDUs) and backup generators, along with lighting, fire and security systems.

Redundancy, security and resiliency are all key issues in the modern data centre. In order to increase resiliency and reduce the risk of failures, redundancy is provided by allowing for multiple sub-systems in the design. It is common practice for example to provide additional air conditioning units in case one of the operating units should suffer a failure, or planned maintenance requires one to be isolated. Regarding power supply, UPS units are supplied in the case of a power failure from the mains; the UPS providing power for the period it takes for backup generators to start up. Security is also high a priority for most data centres due to the sensitivity and business value of the operations that occur within the servers.

Data centres vary in size and may be a room as part of a larger building or be an entire construction in their own right. If they are separate buildings, they will also incorporate support rooms such as offices, meeting rooms and toilets. Figure 2.2 shows what a typical data centre may look like, illustrating the layout of server racks and how several different systems come together to make a successfully operating facility. There are many ways of configuring the layout of a data centre, and these alternative designs will be based upon the method of cooling that is employed.

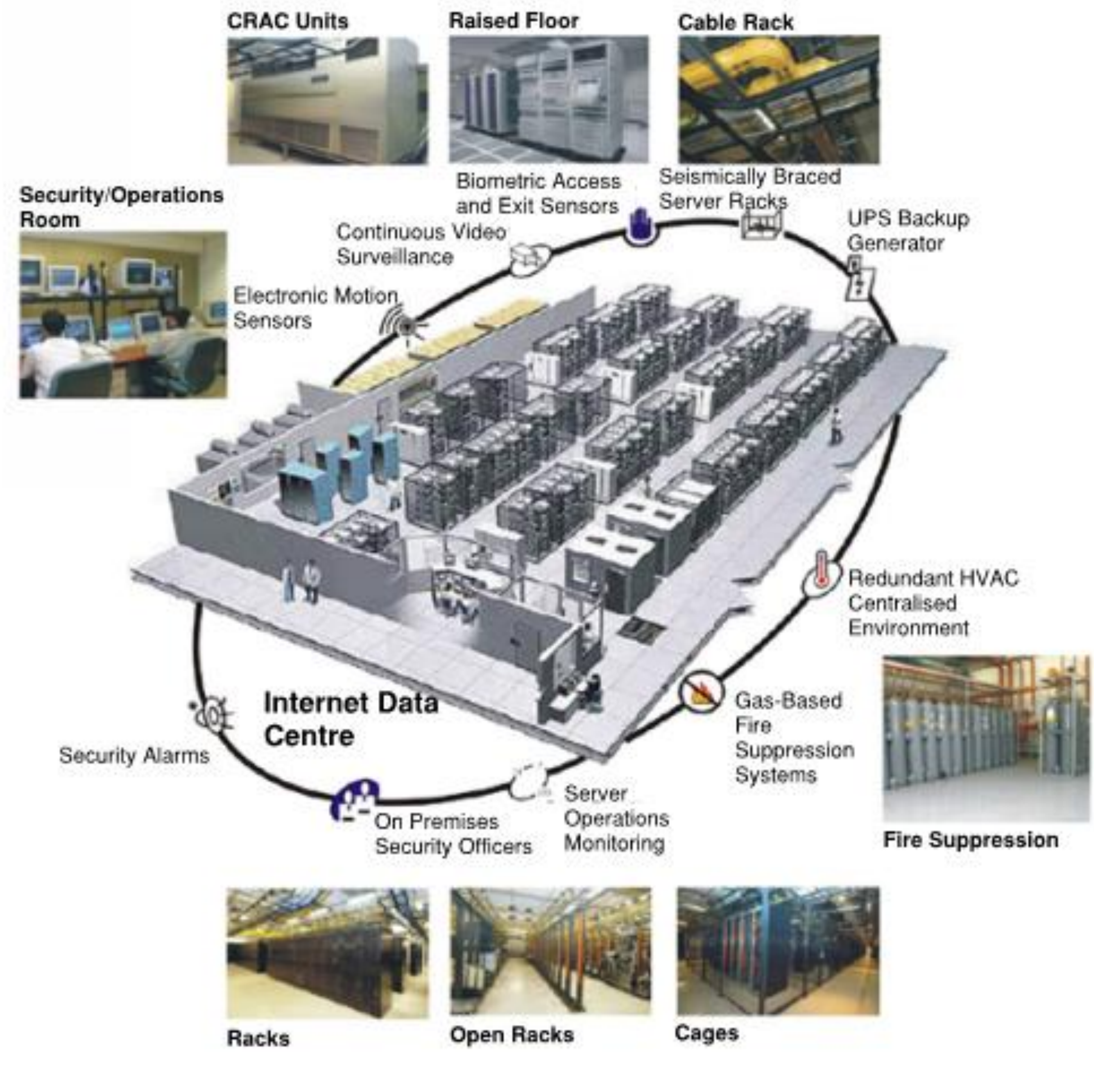


Figure 2.2 A typical data centre layout (sourced from (Cho and Kim, 2011) based upon image from http://www.oriensoft.com/Infrastructure/)

2.3 The Data Centre as a System

Digital infrastructure and the ICT sector are made up of a number of sub-systems, one of which is the data centre facility. Describing how this sub-system operates and what other systems support it, this section explains the scope of research for this thesis.

Access to the internet and communication services is provided through interconnected networks supported by data centres. Physical equipment making up these networks (such as servers, routers and switches) and the way in which they are managed introduces energy consumption and efficiency losses into the wider digital infrastructure system. Therefore, the concept of energy saving should be applied at this high level if overall ICT sector energy use is to be reduced. Many authors in the literature debate the most efficient way of delivering data and communications, with green networks often being used as a term to describe the solutions being offered (Bianzino et al., 2012). Solutions such as virtualization, proportional computing and resource consolidation have all been examined in depth in the literature (Kliazovich et al., 2013, Islam et al., 2015, Klingert et al., 2011, Schulz, 2009). Whilst a brief overview of efficiency in IT equipment and energy proportional computing is given in Chapter 3, an in depth consideration of these solutions is beyond the scope of this thesis.

Systems external to the data centre include energy providers, service providers, equipment manufacturers, design engineers and service users or tenants. These provide supply to and demand for the data centre industry and all play a part in the way these facilities are built up and run. Existing research is available on the part that energy providers have to play in the efficient running of these facilities, with suggestions of initiatives which allow the partnership of data centre energy demand and low carbon energy supply (Agusti-Torra ,2015).

Below this level of networks and supporting services sits the data centre itself and the focus of this thesis. The scope of this research covers the operational level of the data centre by collating the opinions and experiences of colocation providers, owner operators and also engineering consultants. Also sitting at this operational level is the measurement and assessment of energy use. Various tools and methodologies exist for this purpose which cover the physical aspects of the data centre. Current metrics discussed in the literature are examined in this thesis along with a critical evaluation of one key metric that is in use in the industry.

Beyond the operational level sits the technology that enables the data centre to operate, the key component of which is the servers. In the most simplistic of terms, servers enable the conversion of electrical energy into digital, or 'virtual', services. Unfortunately in order to produce the end product (i.e. data, processing, communications), the server consumes electrical energy and converts almost all of this to heat. This thermal energy is a by-product, or in the majority of cases, a waste product that must be removed from the space through the use of cooling and air distribution infrastructure. This along with other necessary supporting infrastructure leads to the data centre becoming a complex integrated system consuming a lot of energy for a given output. Chapter 7 will examine one new way in which it may be possible to save energy at the server level, contributing to an overall reduction in data centre energy consumption.

The scope of work in this thesis covers the operational (by considering the owners and colocation providers along with the metrics that are used to assess energy use) and technology levels (by examining energy use in the server) of the data centre sub-system. By examining both of these, a rounded approach to the research aim set in Chapter 1 can be realised.

2.4 Worldwide Growth in Energy Consumption by Data Centres

In 1965, Gordon Moore presented work on the growth of computing power in ICT equipment leading to the oft quoted 'Moore's Law' (Moore, 1965). Moore's Law predicted that integrated circuits would hold a two fold increase in components (transistors) each year, as had been the trend from 1958 to 1965. In 1975, this prediction was changed to a doubling every two years (Moore, 1975), with this trend holding true since then. Of course this growth cannot be sustained indefinitely; however in the meantime a steadily increasing demand for access to ICT services along with Moore's Law has led to the ICT sector consuming larger amounts of worldwide electricity.

It is a difficult task to isolate the consumption of the ICT sector from the global consumption of electricity, with current assessments having used various methods and assumptions. Difficulties arise mainly due to the lack of data available about the energy consumption of this industry, partly because data centres do not tend to divulge information about their equipment inventories and electricity use. Despite this, there have been a few attempts at quantifying the total world electricity consumption by data centres and the ICT sector as a whole (Huber and Mills, 1999, Walker, 1985, Mitchell-Jackson et al., 2002, Mills, 1999, Malmodin et al., 2010, U.S. Environmental Protection Agency, 2007, Koomey, 2007, Koomey, 2008, Koomey, 2011), with some focussing on the USA alone (Mitchell-Jackson et al., 2002, Kawamoto et al., 2002, U.S. Environmental Protection Agency, 2007).

Of the reports that are available, various methods are employed and assumptions are made leading to a range of results with varying levels of accuracy. One particularly high estimation from 1999 shows that in the United States of America (USA) the consumption of electricity by ICT infrastructure was up at 8%. It was also predicted in the same report that by 2020, around 30-50% of the USA's consumption of electricity will be attributed to the internet and it's supporting infrastructure (Mills, 1999). Electricity consumption figures such as the one already mentioned need to be considered carefully as the boundary conditions of the calculation may not always be clearly specified. The Environmental Energy Technologies Division (EETD) at the Lawrence Berkeley National Laboratory (LBNL) criticised the assumptions made in this early report and concluded that the study may have overestimated the power consumption of equipment assessed. The EETD also mention that the prediction of a significant growth in the consumption of electricity in the near term was over also estimated (LBNL, 2001). It should also be noted that the scope of this prediction for 2020 did not include the electricity consumption of the computers and terminals that are required to use the internet, as the boundaries for the analysis were not set this wide.

Since that 1999 study, estimations have become more accurate as more data has become available. In 2007, the U.S. Environmental Protection Agency (EPA) published a report discussing the growth of data centres in the USA and the associated rising costs of running them. The report also discussed opportunities for increased energy efficiency (U.S. Environmental Protection Agency, 2007). It was estimated in the EPA report that data centres in the USA alone consumed 61 billion kWh of electricity in 2006, equivalent to 1.5% of the country's total electricity consumption for that year and costing in the region of $4.5 billion. This consumption was estimated to have doubled since 2000. Again considering the background of the calculation, these figures were estimated by starting with the installed base of ICT equipment (servers, disk drives and networks in data centres). This was then multiplied by the estimated energy consumption of each device and a factor to take into account the energy used by support equipment such as cooling plant. This figure is much more widely accepted than the high value of 8% estimated in 1999.

In 2010, Malmodin et al also conducted a study on worldwide energy use by data centres. The authors found that they consume 180 TWh of electricity which is equivalent 1% of total world electricity consumption, with the ICT sector as a whole contributing 3.9% (Malmodin et al., 2010).

When it comes to reporting the worldwide electricity used by the ICT sector and data centres, Jonathan Koomey has played a prominent role in the industry. Koomey has so far written two major reports (2007 and 2011) and a paper (2008) on the growth of power consumption in data centres in the USA and across the world.

In the first of these reports, Koomey discusses that there were a lack of studies available at the time of writing regarding the total electricity consumed by servers and associated infrastructure (Koomey, 2007). This lack of availability is said to be down to a lack of data that has been made available for use by companies. It is also down to the nature of the industry to develop rapidly in terms of the technologies used. The data that could be gathered came from IDC, a company tracking the installed base and shipment of servers in the USA and the world. The study by Koomey grouped the installed base of servers into 'volume', 'mid-range', and 'high-end' classifications with each group costing less than $25,000, between $25,000 and $500,000, and more than $500,000 respectively (2007 prices). Within each of these categories, the power consumption of the six most popular units was averaged and this figure was then multiplied by the installed base of servers to find total electricity consumption. The analysis of this data resulted in the following findings;

- the electricity use of servers doubled between 2000 and 2005 across the world and the USA (see Figure 2.3);
- it was found that for the USA, the total power consumed by servers came to 0.6% of the total electricity consumption for the country in 2005 and when the associated infrastructure is included, this increases to 1.2% (this is a close comparison to the 2006 estimate of 1.7% by the U.S. EPA (U.S. Environmental Protection Agency, 2007)); and
- it was suggested in this report that if growth trends continue, then electricity consumption by servers would be 76% higher in 2010 compared to 2005.

Even though the estimations in the report by Koomey may be more accurate than earlier reports from other authors, there are still a few potential sources of inaccuracies identified in the report due to the data that was used. The data for the installed base also includes servers that are not installed in data centres, which depending on the quantities and locations of these servers may skew the results in terms of the amount of cooling needed for these. Another problem with the analysis is that it did not include custom built servers with different operating characteristics which are increasing in popularity, companies such as Google and Facebook are now producing their own hardware (Miller, 2009, Open Compute Project). Assumptions also had to be made regarding the cooling of servers. In order to achieve an estimate, the energy consumption of the cooling infrastructure aspect of the calculation was incorporated by assuming a Power Usage Effectiveness[5] (PUE) of 2 for all of the servers. Actual PUE values vary a lot depending on the data centre concerned and its location. Given this, it is difficult to make an accurate estimation of electricity consumption using an average PUE value.

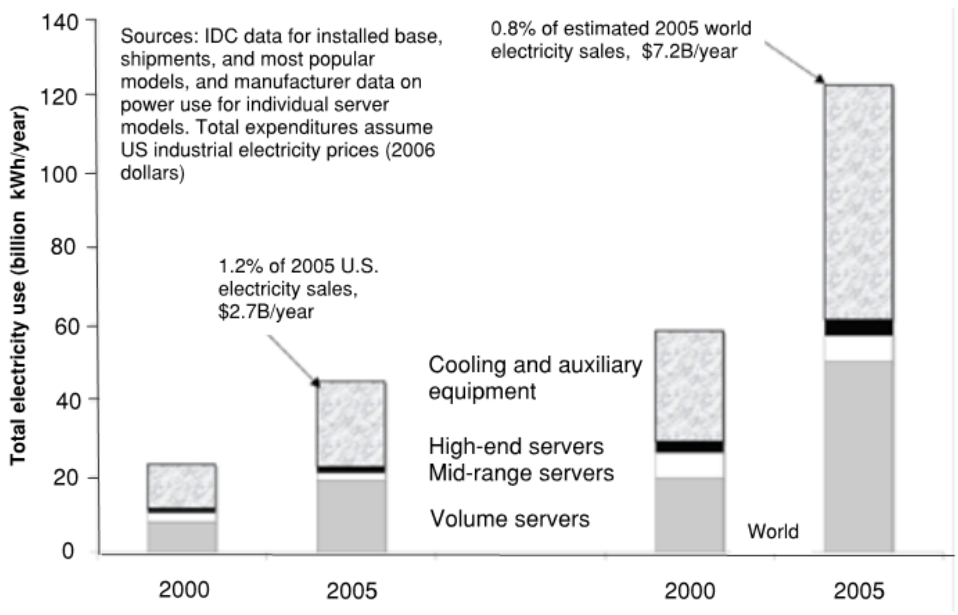


Figure 2.3 Total electricity consumption of servers in the USA and the world for the years 2000 and 2005 [10]

In a 2008 paper by Koomey, it was calculated that a total of 150 billion kWh of electricity was consumed by data centres in 2005 (Koomey, 2008). This consumption consisted of half ICT load and half cooling and power distribution load. When this was compared to the world wide estimated consumption of electricity for that year, it was found that it came to a total of 1% of the world's electrical energy consumption. This study was conducted using data from the Energy Information Administration's *International Energy Outlook* (Energy Information Administration, 2008). Comparing this figure to the figure of 1% total electricity consumption in 2010 from Malmodin et al, it would either appear that there has not been any growth between 2005 and 2010, or there were inaccuracies in the estimated figures.

The latest report from Koomey presents the growth in electricity consumption by data centres from 2005 to 2010. Due to changes in the base data, new assumptions needed

---

5 Further discussion and explanation of this metric can be found in Chapters 3 and 6

to be made. So the author constructed the following four scenarios; 'all trends continued', 'best guess 2007', 'upper bound case', and 'lower bound case'. This report indicates a lower number of servers installed than predicted in the 2007 report, which is mainly due to the economic problems of recent years and also due to the increased usage of virtualisation (which enables servers to operate under higher workloads, reducing the overall number of servers).

This report found the following outcomes from the analysis (Koomey, 2011);

- the electricity consumption of data centres increased by 56% from 2005 to 2010, rather than doubling like it did in the five years previous (this is lower than the 76% growth prediction in Koomey's 2007 study);
- in terms of the percentage of worldwide electricity, data centres consumed 1.1-1.5%. For the USA alone this figure was slightly higher at 1.7% - 2.2% (Koomey, 2011) (which is in line with the 1% estimation provided by Malmodin et al); and
- the growth in electricity consumption by data centres was lower than predicted in the EPA report. This lower than expected increase was attributed in part to the 2008 economic crisis and the increasing use of virtualisation of servers, the actual growth rate was between 36% and 56% (low case and high case scenarios) across the world.

A more recent study by Van Heddeghem found that data centres consumed 270 TWh of electricity in 2012 (around 1.2% of the worldwide total consumption), a value approaching Koomey's upper bound case. The authors in this study also found that the share of worldwide electricity consumption attributed to data centres and the ICT sector is increasing at a rate of nearly 7% per year (Van Heddeghem et al, 2014).

Reports are now generally converging on a figure in the region of 1.5% for worldwide consumption of electricity by data centres.

2.4.1 Climate Change Policy and Data Centres

In 1992 the United Nations Framework Convention on Climate Change (UNFCCC) was formed at the Earth Summit in Rio de Janeiro, coming into force in 1994. The UNFCCC aims to keep atmospheric concentrations of GHGs at a safe level, i.e. at a level which '*would prevent dangerous anthropogenic interference with the climate system*' (IPCC, 2007b). In order to increase the chances to achieving this, the atmospheric concentration of CO_2–eq needs to be stabilized at 450 ppm in order to limit the global increase in temperature to 2°C[6] (any increases above this would lead to significant and irreversible climate change effects) (den Elzen and Meinhausen, 2005). Acting to encourage countries to reduce their emissions, the UNFCCC advises on targets that should be met. In order to set legally binding emissions targets and thus providing a greater incentive to meet them, the Kyoto Protocol was adopted by 37 countries in 1997 commencing in 2005 (United Nations Framework Convention on Climate Change). Under the Kyoto protocol, the UK was legally required to reduce its emissions by 12.5% based on 1990 levels by 2008-2012. In 2013, the UK had met these Kyoto targets, however overall global emissions have still increased.

More recently, the UK Government introduced significantly tougher targets for the UK as part of the Climate Change Act in 2008. The Climate Change Act requires that the UK reduces its GHG emissions by 80% relative to 1990 levels by 2050 (UK Government, 2008). Setting such ambitious targets are necessary if changes in the climate are to be kept to a minimum. Statistics show that levels of GHGs in the UK have been falling since 1990, and are now below the benchmark emissions from 1990 (although they increased slightly in 2012 due to a cheaper price for coal) (Department of Energy and Climate Change, 2013). Whilst this is good news for the UK, it does not mean that efforts should be relaxed especially given that demand for energy is increasing not just in the UK but worldwide. Targets such as those mentioned feed down through industries in individual countries and affect them through their sourcing and consumption of electricity along with the efficient use of energy.

Data centres are affected by these targets due to their high electricity consumption. Over the past few years there has been a greater focus on the data centre industry in terms of its energy consumption and efficiency. This has led to guidance schemes and regulations which directly deal with energy efficiency in data centres, which will be discussed further in Chapter 3.

In terms of climate change policy on a European level, the 20-20-20 Renewable Energy Directive was adopted by the EU in 2008 setting goals for reducing the impacts of climate change. The directive calls for a 20% reduction in GHG emissions, a 20% increase in energy efficiency and also a 20% increase in the use of renewable sources of energy. Directly affecting the data centre industry, this directive means that efficiency needs to be increased significantly. In order to help this, the EU set up the

[6] Relative to pre-industrial levels

Code of Conduct on Data Centre Energy Efficiency (EU CoC), a best practice guideline aimed at decreasing energy use in the data centre (discussed further in Chapter 3).

Trading schemes such as the EU Emissions Trading Scheme (EU ETS) can also be an important piece of policy used for reducing carbon emissions (Kemfert et al., 2006). Whilst the EU ETS has had some difficulties in the past, it has brought results in that it has encouraged generators of energy to reduce their emissions. Whilst at first glance it may seem that data centres would not be covered under this as they are energy consumers rather than generators, all data centres have standby generators. Standby generators are covered in the EU ETS (phase III) if they have a total rated capacity of 7 MW or more. This means that as of the 1st January 2013, most of the larger data centres across Europe will need to pay for carbon credits to account for their emissions from electricity generation.

Other UK policy that affects the data centre industry includes the Carbon Reduction Commitment (CRC) which applies to organisations in the UK that consume more than 6,000 MWh of electricity per year (affecting larger data centres). Organisations affected by this are required to report on their carbon emissions regularly to ensure that they are managing to reduce them over time. Allowances then need to be purchased for any carbon emitted, essentially providing a heavy incentive to reduce the levels of emissions. A recent development however has led to the UK government granting the data centre industry a Climate Change Agreement (CCA), giving the industry a significant discount on the CRC (Jones, 2013). Whilst a discount on the cost is granted under the CCA, this comes with the agreement that the industry will meet stringent energy efficiency targets. By granting the CCA, it means the UK government recognises the importance of the industry and has helped to ensure that it remains attractive to investors looking for ICT services in the UK.

2.5 Maintaining the Data Centre Environment

Electricity consumption in data centres is driven by the electronic equipment, mainly the servers, storage devices, switches and UPS units. Almost all of this use of electrical energy is converted heat by the electronic equipment making precision cooling[7] a necessity within the data centre. It is important that the servers are provided with a

[7] Precision, rather than comfort cooling, is more suited to data centres as it involves closer control of temperature and humidity

suitable working environment as failures can carry extremely costly consequences. Providing the correct operating environment requires precision cooling and air distribution infrastructure which can consume well over 25% of the total energy consumption of the data centre (Kant, 2009, Tsuda et al., 2010, Cho et al., 2009, Pelley et al., 2009, Sun and Lee, 2006). As a significant consumer of energy in the data centre, it is important that the cooling and air distribution infrastructure is considered when trying to reduce energy use.

In order to understand where energy savings can be made and efficiency improved, it is necessary to first discuss how heat is generated in the data centre and the methods employed to remove it from the space. Following this discussion, energy efficient methods of heat removal such as economisation and free cooling are presented and the benefits of different methods are reviewed.

2.5.1 IT Equipment Environmental Requirements

All electrical equipment inherently consumes power, and IT equipment emits almost all of this energy as heat. With 99% of the energy consumption being converted to heat, this results in the IT load contributing over 70% to the total heat load in the data centre (Lu et al., 2011) (the rest of the heat load is a result of UPS units and batteries, power distribution, lighting and people). If this large heat load is not dealt with then it can reduce the lifespan of equipment and lead to increased failures.

As an electric current passes through the semiconductors the majority of the energy associated with that current is converted to heat energy which needs to be dissipated to the external environment. This heat energy, Q, can be determined by calculating the product of the voltage drop, V, across the component, and the current, I, that flows through the component (Remsburg, 2000);

$$Q = VI \tag{2.1}$$

However, the above expression is only applicable if there is a constant voltage and current and so if either one varies with time then equation 2.1 must be used and the power dissipated becomes a mean value, P_{dm};

$$P_{dm} = \frac{1}{t}\int_{t_1}^{t_2} V(t)I(t)dt \tag{2.2}$$

If this heat energy is allowed to build up within the semiconductors, then the lifetime of the equipment reduces along with a severe decrease in the operational reliability. The lifetimes of server chip components are assessed using a power cycling test, where a load is applied and the chip is allowed to reach a maximum temperature. Once this maximum temperature is achieved, the load is turned off and the chip is allowed to cool down to a minimum temperature, after which the cycle begins again (Lutz, 2010). This introduces mechanical stress to the component due to the expansion and contraction of the materials, it is this ability to cope with mechanical stress that determines the lifetime of the component. Over the lifetime of a server in a data centre, continuous cycling between warm and cooler chip temperatures (due to variations in work load demand) reduces reliability due to the mechanical stress. With regards to the effect of fluctuations in server inlet air supply temperatures, Sankar demonstrated that the normal variations in temperature seen in a data centre (5% variation) did not have an impact on the failure rates of the server (Sankar 2012).

Figure 2.4 illustrates the relationship between the failure rate of IT components such as microprocessors and their temperature. It can be seen that once the temperature reaches 70°C and above, the rate of failure of the component increases dramatically.

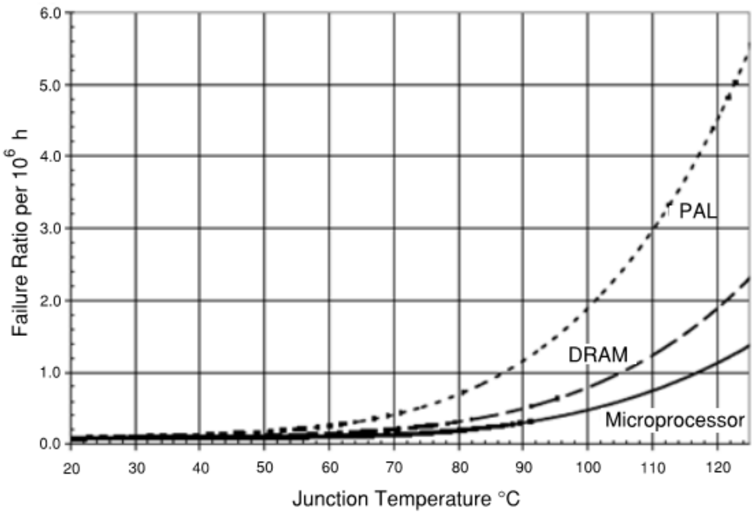


Figure 2.4 Relationship between temperature and component failure rate for microprocessors, DRAM (Dynamic Random Access Memory) and PAL (Programmable Array Logic) (Remsburg, 2000)

In order to reduce the build-up of heat and hence the risk of failure within the server components, the heat needs to be dissipated to the surrounding environment. Heat generated by the servers can be removed by passing cool air over the components, although water cooling is also available by using cold plates (Copeland, 2005) or complete submersion of servers in dielectric fluids (Greenberg et al., 2006, Iceotope, Cooling). In a typical data centre facility, air is delivered to the data hall via an air conditioning system and passes over the server units removing the heat. The air is then removed from the space and is usually rejected to the atmosphere, although in some cases this low grade waste heat can be reused (Miller, 2010).

The heat density of data centres has been increasing over the years due to more powerful equipment and an increasing expansion in the number of servers. Figure 2.5 shows the increase in heat density of certain IT equipment from 1992 to 2010 (Uptime Institute, 2000). This increase in heat density is a problem because it means that there are higher heat loads and therefore more energy is required for cooling. Just how much cooling is required depends upon the conditions that are deemed safe for the equipment to operate under. In order to reduce failure in servers, they need to be maintained in an environment with certain air properties.

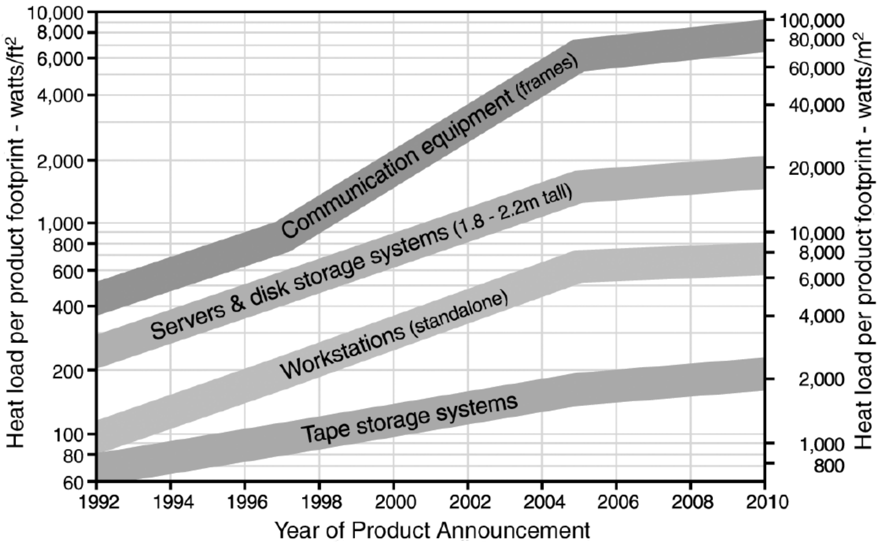


Figure 2.5 Heat density trends in ICT equipment (Uptime Institute, 2000)

The ASHRAE (American Society of Heating, Refrigeration and Air Conditioning Engineers) Technical Committee 9.9 first produced guidelines in 2004 as to the recommended temperature and humidity of air that is to be delivered to servers; further guidelines were then published in 2008, with a later edition in 2011 (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011). The first set of guidelines that were produced aimed to minimise the failure of IT equipment within data centres, however in later editions energy efficiency has become a key focus.

The ASHRAE guidelines set out an operational envelope for temperature and humidity within which it is acceptable to maintain servers and other IT equipment. There are two envelopes within the guidelines, the recommended and the allowable, the latter of which is wider but presents a higher risk of equipment failure. The first envelope has been expanded since its introduction so that supply temperatures within data halls can be increased hence reducing the energy consumption of chillers, it has also meant that free cooling is now possible in more climates. As previously mentioned however, servers cannot operate in conditions above certain temperatures and there are concerns among data centre owners regarding operating at the higher temperatures suggested by ASHRAE (Beaty and Lintner), despite the reduced energy costs that are associated with this wider envelope. This is due to the higher failure risks associated with increased temperatures and the link between server up time[8] (which is highly important for business continuity) and equipment failure rates. Higher equipment failure rates leads to higher economic losses for businesses.

The envelopes recommend the following properties for the temperature and humidity of the air supplied to servers (Figure 2.6 shows these envelopes on a psychrometric chart);

Supply Air Temperature

> The air supplied to servers should be at a temperature range of 18 to 27°C (64.4 to 80.6°F), this is the recommended envelope. Depending on the type of data centre, the allowable envelope extends to a range of 5 to 45°C (41 to 113°F) (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011).

Supply Air Relative Humidity

> The recommended envelope covers a humidity range of 5.5°C dew point to 60% relative humidity and 15°C dew point, the allowable is wider at 20-80% relative humidity or -12°C dew point and 8-90% relative humidity depending on the data centre (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011).

[8] Defined as the length of time during the year that a data centre is operational

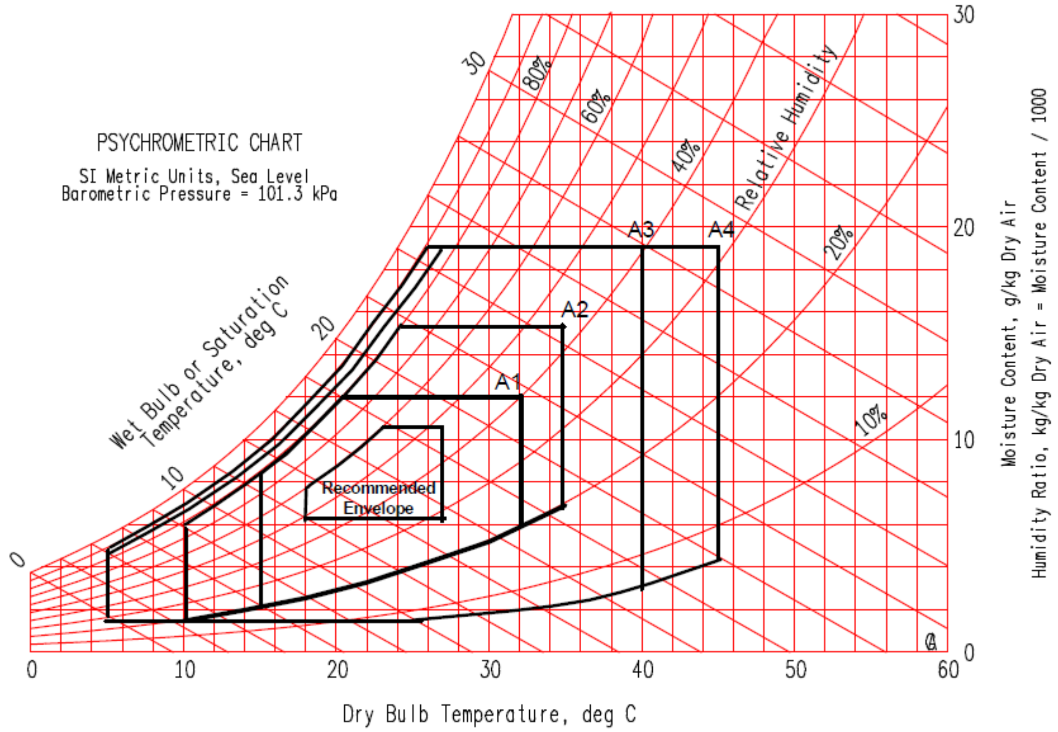


Figure 2.6 Environmental classes within the ASHRAE 2011 Thermal Guidelines for Data Processing Environments (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011)

Whether or not the ambient conditions for a particular geographical location falls within the recommended envelope is driven by climatic conditions. The advantage of the extended allowable envelope is that more locations are able to offer free cooling for data centres as the ambient air properties will fall within this wider range for a greater proportion of the year. Free cooling is a technique whereby compressors are not used as part of the cooling infrastructure. Due to this, a significant amount of energy can be saved as the supply air is cooled by other means such as evaporative cooling. A wider ASHRAE envelope is important for this as free cooling has more reliance on the external temperature to function effectively. Free cooling is discussed in more detail further on in this chapter. Raising the allowable inlet temperature to the server racks also increases the overall energy efficiency of the data centre. Estimates show that a 1°C rise in server inlet temperature can produce an energy saving of 4-5% (42U, 2009). Breen et al have shown through modelling that an improvement of 8% in the coefficient of performance (COP) for the cooling system is achieved for every 5°C increase in server inlet temperature (Breen et al., 2010). Despite the obvious energy saving advantages available with a wider envelope there is still debate as to whether the inlet temperature should be raised or not (42U, 2009), with some studies showing that energy consumption of certain components (such as fans) increases with the raised temperature (Patterson, 2008, Cole, 2011). One of the other issues holding back data centres from extending their operating temperature is that the standards set by some companies are stricter than the ASHRAE guidelines (The Green Grid, 2011).

2.5.2 Heat Removal

The amount of heat that is produced by servers and other ICT equipment within the data centre is at odds with the environment that must be maintained. This requires precision cooling which is described by Emerson as a method to '*strictly regulate atmospheric conditions within a very narrow range*' (Emerson Network Power, 2010). Only this level of cooling can provide the environment required by servers, it is quite different to the comfort cooling of offices for example. This means that it can be a much higher energy intensive system. It also means that designing a cooling and air distribution system for a data centre can be difficult due to the complex temperature distributions and air flows present.

When designing the cooling infrastructure for a data centre, there are many different options to choose from. Each option will have different advantages and disadvantages and there will be plenty of influences over the choice of system. These influences can include location, facility size, facility type, energy consumption targets and costs.

There are several strategies for dealing with the heat load in a data centre. Computer Room Air Conditioning (CRAC) and Air Handling (CRAH) with compression cycles have been used more commonly in the past, whereas current trends are moving towards utilizing free cooling where possible (using outdoor air to remove the heat load rather than a refrigeration system). The recent expansion of the ASHRAE envelope has allowed for more widespread use of less energy intensive methods of cooling as discussed previously.

The following discussion looks at the main strategies employed in terms of cooling systems, which are grouped into the following headings:

- CRAC/CRAH;
- liquid cooled racks;
- air economisation; and
- liquid cooled servers.

CRAC/CRAH based cooling

Traditionally, cooling has been provided in data centres by utilizing a refrigeration system which includes equipment such as CRAC or CRAH units. A CRAC unit is usually situated within the server room itself, using a compressor and evaporator circuit to provide cool supply air. A CRAH unit on the other hand purely delivers rather than conditions air which has been cooled by chilled water. The chilled water is provided by a packaged chiller or other heat rejection equipment situated external to the data centre.

There are several types of CRAC systems that can be used, differing depending on their method of heat rejection. Air cooled, glycol cooled and water cooled are all different configurations of CRAC based cooling. An air cooled CRAC is attached to a dry cooler and involves pumping refrigerant between the two pieces of equipment. It is easy to maintain and has lower costs attached to it. Figure 2.7 illustrates a typical CRAC based system. Warm return air from the data centre is drawn over a cooling coil and returned cool to the data centre. The cooling coil is supplied by a refrigerant which has been through a Carnot cycle refrigeration process.

CRAH units are a feature of a chilled water system, as shown in Figure 2.8. Situated within the server room, they draw warm return air over a chilled water coil, the water for which is provided by a chiller located in a separate room. The chiller then normally rejects this heat to the environment via a cooling tower. This kind of system tends to be installed where the data centre is part of a larger building and shares the cooling infrastructure (Evans, 2012). CRAH systems also tend to be used in data centres with a power rating of over 200kW (Evans, 2012). The CRAH units can be paired with other heat rejection equipment which may offer the opportunity for free cooling, depending on the chilled water return temperatures and ambient air conditions.

CRAC/CRAH systems are used most commonly in data centres; however they are not necessarily the most energy efficient choice. They normally have a refrigerant loop with a compressor which consumes energy, and fans to move the air. A lot of modern data centres are now moving towards compressor-less cooling which enables significant savings in energy use. These systems do however provide a level of security and reliability in that complete control is given over the air temperature and humidity, something which some data centres are keen on having.

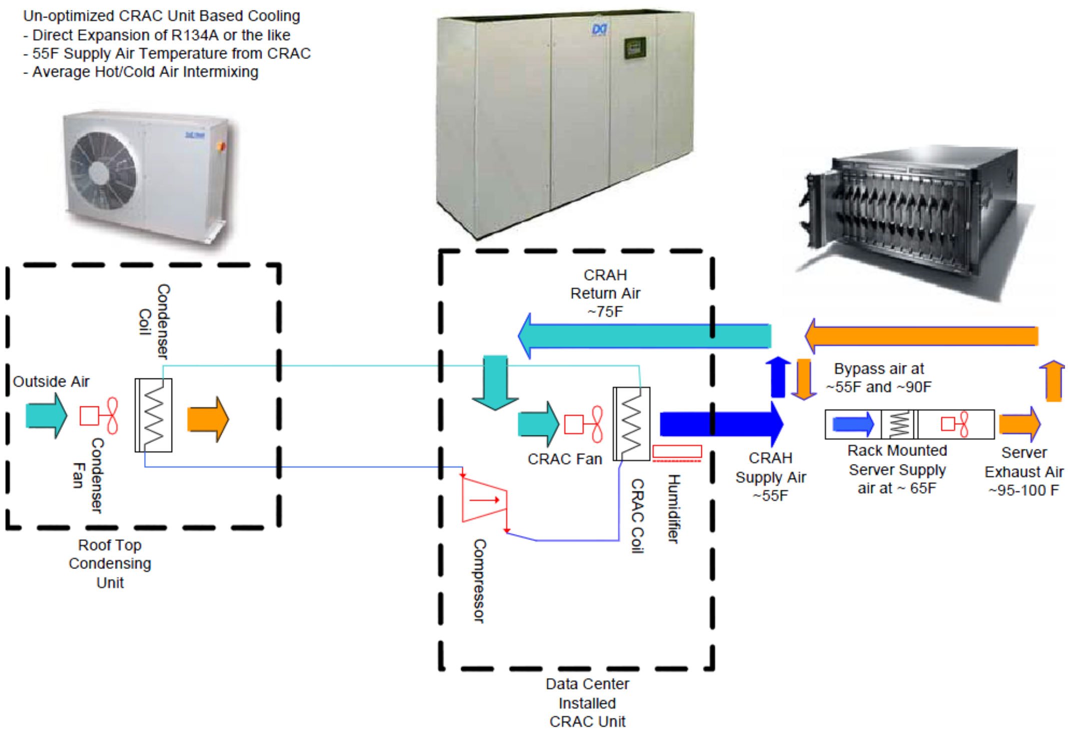


Figure 2.7 Illustration of a CRAC based cooling system (Kennedy, 2009)

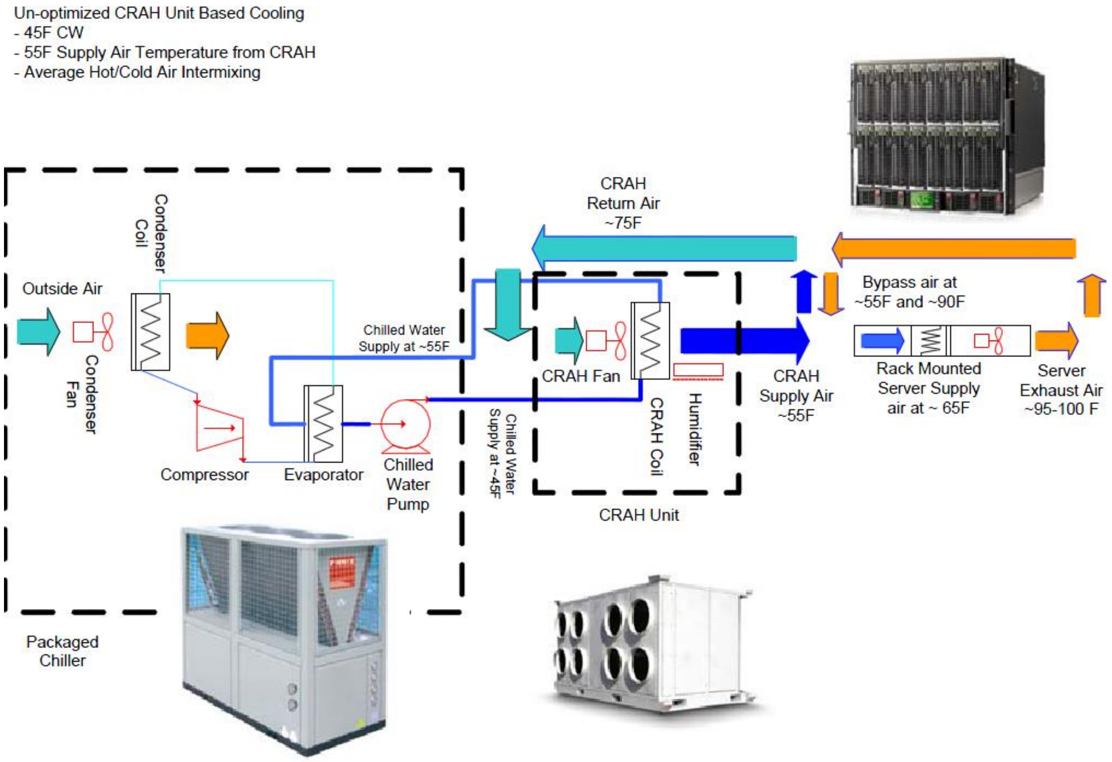


Figure 2.8 Illustration of a CRAH based cooling system (Kennedy, 2009)

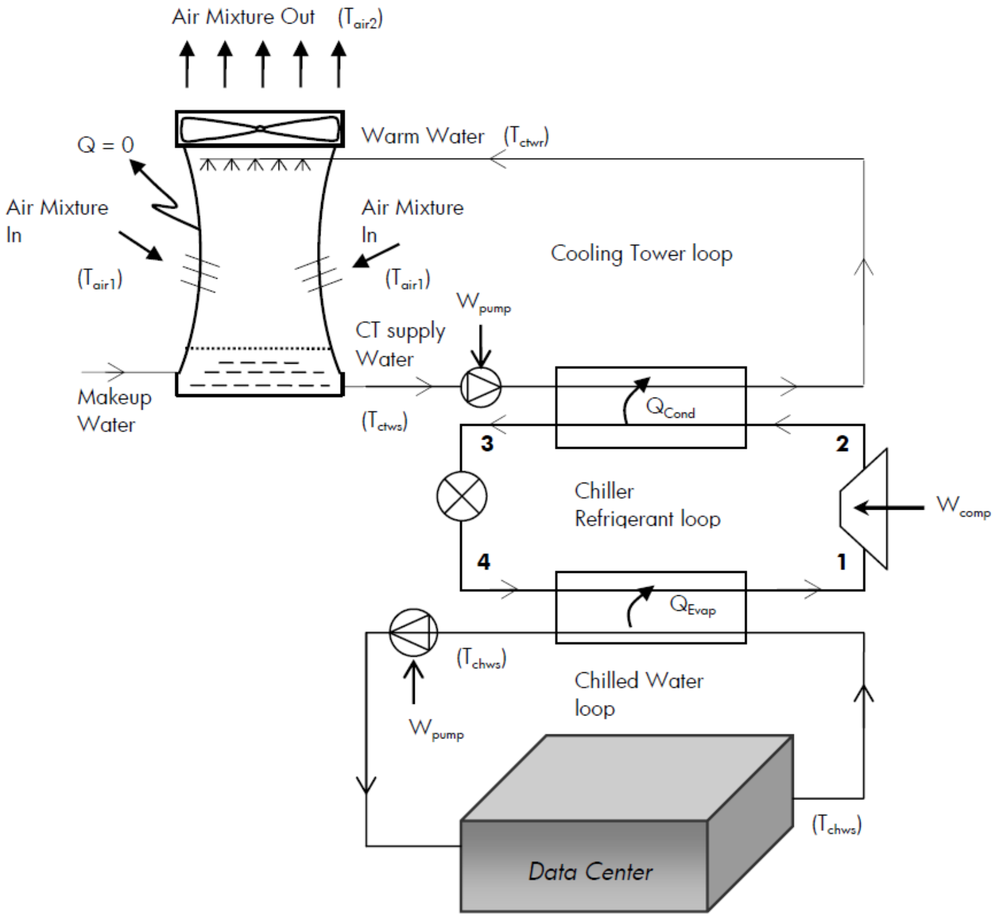


Figure 2.9 Schematic showing the layout of a typical water cooled chiller system for a data centre (Beitelmal and Patel, 2006)

Air Economisation

In regards to cooling, economisation involves using different methods to reduce the energy consumption of the cooling system. Economisation has become known as ‘free cooling’ and allows outdoor ambient air to be delivered to the server room directly, or to be used for cooling water in a chilled water loop so that the data centre environment is cooled indirectly. This negates the need for an energy intensive refrigerant loop and reduces overall energy consumption due to the lack of a compressor. As discussed earlier, this method of cooling has become available to more facilities due to the expansion of the ASHRAE envelope. Free cooling can offer a 75% reduction in energy use compared to chilled water plant, and the reduced use of chillers provides added cost savings through higher reliability and lower maintenance (Greenberg et al., 2006). Other estimates show that the possible savings are higher at 82% (Kennedy, 2009).

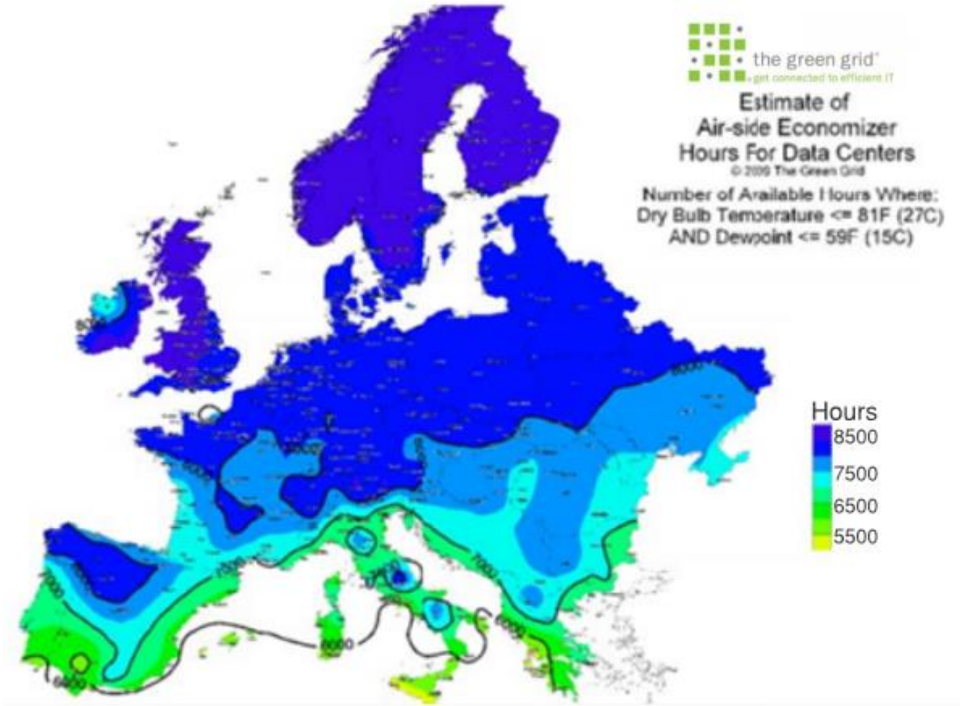


Figure 2.10 Number of free cooling hours available within Europe based on the ASHRAE envelope from 2009

Figure 2.10 shows a map of Europe and indicates how many hours of free cooling are available during one year. Hours of cooling are defined as the number of hours in the year where the temperature is low enough to enable outdoor air to be delivered to the servers without the need of a compressor. The illustration shows the hours available under the 2009 ASHRAE guidelines, clearly illustrating that around 8500 hours a year of free cooling are available across Northern Europe. If the map were to be updated with the more recent 2011 ASHRAE recommended temperatures, then almost all of Europe would be covered by the 8500 hours band of colour. In fact, due to the expansion in the envelope, it is possible for data centres across Europe to use free cooling for over 8500 hours or 97% of the year. This presents a huge opportunity for data centres to decrease their energy usage.

The map above however only indicates the amount of free cooling hours available based upon the dry bulb temperature being lower than 35°C and the dew point being below 21°C. There will be occasions during the year where these two criteria will be met, but there may be times during the year where humidification is required or additional cooling is needed. This is where free cooling systems can exploit the latent heat of evaporation, a method known as evaporative cooling. If water can be evaporated by wetted media or misting jets then the temperature of the air can be reduced, allowing cool air to be provided. This evaporative cooling can be provided directly or indirectly, as illustrated in Figure 2.11 and provides both humidification and cooling. If only humidification is required, then warm return air can be mixed in to bring the temperature back up.

Evaporative cooling can either be delivered directly or indirectly, as shown in Figure 2.11. Indirect cooling allows for less filtration in the system as the outdoor air is not delivered directly into the server room. Direct air supply is a simpler system though as it does not involved heat exchangers which are needed to separate the outdoor and indoor air streams.

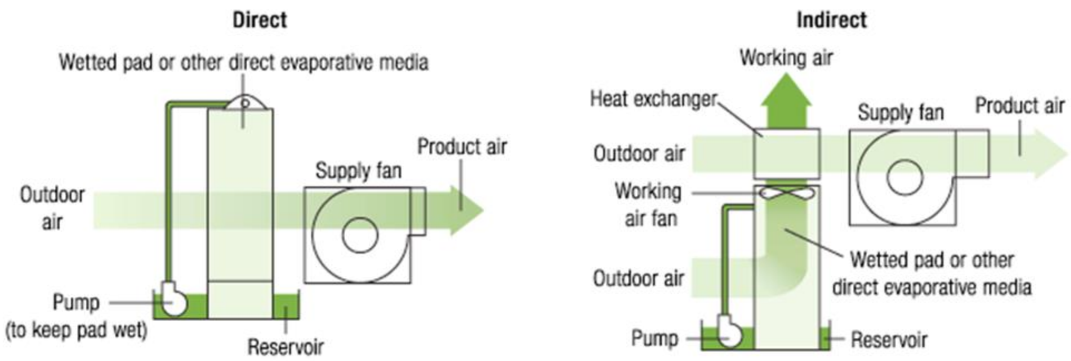


Figure 2.11 Illustration of direct and indirect evaporative cooling (ESource)

Liquid cooled racks

A liquid cooled rack system is still based upon a refrigerant circuit, although unlike with a CRAH system, the units are situated directly in the rows of racks. Due to their position, there is a shorter distance for the air to travel which reduces the energy consumption. Figure 2.9 shows the typical use of a chiller system to deliver chilled water to a data centre, which in turn provides cool air for the servers.

In row coolers also help to distribute the cool air more evenly across the racks of servers (Priyadumkol and Kittichaikarn, 2014). These units tend to be used as additional cooling for areas of higher density IT equipment in the data hall. Cool air from a CRAC/CRAH unit is usually delivered to the servers via a raised floor system or delivered directly into the server room. However this can mean that cold air is not distributed evenly across the rack, the top servers often end up warmer than the bottom ones (Priyadumkol and Kittichaikarn, 2014). Other liquid cooled rack systems involve back door coolers, whereby the air conditioning equipment is attached to the back of each server rack. This is sometimes in the form of a pumped refrigerant system (Evans, 2012), which is a chilled water system like the CRAH however it does not bring the water near the servers, only the refrigerant enters the server room.

Liquid Cooled Servers

As mentioned previously, servers tend to have heat sinks attached to the Central Processing Units (CPUs) so that they aid in removing heat from these components. However, there are two further options which both involve bringing liquid in closer proximity to the servers. These are either using direct liquid cooling with a remote heat sink which is air cooled, or the installation of small pipes within the server with a condenser either located in or away from the server (Chan et al., 2009). These two options offer the ability to reduce the energy consumption of the air distribution system, which can consume a lot of energy as fans require a lot to operate (Smith, 2009, Vogel et al., 2010). By reducing the need for air distribution fans (both external and internal to the server) the overall energy consumption of the data centre can be reduced, due to the power requirements of the fans.

Liquid cooling of servers has been suggested in the past and new developments in technology are making this a promising method of cooling. Servers can either be cooled by cold plates or be completely submersed in dielectric fluids. These offer more efficient cooling than air delivery systems as the latter incurs a lot of losses in energy through the air delivery process. Dielectric fluids also have a much higher specific heat capacity allowing more heat to be stored per unit volume compared to air. It is because of this higher heat capacity that the technology also provides greater opportunities for using heat recovery technology (Tuma, 2010). When using waste heat recovery, liquid cooling provides a much higher grade heat than air would.

Liquid submersion cooling

Liquid submersion cooling involves the components within the server coming into direct contact with a non-conductive liquid. The merits of this technology are discussed by Tuma who investigated the advantages of using open bath submersion cooling (offered by companies such as Green Revolution Cooling (Miller, 2013b) amongst others) in data centres. Advantages include the fact that the quantity of infrastructure is significantly reduced (i.e. no fans, pumps or compressors), added fire protection (through the use of non-electrically conducting liquids) and higher server density (Tuma, 2010). Also, because there is no requirement for fans and heat sinks within the servers, the units can be denser hence saving space within the data centre. In fact, Tuma estimated that rack density could be 2.5 times higher using open bath submersion and a facility density of 10 times higher could be realised (Tuma, 2010).

However, this technology is not without its problems. One of the key issues facing liquid cooling is that each server and component needs to be 'hot swappable', meaning that if one of them is removed from the rack the rest need to remain in operation. Another problem is that there tends to be a loss of fluid due to evaporation, especially in open bath methods, which may have unintended consequences for the environment due to the liquids having a high global warming potential.

Another option for liquid submersion cooling is to have the liquid sealed within the server unit (a technology offered by Iceotope), this means that there is no liquid open to the atmosphere and servers can be more easily maintained (Hopton and Summers, 2013, Chi et al, 2014).

Figure 2.12 Liquid submersion cooling from Green Revolution Cooling (Green Revolution Cooling)

Liquid-to-chip

As an alternative to submerging the whole server unit in liquid, liquid-to-chip or cold plate cooling can also be used. One particular method of doing this is to pipe water through the server where it is then fed into a module attached directly to the CPU. The water flows through the module and takes with it the heat generated by the CPU. This means that the CPU fans and heat sink (which sits on the CPU in direct contact) can be removed, leaving behind the remaining fans to circulate air around the other components (which generate less heat than the CPU). This kind of system could use 90% less energy than a conventional cooling system that uses a compressor (Iyengar et al., 2012).

2.5.3 Air Distribution

As discussed previously, there are several different methods of providing cool air for a data centre and there are also several ways to distribute cool air to the servers and remove the warm return air. Airflow within a data centre is very complex (Fakhim et al., 2011) and there have been many studies assessing different methods of air distribution and their relative energy merits (Fakhim et al., Cho et al., 2009).

The method of air distribution used in any particular data centre will depend upon the amount of space available, as the method chosen may depend upon raised floors, aisle layout or aisle containment. Figure 2.13 shows an illustration of twelve different ways of managing the flow of air in a data centre, as presented by Rasmussen (Rasmussen, 2011). An important consideration when deciding on a layout for the air distribution system is to ensure that supply and return air streams are not allowed to mix as this leads to significant inefficiencies. One way of ensuring that the air streams do not mix is to employ aisle containment. Both hot and cold aisle offer benefits in terms of increased air distribution efficiency, but there is debate over which solution is the best (this is discussed further in Chapter 3).

Chapter 3 discusses in more detail the efficiency gains that can be made through improving air distribution in the data centre. The following section considers the air distribution through the server units themselves, providing the background information for the experimental study in Chapter 7.

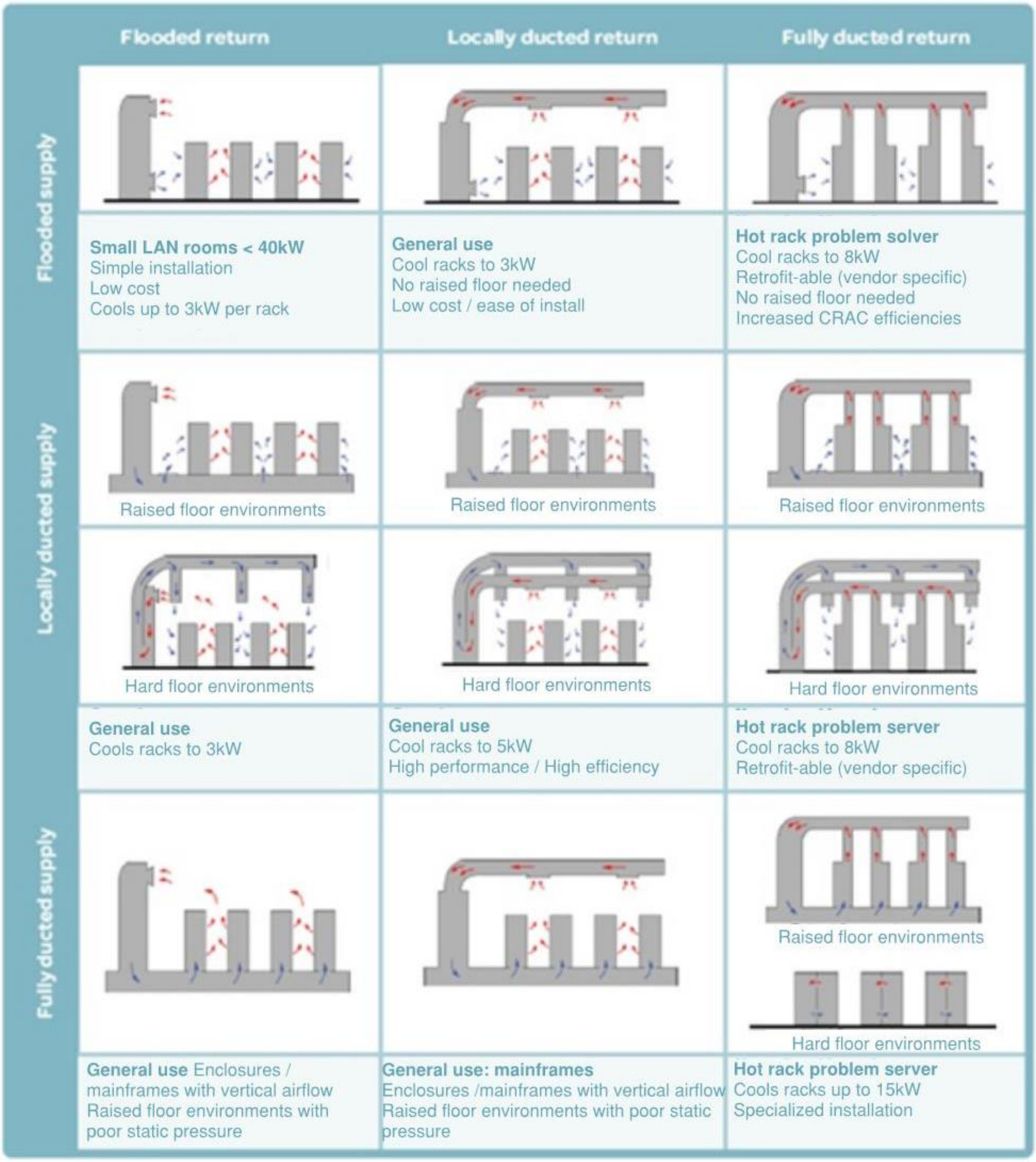


Figure 2.13 Twelve alternate methods of delivering supply and removing return air in a data centre (Rasmussen, 2011)

2.6 A Closer Look at Cooling the Server

A server is a piece of IT hardware that provides services for computers on a network. There can be servers for emails, the internet, gaming and processing of files amongst many others. There are two main types of servers, the traditional server containing all the components needed for its operation, and a blade server that is stripped back and contains the minimum components. A traditional computer server contains IT components such as motherboards, memory and CPUs. Alongside this, they also house support equipment such as ventilation fans and a power supply. Within the data centre, servers are housed in a simple rack that provides a framework for securing several servers stacked on top of one another. A typical server is 44.5 mm deep which is known as 1 U. Typically a server rack will house 42 servers, this being known as a 42 U rack.

An alternative to the traditional server is a blade server. Blade servers are a more compact form of server, whereby components such as cooling, power supply and ventilation are all removed from the inner workings of the unit. These ancillary components are instead located in the rack around the servers providing a service to all the units within the rack. An advantage of using a blade server is that it offers the capability of packing more processing power into a smaller footprint within the data centre. Blade servers do however have their disadvantages. With a higher heat load per server and servers being tied to a specific rack, it is not possible to distribute their heat load more evenly throughout the data centre, as can be done with more traditional servers. This can lead to areas of higher temperatures if the distribution of cool air is even across the racks, or the requirement for extra provision of cool air to the areas with blade server racks. Blade systems are also more complex and require a more specialised installation and operations regime, introducing higher costs for the data centre. They only tend to be economical when a number of units are used, as it is not economical to have a rack of supporting equipment for only a few installed servers. Noise levels are also becoming more of an issue in data centres due to higher density equipment such as blade servers, posing a greater risk of approaching regulated occupational noise limits (ASHRAE, 2009). Every computer server consumes electrical energy, almost all of which is converted to energy in the form of heat. For this reason, servers are fitted internally with small fans to aid in removing the heat from important components. With a typical diameter of less than 150 mm, they operate at a speed of several thousand rpm (Greenblatt et al., 2012). Two categories of axial fans are available, constant and variable speed, the latter of which may either vary in speed continuously or in discrete steps over a varying temperature profile (Bean and Moss, [no date]). Servers tend to use the latter of these two, with the fans responding to the external temperature and the temperature of the internal components. Figure 2.14 is an image of the layout of a typical server including the internal fans. The fans are typically positioned near the CPU and memory as they produce the most heat. The effect of different fan positioning within a server is considered in 2.6.2.

The following text examines the research already available in the literature on internal server fans, with a focus on energy use. This provides a background for the experimental work in Chapter 7.

Figure 2.14 An image showing the inside of a typical server, with the internal fans visible and the CPUs located underneath heat sinks (Ibrahim et al., 2012)

2.6.1 Pressure Loss Across the Server

Server fans provide the means to draw air over the components whilst the central air distribution infrastructure provides the cool air, as mentioned previously. However, as the supplied air passes through the server there is a loss of pressure due to the friction from the components. Therefore, rather than purely cooling the server, the fans also take on the role of recovering the pressure drop (Holahan and Elison, 2007, Smith, 2009). They are also responsible for ensuring that the air is distributed correctly across the server.

The server itself is a very compact unit full of electronic components such as microprocessors, memory and power units. These components cause the pressure drop profile across the server to be very complex, with some areas within the unit being densely packed with components and others more open to air flow. Heat sinks are also provided over the heat generating components, and these contribute the highest pressure loss within the server. Vogel et al found through experimental methods that increasing the fins per inch in a heat sink from 8 to 15 increased the pressure loss by 90% (Vogel et al., 2010). Work by De Lorenzo found a pressure drop of 95 Pa across a processor heat sink within a high density server (at an air flow rate of 5.7 m/s). Further work by the same author uses a pressure drop of 150 Pa across an entire server including the processor heat sink along with the chassis losses and losses from other parts (De Lorenzo, 2002). It is important that the server fans overcome this pressure drop otherwise there will be reduced energy efficiency in the cooling system.

2.6.2 Air Distribution within the Server

De Lorenzo demonstrated through the design of a high density server that it is important not just to consider the pressure and flow in the server, but also the air distribution (De Lorenzo, 2002). Demonstrating that sufficient cooling cannot be achieved by oversupplying the air flow requirements, the authors showed that it is important to deliver it to the correct areas (i.e. areas where significant quantities of heat is generated). It is for this reason that fans are positioned at strategic points within the server.

Certain components within the server, such as the CPU and memory, have the highest requirement for cool air as they produce most of the heat. As such, CPUs are attached (with thermal paste) to a heat sink which is designed to conduct heat away from the processor. Air flowing over the heat sink is then able to convect heat away from the CPU better than just having air flow directly over the CPU. It is possible that closer examination of fan positioning can reveal potential energy savings for the server. Energy savings made in the fans will potentially translate into a significant improvement in overall data centre efficiency. Cooling performance and hence energy savings could be increased by ensuring correct positioning of the fans or by improving the flow of air to the fans.

Grimes and Davies tested a server fan to compare the energy used when it is positioned at the inlet and the outlet to the electronic device being cooled (Grimes and Davies, 2002). Particle Image Velocimetry (PIV) was used to assess the airflow through a contained space over a printed circuit board. Their experiments showed that when the fan is positioned at the outlet of the electronic device, cooling was restricted to the boundary layer within the device. Uniform cooling existed across the circuit board, with higher temperatures in the centre and lower temperatures towards the edge, within the boundary layer. Also contributing to the rise in temperature is the fact that the temperature of the air increases as it flows through, causing higher temperatures to be experienced by the downstream components. With the fan located at the inlet on the other hand, significant mixing occurs and the flow is not as uniform as when the fan is located at the outlet. Less uniform flow and more mixing means that the temperature of the circuit board is not just reduced within the boundary layer, but more evenly across the component. Arranging equipment like this also showed that when the fan is at the inlet, the temperature reduction seems to be independent of the air flow rate across the circuit board (Grimes and Davies, 2002).This may be due to the swirl induced by the fan unit.

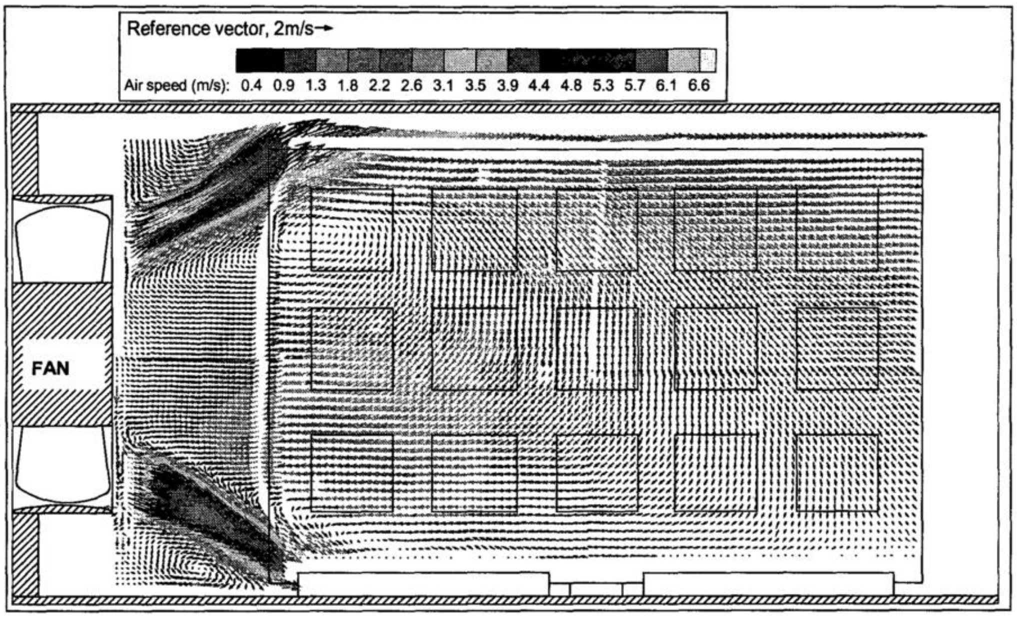


Figure 2.15 PIV measurement showing the flow through the system with the fan mounted at the inlet (Grimes and Davies, 2002)

For the setup in these experiments, the authors found that the best temperature control was achieved by locating the axial fan at the inlet to the electronic device (Grimes and Davies, 2002). Unfortunately, the results in these experiments were not repeated within a server due to the complex air flow caused by objects restricting the flow of air. Even with the fan located at the outlet of the server, it is unlikely that there is uniform flow distribution present due to these obstacles. This may mean that a fan located at the inlet of a server is not as efficient as one located at the outlet in the case of a server. Whilst this study provides evidence that the positioning of fans has been considered, the tests have been done over flat heated plates and do not incorporate the more complex internal geometry of a computer server.

In terms of optimizing the air flow through the server, Vogel et al produced a mock server unit in order to test out the viability of flow optimizers with the aim of reducing energy use. With several sets of fans within one server, the downstream fans tend to notice a lower efficiency as they are taking in air which has come straight from the outlet of the upstream fans (Vogel et al., 2010). This factor was also discussed by Jilesen et al (Jilesen et al., 2006). Jilesen et al recognised the need for increased cooling provision within servers due to the higher levels of processing power that are becoming common. Installing extra server fans was recognised as inefficient as installing them in series means the second fan has to deal with swirl in the air flow caused by the first fan. By introducing a flow control module between the two fans, they were able to demonstrate an improvement in the cooling performance (Jilesen et al., 2006). The exact improvement depends upon the components within the server and the heat sink that is installed. The work done by Jilesen et al helps to highlight the importance of considering the layout of the server when considering performance and air flow. It is not just the server components that affect the air flow, but also any fans that are placed in series. Through conducting a similar experiment, Vogel et al realised a 20% reduction in fan power by placing a flow optimizer between the upstream and downstream fans, which was aimed at reducing disturbance in the air entering the downstream fan (Vogel et al., 2010).

2.7 Performance and Power Consumption of Server Fans

Although server fans may be a very small component of the overall data centre, there are multiple units per server and their energy consumption quickly adds up. Server fans have been proven to consume a significant proportion of the power supplied to the server itself (Smith, 2009, Vogel et al., 2010). Whilst the exact proportion will depend upon the server inlet air temperature, it can actually be as high as 15% as found by Vogel et al (Vogel et al., 2010). Power use for server fans tends to fluctuate depending upon the ambient temperature around the server, as they try to maintain suitable conditions for the critical components (Bean and Moss, [no date]). Other factors can also affect the performance which will be discussed within this section.

Three relationships govern the operation of fans, known as the fan affinity laws. The first law states that the variation in flow rate produced by the fan is directly proportional to the changes in fan speed, as follows;

$$\frac{\dot{V}_2}{\dot{V}_1} = \frac{RPM_2}{RPM_1} \tag{2.3}$$

The second law describes the relationship between the static pressure and fan speed, being that the static pressure varies with the square of fan speed;

$$\frac{P_2}{P_1} = \left(\frac{RPM_2}{RPM_1}\right)^2 \tag{2.4}$$

Finally, the third law states that the power consumed by the fan in order to do work is proportional to the fan speed cubed;

$$\frac{W_2}{W_1} = \left(\frac{RPM_2}{RPM_1}\right)^3 \tag{2.5}$$

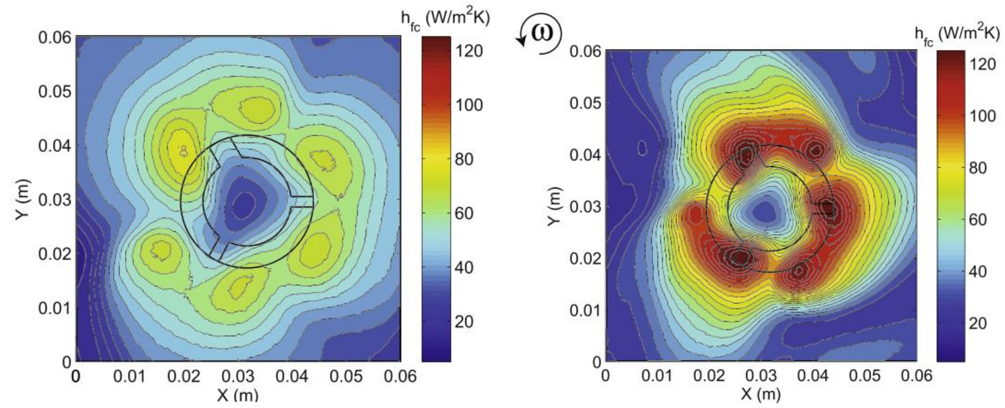
These laws mean that any reductions in fan speed can lead to a substantial reduction in power use due to the cubed law relationship.

The way the fan is powered also affects the overall consumption within the server. A fan can either be 2 or 3 phase and there can be a difference between the power consumption of the two types. Smith found that there is in fact a significant difference in the energy consumption depending on the type of device used. Tests showed an energy saving of 428 kW-hr over 5 years when using a 3 phase rather than a 2 phase fan (Smith, 2009). There is also a noticeable difference in energy consumption when comparing a three phase inverter driven fan to a three phase sine wave fan. The use of an inverter causes 14% more power to be used to deliver the same performance in terms of air pressure and flow (Smith, 2009). This higher energy use is attributed to losses in the inverter, deterioration of motor efficiency and PWM (Pulse Width Modulation) losses[9].

Relating back to correct positioning of the fans within the servers, Stafford et al closely examined the heat transfer properties depending on how close a fan was to the component it was designed to cool (Stafford et al., 2010). The authors performed a study on a 24.6 mm diameter axial fan and its heat transfer effects on a flat heated plate. Whilst this was a miniature axial fan and therefore much smaller than those used in servers, the effect of a change in distance will be the same. During the tests carried out by the authors, the fan's rotational speed was altered along with the distance between the fan and heat plate to investigate what effect this had on the heat transfer.

The study found that the motor supports for the fan had the biggest effect on the heat transfer, causing variations in the heat transfer coefficients. The location of areas of better heat transfer were related to the distance between the fan and the flat surface, and that for higher Reynolds numbers the location of these areas was independent of fan rotational speed.

[9] PWM is a more efficient technique for powering a fan, producing a pulsing DC current that the fan realises as an average voltage



(a) (b)

Figure 2.16 Heat transfer coefficient for a fan speed of 9000 rpm and (a) a fan to plate distance of 15mm, (b) a fan to plate distance of 5mm. contour level: 5 W/m² K (Stafford et al., 2010)

Figure 2.16 illustrates the difference in heat transfer coefficient between a fan which operates at the same speed but at 15 mm or 5 mm away from the heat plate. It demonstrates a significant difference in the amount of heat that is removed from the object by only being 10 mm closer to it. Whilst the authors of this work did not examine the affect that this had on the energy consumption of the fan, it can be approximated that the increased change in pressure across the fan would lead to a higher energy consumption when the fan is situated closer to the heat source. A more in depth study into the energy use of these different positions may demonstrate an optimum balance between energy use and maximising heat transfer through positioning of the fan.

Whilst the fan used in the work by Stafford et al was much smaller than a normal server fan, it still serves to demonstrate that careful positioning of fans within a server is required if their performance is to be optimized and energy consumption decreased.

The fact that server fans need to be so small to fit within the unit is also a source of inefficiency, as larger fans are more efficient at moving air. Holahan et al (2007) measured the energy efficiency of 37 different small fans of varying diameter, the results of which can be seen in Figure 2.17. It demonstrates that fans of a larger diameter exhibit better energy efficiency than those with a smaller diameter. If it were possible to increase the size of the internal server fans, then efficiency improvements could be made. However, server fans are restricted in size by the constraints of the server unit which is of fixed proportions.

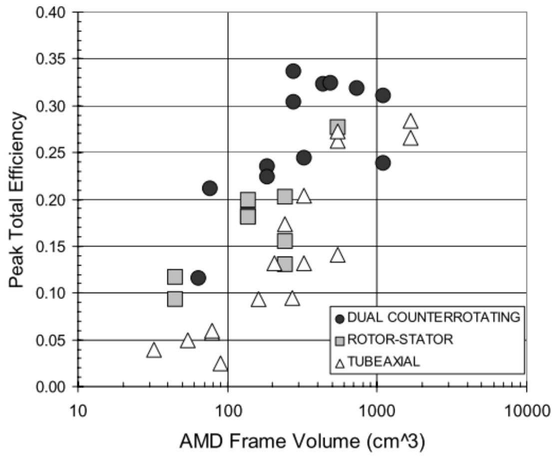


Figure 2.17 Fan peak total efficiency (Holahan and Elison, 2007)

2.7.1 Measuring Fan Performance and Efficiency

Energy efficiency and performance are two separate properties of the fan. Energy efficiency is defined as the useful output power (W_{out}), or the power delivered by the fan (related to the static pressure produced), divided by the total power input (W_{in}) (McQuiston, 2005);

$$\eta = \frac{W_{out}}{W_{in}} \tag{2.6}$$

The performance of a fan on the other hand is related to the air flow and cooling that it provides, it depends on how well it performs its function. Fan performance is normally quantified in terms of the static pressure that it produces within a system. Static pressure is a function of the volumetric flow rate ($\dot{V}$), the input power to the fan (W_{in}) and the rotational speed (ω), along with the properties of the air (density (ρ) and pressure (P_0)) and the physical dimensions of the fan (rotor radius (R) and rotor blade chord length (c)).

$$\Delta P = P_p - P_0 = f_1(\dot{V}, W_{in}, R, c, \omega, \mu, \rho) \tag{2.7}$$

The static pressure is the pressure developed by the fan and hence is equal to the difference between the surrounding air pressure (P_{air}) and the fan pressure (P_f) (Greenblatt et al., 2012). Figure 2.18 illustrates a typical fan curve for an axial fan. The operation point for the fan is at the intersection between the fan curve and system curve, the fan should be designed to operate at this point as this is the most efficient point of operation for that particular system and fan.

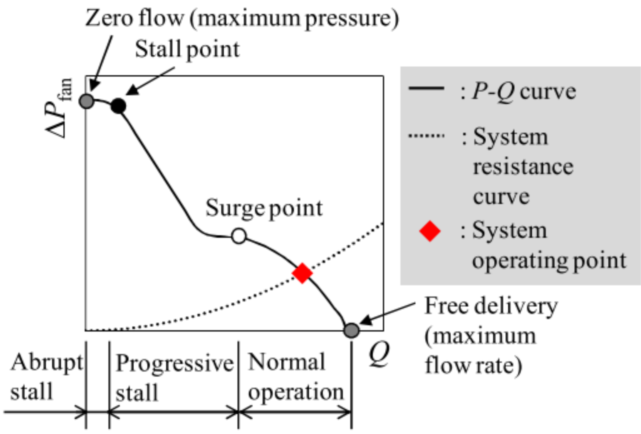


Figure 2.18 Axial fan curve with a typical system resistance curve (Fukue et al., 2011b)

2.7.2 Producing a Fan Curve

Fans curves are assessed by using an air flow bench manufactured to AMCA 210 standards. A flow bench is a sealed container to which a fan or enclosure can be attached to either measure a fan curve or system curve. The flow bench contains an orifice or nozzle through which the air flows. By measuring the pressure drop across the orifice or nozzle, it is possible to calculate the flow rate of air through the flow bench and hence the Device Under Test (DUT). This phenomenon is based upon the Bernoulli equation which states that the difference in pressure between any two points along a pipe will be a function of the difference in elevation (z) and velocity ($\bar{v}$) (hence the flow rate);

$$P_1 - P_2 = \Delta P = \frac{1}{2} P_{air}(\bar{v}_2^2 - \bar{v}_1^2) + \rho_{air} g(z_2 - z_1) \qquad (2.8)$$

By applying Bernoulli's equation across an orifice within a pipe, it is possible to calculate the flow rate through the orifice (see Figure 2.19 for an illustration).

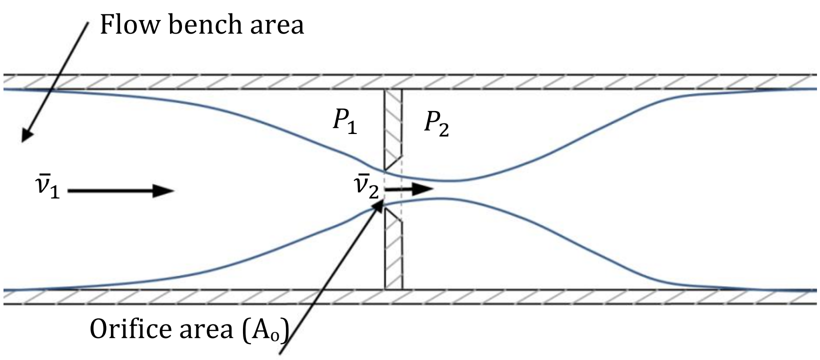


Figure 2.19 Flow through an orifice

Given that there is no change in elevation from one end of the flow bench to the other;

$$\Delta P = \frac{1}{2}\rho(\bar{v}_2^2 - \bar{v}_1^2) \tag{2.9}$$

And given that the flow rate ($\dot{V}$) is a product of the velocity (v) and area (A) of the pipe or orifice where (D_1) is the diameter of the flow bench and (D_2) is the diameter of the flow at its thinnest (a point just beyond the orifice);

$$\Delta P = \frac{1}{2}\rho\frac{4^2\dot{V}^2}{\pi^2}\left(1 - \frac{D_2^4}{D_1^4}\right) \tag{2.10}$$

Simplifying and substituting D_2/D_1 for β;

$$\dot{V} = \frac{\pi D_2^2}{4}\sqrt{\frac{2\Delta P}{\rho(1-\beta^4)}} \tag{2.11}$$

Given that D_2 is unknown, it can be replaced with d_o and a compression factor (C_d), with an expansion factor (ε) being included to account for the properties of the gas flowing through the bench. Therefore;

$$\dot{V} = C_d\varepsilon A_o\sqrt{\frac{2\Delta P}{\rho(1-\beta^4)}} \tag{2.12}$$

Where A_o is the open area of the orifice equal to $\pi d_o^2/4$, where d_o is the diameter of the orifice. Equation 2.12 demonstrates that through the measurement of the change in pressure across the orifice it is possible to find the flow rate through the flow bench and hence the DUT. The values of (C_d) and (ε) need to be calculated as they take into account the compressibility and expansibility of the gas due to the value of (D_2) being unknown. This process requires iteration, as explained in the methodology laid out by Recktenwald (Recktenwald, 2006). This methodology is used for the experimental work in Chapter 7.

Recktenwald also describes in detail the setup of an air flow bench designed to test the performance of fans and the system curve for a specific device or system. The author follows the ASME 210-99 (ANSI/AMCA, 1999) standard by using air flow nozzles (an orifice plate can also be used) to determine the air flow through the flow bench, as illustrated in Figure 2.20. By measuring the pressure drop across the nozzles or orifice it is possible to use Equation 2.12 to calculate the flow rate. Additional measurements such as the static pressure, atmospheric pressure, ambient and internal temperature are all also required as part of the calculation.

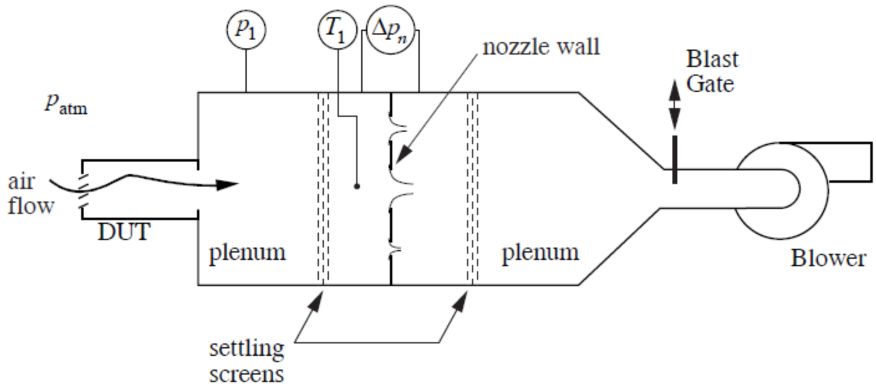


Figure 2.20 Flow bench for measuring a fan or system curve for a Device Under Test (DUT) (Recktenwald, 2006)

Flow bench systems are useful to test the performance of a fan when attempting to make improvements. They are also useful for verifying results that have been simulated, such as those obtained through Computational Fluid Dynamics (CFD, a numerical based method that allows complex fluid flows to be modelled). Liu et al verify a new CFD model for fan testing by using a flow bench fitted with nozzles to measure the performance of a 38 mm x 38 mm fan (Liu et al., 2010). Han et al (Han and Joshi, 2012) use a manufacturer constructed air flow bench which conforms to the ASME standard, using air flow nozzles. Ibrahim et al (Ibrahim et al., 2012) also use a pre manufactured flow bench to test the system curve across a server. Fukue et al (Fukue et al., 2011a) on the other hand conduct tests with a non-ASME standard flow bench and use an orifice plate as the method for measuring air flow.

2.7.3 In Situ versus Manufacturers Stated Performance

Each fan is supplied with a manufacturer's data sheet which should contain details of the fan curve to aid in the design of a system. This fan curve will have been measured and plotted by the manufacturer after testing the fan under laboratory conditions and to a set standard (such as AMCA). However once placed in situ, there will be a difference between the manufacturers fan curve and the actual fan curve that is experienced. Fans within servers are fitted into small spaces and are surrounded by a multitude of components. Being within an enclosed space and having so many objects in close proximity will affect the performance of the fan. Also, once the server is placed within the rack, there are additional obstructions to the air flow that will have an effect (Khankari, 2009).

By testing the effect of enclosure and inlet area on a small axial fan intended for use in servers, Fukue et al found that the performance of a fan differed by about 10% when the fan is placed in an enclosure (Fukue et al., 2011b). This difference in the two curves leads to a difference in performance and hence energy consumption. The work by Fukue et al highlights the need to make in situ tests when considering the performance of a server fan, rather than relying on manufacturer's data. Their results show that the enclosure around the fan may change the operating point of the fan within the server.

Liu and Liu propose a mathematical method through which the in situ fan curve can be predicted given the manufacturers fan curve and one in situ measurement. Following an in situ measurement of the fan curve for an Air Handling Unit (AHU), it can be seen in Figure 2.21 that the authors found it differed from the manufacturers curve (Liu and Liu, 2012). It is therefore important when investigating the performance of a server fan that in situ measurements or a method such as Liu and Liu's is used rather than relying on the manufacturers curve.

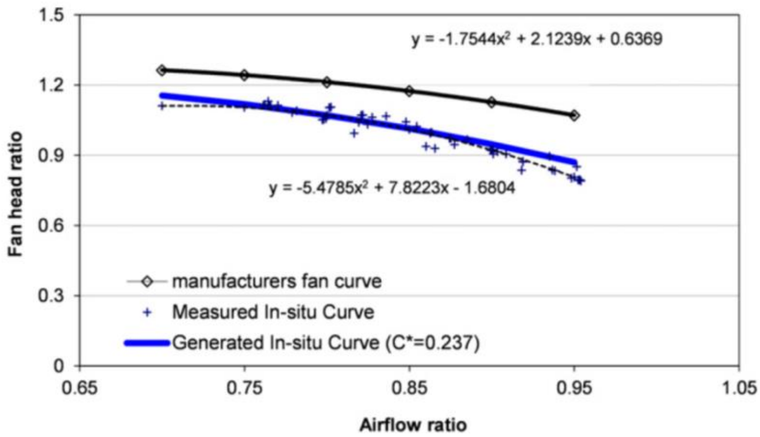


Figure 2.21 Comparison between manufacturers, measured and predicted fan curves for an AHU fan (Liu and Liu, 2012)

With there being multiple fans within each server, the performance that they provide is a combination of each individual fan which differs depending on whether they are positioned in series or parallel. Holahan et al present the theory of multiple fans acting in unison, and propose a model for assessing the performance of a group of fans together which is based upon the fan laws (Holahan and Elison, 2007). Acoustic properties of fans were the focus of this work; with the authors investigating whether it is better (i.e. quieter) to have fewer larger fans rather than more, smaller fans. Whilst they did find that acoustic properties were better switching to larger fans, they did not consider the energy use properties of doing this.

It should be noted that measuring the in situ curve of a server fan is very difficult. Several authors have attempted to find the effects of certain components of the server on the fan, but there are no studies available that take into account the server as a whole unit. For example as discussed earlier, Grimes and Davies (2002) looked at the effect of the positioning of the fan relative to an electronic device and Fukue et al (2011) considered the effect of the enclosure on the fan. Difficulty with assessing the true in situ performance of a server fan comes into play with the technological aspect of this thesis in Chapter 7.

2.7.4 Improving Fan Performance and Energy Efficiency

Within the literature there have been many studies attempting to improve the performance of small axial fans (Smith, 2009, Grimes and Davies, 2002, Stafford et al., 2010, Greenblatt et al., 2012), some with reference to their use within servers. Improving the air distribution across servers has also been considered by some authors (De Lorenzo, 2002, Kumar et al., 2011). In terms of the fans, the following considerations must be taken into account when optimizing them according to Smith (Smith, 2009):

- *The electromechanical conversion* – optimising the design of the electronic drive and motor,
- *The conversion of shaft power to air flow and pressure* – optimizing the aerodynamics of the fan and making sure they are selected properly, and
- *The inlet and outlet of the surrounding equipment* – optimizing the fan inlet and outlet in terms of mechanical configuration.

By optimizing the design of an electronic cooling fan using the three pointers above, Smith showed that it was possible to make power savings of 5 – 15%. This would become significant considering that there are a number of these fans in each server unit, and potentially thousands of servers in a data centre.

Another source of decreased performance in server fans is due to their small size, meaning that they tend to produce air flow with a low Reynolds number. This low Reynolds number means that the fans tend to stall when they are only moderately loaded, causing a drop in performance. In an attempt to address this, Greenblatt *et al* utilise periodic perturbations in order to control the separation of flow from the blades (Greenblatt et al., 2012). By introducing acoustic perturbations, the authors were able to increase the air pressure by 30 – 40%, hence increasing the performance of the fan. Whilst the authors were able to achieve an increased performance, the method used to produce the acoustic perturbations was itself not particularly energy efficient. This is recognised in the authors' work who recommend alternative options for producing them. Figure 2.22 shows the increase in energy efficiency for the fans when acoustic perturbations are used.

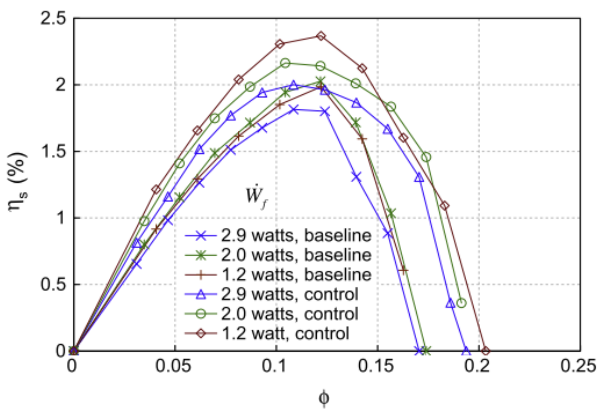


Figure 2.22 Fan efficiency at different power levels for both baseline (fan performance without acoustic perturbations) and controlled cases (fan performance with acoustic perturbations), with a dimensionless fan flow rate on the x axis and the fan static efficiency on the y axis. (Greenblatt et al., 2012)

2.7.5 The Effect of Increased Inlet Temperatures

Reducing energy consumption in the air cooling system is important in order to reduce overall data centre consumption. One way to achieve this is to increase the set point for the cooling equipment (hence increasing the inlet temperature to the servers), as discussed in section 2.5.1. However, this may have a negative effect on the server energy consumption due to the energy consumption of the server fans. As the inlet temperature to the servers increases, the server fans react to this and increase their speed to cope, leading to higher energy consumption. There are several papers which identify or investigate this possible side effect (Bean and Moss, [no date], Muroya et al., 2010, Ahuja, 2012), with the consensus that there is a range of inlet temperatures within which energy is saved overall. However once this range is exceeded, the increase in energy consumed by the servers outweighs the energy saved in the cooling system. Part of the problem behind this is the lack of control that a user has over the speed of the internal server fans. The fan units are controlled by an algorithm written into the software within the server, which is difficult to alter or interfere with (this is a factor which plays into the results of experimental work in Chapter 7).

Muroya et al test a series of servers under experimental conditions to see if there is any difference in power usage when the inlet temperature is increased (Muroya et al., 2010). They find that whilst overall power can be reduced by increasing inlet temperature, this is only applicable over a small temperature range. The authors find that there is in fact an optimal temperature range which depends upon the situation, the data centre and the IT equipment. Figure 2.23 shows the increase in server power consumption as the inlet temperature increases, this is noticed for both conventional and blade servers. Blade servers resulted in a much higher increase in energy usage, although this may be due to the experimental set up. One blade server was placed in a blade enclosure which had one fan included, meaning that there was an overprovision of air flow rate. Figure 2.24 shows the total power consumption for a case model data centre based upon the measurements made on the servers. It shows how there is a maximum temperature at which savings can be made before an increase in power is noted. This is due to the increasing speed of the internal server fans.

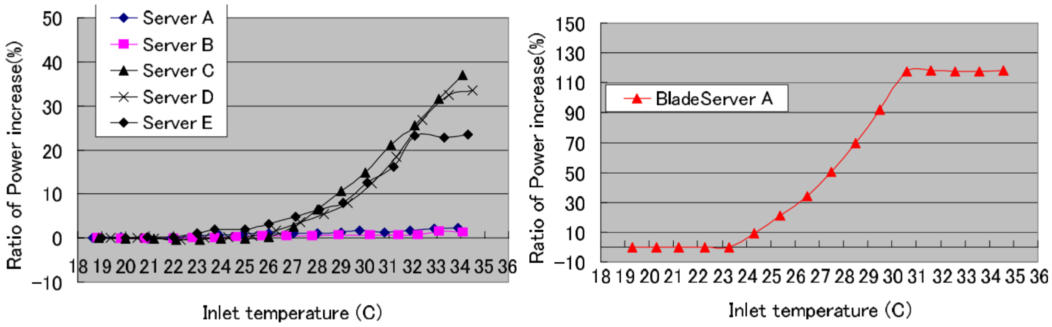


Figure 2.23 Power use percentage increase versus temperature for a conventional server and a blade server (Muroya et al., 2010)

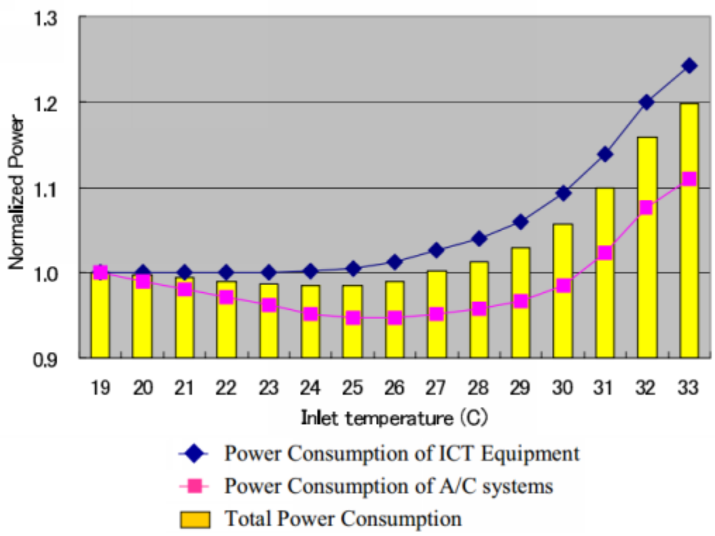


Figure 2.24 Normalised power over a temperature range for a case model data centre (Muroya et al., 2010)

Other work on this problem by Bean et al also tested the hypothesis that there is a decrease in energy consumption when the inlet temperature to the server is increased (Bean and Moss, [no date]). For the several scenarios that were tested, it was found that the highest energy savings were to be made at a temperature of 24 – 27°C. This shows a similar trend to the work done by Muroya et al.

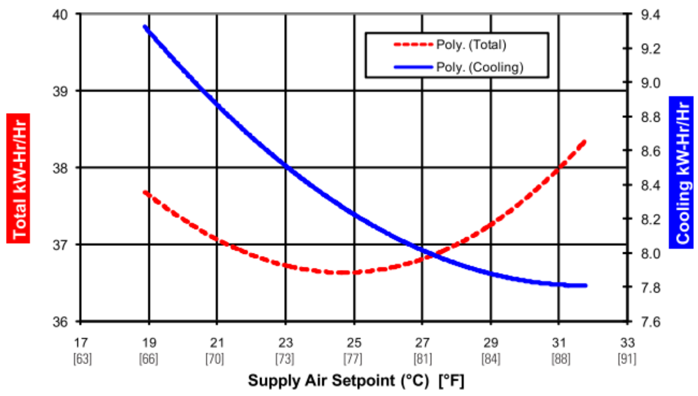


Figure 2.25 Polynomial curves for energy use over an increasing inlet temperature (Bean and Moss, [no date])

2.8 Power and Resiliency in the Data Centre

Power supply, and the reliability of it, is arguably the most important component of a data centre. Data centres have various levels of resiliency built into them in order to ensure that power supply is not lost to the IT equipment. Any complete loss of power can bring enormous consequences in the form of lost money, trade, business and custom. This can have devastating effects for businesses, government and also the general public. A survey by the Ponemon Institute in 2013 (of 584 U.S. organisations) for example found that 71% of respondents rely upon the data centre to be able to generate revenue and to maintain their businesses (Ponemon Institute, 2013). For this reason, it is important that data centres are resilient enough not to lose connection to a power supply. Each data centre will have one or more generators, normally fuelled by diesel. This generator will be automatically switched on in the event of a loss of power from the mains. However, these large generators will normally take at least 10-15 seconds to come up to full power, and so an interim measure is needed to ensure continuous power supply. UPS units along with a bank of batteries are used for this purpose. A UPS unit will supply power to the data centre via a set of batteries for the few seconds that it takes the generator to power up.

Whilst the generator and UPS units will provide resiliency in terms of power supply, data centres also need a system whereby the mechanical failure of equipment is accounted for. Hence facilities will tend to use various levels of redundancy in their infrastructure, especially in the cooling systems. This allows the data centre some level of resiliency should a particular unit fail, it also allows for concurrent maintainability. There are various levels of resiliency and data centres can actually receive certification under the Tier Classification scheme should they wish (a scheme by the Uptime Institute certifying a data centres' level of resiliency).

The problem with resiliency is that it can decrease the overall energy efficiency. Take a typical air handling unit for example, the AHUs operate at higher efficiencies if they are loaded to their full capacity. In a system where two AHUs are provided, unless they operate variable speed drives or operate in a run and standby configuration, then lower efficiencies tend to be noted due to them sharing the load. Obviously resiliency is a key priority for companies, however energy efficiency will often suffer because of this.

2.9 Conclusions and Research Themes

The demand for digital infrastructure services is growing worldwide and data centres are a vital part of the system, providing support to all sectors. The internet and access to data has become an integral part of everyday lives and is a key service for the modern digital world. Data centres have developed rapidly and grown to become complex technological ecosystems which require whole systems of supporting infrastructure. In a world that is striving to reduce carbon emissions, data centres and their services provide a means for other sectors to reduce their emissions by virtualising physical services. Despite the opportunities for reducing emissions that ICT can provide to other sectors, the data centre sector itself is facing rising energy consumption levels and associated increasing levels of carbon emissions.

Estimates put the worldwide electricity consumption by data centres at around 1.5% and with the growth of data[10] reaching 45 ZB in 2020 (Avasant, c2014), this consumption is only likely to grow further. Increasing demand for access to services and data is driving up energy consumption in data centres, which is of course linked to carbon emissions through the use of electricity. If government and worldwide targets for reducing emissions are to be met, then this is a key sector to focus on, especially given the predicted growth in demand.

With a complex combination of electrical, IT and mechanical systems, data centres require a carefully controlled environment to maintain their services and avoid failures. Electrical infrastructure provides the means for the data centre to remain operational should the mains supply be interrupted, and other redundant equipment

---

[10] Both unstructured (transactions, emails and social media data for example) and multi-structured (web applications for example that facilitate interactions between people of machines).

across the infrastructure provides peace of mind should there be any failures. Precision mechanical infrastructure maintains the environment around the IT equipment by providing and distributing cold air to the components that need it. Traditionally, air conditioning systems with compressors have been used although there has been a shift towards free cooling in recent years. Newer forms of cooling such as liquid submersion and liquid-to-chip systems are also being used although they are not yet very common. These newer systems may offer significant energy savings compared to the traditional air conditioning system.

In regards to the main component of the data centre, the servers have within them their own small air distribution system. Each server has within it a group of small axial fans designed to remove heat from the critical components. Whilst doing this, they also aid in recovering pressure losses and make sure that cool air is distributed evenly across the internal components that need it most. In situ performance of these small fans can differ quite a lot from manufacturers stated performance, meaning that in situ measurements are crucial when examining efficiency. Performance can be quantified by producing a fan curve, the data for which can be gathered by using a flow bench constructed to certain standards.

If overall data centre efficiency is to be improved, then fan performance and efficiency is a key area that needs to be considered. Server energy consumption is also mentioned within the interviews results in Chapter 5 as an important consideration for the industry going into the future. Existing research has been conducted into increasing the performance and efficiency of server fans, with various successes and methods. Some authors use experimental work with real fans whereas others consider the theory of fan curves and the fan laws in order to assess performance. CFD is also popular for studying the performance of server fans. However, there are a few studies available that consider the positioning of the fan within the server and the effect this can have on the overall performance. One author (Holahan and Elison, 2007) also referred to efficiency opportunities by switching to fewer, larger fans although this study focussed on acoustic properties and did not go into a lot of detail regarding energy performance.

With the literature, there is limited work looking into more detail about the energy consumption of server fans. Blade servers or the newer liquid cooling technologies do not require small fans to be located within the server; further studies are needed to be able to properly consider the benefits of these technologies. Further work also needs to be done to increase the overall understanding of the benefits of liquid cooling and the removal of server fans (also whether the removal of these fans is indeed possible).

3 Energy Efficiency in Data Centres

3.1 Introduction

Following on from the background to data centres given in Chapter 2, this chapter serves to introduce the concept of energy efficiency in data centres. Chapter 2 discusses the problem of worldwide energy consumption by these facilities, the literature surrounding methods used to reduce this consumption will now be reviewed. Energy efficiency is first defined with a focus on how it applies to the design and operation of data centres. Methods employed in order to achieve higher energy efficiency are then discussed covering technology, behavioural changes and policy implementation. Current metrics used to assess efficiency within the data centre are then examined followed by a review of the literature surrounding barriers to energy efficiency, along with how these may apply specifically to the data centre sector.

3.2 Defining Energy Efficiency

Before energy efficiency in regards to data centres can be discussed in any depth, it is important to introduce the concept of efficiency and define the terms involved. The idea of the efficient use of energy arises due to the fact that processes lose energy thus leading to a lower energy output compared to the initial input. Through the whole process from the use of primary energy to the energy output, losses are incurred due to generation, transportation, storage and finally usage (see Figure 3.1).

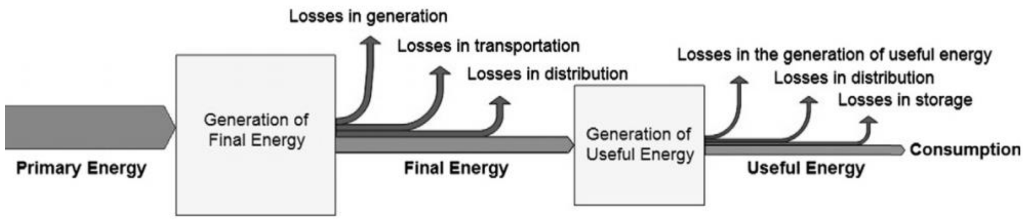


Figure 3.1 Energy losses from generation to use (dos Santos et al., 2013)

There are several definitions of energy efficiency, depending on the situation to which it is being applied. The procedure for measuring energy efficiency or the energy conversion efficiency (η) (Pérez-Lombard et al., 2013, dos Santos et al., 2013) is as follows;

$$\eta = \frac{useful\ energy\ output}{energy\ input} \tag{3.1}$$

Whilst this is the main way of calculating the energy efficiency of a system, the *useful energy output* can be replaced with *useful output* (Patterson, 1996) which is useful for assessing the efficiency of a process where energy is not the focus of study.

An alternative indicator of energy efficiency is to consider what is usually referred to as the energy intensity (*EI*) (Pérez-Lombard et al., 2013);

$$EI = \frac{energy\ input}{service\ output} \tag{3.2}$$

The *EI* is the ratio of resource (energy) input to the service output. Whilst this may seem straight forward, it is often very difficult to assess the service output (Patterson, 1996, Pérez-Lombard et al., 2013). Useful output itself does not have to be in units of energy, it could be production numbers, economic value, or in the case of data centres, ICT services. However when this measure is applied to data centres, significant difficulties arise when attempting to quantify the service output. Data centres are designed and built to provide ICT services, services that are purely virtual. It is therefore highly difficult to assess the output (especially the relevant or useful output), which can lead to difficulties when attempting to assess overall efficiency in the data centre.

Patterson lists four main groups of energy efficiency indicators which are useful for finer definitions; thermodynamic, physical-thermodynamic, economic-thermodynamic and economic (Patterson, 1996). Equation 3.1 is a thermodynamic indicator in that it represents a ratio between energy input and energy output, it is found using measurements of energy use. Physical-thermodynamic indicators however are found using a physical output rather than an energy output. In the case of manufacturing for example this would be the items that are produced, but in the case of data centres the 'physical' output are the ICT services that are provided. This indicator is equivalent to equation 3.2 and whilst it may be more relevant to assessing efficiency, is a very difficult measure to obtain from data centres.

3.2.1 Energy Efficiency Potential

The energy efficiency potential can be defined as the maximum energy efficiency achievable for a certain situation if all efforts were made to increase it. Whilst this may seem straight forward, the International Energy Agency (IEA) have defined several different levels of energy efficiency potential (Energy Information Administration). These different levels are as follows (as summarised by Kreith and Goswami (Kreith and Goswami, 2008)):

- there is a *thermodynamic limit* to the energy efficiency of processes meaning that there is a ***theoretical potential*** for a maximum energy efficiency;
- if all of the most energy efficient technologies that are available at a certain time were utilised then the ***technical potential*** could be realised, this is without economic considerations;
- given a certain set of boundaries such as energy prices, any improvement that can be made within a year is known as the ***market trend potential***;
- the ***economic potential*** is similar to the technical potential, however only the most economically viable and cost effective options are used; and
- only when all of the externalities such as the cost incurred from pollution, global warming and health impacts have been accounted for will the s***ocietal potential*** for energy efficiency be reached.

These definitions highlight the different methods by which energy efficiency can be improved. *Technological potential* is important, especially in the data centre industry. Within the data centre there are several different infrastructure systems including, but certainly not limited to, IT equipment, cooling and power. All of these systems rely heavily on equipment and technology, all of which offer technical energy efficiency potential. If the efficiency of all these systems can be improved, then significant savings could be made. However these improvements can only be made up to a theoretical limit, a limit which is defined by the laws of thermodynamics i.e. the *theoretical potential*. Beyond the technology, further improvements need to be made by considering *market trends* and economic factors. The *economic potential* arises where technology can only be introduced if it is cost effective. Certain measures (often policy driven) can be used to make technological changes more economically viable.

All of these potentials need to be met if a high level of energy efficiency is to be realised, however in reality this may not be possible. Section 3.11 examines this in more detail by considering the barriers against energy efficiency efforts.

3.3 Exergetic Efficiency

Exergy is another theoretical approach that can be used to quantify energy efficiency, being a measure of a systems potential to do work on the external environment or another system. This work must be done in a reversible process, i.e. it must be useful. If energy is lost via an irreversible process then it is said that the exergy is lost or destroyed and the system is therefore less efficient. A system which runs on energy in the form of electricity for example has high exergetic potential as almost all of the energy is transformed into work either in the form of mechanical movement or heat. Based upon the second, rather than first law of thermodynamics, exergy relates to a systems ability to do work hence describing ideal work or the availability of energy. In its simplest form the exergy, or the amount of work (W_{max}) that can be taken from a certain amount of heat (Q), is calculated using the following equation (where T_0 is the fixed temperature of the surroundings);

$$E_x = W_{max} = Q.\left(1 - \frac{T_0}{T}\right) \quad (3.3)$$

The destruction of exergy describes when a system has lost its potential to do work and is therefore another useful measure of the efficient use of energy. If a system has high exergy destruction, then work is being transferred in an irreversible way giving rise to an inefficient system. Exergy analysis is potentially a useful tool in data centre energy use assessment due to its approach of considering the transfer of energy, especially when it comes to considering waste heat production. The use of exergy analysis for data centre is explored further in section 3.5.4.

3.4 Energy Efficiency in Data Centres

As already established, defining energy efficiency as a general concept is an easy task. Conversely, data centres are complex and the outputs are even more so which can lead to difficulties when defining energy efficiency in the data centre. Outputs, or products, produced by a data centre can be represented by the services that they provide i.e. storage of data, network communication and computational processing. Additional to this, a data centre can also be measured on its availability of servers, responsiveness and reliability (U.S. Environmental Protection Agency, 2007). Further discussion will be focussed on measuring efficiency in terms of the output from the IT equipment and the challenges of this in Section 3.5.3.

With a focus on buildings, Cursino dos Santos et al. summarised that there are two environments that need to be assessed for energy efficiency, the macro and the micro (dos Santos et al., 2013). Within the macro environment sits the country as a whole whereas the micro contains buildings, processes and equipment. Whilst the data centre facility sits within the micro environment, it is important to consider the wider macro environment i.e. including the primary source of energy into the assessment.

The complexity of evaluating efficiency within data centres can be realised when it is noted that they cover all three areas within the micro environment. ICT services are critical to the very nature of the data centre, however for these services to work they require equipment, supporting processes and a secure building. If they are to use low carbon energy supplies, the macro scale must also be considered and the energy supply from the country needs to be included.

It is important that energy efficiency is measured across these levels if a full understanding is to be gained of data centre energy efficiency. Being able to measure this accurately and in a suitable manner is important. Discussion around the existing metrics used will be in the following section. This leads into the in depth discussion forming the basis of Chapter 6, where one particular metric is examined in detail. The remainder of the current chapter will focus on the measurement of energy efficiency in data centres, how energy use can be reduced and what barriers may exist preventing further improvements.

3.5 Measuring and Reporting Energy Efficiency

In order to measure and present energy consumption effectively, metrics are needed that are suitable and fit for purpose. A number of metrics have been proposed in the past for measuring efficiency and power consumption within data centres, these are mainly for the micro environment (i.e. on a building and infrastructure level). Most of these metrics are classed as thermodynamic however metrics such as those that consider the ICT services are classed as physical-thermodynamic. This is according to the indicators discussed by Patterson (Patterson, 1996). Key metrics are discussed within this section including energy consumption metrics, thermal management metrics, server efficiency metrics, water or carbon usage metrics and exergy or second law metrics.

Chapter 1 introduced the idea that the data centre is a system within a wider network of digital infrastructure that has several interested external systems or parties including owners, tenants and energy providers. Each of the different groups of metrics discussed in this section have different levels of applicability to these external systems. A metric which measures the efficiency with which air is distributed through the data centre for example only applies to the owner of the facility trying to reduce energy consumption. Such a metric does not notify the energy provider how much energy is being used, or describe to the tenant how fast their information is being processed. It is in this way that metrics can struggle to address the goals of each of these external parties involved in the data centre facility. Energy providers need to know when the peak and lowest demands for the facility will be, tenants need to know that their data is being processed at an acceptable speed within the boundaries of the SLA, and the data centre owner wishes to reduce energy consumption in order to lower bills. These different cross-system goals are difficult to align and no one metric can provide answers to all. This is part of the reason behind the fact that there is a myriad of metrics available, especially when it comes to measuring energy use and efficiency.

3.5.1 Energy Consumption Metrics

The most common form of metric used to assess data centres is one that looks at the energy consumption and efficiency of the operations in a facility, purely a thermodynamic indicator. It is the most useful in terms of making energy and cost savings, and providing an overall analysis of the data centre. The most commonly used energy consumption metric within the industry is the *Power Usage Effectiveness* (PUE).

The Power Usage Effectiveness metric introduced in 2006 (Malone and Belady, 2006) and promoted by the Green Grid (a non-profit organisation of IT professionals) in 2007 (The Green Grid, 2007) has become a de-facto industry standard for reporting the energy efficiency of data centres, being the most commonly used metric (Kant, 2009, Daim et al., 2009, Ruth, 2011). The PUE is useful to present the proportion of power which is actually used to operate the IT equipment with respect to the total power draw of a facility, and is defined as the ratio between the total power usage in the data centre and the power usage by the IT equipment alone (The Green Grid);

$$\text{PUE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}} \qquad (3.4)$$

The PUE is used at the design stage of projects to present a data centres' potential energy efficiency, used post construction to aid in energy costing, and also used for marketing a facility. The Green Grid provide guidelines for this metrics use and state that it should not be used to directly compare one facility to another, also providing guidelines as to how it should be reported. The organisation states that in order for the metric to present a meaningful figure it should be given alongside information regarding how it was recorded. The metric should be reported as $PUE_{a,b}$ where 'a' denotes the location of the energy metering and 'b' denotes the frequency of measurements and the period of time over which they were taken (The Green Grid, 2009). The metering should be taken from the power input to the PDU so that the total power supplied to the data centre can be accounted for. However, when a PUE value is presented it is rarely given alongside these details as recommended by The Green Grid.

Despite the fact that the PUE is commonly used within the industry to assess energy consumption, there are some inherent issues with the metric leading to some criticisms with its use (Kant, 2009). Chapter 6 will assess in more detail the usefulness of this metric.

Whilst the PUE has become widely adopted within the data centre industry, there are many other metrics concerned with energy consumption which have been suggested in the literature and by organizations. One other metric also from The Green Grid is the *Data Centre Infrastructure Efficiency* (DCiE) which is actually the inverse of the PUE, and allows the IT load to be presented as a percentage of the total power draw in the facility. This same metric is discussed by Greenberg et al. but is given the name of *Computer Power Consumption Index*, being defined as the power drawn by the IT equipment divided by the total power required for the IT equipment and supporting infrastructure (Greenberg et al., 2006). This means that it has a maximum value of 1, indicating that 100% of the power input is used to operate the IT equipment, a theoretical maximum as this is not really possible. This metric is easy to read as it clearly shows what percentage of incoming power is used to operate the IT equipment. However, like the PUE it should not be used to directly compare data centres due to differences that there might be in the design of a facility.

Another commonly used metric is *Power Density*, which is measured as the amount of power consumed over a certain area. However, the power density reported depends heavily upon the area that it has been calculated over, and this is not always accurately portrayed. There will be a large difference between the power density measured over the floor area of a whole facility, a data hall, or over a smaller area such as one rack. This can lead to ambiguity over what the actual figure represents, and unless full details are disclosed it can be difficult to use in energy efficiency calculations. Consequently, power density is not a suitable metric to use when comparing different facilities (Sun and Lee, 2006). For one facility in Linköping, Sweden it was found that the power density calculated using the footprint (or area occupied by IT equipment) was 1.15 kW/m^2 whereas over the area of one rack, the power density was 5.0 kW/m^2, clearly a significant difference (Karlsson and Moshfegh, 2005). It would appear that power density is a more suitable metric for assessing the use of space within a data centre, rather than the use of power. Assessing the use of space within a data centre would come into play when considering the economic cost of the physical footprint, both in terms of the servers and the building as a whole.

Mitchell-Jackson et al. reported power density as a key metric but stated that standardisation and further development of metrics is needed (Mitchell-Jackson et al., 2003). The authors suggest that metrics currently lead to confusion due to the way in which they are reported and calculated and that using power density can lead to inaccuracies in results, a lack of definition in the reporting, and also overestimation of the actual power required. Assumption of a power density during the design stages of a facility can provide a result which may differ significantly from the actual power draw of the finished data centre. An assumed power density is often used to calculate an initial PUE when a design engineer is presenting a scheme to their client. Using a power density of 1100 W/m^2 Mitchell Jackson et al calculated an estimated 12.5 MW usage for the data centre under study, which was 14 times the actual amount of power consumed (Mitchell-Jackson et al., 2003).

Sun et al. also developed metrics for assessing energy consumption. After conducting a detailed case study, the authors present energy consumption using a metric termed as the *Relative Energy Effectiveness* for the IT and HVAC (Heating, Ventilation and Air Conditioning) equipment, defined as the IT load over the total facility load and the HVAC load over the IT load respectively (Sun and Lee, 2006), as follows;

$$\text{IT effectiveness} = \frac{\text{IT (kWh)}}{\text{total (kWh)}} \tag{3.5}$$

$$\text{HVAC effectiveness} = \frac{\text{HVAC (kWh)}}{\text{IT (kWh)}} \tag{3.6}$$

The second of these metrics above could actually be classed as the HVAC *in*effectiveness as a higher value indicates a lower efficiency, and equation 3.5 is just the inverse of PUE. Transparency in the language of the metric is important in communicating the efficiency that has been measured, as unclear metrics can lead to incorrect use, reporting, and understanding. Nevertheless, HVAC effectiveness is a useful metric for measuring the amount of energy needed for the cooling system relative to the IT load, and can be used as an indicator for improvements in this sub-system.

Patterson et al propose a newer metric that is similar to the PUE but shifts the focus away from the infrastructure towards the IT systems. This means that the actual computing components (CPUs, memory and storage) within the servers are assessed alongside the supporting infrastructure of the server unit, e.g. the internal server fans (Patterson et al., 2013).

Named the *IT Usage Effectiveness*, the metric is calculated as follows;

$$\text{ITUE} = \frac{\text{Total energy into the IT Equipment}}{\text{Total Energy into the Compute Components}} \tag{3.7}$$

This metric really allows the energy use of the server itself to be considered, something which the PUE cannot do. Given that there has been a lot of focus in recent years on reducing the energy consumption of the supporting infrastructure, it is important that considerations are now being made towards the IT equipment. The same authors also suggest a TUE metric, *Total Usage Effectiveness*, which is the product of ITUE and PUE and gives a better overall picture than the PUE on its own. PUE can often give a false sense of efficiency, as discussed further in Chapter 6. It is possible when comparing two data centres, that the facility with the lowest PUE is not actually the most energy efficient overall. The metric that Patterson et al proposed allows a more meaningful comparison. Table 3.1 (adapted from (Patterson, 2012)) demonstrates the use of this metric and compares it to a PUE result. Using two example data centres, the authors show how a data centre that has a lower PUE and hence looks more efficient on paper actually uses more power to support the computing components than a facility with a higher PUE.

Table 3.1 Example data centre power use and metric results (Patterson, 2012)

	Low Efficiency	*High Efficiency*
	Power Use (MW)	
Total Platform	3.31	2.67
Computing Components	1.98	1.98
Infrastructure	1.99	1.99
Total Site Power	5.3	4.66
	Metric Results	
PUE	1.6	1.74
ITUE	1.67	1.34
TUE	2.67	2.33

The work by Patterson et al clearly shows the need to delve deeper and go beyond simply using the PUE to assess energy use in the data centre, this is reiterated by the findings of this thesis. So many facilities just use the PUE to present their energy use, whereas it is very important to consider the amount of energy required to support the computing components, rather than the server as a whole. Some metrics take this a step further and consider the computing elements of the server; these are discussed in Section 3.5.3.

3.5.2 Thermal Management Metrics

Whilst the metrics discussed in section 3.5.1 allow the energy consumption of the data centre to be assessed, they do not demonstrate the efficiency of the individual infrastructure systems in terms of their ability to deliver the service required. In order to completely assess the efficiency of a data centre, it is not suitable to use the PUE on its own (Dumitru et al., 2011), other systems (the cooling system in particular) also need to be assessed. A significant amount of power in the data centre is consumed by the cooling system and so it is important to be able to measure its efficiency, both in terms of power use and delivery of service. Thermal management metrics are a useful tool to achieve this. The *Coefficient of Performance* (COP) is a common metric used to assess the energy efficiency of a cooling system, illustrating the amount of power required to deliver a certain amount of cooling. It is defined as the following;

$$\mathrm{COP} = \frac{|Q_C|}{W} \tag{3.8}$$

where Q_C is the cooling load dealt with by the system and W represents the power consumed in providing this. A similar metric is used by Mathew et al. who define it as the *Data Centre Cooling Efficiency* (Mathew, 2010). Whilst these two metrics focus on the efficient use of power input into the cooling system (power), it is also important to focus on the use of the other inputs, such as the air used for cooling. Due to this, it is useful to have metrics that assess the efficiency with which air is used to cool the IT equipment in the data centre.

Lu et al. consider the energy consumption of the cooling system within one particular data centre in Finland (Lü et al., 2011). As part of their methodology for assessing the performance of the cooling system, they use a metric termed the *Supply Heat Index* (SHI) as defined in equation 3.9. It is used as an indicator of any mixing of hot and cold air which is occurring in the data hall, this is a metric as first developed by Sharma et al. (Sharma, 2002). The metric has since been used by Fakhim et al. to assess thermal management (Fakhim et al.). A high percentage value indicates that there is a high level of mixing occurring, or that there is an oversupply of cool air to the data hall. This is a useful metric as the cooling system may be efficient in its use of power, but it may be delivering more air to the data hall than is needed to deal with the heat load.

The SHI can be derived through the following;

$$\text{SHI} = \frac{\text{Enthalpy rise due to infiltration in cold aisle}}{\text{Total enthalpy rise at the rack exhaust}} = \frac{\delta \text{Q}}{\text{Q} + \delta \text{Q}} \qquad (3.9)$$

$$Q = mC_{p\,air}\left((T^r_{out}) - (T^r_{in})\right) \qquad (3.10)$$

$$\delta Q = mC_{p\,air}\left((T^r_{in}) - T\right) \qquad (3.11)$$

Where m is the mass flow rate of air through the rack, $C_{p\,air}$ is the specific heat capacity of the air, T_{out} and T_{in} are the average temperatures of the air leaving and entering the rack respectively, and T is the supply air temperature from the cooling system.

Therefore;

$$\text{SHI} = \frac{T_{in} - T}{T_{out} - T} \qquad (3.12)$$

Similar to the derivation above, the metric *Return Heat Index* (RHI) can be calculated which is defined as follows;

$$\text{RHI} = \frac{\text{Total heat extraction from the data hall}}{\text{Total enthalpy rise at the rack exhaust}} = \frac{Q}{Q + \delta Q} \qquad (3.13)$$

It is known that any increase in inlet temperature to the racks will be due to infiltration from the outlet air which needs to be removed from the space hence;

$$\text{SHI} + \text{RHI} = 1 \qquad (3.14)$$

These metrics have been tested by conducting a CFD analysis and they were found to be useful to assess different layouts of server racks (Sharma, 2002). By using the metrics it was possible to show the effect of parameters such as aisle width and ceiling height have on the performance of the data centre.

Another metric known as the *Return Temperature Index* (RTI) was developed by Herrlin (Herrlin, 2007) and demonstrated through a case study by the same author (Herrlin, 2008). This metric enables the data centre owner and/or operator to assess the air-management system in terms of any re-circulation or by-pass air which occurs in the data hall. This provides a good indication of the efficiency or performance of the system which is vital to cooling the servers but consumes a significant proportion of overall energy use in the data centre. The RTI is defined as the following;

$$\text{RTI} = \left[\frac{T_{Return} - T_{Supply}}{\Delta T_{Equip}}\right] 100[\%] \qquad (3.15)$$

The RTI is a very useful metric to incorporate into the energy audit of a data centre as it allows any inefficiency in the air distribution to be assessed. The RTI should return a value of 100%. If the value is above this, then its means that there is recirculation of air leading to hotspots in the data hall, if the value is lower, then there will be air by-passing the servers leading to a waste of energy. A theoretical data centre model was analysed using CFD by Herrlin, and the RTI was calculated (Herrlin, 2008). The author was able to use the metric to assess the airflow in the room and to make adjustments in the system to improve the overall efficiency. One of the problems when assessing a real data centre however is that there will be a different temperature rise across each server due to the variance in workloads.

Previous to the RTI, Herrlin had also put forward a *Rack Cooling Index* (RCI) (Herrlin, 2005) which allows the effectiveness with which the racks are cooled to be measured.

The metric is defined as follows,;

$$\text{RCI} = \left[1 - \frac{\text{Total OverTemp}}{\text{Max Allowable OverTemp}}\right] 100[\%] \qquad (3.16)$$

Total over-temp is defined as the number of servers with an intake temperature that is higher than recommended, and the max allowable over-temp is the total number of intake temperatures that are over the maximum allowable.

This definition is suitable for assessing racks for any temperatures which are supplied with air at the higher end of the temperature range. There is a similar definition which deals with 'under-temperatures' for temperatures at the lower end of the range. An RCI of 100% indicates that all intake temperatures to the racks are less than the recommended temperature used in that particular facility. Any RCI value less than 100% shows that there is at least one temperature measurement which is over the maximum recommended temperature.

Van Gilder et al. put forward a metric termed the *Capture Index* (CI), which is also a cooling system performance metric. The CI was developed to aid in assessing the cooling performance by considering the flow of cool supply and hot return air in the data hall (VanGilder, 2007).

Two variations on the metric were defined, one each for hot aisles and cold aisles;

$$\text{Cold Aisle CI} = \text{'fraction of air ingested by the rack that originates from local cooling resources'} \quad (3.17)$$

$$\text{Hot aisle CI} = \text{'fraction of air exhausted by a rack that is captured by local extracts'} \quad (3.18)$$

These quantities can be measured by conducting CFD analysis and used to see where improvements can be made in the layout of the racks and cooling system. The CI is useful for facility owners and operators to see where air is not being effectively delivered to and taken from the data hall. However, in reality, this may be difficult to measure without significant monitoring within the data hall and so may not be practical for most data centres.

3.5.3 Server and IT Efficiency Metrics

Thermal management metrics consider the infrastructure which supports the IT equipment and power consumption metrics look at the total power usage in a data centre. However if a true efficiency is to be arrived at, then the IT equipment (servers in particular) needs to be looked at closer in terms of the efficiency of the servers, both at using power and processing data. This is an area of data centre metrics which has not received as much attention, despite its importance in achieving a highly efficient facility.

One metric suggested by the Green Grid addresses this issue by calculating the *Data Centre Compute Efficiency* (DCcE) assessing the proportion of time a server spends operating under its primary function (The Green Grid, 2010b). There is also the *Server Efficiency Rating Tool* (SERT) which has been developed by the Standard Performance Evaluation Corporation (SPEC) to assess the relationship between the operation of servers and their power consumption (Tricker and Lange, 2011).

Measurement of server efficiency is a key step towards reducing energy demand in a data centre, however it will be important in the future to be able to integrate this with the efficiency of the supporting infrastructure. A white paper released in 2012 aimed to do just this by suggesting a metric entitled the *Fixed to Variable Energy Ratio* (FVER) defined as the following (Newcombe et al., 2012);

$$\text{FVER} = 1 + \frac{Fixed\ Energy}{Variable\ Energy} \quad (3.19)$$

The FVER metric assesses the fixed energy within a data centre (that which is a fixed load such as idle server power use) and the variable energy use (load which changes depending on the useful work delivered by the servers). One of the problems with servers is that even when they are doing no useful work they still consume energy, which is known as idle power. This power can be a significant percentage of the power used when working at full utilisation and so it is important to be able to recognise this fixed load and reduce it. The FVER metric attempts to isolate the energy use that varies with the useful output by the servers, enabling operators to increase this relative to the fixed energy use. For example, a FVER of 10 means that 90% of the total energy consumption is fixed and is therefore not related to the useful output from the servers. Ideally, if the servers are doing no useful work, then the energy consumption should reduce dramatically. Therefore, an ideal FVER would be around 1 meaning that 100% of the energy consumed is related to the useful output delivered.

The FVER metric relies on the operator weighting data centre tasks in terms of their importance to the purpose of the data centre. If the main purpose of the business is to process financial transactions, then these will be weighted the highest with functions such as emails given less. A graph is produced which illustrates the relationship between useful computing hours and the energy consumed by the server and the data centre as a whole. The intercept of the graph would ideally be zero power use when there is nothing being done in the servers, however this zero value is not possible and most often the value is quite a high proportion of the power used when the data centre is doing 100% useful work. The use of FVER takes more input than standard energy metrics, however it gives a much more useful output as it helps to quantify energy used relative to useful work done. This metric should be useful in identifying the fixed energy use in the data centre, a problem that the most commonly used metric (the PUE) cannot address.

3.5.4 Exergy Analysis and Second Law Metrics

Exergy analysis is a way to measure the ability of a system to do work, as introduced earlier. If exergy is lost then a system has a lower efficiency as work is carried out in an irreversible way. Exergy analysis can be applied to data centres and many authors have already done so (Lettieri et al., 2009, Wemhoff et al., 2014, Mcallister et al., 2008). So far, the metrics discussed in this work have been based upon the first law of thermodynamics as they consider the efficiency with which energy is used. Second law metrics on the other hand consider how this energy is transferred, which is where exergy analysis and metrics based upon this are useful. Lettieri et al provide a useful comparison between the use of metrics based on the first and second laws of thermodynamics (Lettieri et al., 2009). They compare several metrics to a life cycle exergy metric for use on servers within data centre. Life cycle exergy analysis is concluded to be better than a straight forward energy analysis as it more thoroughly assesses environmental impact (Lettieri et al., 2009).

Wemhoff et al also provide a comparison between exergy and first law metrics by carrying out an exergy analysis on a rear door heat exchanger (Wemhoff et al., 2014). They expand upon work by Almoli et al who investigated the efficiency of these systems through CFD analysis (Almoli et al., 2012). Whilst Almoli et al demonstrate that the use of these systems can provide efficiency savings (Almoli et al., 2012), Wemhoff et al bring the additional finding that exergy destruction is lower when using rear door heat exchangers. A lower level of exergy destruction means that there is more available work when using the rear door heat exchanger, providing the opportunity for greater reuse of heat elsewhere. Wemhoff et al conduct their exergy analysis through the use of their own analysis code, something which McAllister et al also do (Mcallister et al., 2008). However, McAllister et al realise the need for this kind of analysis to be able to be carried out within live working data centres if this methodology is to be adopted by the industry. McAllister et al therefore develop an exergy analysis program with the aim of making it easy to set up and for calculations to be solved within a reasonable time when carried out on a typical workstation. The authors decided this was an important factor in developing any analysis software as for it to be truly useful, it needs to be practical for the data centre industry to use within their facilities, without too many cost or time penalties.

Exergy is a useful form of energy analysis for data centres, although the process of measuring and calculating it is much more complicated than using more simple first law metrics. It requires numerous measurements to be taken within the live facility and then compute time and power (not dissimilar to CFD) to process these measurements. Exergy analysis would benefit from software such as that designed to make CFD more accessible to the industry, and applications such as that written by McAllister et al. Exergy analysis is a very useful tool when assessing the potential to reuse waste heat produced by the data centre as it can provide a case for reusing, rather than rejecting the heat.

3.5.5 Carbon and Water Usage Effectiveness

Whilst energy usage in the data centre is an important issue, it is also important to consider the use of other resources if a whole picture of usage within a facility is to be presented. The other main supply aside from electricity which needs to be considered is water, as data centres can use a significant amount depending on the cooling system employed. Carbon emissions also need to be accounted for to see the effect that usage of electricity is having on concentrations in the atmosphere.

Besides the PUE, the Green Grid have also proposed other metrics for the purpose of assessing data centres including the *Carbon Usage Effectiveness* (CUE) and *Water Usage Effectiveness* (WUE). Like the PUE, both of these metrics use the IT energy consumption as a reference. In the case of CUE this means that it does not take into account the carbon emissions associated with the IT equipment. The CUE uses the PUE value in conjunction with a Carbon Emissions Factor (CEF) in order to report the carbon emissions per unit of power used for the ICT equipment (The Green Grid). The CEF is the $kgCO_2eq$ emitted per kWh of electricity and depends on the region where the data centre draws its power from, as it depends on the electricity generation mix in the grid system. Whilst this can be an effective metric when attempting to reduce the emissions from a data centre, it is more useful to look at the actual tCO_2 emitted from the data centre over a year. This can be calculated by the CEF and the total facility power consumption. This measurement does not of course take into account the emissions resulting from manufacture of equipment, construction of the facility and the future use or demolition of the facility, otherwise known as lifecycle emissions. In order for full lifecycle emissions to be measured, a Life Cycle Analysis (LCA) needs to be carried out on the facility (LCA is discussed in the next section). This is no easy task as the emissions from the manufacture of each item of plant associated with the running and support of the IT equipment would need to be quantified and added to the emissions created during its operation. A full LCA of a data centre would also include the carbon footprint of the servers and the building fabric.

Another issue with the CUE is that it also does not take into account the CO_2 emitted from other sources in the operation of the data centre such as water usage, as it only takes into account the electricity use of the supporting infrastructure. The use of water does have an associated CEF and should really be included in the CUE if the true carbon emissions of the facility are to be calculated. Whilst the WUE metric has been suggested by The Green Grid, this does not account for the carbon emissions associated with this use, although this could simply be multiplied with a CEF for water usage.

3.6 Other Methods of Assessing Data Centres

Metrics are a useful tool to present the energy use of and efficiency within the data centre, however they only tend to provide a snapshot view of what is going on. Measured through an energy auditing process, most of the available metrics focus on one particular aspect of the data centre. There are schemes available however which provide a more holistic view of energy use and sustainability of the data centre.

The *BREEAM* (Building Research Establishment Environmental Assessment Method) *Code for Sustainable Data Centres* for example provides a set of requirements which the building must meet as part of a tiered award system, which certifies data centres based on their sustainability credentials (BREEAM, 2010). A successful scheme which has been running since 1990, it has recently developed a specialised set of requirements for data centres. This allows the designers, owners and operators of data centres to achieve ratings spanning from unclassified up to outstanding which can be done by earning credits for achieving certain requirements. The credits involved with the BREEAM Code for Sustainable Data Centres include considering the reduction of CO_2 emissions, water usage, considerate construction and recycling waste amongst many others. In terms of the energy requirements, BREEAM offers a total of 15 credits for reducing the CO_2 emissions from the data centre. The actual amount of credits awarded depends on the PUE value achieved and the CO_2 index, the full amount is awarded for facilities with a PUE of 1.22 and a CO_2 index of 25 or less (BREEAM, 2010). This scheme allows facilities to be compared in terms of their BREEAM score, and encourages competitiveness in reducing energy consumption within the data centre industry.

Similar to the BREEAM scheme, *The Leadership in Energy and Environmental Design* (LEED) rating system from the U.S. Green Building Council also works on the basis that credits are gained and tallied to receive a rating for a building. In the past, data centres have been awarded LEED ratings however these have been based upon a system set up for different buildings. A new document (LEED 2012) was released in November 2012 and contains a new rating system specifically for data centres (Council, [no date]). Within Europe, a specific scheme for assessing data centres was setup in order to encourage data centre owners and operators to reduce the energy consumption of their facilities. The EU CoC is a voluntary scheme that provides a list of best practice guidelines, data centres can enlist as participants if they meet the required guidelines. Engineering practices signing up as endorsers where they have the role of encouraging their clients to follow the guidelines (European Commission, 2008).

Besides rating schemes for the whole building, there are also smaller schemes which focus on the equipment within the facility. *The Energy Star rating* system is one such scheme by which an energy efficiency rating can be applied to a piece of IT equipment allowing buyers to select the most efficient equipment for their data centre (Sullivan, 2010). One of the requirements of the EU CoC is to use Energy Star rated products in data centres once they are widely available.

3.6.1 Life Cycle Analysis and Embodied Carbon

In terms of looking at the overall environmental impact of a data centre, LCA is one option which has not yet been fully explored by this industry, although there has been some work done (Whitehead, 2012, Meza et al., 2010, Shah et al., 2012). It is a process through which a complete carbon footprint of a system can be measured. For data centres, a carbon footprint should measure the total amount of emissions resulting from the construction (including the sourcing of materials and carbon embedded in the equipment), operation and demolition or re-use of the facility.

Several studies have shown that the server is the constituent part of the data centre with the largest environmental impact, either through its electricity use or its material makeup. Meza et al presented the breakdown of a server in terms of its constituent materials compared to the infrastructure that runs it and its operational impacts (the latter two of these were measured assuming a PUE of 1.6 and a power utilisation of 25%). It clearly demonstrates that the carbon footprint of the data centre is heavily influenced by the materials used within the servers, and that if the overall environmental impacts are to be reduced then this is a key area to consider (see Figure 3.2). This impact of the constituent parts of the server can be even higher over the lifetime of the data centre as IT equipment is replaced every 3-5 years (BSI, 2014). This is where operational procedures can affect the overall footprint.

Whitehead et al demonstrated in their LCA study that the IT equipment contributes 80% of the embodied impacts, and high refresh rates are a significant part of this (Whitehead, 2012). The technology behind IT equipment changes rapidly as time goes on, leading to businesses and data centre owners requiring new equipment frequently. Whilst high refresh rates encourage higher efficiency due to newer, more energy efficient servers being installed, conversely this increases the overall lifetime impacts of a data centre, by increasing the operational carbon footprint as shown by Whitehead et al.

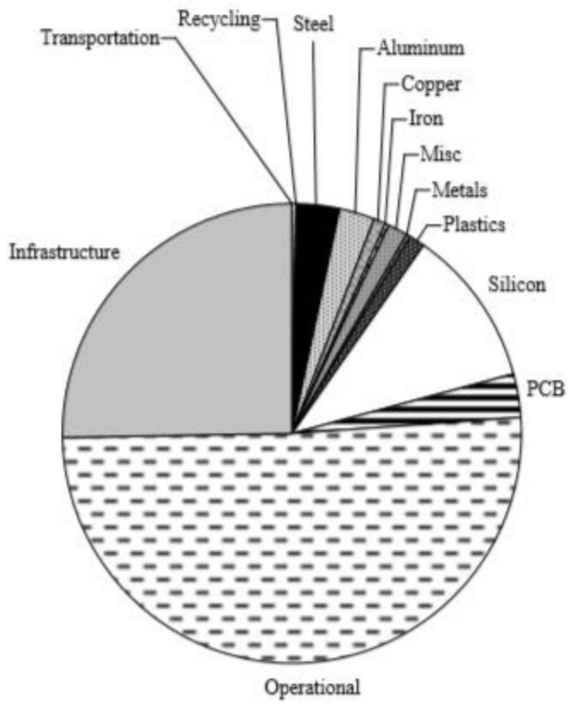


Figure 3.2 Proportions of different materials that make up a physical server (Meza et al., 2010)

Meza et al use an exergy based approach to assess the lifecycle of a data centre, although they exclude the manufacturing and building of the power and cooling infrastructure (Meza et al., 2010). They note that the embodied exergy of these components is small compared to the total embedded exergy consumption of the servers that reside in the facility. Concluding that the server has the biggest impact on the lifecycle impacts of the data centre, the authors suggest an entirely new way of situating computing components in the data centre. They suggest a design whereby the case of the server is negated, with circuit boards being mounted on a perforated spine which allows air flow across them. Enabling common components to be shared across the facility, this would dramatically reduce the amount of sheet metal used and have an impact on the total embedded exergy.

Research by Shah et al showed slightly different results in that the greatest contributor to environmental impact is the electricity consumed by the data centre, followed by the embodied energy in the IT equipment. Shah et al model the environmental impacts of a data centre that consumed on average 1.8 MW IT load (with a total capacity of 3 MW), a 1.3 MW cooling load and a power delivery efficiency of 93%. Figure 3.3 shows the relative impacts of different components of the data centre system on climate change, ecosystem quality, human health and resource depletion. It can be seen that the consumption of electricity by the IT equipment has the highest impact in three of those categories, with the fourth category (ecosystem quality) being affected the most by the material use in and production of servers.

Research has shown that the server is the highest contributor to both data centre electricity consumption and overall lifetime environmental impacts. Any improvements in server energy use would therefore be beneficial, especially if that were to be paired with a reduction in material use and environmental impact.

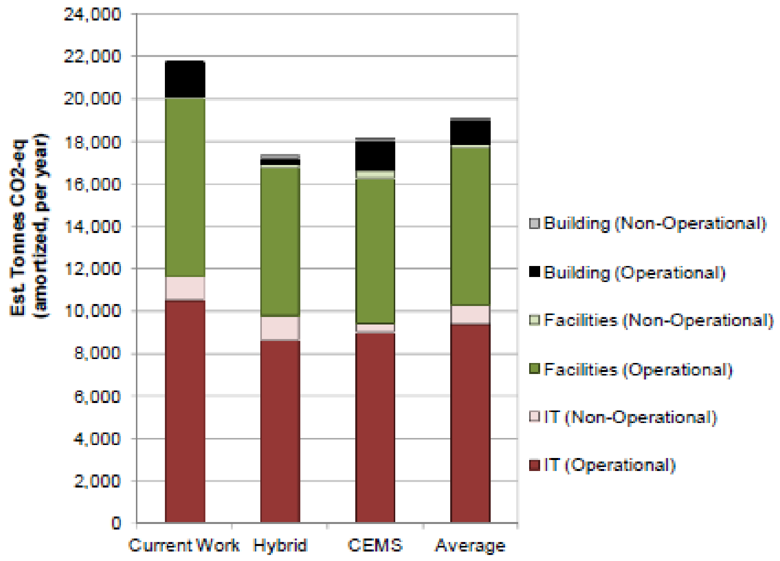


Figure 3.3 Carbon emissions from the IT equipment, facilities and the data centre building

LCA is a useful tool to assess the environmental impacts of data centres. It provides clear evidence of which components of the system the industry should be focussing on during efforts to reduce carbon footprints, not just energy use. It is clear from existing studies that the IT equipment presents the biggest impacts, with the materials of its constituent parts being the reason behind this followed by their large consumption of electricity.

3.7 Energy Consumption Case Studies

The metric most commonly used within the industry to measure energy consumption is the PUE, with a value of 1 being perfect although theoretically impossible (this metric is discussed further in Chapter 6). Whilst there is currently not enough data to illustrate the PUE of data centres on a world-wide scale (Daim et al., 2009), there have been some smaller studies completed. The Green Grid conducted a survey in 2011 with 115 respondents, 70% of which reported their PUE with an overall average value of 1.69 (The Green Grid, 2011). A similar study (using a metric termed *Computer Power*

Consumption Index equivalent to 1/PUE) for 22 data centres found PUE values between 1.33 and 3 with an average of 2.04 (Greenberg et al., 2006). However there is no indication of the locality of the facilities, the scale of the operations within them, or how the energy consumption assessment was carried out in either of these studies.

A more in depth analysis of energy consumption was provided in (Karlsson and Moshfegh, 2005) which assesses the energy consumption of a small data centre in Linköping, Sweden. It was found that 71% of the power supplied was used to operate the central chiller plant and the CRAC units, with only 29% being used by the IT equipment. Although a PUE was not reported, the figures given would suggest a value of around 3.4, indicating a very low efficiency. Inefficiency was reported as being due to an oversized air conditioning system, inadequate use of the cool air by the racks, and the fact that air conditioning units may be cooling air already chilled by other units due to a difference in the set temperatures. The energy consumption assessment of this data centre was however only conducted during the coldest period of the year, February. PUE should be an annualised figure and so should take into account any changes in ambient conditions during the year, rather than just focussing on one particular season in the year. A study conducted in 2006 (Sun and Lee, 2006) measured the energy consumption of two data centres in Singapore at intervals of 10 minutes for the duration of a week and presented detailed information of energy consumption values for the IT load, the HVAC (heating ventilation and air conditioning) systems, UPS losses and lighting. Due to the local climate experiencing relatively constant temperatures, the data could be extrapolated to represent a whole year. Despite this ability to measure a shorter period of time, this may still not be long enough to measure any variations in the IT load.

A study carried out over the period of 1 year at a data centre in Helsinki, Finland showed a variation in PUE based upon the variation in IT load and the variation in CRAC energy usage (Lu et al., 2011). The authors gave detailed information about the cooling system equipment power usage, alongside information about the footprint of the data centre. Lu et al report that on average the PUE was 1.2 in the winter and 1.5 in the summer.

A similar study was carried out at a data centre in Silicon Valley, California, USA (Mitchell-Jackson et al., 2003). The power requirement for the IT equipment was only taken over a short period of one month (January) from the power distribution units (PDUs), giving a power density for the IT equipment of 170 W/m^2 (the power consumed by the IT equipment divided by the floor space occupied by active computer rooms). In order to calculate the overall power consumption, other loads such as lighting, fire systems, backup generators, and losses from power distribution were factored in. Assumptions still had to be made during the calculation though, such as assuming that 5 of the 22 CRAC units were able to dehumidify and reheat, and assumptions about the power losses from UPS and PDUs. The authors found the total energy consumption of the facility was 907 kW and the IT load was 432 kW, resulting in a PUE of around 2.1 (although at the time of assessment, the racks in the facility were being used at a third of their full capacity). An overall value for the power density of 355 W/m^2 was presented in terms of the total power consumption divided by the floor area of the computer room.

3.8 Increasing Energy Efficiency in the Data Centre

Chapter 1 introduced the global issue of greenhouse gas emissions and climate change along with how energy consumption relates to this problem. So far, the energy use of data centres has been discussed in Chapter 2 and this chapter has introduced the myriad of different metrics that can be used to assess these facilities. In section 3.11, the problems of introducing more energy efficient technologies and policies within a data centre will be discussed by considering the literature surrounding barriers to energy efficiency. Before barriers can be discussed however, it is necessary to introduce the main areas where or methods by which energy efficiency can be increased.

Currently in the literature, there are many authors and works discussing different taxonomies for increasing energy efficiency (in all sectors, not just data centres) and the barriers that act against this. McKenna identifies three main approaches by which energy efficiency can be improved[11]; behavioural, technological and policy-related measures (McKenna, 2009). Technological measures involve the introduction of new equipment with the direct aim of increasing efficiency, the replacement of equipment leading to indirect increases in efficiency and the use of a new process. Behavioural measures are concerned with targeting management methods and attitudes in order to encourage higher energy efficiency. Whilst the technology may be at peak efficiency, poor management of the processes and a lack of knowledge on behalf of staff can cause a drop in potential efficiency. Lastly, policy-related measures can be used to encourage technological and behavioural changes. These can include the introduction of

---

[11] In this case, the focus was on energy efficiency within industry.

legislation, standards and also economic schemes and incentives. Whilst McKenna discussed these three measures in terms of industrial energy efficiency, they can also be directly applied to data centres.

According to a report by the U.S. EPA there are three levels of improvements that can be made in data centres; improved operation, best practice and state-of-the-art (U.S. Environmental Protection Agency, 2007). Each of these offer increasing energy savings and cover all three areas as listed by McKenna; technological, behavioural, and policy-related measures. These are all now discussed in more detail as they apply to data centres.

3.8.1 Technological Measures

One of the main ways that energy efficiency can be increased is through the use of efficient equipment and the technologies that can enable such a change. By their very nature data centres are driven by the technology within them. They are full of numerous pieces of equipment and systems, all of which consume electricity. If every individual piece of equipment and each system of infrastructure were designed and selected to be the most energy efficient for the particular situation it is used for, then there would be the potential for significant energy savings. Due to the fast moving pace of IT technology, the data centre sector is also inherently fast moving. There is the opportunity to introduce newer equipment at a much higher rate than other industries, especially regarding the IT equipment. There are numerous technological changes that can be made in a data centre to reduce energy consumption including;

- **cooling and air distribution**; examples include *aisle containment*, *free cooling* and also the *reuse of heat* either within or externally to the data centre;
- **power supply**; *power supply network* and resiliency including *UPS units* and *backup generators*;
- **lighting and other systems**; installation and use of *LEDs* and *PIR*[12] *sensors* for lighting, reducing other small systems power through use of *efficient equipment*;
- **IT hardware and software**; efficient *IT hardware* but a focus is also needed on the efficiency of the *software* as these are both the core functions in the data centre; and
- **introduction of new schemes** such as *renewable energy* and the *use of, rather than rejection of heat* from the IT equipment.

[12] Passive Infrared Sensor, allowing lights to be turned on and off automatically

3.8.1.1 Cooling and Air Distribution

Cooling and air distribution systems are a crucial part of the data centre. Without these systems, the IT equipment (especially servers) would not be able to function as their internal temperatures would be far too high for them to operate. This vital piece of infrastructure can however consume a significant proportion of the electricity supplied to the data centre, as discussed in Chapter 2. This is not an insoluble problem though as there are numerous options and technologies that are available which can reduce the energy consumption of the cooing and air distribution systems.

Whilst data centre owners and operators may focus on using more energy efficient plant to provide cooling, huge energy savings can also be made by practicing good air flow management, a point frequently overlooked (Slaters, 2009). The benefits of carefully provisioned air distribution within data centres is well discussed within the literature. Several authors have conducted energy audits on data centres and provided evidence that better air distribution can lead to savings in energy. Fakhim et al took an existing inefficient data centre and used CFD to identify problem areas in the air distribution system; they then investigated ways in which it could be improved. Through further CFD analysis they demonstrated that improvements could be made by using blanking plates, altering the layout of the CRAC units, employing cold aisle containment and using ceiling warm air return ducting. Of these different options, they found that better placement of the CRAC units had the best impact in terms of the efficiency of delivering cold air to the servers (Fakhim et al.).

Separating the supply and return air streams increases efficiency significantly and aisle containment is the main method of doing this. Chapter 2 introduced the idea of aisle containment, a method used to ensure that hot and cold air streams do not mix with each other. This offers a much greater efficiency as cool air is not heated up by warm return air from the server. Kumar also found that careful placement of servers within the data centre would also reduce the energy use of the air distribution system. They found that by having racks with uniformly loaded servers along with clustering the servers meant that better air flow management could be achieved (Kumar et al., 2011).

Another major way of saving energy in the data centre is to consider the energy use of the cooling system, one particular method of saving energy is to utilize a process known as '*free cooling*' or economisation. By negating the need for a compressor, free cooling technology consumes significantly less energy than more traditional cooling systems. There are several different types of system that offer free cooling which can be grouped as either waterside, airside or a heat pipe system, although the latter tends to be less common. Airside and waterside systems can be further categorised as either direct or indirect airside cooled, direct water cooled, air cooled or a cooling tower system (Zhang et al., 2014). Whilst heat pipe systems are less utilised, a review of the literature by Zhang shows that they have potential because of their ability to cope with transferring heat even at lower temperature differences (Zhang et al., 2014). The use of heat pipes in the data centre as part of an indirect cooling system is still relatively new. Jouhara et al suggest that further research into this technology is needed, research which engages properly with members of the data centre industry to ensure suitability (Jouhara and Meskimmon, 2014).

The Green Grid conducted a survey in 2011 on the use of economizers in data centres. It was found that the use of economizers is increasing, and that they are well known by decision makers in the industry. The survey had 115 responses from participants who were facility managers or had responsibility for ICT and facilities in data centres over 2,500 square feet (half of the facilities were over 25,000 sq ft). With 90% of the respondents situated in the USA it was possible to see the relationship between geographical position and the use of economizers. It could be seen that due to the local climate (which has a higher humidity and temperature), none of the facilities located in the southern most states in the USA used economizers. It was found that 49% of respondents used economizers in their facility and that this use was more common in the larger data centres (64% of facilities ≥ 50,000 sq ft used economizers whereas only 10% of facilities between 2,500 and 4,999 sq ft did). The survey results indicated that on average the respondents could operate their economizers for 4,724 hours a year however on average only 80% of these hours were actually spent using the economizer. One of the main reasons for this underutilization is that the standards set by the particular companies that run the facilities are stricter than the ASHRAE recommended operating range. Other concerns were raised over air pollutants, humidity and maintenance costs. 38% of respondents reported that although they had considered an economizer, it was not possible due to the process of retrofitting being too difficult. 28% of respondents reported an energy cost saving of over 25%. The survey also collated PUE values from the respondents, 70% of which provided a PUE with an average of 1.69. It was concluded that there was no significant statistical difference in the PUE value between using an economizer and not. This is a valuable survey which records the use of economizers and the decisions made as to not have an economizer. It provides a good insight into the decisions made by facility managers in the adoption of more energy efficient systems, and their willingness to operate outside the ASHRAE recommended envelope (The Green Grid, 2011).

An Energy Comparison between Different Systems

The methods of reducing energy consumption in the cooling and air distribution systems mentioned above are only a few of what is currently available. There are many different methods through which energy can be saved, and each one is suited to different data centres due to the equipment and layout requirements, they also all offer different levels of savings. Table 3.2 is adapted from work by Kennedy who provided a summary of the energy savings potential of each different cooling system (Kennedy, 2009). The author used averaged energy consumption values for different methods of cooling in order to find relative energy savings for each one to allow a comparison to be made. Whilst there are no real details given in the study regarding the calculation of these values, the results are valuable in order to see a basic comparison between cooling systems. It is clear that the choice of cooling system can have a dramatic effect on the overall energy consumption of the data centre. When employing a more traditional CRAC or CRAH based cooling system, energy savings can be made by implementing aisle containment of by increasing the chilled water temperature. However, these savings are only small. If a real impact is to be made, then an alternative method of cooling is required. According to Kennedy, the most savings can be made by switching to liquid cooled servers, an option that could present a saving of 82% compared to a traditional system.

It must be remembered however, that whilst some systems can provide significant savings (48% in the case of air economisation (Kennedy, 2009)), implementing them is not always possible. Air side economising for example requires a large amount of space, and is easier to use in a new build data centre rather than being installed as a retrofitted system.

Table 3.2 Comparison of the energy savings possible through using different cooling systems*

System	*Energy savings compared to standard CRAC/CRAH system*
CRAC/CRAH	• Implementing aisle containment can save **7.3%** • If the chiller water temperature is increased, this can save **18%**
Liquid cooled racks	• These can produce savings of **40%** • If free cooling is employed this can be increased to **49%** • If evaporative free cooling is used then this can save **55%**
Active/passive rear door cooled racks	• Active systems can save **57%** • Passive systems can save **66%**
Air economisation	• This can save **48%**
Liquid cooled servers	• This system is the most beneficial, saving **82%**

*Note: the figures in this table are adapted from (Kennedy, 2009).

3.8.1.2 Power Supply and Alternative Sources of Energy

As explained previously in Chapter 2, in order for the data centre to have a certain level of resiliency, UPS units are used to provide security in the event of a mains power failure. The UPS unit provides power to the data centre whilst the back-up generators are starting up, and provide a better quality of power supply when running off the mains. However these units introduce a loss into the power supply system as they have an associated efficiency (California Energy Commission, 2008). There are several different types of UPS unit available for use in data centres, with each one offering different levels of energy efficiency, although with all units the highest efficiency is only achieved when the unit is fully loaded. A delta conversion unit for example is only 93-94% efficient at 25% load but at 100% load it can reach 97% efficiency (California Energy Commission, 2008), a significant difference when considering the amount of power flowing through these units.

Alongside improvements in the energy efficiency of the power supply network within the data centre, more recently the issue of where energy is sourced has become more of an issue within the industry. Large data centres have in the past been criticised by environmental campaign groups and the media if they source their electricity from coal fired power plants (Greenpeace, 2011). Dos Santos et al. discuss the problem of primary energy source when making efforts to be more efficient (dos Santos et al., 2013), they mention that by not including primary energy sources in energy audits it can distort the true efficiency of the system.

There are several examples of data centres choosing alternative energy supplies, such as on site generation (Judge, 2013), combined cooling, heat and power (Xu and Qu, 2013), renewable energy (Sverdlik, 2013b) and fuel cells (Sverdlik, 2013a). Alternative and renewable energy sources for data centres will become more important as the limit of energy efficiency improvements is reached. Once energy efficiency is at its peak for all systems within the data centre, greener sources of energy will need to be sought if further environmental impacts are to be reduced. Figure 3.4 illustrates this by showing that it would be possible to reduce the carbon footprint of a typical data centre in the USA by 88% if best practice energy efficient changes are made. Further savings of an additional 10% would then be possible by changing to alternative energy sources (Masanet et al., 2013).

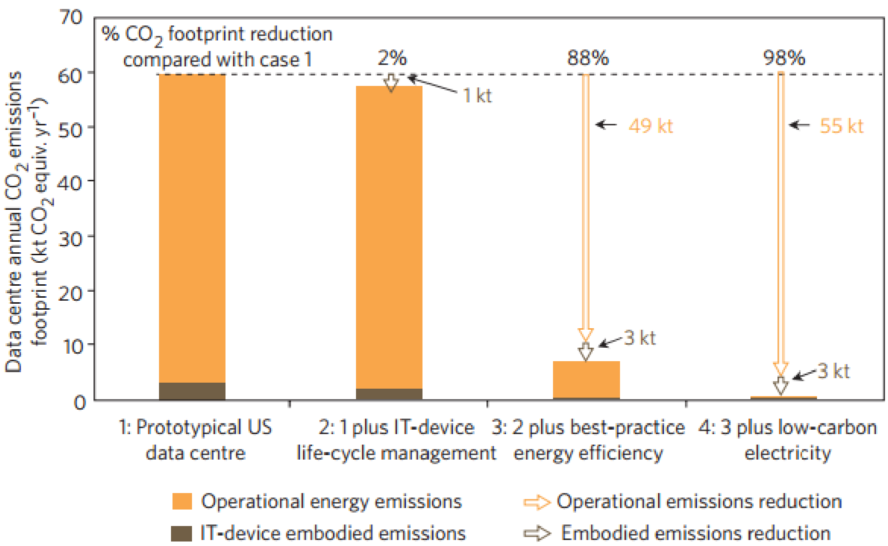


Figure 3.4 Possible carbon footprint reductions for a typical data centre in the USA, with IT device management, efficiency best practice and low carbon energy source options (Masanet et al., 2013)

3.8.1.3 IT Equipment and Energy Proportional Computing

A lot of focus has been cast on the supporting infrastructure in the industry's efforts towards reducing energy consumption, however as the largest consumer of energy within the data centre it is important that the IT equipment is also considered. Figure

3.5 shows the power consumption of each component of the data centre relative to the IT equipment, it is clear that if significant savings are to be made then reductions must be made within the IT systems themselves.

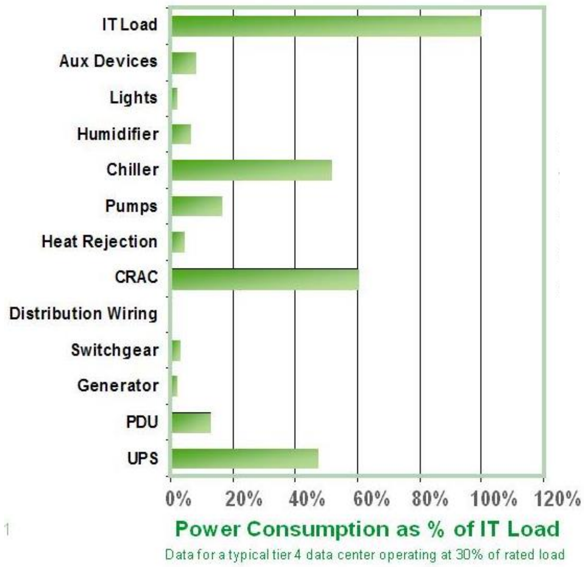


Figure 3.5 Power consumption of equipment within a typical data centre as a percentage of IT load (Auvil, 2013)

With regards to energy efficiency within the IT systems, there are two areas that can be considered. Hardware efficiency involves looking at the energy use of the server unit itself and the components within it, in terms of how much energy it requires to operate (or it can also involve reducing the amount of redundant equipment). Software efficiency then concerns the efficiency with which the IT hardware can carry out its required functions. Both of these efficiencies will have an impact on the overall efficiency of the data centre. Furthermore, the method through which IT resource is used within the data centre adds a further area for efficiency improvements, techniques to improve this are known as energy aware or energy proportional computing.

In terms of the efficiency of the server hardware, there have been efforts in the past to encourage the sale of more efficient equipment. Energy Star for example is a system which rates the server based upon its energy efficiency, an Energy Star certified server will use just under half the amount of energy needed to run a less efficient, none rated server over a year (Energy Star). Energy Star rated servers come with features such as efficient power supplies to reduce conversion losses, the capability to measure parameters such as air temperature and are also built with efficient components (Office of Energy Efficiency and Renewable Energy, 2014). Energy Star servers are tested for their power consumption over different levels of CPU utilisation, the results of which are published by SPEC[13]. This enables better comparisons between servers and allows buyers to weigh up energy use and performance of their equipment before purchasing.

Whilst efficient hardware is important, efficiency within the software can provide significant overall energy savings for the data centre, particularly when it comes to resource allocation. Klingert et al list virtualisation, energy-aware work load allocation and batch processing as all important measures to create an energy aware data centre (Klingert et al., 2011). Virtualisation is a method through which the capabilities of several virtual servers are placed within one physical server, thus saving physical space within the data centre. However, this can lead to hot spots within the room as some servers will have increased their workload. Islam et al put forward a solution to this by suggesting a temperature aware scheduling algorithm. Their algorithm proactively decides how to allocate loads to servers based on predicted temperatures, rather than reactively making a change once temperatures have increased. The algorithm also enables IT equipment that is not needed to shut down and to only consume power when they are required to do work. Enabling servers to shut down offers the potential to save large amounts of energy, as they consume a good proportion of their maximum power rating when doing no useful computing work. This method of management can be problematic though as upon turning the servers back on they do not have the ability to operate immediately and can take a considerable amount of time to become operational. Kliazovich et al consider the benefits of a job concentration scheduler which allows servers to be completely turned off when not being used, the management of scheduling allows for the length of time it takes for a server to switch back on again (Kliazovich et al., 2013).

Turning off idle servers as part of their resourcing algorithm, Islam et al have shown this to be beneficial in terms of energy saving (Islam et al., 2015). Figure 3.6 shows a comparison between three different resourcing algorithms; EQL, which distributes work evenly across the data centre; TempU, a temperature unaware resourcing algorithm and DREAM, the authors' algorithm which is temperature aware. It demonstrates a clear saving in energy when the workload is at lower percentages of maximum capacity, which is due to the algorithm switching off IT equipment when it is not needed.

[13] Results are published for public use at www.spec.org/power

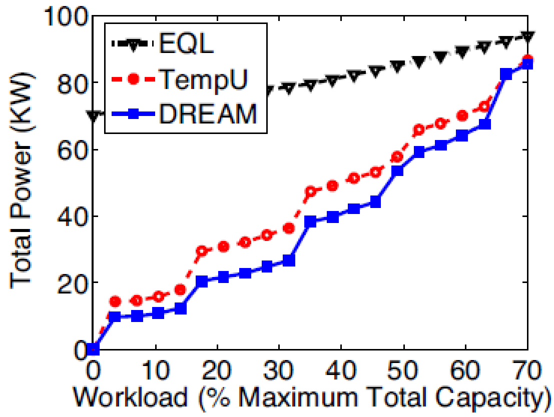


Figure 3.6 Power consumption of IT equipment for a given workload percentage (Islam et al., 2015)

Through the use of their DREAM algorithm, the authors have shown that compared to an algorithm that is not temperature aware, theirs reduced power consumption by up to 33%, and showed an improvement over equal load distribution of up to 86% (Islam et al., 2015) (Figure 3.7).

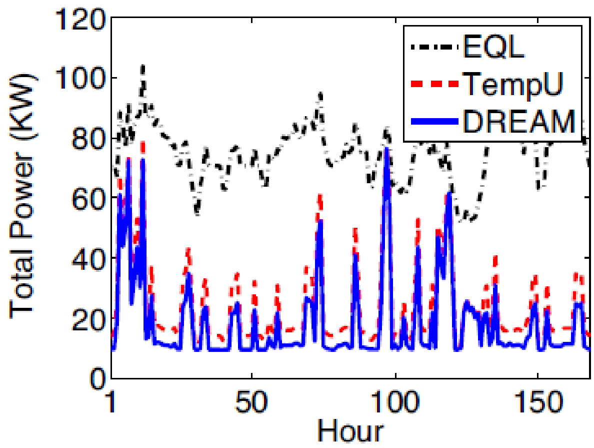


Figure 3.7 Power consumption of servers under different resourcing algorithms (Islam et al., 2015)

Other algorithms for resource management and energy-aware or energy proportional scheduling in data centres have been widely discussed within the literature (Gao et al., 2014, Beloglazov et al., 2012, Garg et al., 2011, Banerjee et al., 2011, Guzek et al., 2014, Sun et al.). Whilst they offer differing levels of energy savings, some authors recognise that it is important not to disrupt the Quality of Service (QoS) whilst doing so (Islam et al., 2015, Lent, 2015, Zheng et al., 2011, Beloglazov et al., 2012). Disrupting the QoS would lead to a breach of the SLA, which could happen if computing times were slower due to the resourcing method used.

Resourcing algorithms whether temperature or energy aware have the potential to offer significant energy savings, savings which are available at a lower cost compared to those that require physical changes in the data centre. An algorithm suggested by Bodenstein et al has the potential to reduce energy use by 40%, the authors suggesting that this is a short term solution that is also inexpensive whilst delivering a cost saving (Bodenstein et al., 2012). Most physical changes needed to reduce energy use require large capital investments and sourcing funding for this can be difficult for many facilities.

3.8.1.4 The Reuse of Waste Heat

Although not strictly a method of enabling greater energy efficiency or reducing energy use, the reuse of waste heat from the data centre has been suggested as a way of reducing their overall impact on the environment. It has been suggested that the heat generated by IT equipment could be utilized for other purposes (such as space heating), rather than just rejecting it to the external environment (Miller, 2010). The reuse of waste heat from the data centre is very limited in real life examples, although the technology is gaining ground within the industry and is recognised as a best practice by the EU CoC.

Several authors have suggested different methods for capturing waste heat so that it can be used to good purpose elsewhere (Ebrahimi et al., 2014, Haywood et al., 2012, Kim and Gonzalez, Zachary Woodruff et al., 2014, Zimmermann et al., 2012). Ebrahimi et al review each available low grade waste heat recovery technique including district heating, absorption cooling, colocation options and thermoelectrics (Ebrahimi et al., 2014). Low grade heat is provided by heat sources with temperatures in the range of 30-90°C, with the higher range being upwards of 60° (Fang et al 2015). High grade heat, above 90°C, offers a reasonably easy source of energy to reuse of which steam is one such example. Low grade heat is much more difficult to use because of the lower exergy value of the source. Waste heat from a data centre for example (using air as the cooling medium) will have a grade of heat within the lower range, 30-60°C. With a much lower level of exergy available, there is less incentive to reuse this source of heat elsewhere. Also, because of this low exergy level, it is more advantageous if the heat can be reused close to the site it was generated. Given this, data centres that do reuse heat tend to be located close by to the location where it is required. One suggested solution to this is from Woodruff et al who suggest an approach which involves dividing a large data centre into several nodes which can then supply waste heat to separate buildings, an approach they have named Environmentally Opportunistic Computing (Zachary

Woodruff et al., 2014). This approach means that rather than having one large isolated data centre, an organisation divides this into nodes and integrates them within the organisations buildings.

Alternative forms of cooling such as liquid cooling strategies employed in the data centre can allow for a higher grade of waste heat to be produced (Ebrahimi et al., 2014, Zimmermann et al., 2012), which opens up a wider range of available methods by which this waste heat can be utilized. Zimmerman et al for example demonstrate the value of reusing the waste heat from a data centre that uses hot water cooling (Zimmerman et al., 2012). Using warmer water to cool the servers allows for a higher temperature at the outlet and hence a higher grade of waste heat. Findings from their work show that the technology provides a greater potential to reuse the heat elsewhere, as the exergy results demonstrate a higher grade of waste heat being produced. Through the use of exergy analysis the authors show there is more value in using the heat for space heating than there is to reuse it to do mechanical work. If the reuse of heat from the data centre is to be further encouraged in the future, more metrics based upon exergy would be useful, although they do of course need to be user friendly.

3.8.2 Behavioural Measures

Behavioural measures are important in the effort to reduce the environmental impact of data centres. Even if every efficiency method possible were to be made within the technology side of the data centre, inefficiencies could still occur due to the management of the facility. Behavioural measures can include the way that the equipment in the facility is operated and maintained, or whether or not an energy management and auditing procedure is in place. Training of staff can also be a key behaviour for ensuring that the infrastructure is used in the most efficient manner (Birchall, 2011).

One of the major problems facing data centres is that there is often a disconnect between the department responsible for the energy bills and the IT department. This can often mean that whilst consideration is paid towards improving the infrastructure in terms of energy use, the IT department can be more concerned with ensuring that the best equipment is always available regardless of the energy costs. The EU CoC attempts to address this in one of the clauses (Section 3.1 of the code, Involvement of Organisational Groups (European Commission, 2014)), by requiring that participants create an approval board. This board must contain people from software, IT and M&E (Mechanical and Electrical) disciplines. If any changes are to be made in the data centre, they must meet the approval of this board which should in theory reduce the problems of lack of communications between teams. There can also be problems with some organisations being more concerned with their availability and reducing the risks of failure than they are about the amount of money they pay for energy use in their facility. Behavioural aspects of energy efficiency improvement will be examined further in the qualitative research within Chapter 5.

3.8.3 Policy-related Measures

There are several codes and regulations which cover the improvement of energy efficiency in data centres, with the number of these increasing over the last several years. In December 2006, the USA government passed an act known as Public Law 109-431 in order *'to study and promote the use of energy efficient computer servers in the United States'* (U.S. Environmental Protection Agency, 2007). This led to the production of a report written by the U.S. EPA to summarize the trends in growth of data centres and had an aim of investigating how energy efficiency could be implemented to reduce the costs associated with running computer servers. So whilst the focus of this report was on increasing the energy efficiency of data centres across the USA, this was for monetary reasons rather than reducing carbon emissions. This is a common driver for reducing energy consumption as prices for fuel and electricity rise across the globe and one major benefit of doing so is that emissions of CO_2 along with other GHGs are reduced.

The following is a summarised (although not exhaustive) list of policies, regulations and schemes released which aid data centre designers, owners and operators to reduce their energy consumption. These cover the full range of aspects of data centres from design and construction, sourcing of hardware, to monitoring energy use once the facility is in operation;

ASHRAE Technical Committee 9.9

Introduced in Chapter 2, the ASHRAE guidelines on recommended environmental requirements for data centres has become a standard which all facilities follow. By providing recommended envelopes for the temperature of the supply air within the data centre, they are pushing the limits regarding how warm the supply air can be. As previously discussed, by expanding the envelope more regions around the world are able to safely use free cooling, which has contributed to significant energy savings. These guidelines from ASHRAE are a good example of how a policy related measure can enable the adoption of more energy efficient cooling technology.

ISO 14001

ISO 14001 is a voluntary international standard regarding energy management systems within a company or organisation. As a voluntary standard, ISO 14001 tends to have the effect of encouraging environmental improvements in other organisations once one has adopted it. Arimura et al studied the effect that this standard had in the industrial sectors and found that if an industrial facility adopts the standard, then it can encourage environmental improvements on behalf of their supply chain (Arimura et al., 2011). By encouraging a facility to examine the environmental credentials of its supply chain, it means that companies down the supply chain also have to reconsider their impacts. Within the data centre industry, facilities need to compete for clients and individual companies can stand out from the rest if they have ISO 14001 accreditation. This encourages better energy management within facilities.

EN 50600

Published in 2012, this standard sets out a series of documents aimed at aiding various participants of the data centre industry including owners, managers, consultants, architects, suppliers and maintainers. Contained within this standard is a section on the design principles for energy efficiency within the data centre. The standard encourages designers to use the best practices as listed by the EU CoC when designing a new facility. This standard lists two drivers for achieving higher energy efficiency, firstly that there is *'a need to reduce the cost of operations'* and secondly that there is a need to support *'local and regional initiatives'* (The British Standards Institution, 2012). Again, reduction in costs is placed as a higher priority to reducing emissions and other environmental issues.

Climate Change Agreement

After the Climate Change Levy (CCL)[14] was introduced in the UK, CCAs were created for the most energy intensive industries which would be most affected by the carbon tax. This addition was due to the possibility that some industries would struggle to maintain their competitiveness with other countries due to this tax, causing a loss in business or migration of business. A CCA entitles an industry to have a 65% reduction on the levy or tax, provided that it meets strict carbon reduction targets, from April 2013 this was increased to an 80% reduction (Department of Energy and Climate

[14] A piece of legislation falling under the umbrella of the Climate Change Act 2008, it encourages energy efficiency improvements through a carbon tax and is applied to emitters using more than 6,000 MWh of electricity per year.

Change, 2011). As of 2011, there were 54 sectors entitled to this reduction through this scheme including representatives from the food and drink sector, the agricultural sector and manufacturing sectors. From 2014, the data centre industry has been granted its own CCA by the government, a recognition of the importance of this industry. Whilst the data centre industry CCA does reduce the amount of carbon taxes payable by the largest facilities, it does come with requirements. An organisation under the agreement of a CCA must measure and report energy use and also carbon emissions, and have targets that have to be met over set periods.

EU Code of Conduct on Data Centre Energy Efficiency, BREEAM, LEED and CEEDA

Best practice guidelines tend to be a more common method of encouraging higher energy efficiency and less energy use within data centres. Schemes such as the EU CoC, BREEAM, LEED and CEEDA (Certified Energy Efficiency Data Centre Award) all offer guidelines which if followed lead to a more environmentally friendly facility. The latter of these three offers a rating system where different levels of awards can be given depending on the amount of effort made to reduce energy use and increase efficiency. The EU CoC on the other hand simply provides participant status, or endorser. The reception and use of such schemes will be addressed in more detail as part of the qualitative study presented in Chapter 5.

Service Level Agreements

All data centres enter into a Service Level Agreement (SLA) with their clients, this acts as a contract between the two parties and details the service that the data centre agrees to provide. Written within it is information such as how the environment around the servers is maintained and details such as supply temperatures. An SLA also details the penalties that must be paid by the data centre should they not deliver the service as promised. In order to encourage better environmental credentials, Klingert et al suggest the use of a GreenSLA (Klingert et al., 2011) which they define as an SLA with a focus on operating a more *'eco-efficient'* data centre. Having an agreement between the data centre and its clients as to the environmental impacts of the facility may be a good way to drive further savings. If both data centre and client are required to conform to an agreement to reduce energy use, then it may work better than just the data centre having to make an effort. Their business is driven by their clients and unless the clients are interested in (or required to think about) energy use, then it is difficult for the data centre to justify changes. This will be addressed in more detail within Chapter 5.

3.9 Uptake of Energy Efficiency Measures

It is clear from the discussion above that there are many opportunities to increase the energy efficiency of data centres and reduce their energy use. From the mechanical and electrical support infrastructure down to the servers themselves, there are a range of different options for decreasing energy use. New data centre facilities tend to be built with the latest energy saving technologies built into them; however this can be more difficult when considering older facilities. Finding out how many data centres worldwide are adopting these energy efficiency strategies is important so that inefficiencies can be targeted. The Uptime Institute have done a lot of work in surveying the industry on a range of issues, including energy use and efficiency. Each year they survey a number of data centres and collate the results to show how the industry is addressing its energy use. The 2013 report by the institute presented results from 1,000 data centre owners and operators, who were located worldwide.

One of the main things that the Uptime Institute surveys look at is the adoption of efficient technologies in the data centre. Figure 3.8 shows the adoption of efficient cooling technologies for data centres which are running either under 1,000 or over 5,000 servers. It appears that the more servers a data centre is running, the more likely they are to adopt energy efficient technologies; this is likely because they will have a larger power bill. A larger data centre may also have a more diverse workforce which may include energy manager roles. Capital costs of projects also play a part in how many facilities have adopted certain technologies.

More data centres have adopted aisle containment, as this is easier to implement in existing facilities than it is to install a new cooling system. The Uptime Institute argued in their 2012 report that larger data centres are more likely to have adopted efficient technologies and strategies because they have more resources available to them than smaller organisations (Stansberry and Kudritzki, 2012). Resources include money, knowledge and staff, all of which make it more likely that the data centre will have adopted more energy efficient technology and practices.

Reporting energy use is also a focus for the annual Uptime Institute survey. The results of the 2013 survey showed that smaller data centres don't tend to use the PUE metric to report energy use and in fact 34% of the respondents don't use it. They also reported some misuse of the metric with 4% of respondents giving a PUE of less than one (an impossible figure). 90% of respondents whose data centres contained over 5,000 servers use the PUE metric. This perhaps shows that larger facilities tend to focus more on their energy use than smaller ones. Just over half of enterprise data centres surveyed only report performance quarterly or less, or never. The results show that the majority of data centres report their performance monthly to management (Figure 3.9).

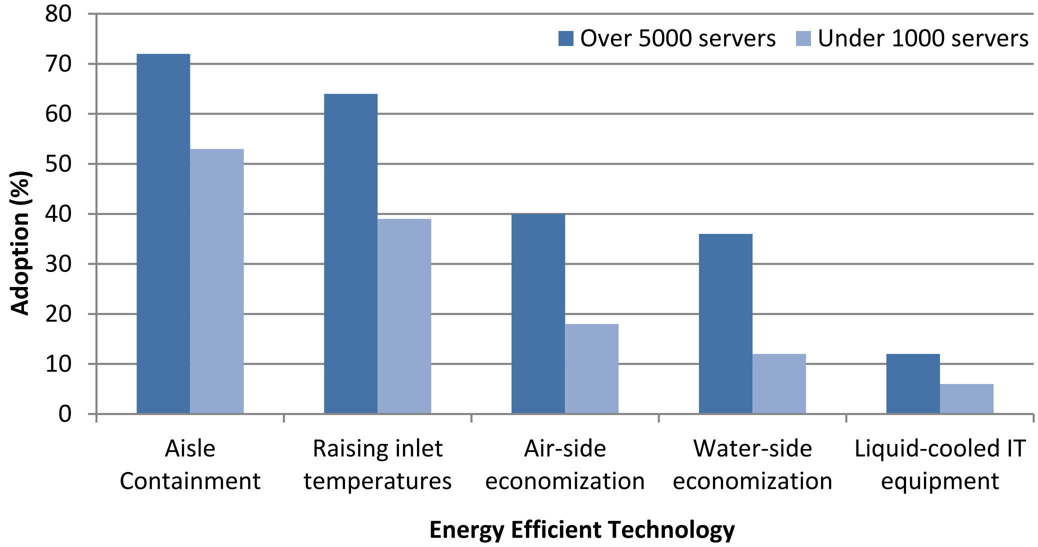


Figure 3.8 Energy efficient technology adoption in data centres with either over 5,000 servers or over 1,000 servers, adapted from (Stansberry, 2013)

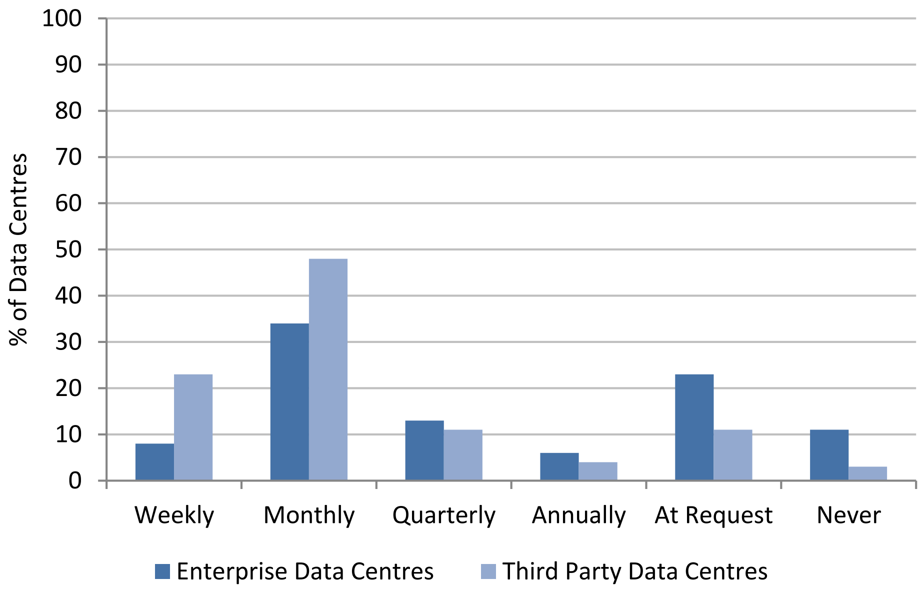


Figure 3.9 Percentage of data centres that report performance or cost metrics to high level management, adapted from (Stansberry, 2013)

The Uptime Institute surveys are valuable to the industry as they provide a good perspective on what efforts the industry has made to reduce energy consumption.

However, the results above are only reported for data centre owners and operators and do not delve deeper into the reasoning behind some of the issues. Reporting on the uptake of energy efficient technologies is beneficial, but improvements can only be made if the reasons behind not implementing them are discovered. Further research needs to be done into why the use of energy efficient technologies is not more widespread. The Uptime Institute surveys also do not go into the opinions of individual professionals within the industry. Gathering opinions on the different issues is also important to be able to provide incentives or interventions to drive further energy savings.

3.10 The Energy Efficiency Gap

Given the availability of different technologies and policies for decreasing energy use and the fact that uptake of these is not widespread, there is a gap between the opportunity for reductions and actual reductions. Just as in other sectors, this means that there is what has been called an *energy efficiency gap* (Sorrell et al., 2000, Jaffe and Stavins, 1994) in the data centre industry.

The term energy efficiency gap was first defined by Jaffe and Stavins in 1994 in order to describe the '*paradox of gradual diffusion of apparently cost-effective energy-efficiency technologies*' (Jaffe and Stavins, 1994). Jaffe and Stavins explain that this gap occurs when technologies which appear to be cost effective have very limited success on the market. Market barriers are cited to be the cause of this, i.e. problems within the market are acting as a barrier against the widespread use of energy efficient technology.

Whilst market barriers for a wide range of sectors have been examined thoroughly in the literature, it is also known that several other barriers exist contributing towards the efficiency gap. This has in turn led the energy efficiency gap to become broader in its definition. One definition of the efficiency gap given by Sorrell et al. is as follows;

> The efficiency gap is *'a gap between the opportunities for cost-effective energy efficiency investment identified in energy models and the levels actually seen in practice'* (Sorrell et al., 2000).

In 2012, Backlund et al. took the concept further and introduced the idea of an extended energy efficiency gap, adding a further definition to the term. Backlund's extended efficiency gap is based upon the fact that further efficiency gains could be realised if energy management practices were successfully employed. This definition brings an added social factor that is integrated into the problem alongside the market and technological factors. It demonstrates that in order for an efficiency gap to be reduced, economical, technological and social factors need to be accounted for and considered together. Figure 3.10 illustrates Backlund's concept that once any barriers preventing technological diffusion have been overcome, barriers to energy management also need to be considered in order to achieve the extended energy efficiency potential.

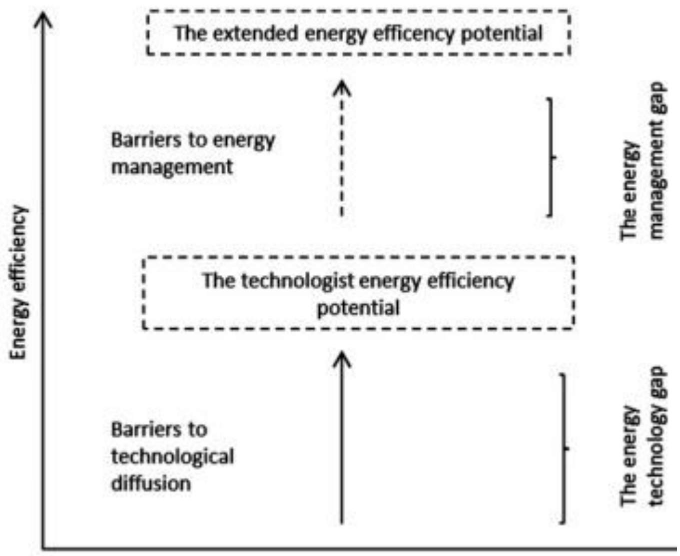


Figure 3.10 The extended energy efficiency gap (Backlund et al., 2012)

Buckland's definition of the extended efficiency gap is well suited to the data centre sector. As Koomey argued, a lot of the inefficiencies in data centres are due to a lack of best practice management, which is to say that they are *'institutional, not technical'* (Miller, 2013a). Therefore, when considering the barriers against energy efficiency in the discussion that follows, the energy efficiency gap will be considered as extended to include management practices. Even though it has been said that institutional problems cause a lot of the inefficiencies in a data centre, they may not be totally unrelated to the technical issues. A lack of technical expertise or lack of access to equipment (such as metering and data collection) would have an effect on the management of the data centre. Without access to a good level of metering in the data centre it is not possible to monitor energy use and therefore good management practices may not arise in order to reduce that energy consumption. Good energy management practice needs to be backed up with the required technology in order to support it. So whilst a lot of the problems of inefficiencies may be due to institutional rather than technical issues, the two can be very much interrelated.

3.11 Barriers to Energy Efficiency

The presence of an energy efficiency gap within any sector would suggest that there are barriers acting against the effort to increase energy efficiency. These barriers range from economic and technological to managerial and include many others. The following discussion takes a closer look at the literature surrounding energy efficiency barriers, also with reference to sectors outside the data centre industry. The main barriers are discussed along with the methods used to assess them and what can be done to overcome them. How these may apply to the data centre industry will then also be outlined.

3.11.1 Understanding Barriers Can Enable Change

If energy efficiency is to be increased, then it is very important to understand the barriers that act against any efforts. Understanding these barriers allows steps to be made to overcome them. Interventions can be made to encourage the uptake of energy efficiency, and these are mainly in the form of policy or market incentives. Policy interventions can cover pricing mechanisms, regulation, financial incentives, promotional mechanisms, organisational development and financial remediation. Market incentives can include funding schemes which provide money for investing in certain technologies, or they can be in the form of tax relief if energy use or carbon emissions are reduced.

3.11.2 Research Opportunities

Before going into detail into the barriers that do exist, either for the data centre industry or others, a summary is now provided in terms of the research opportunities. The first objective of this thesis is to investigate opinions about and barriers towards energy efficiency in data centres. Chapter 4 details the methodology behind this section of research, explaining that the literature is only examined in detail once the research has been carried out. Therefore, research opportunities and gaps in the literature are first discussed as a background to the work in Chapter 5, before heading on to discuss the specific barriers in more detail.

Currently, there is a lot of research available within the literature surrounding energy efficiency barriers. Stover et al. mentioned that market failures as a barrier have already been studied extensively and a lot of policies have been targeted towards this (Stover et al., 2013). This has meant that a lot of research has focussed on the economic side of encouraging further energy efficiency improvements, rather than any other aspects. Additional to this, Backlund et al. talk about how research has focussed heavily on the technological side of increasing energy efficiency and not so much on energy management practices (Backlund et al., 2012).

Cagno et al. suggest a detailed model for assessing the barriers, using previous research available in the literature. Part of their conclusions however was that more research was needed into barriers that are specific to certain sectors. They recommend that further research is needed into their barrier taxonomy by specific sectors, technologies and the size of the firm (Cagno et al., 2013).

Given this, the qualitative research work in Chapter 5 will cover the following:

- investigation of barriers in a specific sector, the data centre industry; and
- a focus on technological and behavioural barriers rather than market failures.

As discussed earlier, the Uptime Institute conduct annual surveys into the state of the industry and publish figures regarding energy efficiency. However the figures only cover data centre owners and operators, they also do not go into detail regarding the opinion of the industry. Therefore, Chapter 5 will also cover:

- qualitative research covering owners, operators, colocation providers and consultants; and
- research into the opinions of the industry regarding energy efficiency issue.

3.11.3 Taxonomies and the Main Barriers to Energy Efficiency

When assessing barriers, authors present them as a taxonomy and classify the barriers into different groups. Each author suggests different taxonomies for these barriers and they can vary from author to author, depending upon which sector they are assessing and what their main focus is. A taxonomy can however be applied to a different sector that it was originally proposed for, in order to test its relevance and to cross reference barriers.

A number of authors split barriers into several main categories. Sorrell et al for example (in their investigation into efficiency barriers within organisations) grouped barriers into three different perspectives, economic, behavioural and organisational (Sorrell et al., 2000). Chai and Yeo extended their definitions and group barriers into five categories; economic non-market failure or market barriers, economic market failure, behavioural/institutional, organizational and physical constraints (Chai and Yeo, 2012). Sudhakara Reddy took a slightly different approach and introduced a new taxonomy classifying the barriers as micro, meso and macro. Micro barriers occur at a low level (i.e. design stages), meso barriers happen at the organisational level and macro barriers act at the highest level (i.e. state, market and civil society) (Sudhakara Reddy, 2013). Whilst there are many ways of presenting these barriers, they tend to be mainly placed under the same broad headings, economic, behavioural and organisational.

Due to the variation in the way that barriers are defined, Cagno et al argued that better definition is required so as to avoid any overlapping or interaction. The authors say that a clearer view is needed to distinguish between real and perceived barriers to enable policy to be more effective (Cagno et al., 2013). Within their taxonomy, they group barriers into two distinct origins, *external* and *internal* which are then divided into several areas that are further subdivided into barriers. This method of identifying barriers provides an easier way to address issues and to overcome them. Being able to identify easily whether the barrier is external or internal to an organisation or process is essential if proper measures are to be created to overcome them.

Trianni et al (Trianni et al., 2013) use the taxonomy put forward by Cagno et al (Cagno et al., 2013) to place empirical values on barrier analysis. They form a questionnaire which asks questions aimed to address each of the barriers in the Cagno taxonomy. They then use ratings such as the Likert scale[15] to rate the answers from the respondents. This allows the authors to place an empirical value on the barriers included, hence providing a rank of barriers in terms of importance and influence. This is important because it allows the authors to identify the most important barriers, in this case economic and information barriers in order to see where action needs to be taken.

Chai and Yeo discuss the fact that some researchers have attempted to rank barriers in relation to their subject of study. However, they argue that the results from this would not be able to be applied to other sectors or geographical areas (Chai and Yeo, 2012). However, whilst this may be the case, it could be argued that specific industries or technologies should be looked at like this as each one will have their own boundaries and each situation is different. The framework put forward by Chai and Yeo is specific to one industry, so the authors write that their taxonomy needs to be tested and trialled for other industries. Whilst it is true that some barriers cross over between industries

[15] A scale often used in surveys whereby the participant can select an answer that sits on a visual scale between (for example) satisfied or un-satisfied.

and that taxonomies may be able to be applied to more than one sector, it is important to consider the individuality and needs of each industry.

Whilst there are differences between the ways that barriers are presented and assessed, in the main authors investigating energy efficiency barriers tend to be in agreement as to the main ones. In general, the barriers have been studied in depth within the related literature and are well understood. In 1990 Hirst et al. predicted that over the two decades following 1990, the USA would only realise half of its potential to increase energy efficiency (Hirst and Brown, 1990). Causing what has become known as the energy efficiency gap, structural and behavioural barriers were listed by the authors as being the main two groups of barriers acting against implementation of energy efficient methods and technologies. Within these two cited behaviours, standards, attitudes, perceived risk and information gaps amongst others were also mentioned.

Stover and Sachs report on '*cryptic barriers*' affecting energy efficiency, which they define as being barriers that are unrecognized. These barriers are often tied into building standards rather than being connected to market failures (Stover et al., 2013). After identifying several of these cryptic barriers, the authors focus on examining some in detail, one of which serves as a good example demonstrating how barriers affect the effort to increase efficiency. The authors examined the reasons behind why plastic ducts are not being used in residential situations. Plastic ducts can increase energy efficiency because of the presence of a very smooth internal surface, they are also easy to fit and have minimum risk of leakage. All of these factors would be beneficial if plastic ducting were to be used, however the authors discuss that because they are not being used, then there must be something preventing or discouraging their use. The main barrier that they identify is the flammability of plastic, as ductwork can act as a conduit along which smoke and flames can travel. This is labelled as a barrier because code restrictions are unclear about the use of plastic ductwork for residential purposes. Whilst the authors label this as a cryptic barrier it could, as a definition, fall under the area of policy as a change in policy or standards would enable the use of these ducts. Mechanisms such as technological (i.e. the use of fire proof materials) and policy changes are suggested as a means to remove the barrier to the use of plastic ductwork. The authors also point out that even if such mechanisms were put in place, the ductwork would need to gain acceptance from the industry and also consumers. This is an important point to remember, as acceptance can act as a barrier itself.

3.11.4 Barriers Specific to Data Centres

With regards to barriers that are specific to the data centre industry, there have been some studies into this area. Loper and Parr state that the cost of implementing energy efficient technologies is one of the biggest barriers to their use in data centres (Loper and Parr, 2007). They quote the energy costs of data centres being a '*relatively small portion of overall costs*', especially if the facility is only small (Loper and Parr, 2007). This is in reference to the fact that the performance of a data centre is not measured on its energy efficiency, rather its ability to deliver secure and reliable IT services. Budgets set aside for innovation and investment in new technologies can also be a problem. Research shows that IT companies only tend to spend 20% of their budgets on planning and innovation, most of the budget is spent on just keeping exiting operations running (Crane, 2013).

Aside from costs, Loper and Parr also refer to the problem of data centre managers not necessarily being responsible for paying the power bill for their facility (Loper and Parr, 2007). This was reflected in the 2012 Uptime Institute survey which found that whilst 57% of respondents considered energy use reduction to be very important, they also found a disconnect between recognising this importance and acting upon it (Stansberry and Kudritzki, 2012). The survey found that this disconnect comes mainly from the same point that Loper and Parr raised, i.e. whilst it is the IT department within the company using the energy, they are not the ones to normally pay for it. Likewise, the department paying the energy bill does not necessarily have an understanding of the opportunities to save energy. In the following 2013 Uptime Institute survey, it was found that the power bill for a data centre is paid for by the facilities team in 80% of cases and by the IT in 16% of cases. Highlighting the problem of energy not necessarily being a priority considered by IT departments (Stansberry, 2013). This disconnect between priorities was also recognised in the results of another questionnaire which was handed out to IT facility managers (Big Switch Projects, 2004). In this survey (given to data centres within commercial offices, and completed by IT and facilities managers), the main barriers to energy efficiency given as the following:

- availability;
- downtime risks;
- equipment damage risks;
- lengthy set up times;
- IT technology constantly increasing in heat load and power consumption;
- standard practices in the industry (such as a 20°C set point);

- poor design;
- costs;
- lifetime of equipment; and
- lack of knowledge about energy bills.

These were summarised into the following four major barriers:

- risk aversion;
- standard industry practices;
- technology; and
- a disconnect between energy costs and operation.

This report also highlighted the difference in priorities between different parties. The majority of IT managers interviewed cited availability as a top priority, but facilities managers listed it as a barrier to achieving better energy efficiency. The priority for the IT manager is to deliver a constant and reliable service, whereas the facilities manager will be responsible for the energy bill, the two are essentially at odds with each other. The other problem is that the person responsible for ordering new equipment is not necessarily the person paying the energy bills meaning that energy efficiency may not affect their choices (Golden, 2013). This type of barrier has been well known for a long time, in 1980 Blumstein et al termed it as '*misplaced incentives*' defining this as the person benefiting from a reduction in energy use is not always the person reducing it (Blumstein et al., 1980).

Another big priority for data centres is to ensure that the IT equipment remains up to date. With technology constantly improving and new hardware becoming available it remains difficult to continually update the infrastructure to keep up. As the infrastructure within the data centre (e.g. the cooling system) has a longer operational life than the IT equipment, this can lead to a disparity between the requirements of the IT equipment and what is supplied by the cooling system (Loper and Parr, 2007). Hence inefficiencies arise due to this disparity. Quickly changing technology was also identified as an issue in the 2007 EPA Report to Congress who also cited the following as barriers (U.S. Environmental Protection Agency, 2007):

- first cost;
- split incentives;
- risk aversion;
- learning curve;
- quickly changing technology; and
- lack of energy monitoring.

Learning curves and access to information is a barrier which has also been highlighted by other authors (Energy Star, 2012, Energizing Tech Blog, 2010), this covers a wide area including a lack of knowledge about available technology and also energy issues. Knowledge about the issues comes from keeping up to date with the latest technology offerings, but also making sure that staff are trained in efficient operation of equipment and the issues with energy use. Another way that information and knowledge is passed on in the industry is through organisations learning from others that have adopted new technologies and operational processes (Loper and Parr, 2007), perhaps through attendance at industry conferences or publication of white papers. However, competitive advantages gained through innovation are unlikely to be shared in detail.

Besides the problems with communication and information transfer, other barriers within the data centre industry have been cited as legacy facilities (Soderman, Unknown), small server rooms (Miller, 2013a), data collection (Germain-Renaud et al., 2011), high capital costs (Energy Star, 2012) and risk aversion (Loper and Parr, 2007). Within the industry, there are a lot of legacy data centres and also smaller server rooms within buildings. Miller argues for a move towards the cloud (large centralised data centres) which would remove the need for small inefficient in house server rooms (Miller, 2013a).

Risk aversion is an important factor when examining energy use improvements, as it is often noted that reducing the risk of outages is a much higher and more crucial imperative for data centres rather than energy efficiency. An inefficient data centre will have a much smaller impact than the loss of power and data. Difficulties with monitoring and collecting data for each individual component of a data centre, as discussed by Germain-Renauld, causes problems with gathering accurate information on energy use (Germain-Renaud et al., 2011). Installation of energy monitoring systems such as DCIM (Data Centre Infrastructure Management) is becoming more common; however the setup of this system is still expensive. This problem with data collection was brought up in a workshop where industry members were asked to identify problems they had encountered when attempting to quantify and assess energy use in their facility. Several issues were identified, including the following (U.S. Department of Energy and U.S. Envrionmental Protection Agency, 2008):

- access to data regarding the efficiency of infrastructure equipment;
- converting technical metrics to business metrics and incentives;
- a lack of instrumentation for data collection;

- a lack of a standard definition for data centre productivity (also a lack of definition for useful work); and
- computing output is not accounted for nor are energy sources.

All of these problems make it difficult for a data centre to quantify things such as useful work done or the productivity of the facility, to link energy metrics with business benefits and to get a true picture of energy use and overall environmental impact.

In order to overcome these barriers within the industry, Loper and Parr identify the following policy recommendations (Loper and Parr, 2007):

- data centre energy use metering;
- measurement of energy performance;
- energy performance standards;
- building codes for data centres;
- financial incentives;
- research into metrics, efficient coding and the benefits of '*better applications management*';
- implementation of government programs; and
- awareness of information.

However, taking into account the work by Germain-Renauld (2011) work needs to be done regarding the measurement and presentation of energy use before certain barriers can be overcome. Therefore, with the suggestion for measurement of performance and performance standards, it is also important to ensure that the right metrics and available information are in place first.

3.12 Conclusions and Research Themes

The current chapter has reviewed the literature surrounding energy efficiency within the data centre. Following on from the background literature reviewed in Chapter 2, this chapter has discussed the measurement and presentation of energy use, methods through which energy use can be reduced and finally, barriers which may be preventing further energy savings. The following concludes the findings from this literature review, with an emphasis on metrics and barriers which will form the focus for this thesis.

3.12.1 Measuring and Presenting Energy Use and Efficiency

A review of the literature has revealed that there is a large array of metrics available for assessing various aspects of energy use and efficiency as well as thermal properties.

Despite this wide range of availability, it would appear that most of these tend to be used for academic rather than industrial purposes (this is reflected in the interview participants responses discussed in Chapter 5). Newer metrics tend to focus on the need to more closely assess the IT equipment itself within the data centre. However, difficulties arise here when attempting to assess the useful work that servers carry out.

Within the industry, the most commonly used metric is the PUE metric, a measurement which allows the user to find the total energy use of the data centre as a proportion of the IT energy use. Some authors have attempted to conduct studies into worldwide figures for PUE measurements; however this has been a difficult task due to a lack of available data. Some have also briefly discussed the fact that there may be a few problems with the use of this metric.

Part of this review into energy consumption by data centres has also highlighted a lack of extensive energy consumption measurements that run for an annual cycle. The review shows a lack of detail in some studies for calculation of the energy efficiency in a data centre. Many authors use some form of power density, unfortunately this metric does not allow two facilities to be compared if they have different floor areas and IT loads, and it also does not directly show the energy efficiency of a data centre. PUE has been reported in a limited number of papers, but it is not necessarily critically assessed in terms of how useful it is as a metric and any problems there may be with its use.

There is a lack of literature available which describes the method of calculating the PUE in the detail needed to reproduce the calculations leading to its stated value. Any values that are presented are also subject to strict privacy measures due to the sensitive nature of the industry, and the tendency of companies to be somewhat elusive about their designs means that data centres often have their power consumption kept hidden from the public domain (Mitchell-Jackson et al., 2003). If energy use and efficiency metrics are to be improved, it is important that reporting is transparent. Chapter 6 will therefore delve further into the issue of lack of access to data and the problems that there may be with using the PUE metric. By using open source information, the case study described in Chapter 6 attempts to measure a PUE value with limited access to data. It has also been found that there is a general lack of detailed energy consumption studies available in the literature, and there are not many that encompass a whole year. The study in Chapter 6 therefore takes into account changes in the weather over a year and provides an energy use measurement for an annual cycle.

Regarding other schemes that are available (such as the EU CoC and BREEAM), it is unclear as to which of these are actually used in the industry and how they are received.

Therefore, in order to find out more about this, the interviews conducted within Chapter 5 will include questions asking about the interviewees' use of such schemes.

3.12.2 Barriers to Energy Efficiency in the Data Centre Industry

This chapter has introduced the idea of the energy efficiency gap, a phenomenon whereby although there are plenty of efficient technologies and measures available, certain industries are still not as efficient as they could be. Barriers act to prevent further efficiency improvements and decreased energy use. There are a significant amount of detailed studies existing within the literature based upon barriers to energy efficiency within various sectors. Several different taxonomies have been suggested by many authors, some focussing on certain industries or geographical areas. Some authors have said that further research needs to be done in terms of closer consideration of certain industries.

A number of authors have suggested different barriers in the data centre industry (although this available research is somewhat limited) such as upfront costs, a disconnect between the people buying equipment and those responsible for the energy bill, and a lack of skills and knowledge. Despite this, investigations into the barriers that may be preventing further energy efficiency savings in the data centre industry are currently limited. This may be because of the closed nature of the industry; data centres operators and owners tend not to share very detailed information about their infrastructure or energy usage. Having said this, in recent years there has been a trend to become more competitive over energy use and this is helping to drive more openness in the industry. This can be seen in particular for the large companies that own their facilities (e.g. Google and Facebook), who pay particular attention to advertising their use of energy efficient infrastructure.

There have been some major studies into the level of energy efficiency of data centres throughout the world (although mainly in the USA); however none of these delve into the reasons why certain efforts have not been made. They also do not necessarily include surveying all actors directly involved with the data centre industry. Therefore, the research within this thesis aims to find out if there are any other barriers which need to be considered. Another key aspect of this research will be to gather opinions from several different participants within the industry including colocation providers, owner operators and consulting engineers.

Some authors working in the area of barriers to energy efficiency in other industries have also pointed out the need for a focus on particular technologies, in terms of the reception of the industry to them and the barriers that may be preventing their use. Therefore, the research within this thesis will also cover the opinions and reception of a certain energy saving technology. Chapter 4 explains the qualitative methodology that will be used for this investigation, with the results being presented in Chapter 5.

3.12.3 Summary

Energy efficiency has been identified as a key issue within the data centre industry today. As the demand for data centre services increases and IT devices become more powerful, it is crucial that measures are taken to increase the energy efficiency of these facilities. Technological, behavioural and policy-related measures can all be taken to help increase energy efficiency levels, however barriers exist to prevent these from happening or being successful.

Whilst the qualitative methodology described in Chapter 4 will cover barriers, it will also gather general opinion from the industry regarding their attitudes towards energy efficiency and the efforts they have made to reduce energy use. This is important as whilst the literature review shows that the issues surrounding energy efficiency are well understood from an academic sense, it is important to also find out whether they are understood in the industry. Other gaps in the literature include what efforts have been made to reduce energy use and what the drivers behind this have been, research also needs to be done to find out how much of an influence the users of data centres have over their efficiency. Whilst the Uptime Institute has gone someway to find this out by carrying out a survey[16] which included a large number of participants, the interviews in this work will be able to uncover information in more detail.

[16] It should be noted that this survey was not part of the peer reviewed open literature and should therefore be considered with caution.

4 Qualitative Research Methodology

The present chapter describes the qualitative research methodology used to obtain the results discussed in Chapter 5. Chapter 5 addresses the first objective of this thesis by collating opinions and views from industry professionals on various aspects of energy use and efficiency in data centres, with a focus on the barriers they face. This chapter discusses the ethical considerations, the methodology employed and the design of the research interviews.

4.1 Introduction

The first objective of this work is to interview professionals in the data centre industry about issues surrounding energy efficiency. Part of this objective is to try and find out the barriers that may be acting within the industry, making it difficult for further savings in energy use to be made. An important overall aim of this research is to gather information from different professional members of the industry, to see how opinion varies from one to the other, depending on what their specific role is in the industry. There will also be a focus on the industry opinions about a certain set of technologies that can help to reduce energy use in the data centre. It is identified that qualitative research methods are more appropriate than quantitative methods for this study as the aims are to investigate opinions and knowledge rather than statistical figures. Interviews are the most suited to this as they allow in depth discussion of the issues and the questions are open ended. It also allows the discussion to evolve as the interview takes place, whilst making sure not to digress too much.

Whilst there is already significant research into the barriers to energy efficiency, little of this focuses specifically upon the data centre industry as discussed previously in Chapter 3. There has also been little research carried out on the opinions of industry professionals about specific technologies. It is important that these barriers and opinions can be understood; otherwise further steps cannot be taken to reduce energy consumption. From the discussion in Chapter 3 it is clear that there are plenty of options, methodologies and technologies available to increase energy efficiency within the data centre. Despite this, the industry is still facing issues of increasing energy consumption. Whilst increasing demand for access to services such as the internet will play a big part of this, it is clear that energy efficiency methods are not being used to their fullest capabilities. Due to this, it is important that the issues holding back these additional savings are investigated and discussed.

4.2 Qualitative Research Aims

The methodology introduced in this chapter addresses the first objective of the thesis as set out in Chapter 1;

> *'Investigate the industry opinions of, attitudes towards, and efforts made relating to energy use in the data centre. Find out if any barriers exist that may be preventing increased savings. '*

Through using the qualitative research methodology described here, Chapter 5 addresses the above aim. Opinions of the industry when it comes to improving energy efficiency in data centres are investigated by holding interviews with data centre professionals. These interviews have been used to gather opinions and perspectives from different professionals within the industry, to paint a better overall picture of the industry efforts to reduce energy use, gather opinions about energy efficiency and discover any barriers that may be preventing further savings.

The main questions to be answered in this study are the following;

- What is the participant's view on improving energy efficiency in data centres?
- What are the participant's motivations for increasing energy efficiency in this industry?
- With regards to liquid submersion cooling in particular;
    - What is the industry knowledge/opinion of the technology
    - What incentives would the industry need in order to consider the technology?

This study is unique in that it will take into account the viewpoint of engineering consultancies alongside owners/operators and also aims to place some focus on one particular technology. The proposed outcomes of this study are as follows;

- A synopsis of the industry opinion of this emerging technology,
- A comparison of opinions between different parties within the industry regarding energy efficiency in the data centre,
- An insight into how much influence engineering consultancies have over energy efficiency in new or retrofitted data centres, and how the tenant/landlord relationship in colocation data centres affects the drive to reduce energy use, and
- Recommendations for how to overcome barriers or how to encourage use of energy efficient technology.

Overall, the study is concerned with understanding how much the industry knows about issues surrounding energy efficiency and how far they have gone to deal with energy use. By interviewing representatives from within the industry, this investigation will find out what they perceive to be good efforts towards reducing energy use. It will measure their understanding and find out the reasons behind their efforts.

4.2.1 Ethical Considerations and Participant Recruitment

It was important from the outset of this work that the ethical implications were fully considered. Confidentiality was of course a very important consideration throughout the research process. The data centre industry tends not to be very open about details of their facilities due to the nature of high security in them. Therefore it was of high importance to explain to the participants that their data would be kept confidential and anonymous. Interview data was collected for analysis using an audio recording device (Zoom Handy Recorder H2) so that it could be transcribed at a later date. Once each interview was completed, the audio file was removed from the memory card within the recorder and kept securely on a password protected computer. In this way, care was taken to ensure that each audio file did not remain on the portable device for longer than absolutely necessary. In order to maintain anonymity, it was also important to ensure that each interview recording was not attributable to a personal or company name. Therefore once the audio files were removed from the audio recording device, they were given names such as Colo_1 or Con_3 so that individual or company names were not used. Whilst all participants were reminded at the beginning of the interview that any information given would remain anonymous and that they need not mention any names, some chose to mention names which would attribute the data either to themselves or their company. Therefore, during the transcription process any names that were mentioned were blanked out so that no one could be identified.

In terms of searching for and recruiting participants, they were found both through existing contacts and also by searching for relevant companies on the internet. It is recognised that by recruiting participants through existing contacts, some amount of bias may be introduced into the study. This would be because existing contacts may already have a strong interest in issues surrounding energy efficiency in data centres. However, most of the existing contacts were for colocation providers, and they still provided interesting information regarding the attitudes and efforts of their customers. Also, it still allowed for comparisons of the level of efforts made between them, and the difficulties which they faced.

Initially, participants were contacted with an email asking their interest in joining in with the study. Attached to this email were an information letter and a consent form to take part in the research (an example letter can be found in Appendix B). The initial invitation email briefly described the research project and what would be required from the participant. The information letter contained further details about the project and exactly what would be required from them for the interviews.

Appointments to hold the interview were generally arranged via email, although some were done via the telephone as it was easier for the participant. Before each interview took place, the participant was asked to sign a consent form to confirm that they understood the process, that they understood the fact that everything would remain confidential and also that they consented to the data being used for research purposes. Before the interview commenced, the interviewer made sure that the participant had read the information letter, and that they fully understood everything involved in the process. The interviews were conducted during working hours either at the participants place of work or over the telephone. The choice was left up to the participant so that they were comfortable with the arrangements.

4.3 Research Methodology and the Application of Grounded Theory

There are a few different methodologies that can be used to create the data collection and analysis process. Methods can vary from beginning with an extensive literature review which then influences the way the research is conducted, to collecting the data first before developing any theories and doing a literature review to compare the results[17]. This latter approach allows the data, rather than the literature, to have a larger influence on the final theory and results. After an initial literature review in the beginning of section 3.11 considering the gaps and opportunities for research, it was decided that the qualitative data should be collected and analysed before moving onto a theory from the results and a more detailed review. This more detailed review can be found in the rest of section 3.11 in Chapter 3.

This method of qualitative research can best be described by *grounded theory*. Grounded theory is a methodology that is used to design and analyse qualitative data,

---

[17] An in depth discussion of the merits of each qualitative data analysis methodology is outside the scope of this thesis. There is already a lot of literature available on the subject of methodologies for qualitative research.

it can be defined as *'a systematic theory that is* grounded *in, or based on, the observations'* (Schutt, 2012). This methodology for assessing qualitative data was devised by Glaser and Strauss in 1967 as a way to ensure heavy involvement of the data in the formulation of a theory, this was initially applied to the field of sociology (Glaser and Strauss, 1967). The main features of this theory are that it allows for *constant comparative analysis* where results can influence the next round of data collection and *theoretical sampling* whereby the sample size is not fixed and is allowed to saturate.

4.3.1 Theoretical Sampling and Sample Saturation

Theoretical sampling means that the sample size is not decided as part of the research design (Marshall, 1996, Glaser and Strauss, 1967), the exact number of participants is only known at the end and depends upon the data collected as the research continues. Sample size normally tends to be a lot smaller in qualitative research than that for a quantitative study (Mason, 2010) although exact recommendations for size tends to be vague in the literature. In 2006, Guest et al reported that at that time, there was no literature available giving a robust method of arriving at a suitable sample size. Articles that were available tended to be vague and there were no published guidelines available (Guest et al., 2006). However, what is known is that qualitative research is a time consuming process and the advice is that sample size should be kept to the minimum level required to find out opinions. The sample needs to be large enough to ensure that enough opinions are gathered to make a good assessment, but not too large that opinions are simply repeated. Saturation of the sample is a concept used in grounded theory to give the researcher an idea of when the sample size is large enough. Mason describes the sample to be saturated when *'new data does not shed any further light on the issue under investigation'* (Mason, 2010). This means that rather than having a predetermined sample size, the research simply stops once no new information is found from the gathered data, or themes are repeated (Marshall, 1996). Whilst this method of sampling is widely accepted, there is some debate as to what is the most suitable and robust sample size for qualitative research interviews. Guest et al argue that the concept of saturation does not provide enough guidance to the sample size, especially at the initial stages of planning the research (Guest et al., 2006). It was decided at the beginning stages of this research that interviews would need to take place with around 5 of each representative within the industry, so 15 interviews in total[18]. The principle of saturation would be applied in that if the data were to saturate before this, then no more were needed and the sample size would continue to expand until enough saturation in the data was achieved.

In the case of this research study, the sample of participants could not be selected randomly due to the nature of the subject area; each participant had to be knowledgeable about the topic in hand. Therefore specific companies were targeted to take part in the research, something for which grounded theory allows. The participants in this study were recruited because they had specific knowledge, which can be known as a key informant sample (Marshall, 1996).

4.3.2 Constant Comparative Analysis and Coding

Grounded theory allows an iterative process whereby data collection and analysis are carried out concurrently (Zafeiriou, 2000), also known as the *constant comparative method* (Glaser and Strauss, 1967). This way, a theory is developed as the data is collected rather than being assumed at the start of the study. Glaser and Strauss write that this is used in conjunction with the theoretical sampling, this allows for an iterative research process and one which is constantly developing (Glaser and Strauss, 1967). This method of analysis means that the process of *coding* can begin whilst data is still being collected and also can continue once analysis of the data begins.

Coding is the initial stage of qualitative data analysis and is a method employed to interpret the data. Essentially it is a means to categorise the transcribed interview data and to aid in the development of a theory (Flick, 2009). Codes will be used several times when looking at a large set of data such as interview transcripts, and simultaneous coding can occur where two or more codes are applied to a single reference within the text. This coding analysis is an iterative process and whilst they may change as more data is added, the codes and categories tend to become more refined over time (Saldana, 2009).

Coding is a staged process and it generally involves three stages as illustrated in Figure 4.1, although there is some debate as to which those three stages are (Walker and Myrick, 2006). A code itself tends to be a single word or short phrase that represents a portion of the data (Saldana, 2009).

[18] In the end, 6 colocation providers, 3 owner operators and 2 consultants were interviewed. The sample did saturate at this point. Difficulties were faced when trying to contact more consultants to take part in the research.

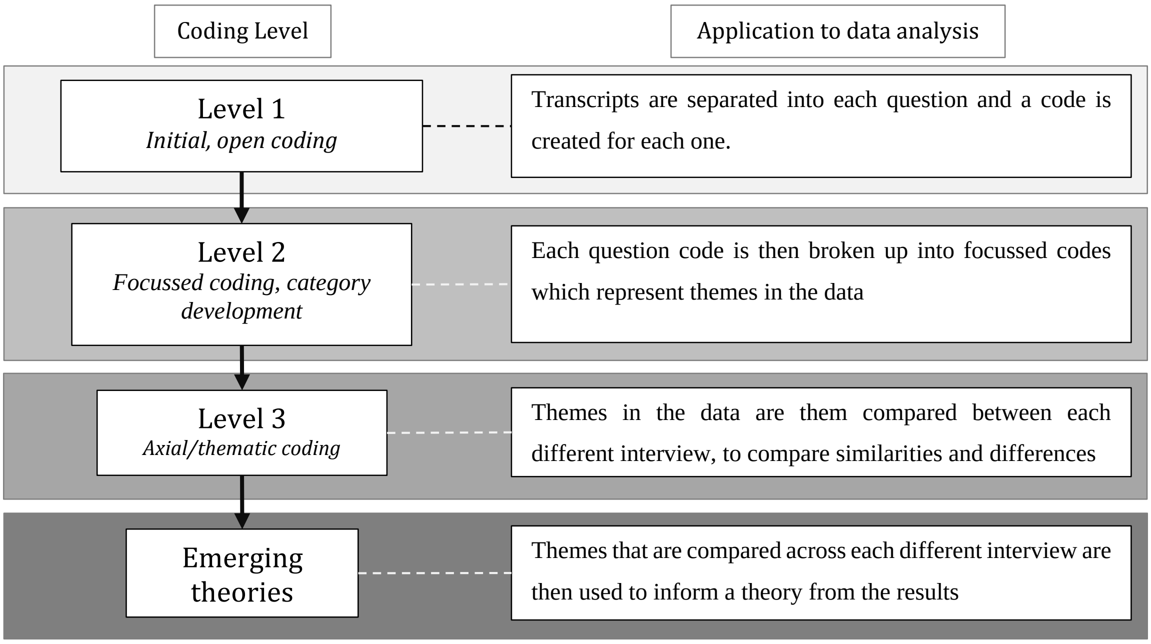


Figure 4.1 Coding categories according to Hahn and how they apply to this research (Hahn, [no date])

The first stage of the coding process will be to separate each different question within the interview transcript and give each of those a code. Once answers to each question have been separated, focussed coding will be used to identify themes within the data. Themes may reoccur across several answers depending on what was discussed by the participant. Finally, axial coding is applied whereby the themes arising from focussed coding are linked together. This method of coding allows the data to be broken down into main and then sub-themes before allowing cross referencing between each interview.

The level on which the transcribed material is coded depends upon several factors including the nature of the research question, the data being analysed, and also the stage of the research being carried out (Flick, 2009). This research does not require such a level of detail as coding every word in the transcription, and it may not require every sentence to be done. The approach taken was to remain flexible and interpret each sentence or paragraph as necessary. The level of coding detail may also decrease as the number of interviews increases, as common themes are likely to arise.

Another important element of grounded theory is the use of memos. Memos are used alongside the coding process to add thoughts to the interview data as it is being analysed. This allows the researchers thoughts to be recorded as the data is processed. Memos which have been written for the data gathered in this research have informed the data analysis in Chapter 5 and also the alterations made to the interviews as the research progressed (see 4.4.4).

4.3.3 Process Model

The process of designing interview questions, collecting the data, processing the data and then collating the results is a continuous one when employing Grounded Theory. Each stage of the process feeds into the next and changes in one can affect the other. It is in this way that it is an iterative process, a process that can be altered as the research continues and evolves.

Strauss provides a useful illustration of the process, or coding paradigm, that qualitative research follows under grounded theory (Strauss, 1987). Figure 4.2 demonstrates how each phase of the research feeds into the next, but each one can be influenced by the other leading to a circular and interlinked process that allows the research design to develop alongside emerging theories.

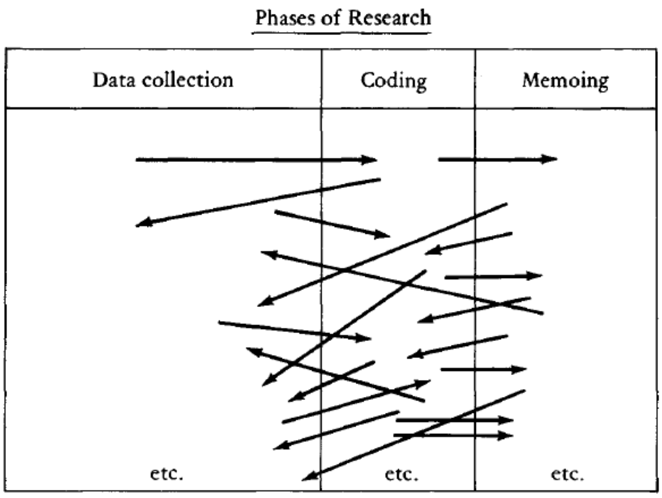


Figure 4.2 Coding paradigm for grounded theory as illustrated by Strauss (Strauss, 1987)

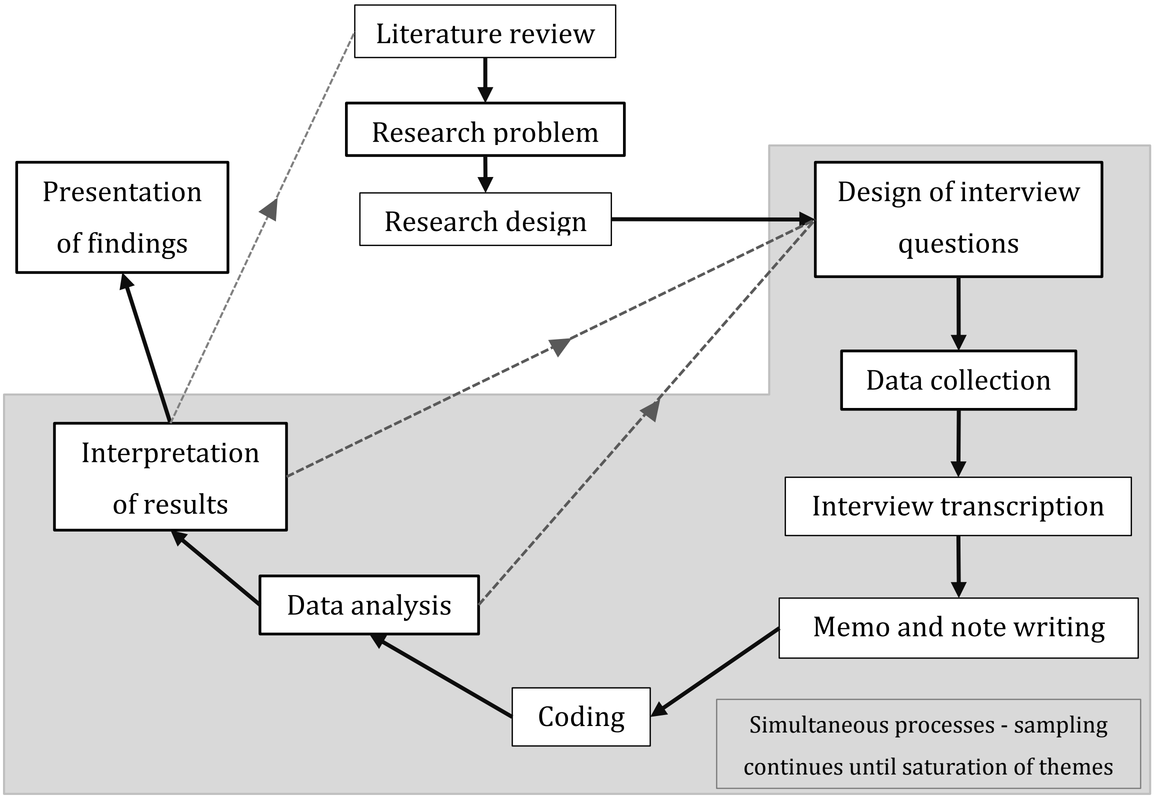


Figure 4.3 Research process incorporating grounded theory

Figure 4.3 illustrates the process that this research study follows. Based upon other diagrams from other authors, it shows the stages of research. A lot of authors depict the research process using Grounded Theory as a continuous and circular process. Whilst this is true, the interpretation of the process here allows for feedbacks between each stage and includes the beginning and end phases which do not follow a circular pattern. The stages of the process that are within the grey area are all conducted simultaneously. Grey dashed lines indicate points at which different stages feed into one another. Once the interviews and data collection commence, the interview questions can be redesigned and altered if necessary should new queries come up in discussions. The research can be designed once an initial literature review is carried out; the literature review can then expand as the results are analysed and interpreted. This relates to the point made earlier about not doing a detailed literature review to start with to avoid informing the results of the research. Throughout this continuous process, the sample size remains flexible and increases until a point in the interpretation of the results where it is decided that enough data is collected to warrant sample saturation (see 4.3.1). This will only be clear once the majority of the interpretation of results is completed. Once this is finished, the results were written up and concluded.

4.4 Development of the Interview Structure

Again there are varying ways in which an interview can be written and many different forms of interview questions. These can range from a conversational interview and open-ended questions to a guided interview and closed-end questions from which the participant must choose the most appropriate answer. The format of questioning which is chosen must be suited to the interview purpose as it influences the quality and type of answers that are gained from the participants (Keats, 2000). An interview with closed-end questions would not be suitable for this study as it would not allow discussion but only allow for specific answers which would limit the results. Whilst it would allow for some standardisation between the answers given by each participant, it will limit the answers that they can give. On the other hand, a completely informal and conversational interview would have no common structure between each participant and would make data analysis difficult. Therefore it was decided that a semi-structured interview consisting of open ended questions would be the most suitable option for this study.

4.4.1 Interview Aims

The aim of the interviews was to question the participants on three main themes:

Opinions - investigating the participants opinions of energy efficiency and consumption in the data centre industry;

Efforts - finding out the lengths to which the participant has addressed the issue of energy use in data centres, and

Barriers - problems they may have had in trying to implement energy saving methods.

To ensure that each of these three main themes were brought up in the interview, a script was written with main and sub questions (Appendix A). Within each section of the interview there were a set of main questions to be asked, sub questions were provided should the respondent not give a full answer as expected. Although the interview was scripted, it was not followed strictly and conversation was allowed to flow naturally i.e. the questions were not always asked in the same order from one interview to the next, and not all sub questions were needed.

4.4.2 Avoiding Bias and Leading Questions

When writing interview questions it is important to avoid bias in the wording used, the interviewee should avoid leading the participant to an answer. The interview scripts were written with this in mind, however due to the nature of the study it was impossible to avoid all bias and leading questions. From the outset of inviting participants to be involved in the study, they were aware that it is focussed on energy efficiency and energy use in data centres. Add to this the fact that the participants are representing their companies, those who are not interested in the issues or have some stake in them are not likely to get involved in the research. There were also some difficulties in writing questions that would both lead to good quality answers and were not leading, this particular point was mostly concerned with the section on barriers on the interview. This problem led to several rewrites of the relevant questions, as discussed further in 4.4.4.

4.4.3 Interview Structure

The interviews were split into five sections, introductions, energy use and efficiency, barriers, policies and liquid cooling technologies (A full interview transcript example can be found in Appendix C). The wording for each question along with a description of the purpose of each one is given alongside the data analysis in Chapter 5. The following is a brief description of each section of the interviews;

Interview opening - Before the audio recording device was turned on, the importance of confidentiality and anonymity in the research was reiterated to the participant and they were asked whether they had read the information sheet. They were also asked to sign the consent form if they had not already done so. The participant was then given a chance to ask any questions about the interview process or the research project. Once these had been covered, the recording was started.

Section 1 - **Introduction** - First of all, an introduction to the participant and their company was made. Introductions involved questions based upon the respondents company and their role in the data centre industry, the facilities that they ran and their job description. This served two purposes, firstly to establish basic information to provide a background to the respondent, and secondly, to build a rapport between interviewer and respondent. This is important in order to make both feel more at ease before moving onto more in depth topics.

Section 2 - **Energy use and efficiency -** Secondly, questions were asked regarding energy use and efficiency. This covered questions asking what efforts (if any) they have made to increase efficiency and decrease energy use. This section was important to get an idea about the level of changes they had made (if any) and what they perceive to be the best methods to implement. Questions within section 2 were different depending on the participant and whether they were a colocation provider, consultant engineer or owner operator. It was also important in this section not to just ask the participant if their data centre practices are energy efficient, this may lead to them giving '*socially acceptable responses*' as identified by Trianni et al in their study on barriers to energy efficiency (Trianni et al., 2013). Instead, they will be asked about their approach towards energy efficiency and their opinions of the issues.

Section 3 - **Barriers -** This section of questions was concerned with the respondent giving information about any barriers that were preventing them from introducing methods to increase energy efficiency. Questions also addressed any problems that they had implementing anything in the past. This section went through quite a few alterations as the interviews continued, as the questions were not garnering detailed answers, this will be discussed further in 4.4.4.

Section 4 **- Policy -** Policy and regulation are important tools for enabling reductions in energy use. Therefore, participants were questioned on their awareness of them (mainly the EU CoC), whether or not they used them, and their usefulness. Care needed to be taken here not just to question the participant on their awareness and knowledge; Trianni et al suggest that this can be seen as evaluating the respondent themselves. Instead, they suggest that participants should be questioned on their practices and behaviours (Trianni et al., 2013). However, for the current study it would be both interesting and useful to find out the participants awareness. Care was therefore taken over the wording of questions in this section.

Section 5 - **Liquid cooling -** Finally, the respondents were asked their opinion about liquid cooling technologies and then their thoughts on the future of the industry.

Interview closing - The audio recording was stopped and the participant was thanked for their time. They were also given another chance to ask any questions about the research and it was made clear that they could get in contact if they had any questions in the future.

4.4.4 Alterations during Data Collection

As interviews were conducted and transcribed, it was important that the script remained flexible to changes. Several questions were reworded throughout the research process and some were dropped from the interviews completely. The process was treated as an iterative one whereby after each interview alterations that were needed were made before the next. This allowed the interview script to develop as the research was carried out.

A pilot interview (the transcript for which can be found in Appendix C) was carried out initially and although it went well, changes needed to be made to the wording of some questions. A couple of questions were also left out from subsequent interviews. After the first set of four interviews, the questions were reviewed and adjusted depending on the data collected. Re-wording of the questions was not extensive; it was more to fine tune them to allow them to be better understood by the participant, and to garner better quality answers. New questions were also added to address themes which arose as a result of questioning in the first interviews; this was mainly more questions about the participant's clients views on energy issues.

After the first few interviews, it was also realised that it was important to distinguish between energy efficiency and energy use. This improved the discussion and allowed the respondents to talk more about reduced energy use rather than just efforts to increase efficiency.

Section 3 of the interviews concerning barriers to energy efficiency proved to be a difficult part of the interview. In order to avoid a leading question, the question was written with as little information as possible to allow the participants to suggest what barriers they thought were acting. However, it was found that the question '*thinking about the measures or policies you may have put into place, did you encounter any difficulties*' resulted in answers that did not touch upon barriers. Therefore after a number of initial interviews, this question was expanded to include a statement on barriers and listed a few examples such as cost, knowledge, and technology. These specific examples were chosen as a result of answers given in previous interviews. In their study on barriers to energy efficiency in manufacturing, Trianni et al also found this a difficult topic to assess with interview techniques. They found that with questions surrounding barriers participants tended to give their answers in the form of opinion and behaviour, leading to data that is hard to rank (Trianni et al., 2013).

Difficulties were also encountered when trying to ask questions regarding liquid cooling in section 5. It was not always clear to the respondents what the aim of the question was, indicating that it needed to be improved. In order to achieve a clearer question, additional text was added to the script regarding the opinion of their clients about which technologies were used in the data centre. The aim of the initial question was to find out if anyone had issues with bringing water into the data centre, near to the servers. The main point was to find out reception of these technologies and also whether clients were interested in the technologies used or not.

4.5 Transcribing and Analysing the Interview Data

Transcription of the interviews is an important step in the analysis of the data gathered from participants. Transcription of interview audio files only needs to be as detailed and accurate as needed by the nature of the research (Strauss, 1987). For the purposes of this research, the interviews were transcribed word for word but auditory noises such as ‘um’ and ‘ah’ and also repeated words were not transcribed. It was decided that this level of detail would not be needed in this case, as the research is not aiming to analyse the way in which participants respond to the questions. Pauses were also not included in the transcription process. Once the transcription was complete, a second check was done to ensure accuracy and to make sure no passages were missed.

The software package NVivo was used to aid in the transcription and coding of the interview data (a sample of which is illustrated in Figure 4.5). The ability of the software to slow down audio playback made transcribing the interviews much easier. NVivo was also useful during the coding stages of the research, codes are known as nodes within the software. Each passage within the transcripts was assigned codes which could then be arranged and sorted into lists within the software. Common codes across each interview could be identified and passages which have common codes could be read alongside each other. Interview transcripts were first assigned codes to separate each question and also to allow these to be read side by side within NVivo. Once this was completed, each question was then coded depending on the themes that arose within the answer, some codes spread across more than one question which was visible in the software. Within each question, the individual codes were grouped into categories with similar themes.

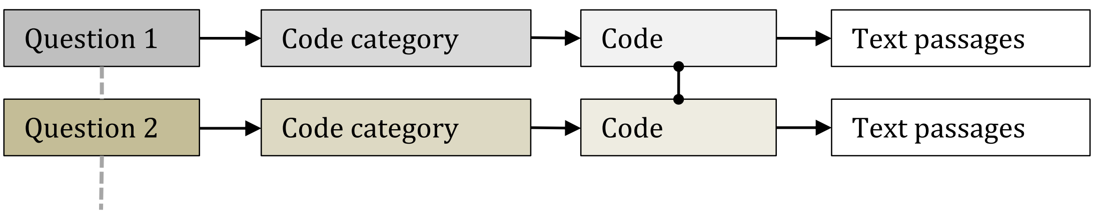


Figure 4.4 Coding break down of questions, often with crossover of codes between questions

Once the coding using NVivo was completed, the codes were placed into tables to display the quantity of each theme that arose (see Chapter 5 for the details of this). Phrases that represented each code and category were then interpreted in order to form a theory from the data. This theory will then be the conclusions and results from the qualitative study and helps to form the conclusions and recommendations in Chapter 8.

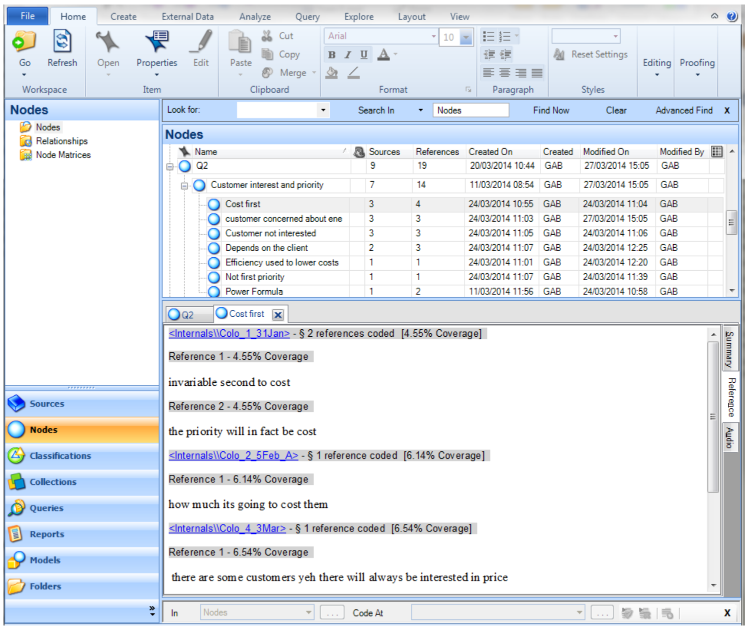


Figure 4.5 Image of NVivo coding showing ability to compare statements between interviews

4.6 Summary

This chapter has introduced the qualitative research methodology that has been followed in order to find out opinions about energy efficiency issues within the data centre industry. Use of this methodology will allow the best data to be gathered on opinions of the industry and barriers to energy efficiency. It has been identified that the use of grounded theory would be the most applicable to this study as it allows the interview script to remain flexible throughout the research process. It also allows participants to be targeted based upon their profession and knowledge, an important part of this research. The grounded theory approach also allows for flexibility in the sample size, relying on saturation in the data to determine how many participants are required. Being able to interpret the boundaries from the data as it is collected and adjusting the sample whilst elaborating on the initial theory (Marshall, 1996) is also an important component.

The following chapter discusses the findings from the interviews in detail including providing a detailed account of the coding procedure in NVivo.

5 Industry Views on Energy Efficiency in Data Centres

This chapter presents the results from the qualitative study into the opinions of the data centre industry when it comes to issues surrounding energy efficiency. Previously, Chapter 4 described the qualitative research methodology and explained the process which would be followed to analyse the interview data. Detailed discussion of the data follows in this chapter.

The current chapter is presented in the following format. Each question is written out in its general form[19] followed by an in depth discussion of the responses from participants. It should be noted that the wording for each question along with the exact order in which they were asked were not the same from one interview to the next. The reasoning behind this is that the interviews were treated as a conversation and to a certain extent the response to one question informed which question was asked next, and how it was phrased. The justification behind this is explained in Chapter 4.

In order to differentiate between each interview whilst maintaining anonymity in the data, the following identifications have been used:

$$\underbrace{\text{Con/Colo/OO}}_{\text{Participant}}\_\underbrace{1}_{\text{Interview number}}$$

Con represents a participant that works for an engineering consultancy, *Colo* represents one that works for a colocation provider, and *OO* represents one that works for an owner operator. To differentiate between interviews with the same type of participant, a number is provided on the end of the interview identification.

The structure of the present chapter has been adopted from Wallace, whose thesis used qualitative research in the form of interviews to consider the effect of foot printing and personal allowances on the effort to reduce household carbon emissions (Wallace, 2009). Their approach of displaying the main and sub-themes in tables before moving onto detailed discussion has a clear and descriptive layout. The tables have been adjusted for this research to enable different groups of participants to be compared to

[19] General form is mentioned here because the interview script was not kept to strictly meaning that the wording of questions varied slightly between interviews. Care was taken to ensure that the questions were near exactly the same though.

one another. Before discussing the themes that arose from the data, the demographics of the participants are introduced.

5.1 Demographics of the Participants

Question 1 served as an introduction to both the participant and their company, with the aim of providing background information on each participant and easing both participant and interviewer into the interview process. Question 1 was split into two parts which were worded as follows:

> ***Q1.a:*** *Could you give me a brief description of what your company does and the role that it plays in the data centre industry?*
>
> ***Q1.b:*** *What is your role in the company?*

The first part, Q1.a, gives the participant the opportunity to describe the role that their company plays in the data centre industry, aiming to provide background information about their company and the services that it provides. The question was also designed to obtain certain pieces of information from the participant, however due to it being quite an open question it led to various levels of detail in the responses from participants. Therefore, further sub-questions were written into the interview script in case the participant did not provide all the details required. These sub-questions varied depending on whether the participant was a colocation provider, an owner-operator or an engineering consultant. For a Colo or OO participant, the sub questions addressed the following details; the purpose of the company data centre(s), how many data centres they owned, the services that they provide, the power rating of the data centre(s), how old the data centre(s) is, and whether or not the data centre(s) is running at full capacity. For Con participants, the sub questions were slightly different as follows; whether or not they are solely involved in data centres, how long they have been involved in data centres, what level of involvement they have in a project, and what their scope of design is.

The second part of the question, Q1.b, was aimed at gathering information about the position and role of the participant in their respective companies. This provided information about their relevance to the interview along with an indication of their area and level of expertise in the subject area.

The information obtained from Question 1 is represented in Table 5.1 which summarises the services that each of the participant's companies provides and the role of the participant in that company. Table 5.2 then provides details about the data centres owned and operated by the Colo and OO groups of participants.

Table 5.1 Participant role and company description

Interview Identification	*Company Role*	*Participant Role*
	Colocation Providers	
Colo_1	Colocation data centre, SME	Project Director and Operations Manager
Colo_2	Colocation data centre services	Head of Operations
Colo_3	Colocation data centre	Technical Director
Colo_4	Colocation services, managed services	Technical Director
Colo_5	Colocation data centre	Data Centre Operations Manager (participant 1), Customer Services Manager (participant 2)
Colo_6	Managed service office environments, considering moving into the colocation marketplace	General Manager
	Owner Operators	
OO_1	University, education	Data Centre Manager
OO_2	University, education	Data Centre Manager
OO_3	University, education	Head of Service Delivery and Infrastructure (participant 1), Core Infrastructure Manager (participant 2)
	Consulting Engineers	
Con_1	IT and data centre consulting and project management	Project Manager, IT Consultant
Con_2	Data centre design and built consultancy, facilities management	Senior Manager, Design Team

Whilst the companies that each participant represents fall into one of the three main categories, Colo, OO and Con, the services that each company provides varies slightly. This demonstrates the varied nature of the industry in that not one company will be directly comparable. Even between the Colo participants, each company had different clients within their data centre leading to them having different issues to deal with. For example, the main tenant of Colo_5 is a local authority council that have environmental and ethical obligations and so are more interested in energy use. In contrast to this, other participants have financial tenants who are more concerned with risk aversion than they are about saving energy.

The participants for each interview tended to have roles with significant responsibility and they were all very knowledgeable both about the subject area and their company. However it was noted that there was a difference in answers given depending on the exact role of the participant within their particular company. This difference was most obvious when comparing answers given by technical directors and operations managers. Whilst technical directors tended to be very knowledgeable about their systems, they didn't give detailed answers to questions about their company, for example when they were asked about difficulties and policy. This is in comparison to operations managers and data centre managers who tended to 'sell' their facilities and company hence giving more detail in their answer. Take one answer from the interview with Colo_3, a technical director for example:

> Interviewer: Could you just give me a quick description of what it is your company does and what it's role is in the data centre industry?
>
> Colo_3: We manage a colocation data centre, that's it really, we rent racks to people
>
> Interviewer: And is it just the one facility
>
> Colo_3: We've got one building with two data halls

This is in comparison to Colo_1 for example where for the same question the participant (operations manager) gave an answer that was nearly 200 words long. Whilst the information given in both participants' answers are useful and include all the necessary information, Colo_1 provided a much more detailed account.

Despite this difference in the level of detail given by participants, it was generally found that all had enough knowledge to be able to answer all of the questions. Only in a couple of cases were the participants unable to give a full answer, this was mainly surrounding the questions asked to technical directors focussing on policy. As their background was more on the technology and infrastructure side of the business, they tended to know less about customer interest and use of policies guidelines within their data centre. Table 5.2 provides the details of the OO and Colo participant's data centre facilities. It provides background information to consider when the participants are talking about the efforts they have made to improve energy efficiency.

Table 5.2 Details of data centres owned by Colo and OO participant companies

Interview Identification	*Age of data centre(s)*	*Load in data centre(s)*	*Capacity of the data centre(s)*
Colo_1	7 (DC1*), 3 (DC2)	90 kW (DC1), 60 kW (DC2)	250 kW (DC1), 250 kW expanding to 1 MW (DC2)
Colo_2	13 (DC1), 10/6/4 depending on installation (DC2), 4 (DC3)	The under used data centre has 1 MVA load	Two are almost full (one at 7.6 MVA), one is under used (7.5 MVA)
Colo_3	7 (DC1), 1 (DC2)	650 kW for both	2 MW for both, 85% full (DC1), 10% full (DC2)
Colo_4	9 (DC1), 5 (DC2), 2 (DC3)	1 MW	DC1 and 2 at full capacity, the rest expanding
Colo_5	5	1.4 MW	Full capacity
Colo_6	9 (DC space), 13 (building)	640 kVA	750 kVA
OO_1	25-30 (old buildings, but undergone transformation),	250 kW	-
OO_2	17 (refurbished 4 years ago) (DC1), 7 (but with new infrastructure) (DC2), 11 (refurbished 4 years ago) (DC3)	1.7 MW across all	Very full (DC1), 25% full and 30% for the others
OO_3	40 (DC1), 10 (DC2)	700 kW (DC1), 4 MW (DC2)	Running at 190 kW (DC1), very full (DC2)

*DC = data centre

5.2 Qualitative Themes from the Interviews

Before the interview data was considered in detail, the transcripts were coded to find themes within the text. As discussed previously in Chapter 4, the transcripts were first coded to separate each question before more detailed coding was carried out. After coding each question in detail and then categorising these codes several main themes appeared, some of which spanned several questions. Table 5.3 lists the main themes that arose through the coding process. It is recognised that the main themes as listed in Table 5.3 have arisen because of the subject matter in each question that was asked. Therefore, it was predicted before the interviews took place that these main themes would appear. Beyond these predicted themes, there were several themes that appeared from the data analysis which could be grouped together under particular main themes. These have been categorised as sub-themes and are presented in Tables 7.4 - 7.11.

The exact quantity of sub-themes within the main themes vary depending on the role of the participant (Colo, Con or OO). This perhaps helps to illustrate the difference between participants in terms of what meant most to them during the interview, for example efficiency measures, difficulties[20] and the future of the industry. It is also noted that the transcripts from the Colo and OO participants tend to have more themes than the consultant engineers. Con participants were predicted to have less themes in the data as they do not own data centre facilities.

Difficulties and overcoming problems, *energy efficiency measures*, *future of the industry* and *customer interest* had the most sub-themes arising from the data with 14, 9, 8 and 7 sub-themes respectively. A higher number of sub-themes indicate a more detailed discussion with the participant and it also likely indicates that there may be more issues associated with that particular theme. It is also clear that the Colo participants generated more sub-themes than the other two groups of participants.

[20] Difficulties covers any problems that the data centre may have faced when trying to improve their energy use or introduce new systems

Table 5.3 Main themes from the interview data[21]

Main Theme	No. of sub-themes			
	Colocation Provider	Owner Operator	Consultant Engineers	Total No. of sub-themes
Difficulties and overcoming problems	12	6	5	14
Energy efficiency measures	9	6	4	9
Future of the industry	8	6	8	8
Customer interest	5	3	0	7
Priority of energy	4	2	0	7
Incentives	4	2	3	4
Monitoring and metrics	3	3	3	4
Policy	4	2	3	4

Each table documenting the sub-themes has been split into the three groups of participants. This is an important step as it will allow the views of each different group to be compared to one another, addressing one of the aims of the research as set out in Chapter 6. Some sub-themes are repeated across the main themes. These are marked in the tables with †.

5.3 Detailed Analysis of the Qualitative Data

The following section presents the findings from the interview data in detail, with the following format. The analysis is broken up into sections depending on the subject of the question being asked in the interview, it follows roughly the same order that the questions were asked in. Each question is written out in its general form along with a description of its aim. Following this, the themes arising from the data are discussed along with a selection of quotes from the participants.

[21] Questions from the interviewer, along with passages that just contained speech from the interviewer were not coded so as to avoid large numbers of themes that are just based on the wording of the questions.

5.3.1 Attitudes towards Energy Efficiency

Question 2 aimed to find out what the attitudes of each participant were towards energy efficiency issues within data centres. Split into two parts depending on the role of the company that the participant worked for, Question 2 was worded as follows:

> **Q2.a** (for Colo and OO): *What is the company approach and philosophy regarding energy efficiency within the data centre, how does it prioritise against other factors?*
>
> **Q2.a** (for Con): *What is the company approach when it comes to energy efficiency, does energy efficiency and low energy consumption play a key role in the design of a new data centre or retrofit?*
>
> **Q2.b** (for Colo): *Do your clients request energy efficiency when considering your services?*
>
> **Q2.b** (for Con): *Do your clients actively seek out energy efficient designs, if not, then do you encourage such designs?*

The aim of this question was twofold. Firstly to find out what the attitudes of the participant's company was towards energy efficiency issues, and secondly to find out the attitudes of their clients. What type of clients these were depended on whether the participant was a Colo or Con. Clients of a Colo participant are defined as users of a data centre, whereas the clients of a Con participant are defined as either owners of a data centre (and looking for auditing services or refurbishment) or those who were looking to have a data centre facility designed and built. OO participants do not have clients as they both own and operate their own facility.

Answers to this question led to the main themes of *customer interest* and *priority of energy* which both have 7 sub-themes. These two themes are now discussed in detail.

Priority of energy

Generally it was found amongst the participants that energy efficiency and the consideration of energy use was on their list of priorities, but there was variation in the level and importance of it compared to other priorities. There were also several reasons given for increasing energy efficiency when this question was asked, leading to the sub-themes listed in Table 5.4. It is clear from Table 5.4 that costs are a key issue considered when thinking about energy efficiency, both of the most common two sub-themes are centred around cost. This is either for the data centre itself, or the costs passed onto clients.

Table 5.4 Sub-themes under the main theme of *'priority of energy'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
Efficiency used to lower cost†	4	0	0	4	5	0	0	5
Cost for client	2	0	0	2	2	0	0	2
Not a priority	0	1	0	1	0	2	0	2
Efficiency for environmental reasons	1	0	1	2	1	0	1	2
Marketing†	1	0	0	1	1	0	0	1
Depends on the client†	0	1	0	1	0	1	0	1

For the OO participants, the consensus was that energy efficiency is a big consideration in the operation of their data centres. This may be because both the participants were representing universities who have sustainability and estate decarbonisation targets which need to be met. OO_2 remarked that *'we have sustainability key performance indicators, KPIs, for the university as a whole and that covers the estate and it also covers the IT'*. Whilst universities have decarbonisation targets, they are also responsible for the associated power costs and will have a driver to reduce these, just like OOs in other sectors will be keen to reduce their costs.

In terms of the Con participants, they were also asked if energy efficiency was a priority to them. The two participants that were interviewed gave very different answers. The discussion with Con_1 went as follows:

> Interviewer: … what is your approach when it comes to energy efficiency, is it something that plays a key role in your projects?
> Con_1: At this point in time, I'd say no. It's something more often after thought … however my opinion is that shouldn't be the case, it should be more around, we should always have an element of energy efficiency in there. Obviously running the business is the last critical component and obviously being energy efficient is the other one, so I think it's not high on the agenda as such.

Whilst this participant stated that energy efficiency tends to be an after-thought, when asked the same question, Con_2 said that energy efficiency is one of the first things that are brought up with new clients. They stated that they would *'push energy efficiency where possible'* but that *'there may be restrictions from a client brief point of view'* that would take them down a less energy efficient route in the design. From a consultant's point of view, they have to stick to their clients wishes for design and whilst they can push for energy efficiency, ultimately it comes down to the client's point of view. This demonstrates the variation in the focus on energy efficiency depending on the consultant engineer. However, it should be noted that they are from two very different consultancy firms. Con_1 was from an IT consultancy whereas Con_2 worked for a data centre design and build firm who focus on the infrastructure.

In terms of the Colo participants, a couple of them took the opportunity at this stage of the interview to explain their use of a power recovery formula to estimate costs for clients. The power recovery formula, used by colocation providers, is a mechanism through which the company can '*get back all of the costs of power for operating the data centre by the end of the year'* (Colo_2). A component of the costs for a colocation client is the cost of the power that they use. This will normally be the power they directly use through their IT equipment plus a share of the building power use which includes the power use of systems like cooling, lighting and security. This leads to an overall service cost for the customer. It is in this way that the colocation provider can *'recover the costs of power used in the data centre from the customer as well as their service charge and rent…as well as their rent we can charge for the power they use'* (Colo_2).

The fact that the customer pays for the energy that they use leads the colocation operator to want to focus on their energy use, as one participant put it; '*obviously energy is a direct cost to us and the less we can pay as a direct cost, the less we have to charge our customers for racks'* (Colo_3). Passing on their energy bills means that colocation providers may make energy use and efficiency a priority if they can reflect this by lowering their costs to clients, thereby providing a marketing incentive for them as a company. One participant stated that they did *'indeed'* (Colo_4) use energy efficiency as a marketing measure[22]. However, it also presents difficulties for the colocation provider in that because they do not pay the energy bill themselves, unless they can see a benefit through increased custom then it is not worth them making expensive changes to their facility.

[22] Marketing also came up as a sub-theme under the main theme of *incentives* (Table 5.8).

Whilst the majority of the colocation providers stated that if energy was a priority, it was in order to keep operating costs down, one also stated that power is *'extremely important'* both from a financial but also from *'an environmental point of view'* (Colo_4). This was unusual as they were the only participant to mention environmental factors as a reason to increase energy efficiency. Environmental reasons were also mentioned as an incentive to reduce energy use by the same Colo participant, one Con participant did also mentioned this as an incentive.

Customer interest

When investigating the level of interest in energy efficiency that the participants had, it was also important to gauge the view of their customers. As businesses, it is important that the data centres meet the needs of potential tenants or owners. For this reason, they may have the biggest influence over how much effort is made towards increasing energy efficiency. Q2.b aimed to investigate whether this is the case, and whether users of data centres are interested in energy issues. Q2.b only applied to Colo and Con participants, with their clients being IT users and data centre operators (these may be either owner operators or colocation providers) respectively.

Generally, it was found that clients of Colo participants do enquire about energy efficiency when thinking about purchasing their services, however it is not normally high on their list of priorities. It can be seen from Table 5.5 that there is an equal split between customers that are concerned about energy efficiency and those that are not. However this balance shifts depending on whether they were a Colo or Con participant. It was found that the sub-theme of *customer not interested* arose more from discussions with Con participants than Colo ones. Only three of the Colo participants interviews actually said that their customers were concerned about energy efficiency.

Although Q2.b was worded slightly differently in each case, the general wording of whether or not clients are interested in energy efficiency and if they request it or not led to a few different responses. Colo_1 stated that whilst their customers are interested in energy efficiency, it is '*not the number one priority'*. Colo_2 reported that clients are '*interested at the start of the process and are sort of selecting who to go to'*, indicating that efficiency is used by clients as a means to differentiate between providers. Relating back to Q2.a which demonstrated that some participants of the research would use efficiency as a marketing method, the above statement from Colo_2 shows why this can be the case.

When asked if customers are concerned about efficiency, *'I think customers are concerned'* was the response of Colo_3 and when asked whether customers enquired about energy use and efficiency, Colo_4 replied that *'yes, in a lot of cases they do actually'*. Despite this positive response, Colo_4 also mentioned that some customers are not that interested by giving the example of customers bringing their own IT equipment into the data centre. By bringing in their own equipment, the colocation provider has no control over the efficiency of customers IT equipment. According to Colo_4, some will bring in IT equipment that is old and inefficient as they are not willing to make the investment in newer, more efficient equipment. Consequently, whilst it appears that some customers in data centres are concerned about energy efficiency (even this is only due to the cost of the energy bill), some are not willing to make investments in newer equipment that might bring the energy bill down further.

For Colo_5, interest in efficiency very much depended upon the customer, their two main customers being a local council and an IT systems provider. The local council were '*interested from an environmental point of view'* however the IT provider was more interested in paying for space in *'an industry accredited data centre unit so they wouldn't ask for, you know, what's your PUE rating and DCIE etcetera'*. Having a variety of customers hosted within their data centre is one of the problems that colocation providers face when trying to increase energy efficiency. They need agreement from all of their customers to be able to make changes to their facility, especially when it will require the customer to pay for a share of the capital costs of improvements. Difficulties like this are discussed further in section 5.3.5.

In terms of the reasoning behind client interest in energy efficiency, it appears that cost is the major driver, although greater costs will stand in the way of better energy efficiency. As one participant put it, energy efficiency is *'almost invariably second to cost'*. This means that if the pursuit of greater efficiency puts a higher price tag on the service being provided, then they are unlikely to choose it as an option despite their interest in energy efficiency. This presents a difficulty for the colocation provider. They need to remain competitive in terms of cost, but also maintain their marketable image as an energy efficient provider. However higher energy efficiency invariably leads to a higher cost, meaning that there is a balance that needs to be met between cost and efficiency.

A similar response was noted from the consulting engineering participants for this question, although it really depends upon who the client is and who they are talking to at the time. Con_1 said that energy efficiency is not something that their clients actively seek out, '*the reason being is one, it depends on who you speak to in an organisation'*. The implication of this is that depending on who they talk to, they may have different agendas and requirements even if they represent the same client. Con_2 also agreed that efficiency can be a priority, but it depends on the client. They mentioned that energy efficiency tends to be *'somewhere within every single clients list of priorities, not necessarily at the top'*. It very much depends upon the cost of achieving that higher level of energy efficiency.

Table 5.5 Sub-themes under the main theme of *'customer interest'*

Sub-theme	No. of interviews			No. of occurrences		
	Colo	Con	Total	Colo	Con	Total
Cost first	3	0	3	4	0	4
Customer concerned	3	0	3	3	0	3
Customer not interested	1	2	3	1	2	3
Depends on the client†	0	2	2	0	3	3
Use of power formula	2	0	2	2	0	2
Efficiency not first priority	0	1	1	0	1	1
Efficiency used to lower cost†	1	0	1	1	0	1

In summary of Q2.a and Q2.b it is noticeable that energy efficiency and energy use is on everybody's agenda, although how high it features on the individual agendas varies from one company to another. Efficiency will almost invariably lose out to costs and of course companies need to know that the data centre they are using or investing money in is secure and reliable so these are also high priorities. Colocation providers face difficulties engaging their customers' interest in efficiency as they are reluctant to pay their share in the capital costs (see section 5.3.5).

For owner operators, energy efficiency seems to be a higher priority; this is most likely because they are paying their own power bill. Consulting engineers will tend to offer several different designs to a client, however it often happens that the customer will chose the cheaper or more resilient offering over the more energy efficient option.

5.3.2 Measuring and Monitoring Energy Use

During the interviews, the participants were asked about monitoring and metrics to find out what level of energy monitoring they carry out. Part of the idea for this question was to ask what metrics they use, in order to bring up discussion about metrics. From the literature it has been shown that there may be issues with the PUE metric. Whilst participants were not directly asked what they thought of the PUE metric, it was interesting to see which ones mentioned it without a prompt. This question was worded as follows;

For Colo and OO: ***Q3:*** *Can you tell me about any monitoring of energy use that you may carry out?*

For Con: ***Q3:*** *Are you involved in energy assessment of existing data centres?*

The main purpose of this question was to discover the importance that participants place on the measuring and monitoring of energy use. Some may simply use power bills, whilst others may have invested a lot of money in software for the purpose. It was also designed to find out what metrics they use, and was a chance for them to give an opinion on these.

All Colo and OO participants stated that they do use some form of measuring or monitoring of energy use within their data centres. Out of the three Colo participants that mentioned the specifics of how they measure power, two said that they measured the power use of individual racks giving them a very detailed picture of power use. The third participant said that they had recently '*realised the importance of metering the overhead power'* (OO_3) and so they have now improved the level of detail in their energy use measurements. OO_1 talked about the fact that they had many monitoring points within the data centre monitoring the environmental aspects. They also use SNMP[23] protocols to log how much energy each server is using. When discussing what measurement of energy use the participants took, some also mentioned that they had monitoring systems specifically for that purpose. These monitoring systems tended to be software specifically designed to monitor power.

Participants were also asked if they used any particular metrics to represent their energy use or efficiency in their data centre. PUE was mentioned by almost all participants, either at this stage of the interview or when discussing their data centre

---

[23] Simple Network Management Protocol, this allows enquiries to be made into a set of parameters within the server (e.g. temperature and power use).

facilities. Several mentioned it when describing their facility, as a general descriptive term to identify their data centre. PUE was mentioned as being a useful educational tool for Colo_2 who stated that they have recently begun to measure overhead power as a result of the PUE becoming '*a more interesting topic'*.

Con_2 discussed the fact that the PUE metric is useful to them to encourage customers and to give them a mark to improve on, it also helps them to demonstrate improvements after changes have been made in existing facilities. Colo_3 simply said that PUE does not make any sense to them. They were not clear on their reasoning behind this, but they did imply that this was from the point of view of using it. Rather than using the PUE, they simply measured the power usage of the data centre and used this as a marker.

The interviews have helped to show that the PUE metric is commonly used, but the way in which it was mentioned during the interview was mainly as a descriptive marketing term. A couple of participants remarked that it needs to be used carefully. Con_2 said that they do use the metric but noted that they are '*very careful about how we use that PUE from a comparative point of view across different facilities*'.

From Table 5.6 it can be seen that the PUE was the only metric mentioned, showing that either the participants simply did not use others or that they were not aware of them. There are a lot of metrics available in the literature for data centres (see Chapter 3), however they do not appear to be in common use within the industry.

Chapter 6 takes this theme of metrics forward by critically assessing the PUE. A case study is carried out in order to demonstrate some of the problems with the use of this metric for assessing overall energy efficiency.

Table 5.6 Sub-themes under the main theme of *'monitoring and metrics'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
PUE	4	2	2	8	7	2	3	12
Monitoring system	5	0	2	7	6	0	4	10
Measuring	3	0	2	5	3	0	4	7
Auditing	0	2	0	2	0	2	0	2

5.3.3 Efforts Made Towards Reducing Energy Consumption

Question 4 was intended to allow the participant to discuss the efforts that they may have made in order to reduce energy consumption in the data centre. In the case of Colo and OO participants, this involved discussing the technologies they have introduced or changes they have made in their facilities. In terms of Con participants, the question was rephrased to ask what suggestions they normally make to clients. Question 4 was worded as follows;

> *For Colo and OO: **Q4.a:** What measures or steps has the company taken to increase energy efficiency or reduce energy consumption in your data centres?*
>
> *For Colo and OO: **Q4.b:** Are there any measures that you plan to implement in the future?*
>
> *For Con: **Q4:** What kind of recommendations do you normally make for clients to decrease their energy use?*

This question was designed to find out the measures and changes that participants had adopted with the aim of reducing energy consumption. The question was left open ended so that they had the opportunity to state the things that they consider to be the most important efforts. This will show which methods are considered the most useful, it will also show which ones people have been able to do already. Q4.a and Q4.b for Colo and OO participants also had several sub questions to address the following; the success of any measures used and whether they were worth undertaking, whether they have seen or are predicted to see a return on their investment, if there were any measures which they had not been able to use for any reason and whether there were any measures they plan to implement in the future.

All participants were also asked whether any managerial measures were used alongside technological methods. This has been defined as measures such as temperature changes, energy management policies and air flow management, as against the installation of new infrastructure systems and equipment.

Table 5.7 Sub-themes under the main theme of *'energy efficiency measures'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
Energy efficiency measures	5	2	3	10	18	7	15	40
Managerial changes	4	2	2	8	11	7	5	23
Limitations	4	0	2	6	10	0	5	15
Future plans	5	1	3	9	8	1	4	13
Beneficial	3	2	1	6	4	0	3	8
Payback	3	0	2	5	4	0	3	7
Costs†	1	1	2	4	2	1	3	6
SLA†	2	0	0	2	4	0	0	4
Research	3	0	0	3	3	0	0	3

Table 5.7 summarises the sub-themes that were categorised under the main theme of *energy efficiency measures*. The majority of these themes arose as a result of Question 4. Whilst this question was aimed at finding out what measures they had used, it was not aimed at finding out every single measure they had put in place. The idea was to find out which ones they would mention and hence regard as the most important to them. It was also designed to see which would be the most common answer that is given. This answer could reflect which measures are recognised as the best or the most commonly accepted measures. It was important to notice how many participants would mention the measures that are easier to implement such as blanking and air management.

It was also important to find out how many participants would mention managerial measures to increase efficiency. This was asked as part of the question. It has been shown that managerial efforts such as temperature changes and the way the data centre is run such as staff training can have an effect on the overall energy use and efficiency. It was important to gauge whether any participants saw this as an important thing to do alongside ensuring that the infrastructure is efficient.

There were many sub-themes from this question with the top three being *energy efficiency measures, managerial changes* and *limitations.* This was foreseen as the questions directly addressed these topics. The amount of occurrences for the theme *energy efficiency measures* though does serve to demonstrate the vast array of measures that data centres can take to improve their energy efficiency.

Energy efficiency measures

When questioned about the measures participants have taken to improve energy efficiency or reduce energy use, there was some variation in the level of detail given. Some participants gave detailed accounts of the changes that they had made, whereas others just mentioned the major alterations and measures that were used. The measures that were mentioned give an indication of the level of importance that is placed on them by the participant.

Figure 5.1 breaks down the sub-theme of *energy efficiency measures* into the specific measures that interview participants mentioned they had used or were planning on implementing. What is immediately clear from the diagram is that the focus is very much on the cooling and air distribution systems, i.e. the infrastructure that maintains the environment of the data centre.

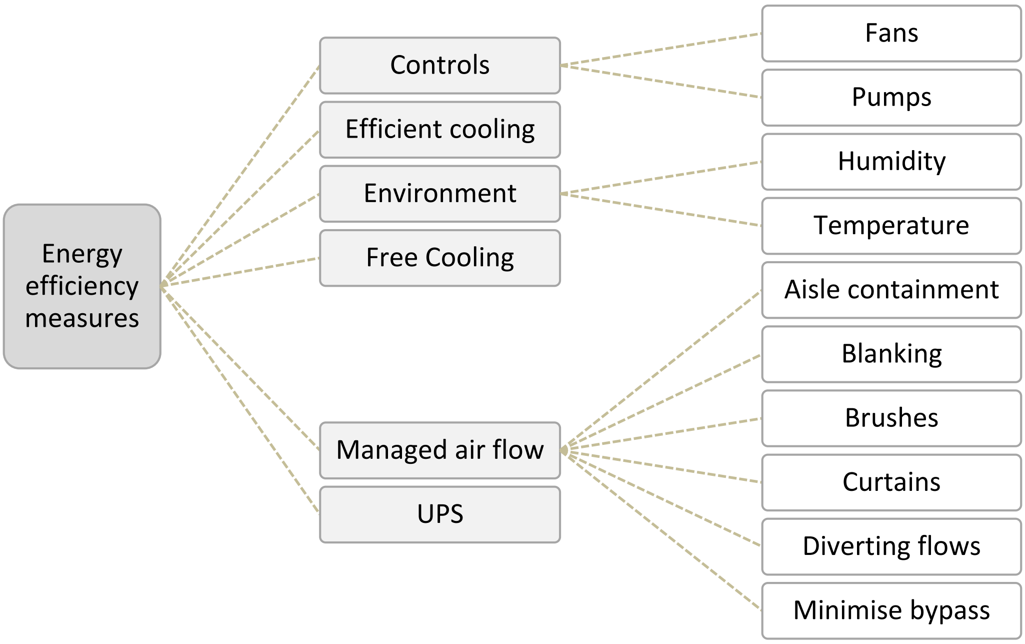


Figure 5.1 Specific measures under the sub- theme of '*energy efficiency measures*'

Free cooling was mentioned by all of the Colo participants highlighting this technology's popularity as a means to reduce energy use within the data centre. The EU CoC lists free cooling as a very beneficial technology in terms of reducing energy use. Despite its potential the technology is only optional for participants in the EU CoC due to the disruption that would be caused should it be installed into an existing data centre. It is therefore a technology which is suggested to be addressed once other changes have been made, changes which offer return for the amount of money invested and within a more immediate time scale (European Commission, 2014). Some participants mentioned smaller measures that had been taken to improve efficiency, mainly applying to air distribution and management within the data centre. Brushes, aperture strips, curtains and blanking plates were all mentioned by various participants. These smaller measures can be seen as ‘low hanging fruit’ as they are easier to implement and are available at a relatively low cost. Con_2 mentioned that when they are working on a project, they suggest making initial changes by focussing on *‘the low hanging fruit’*, which to them means ‘*introducing blanking panels’* and similar simple technologies. These options are cheaper to implement than installing new cooling equipment, and they can provide good returns on investment in terms of energy saving. The ease with which these modifications can be adopted is reflected in the fact that the above changes are compulsory for participants of the EU CoC, although they do not offer as high a level of points as technologies such as free cooling. Blanking plates are the simplest change to make in the data centre, with a relatively high score in the EU CoC. Aperture strips on the other hand are only compulsory should a refurbishment or retrofit of the data centre be carried out, as to install them would mean major disruption within the server room.

A lot of participants discussed the fact that they utilise aisle containment, although it is clear that this technology is difficult to use in some circumstances. The level and complexity of containment varied between participants as some used butchers curtains whereas others had full aisle containment. What they are able to achieve however very much depends upon the situation that they are in. An old data centre with an irregular layout of equipment for example will find it more difficult to implement aisle containment than a brand new facility which include is as part of the design. The EU CoC suggests that whilst aisle containment is very beneficial for reducing energy use, it is only compulsory for a new build data centre or one that is undergoing a retrofit. This is due to the difficulties that can be had when trying to arrange aisle containment within an existing facility. Some participants mentioned controlling temperatures and humidity in the data centre, which is a compulsory section of the EU CoC, offering immediate savings for low economical investment.

Installation of more efficient UPS units is also something that was mentioned by a couple of participants. Chapter 4 demonstrated the fact that power distribution can have a significant effect on the overall energy efficiency of a data centre, so replacement of UPS units can achieve sizeable energy savings. Colo_4 mentioned that they have installed new *'cutting edge UPS systems [with] 98% efficiency'* and OO_2 managed to take their efficiency a step further by introducing *'99% UPSs which reduced [their] electrical losses down to less than 1%'*. OO_2 had achieved this high level of UPS efficiency by moving to a transformer-less unit.

Managerial changes

Participants were also asked if they had made any managerial changes in order to achieve better energy efficiency. Managerial changes in this case have been defined as anything that doesn't involve the installation of new equipment such as temperature alterations, management practices in the data centre, energy policies, education and staff training. This definition was explained to the participants.

Colo_1 said that they used both technological and managerial measures to achieve better energy efficiency in their data centre, however they gave no further details on the latter of these. Colo_2 gave a more detailed response and explained that they have been '*tuning the data room'* in order to minimise bypass air flow. Part of the solution to this was to consider how they operated the valves in the chilled water supply to allow them to reduce fan speed. A few participants mentioned changing the environmental conditions in the room including OO_1 who commented on the fact that they have '*raised the set points constantly over the last three, four years. Originally the data centres had a set point at something in the region of 19 or 20 degrees, over the last three years we've actually raised that to 27*'. They mentioned that this was in order to comply with the ASHRAE recommendations, which has enabled them to reduce their energy consumption.

Con_2 discussed the fact that when working with a client, they tend to introduce improved controls for existing equipment and plant in the data centre. This allows for more efficient management of the equipment already in place. They then say that they start '*looking at increasing the inlet temperatures within their facility'* in order to make more savings. In order to find out if they just suggest technological measures, Con_2 was asked the following;

> Interviewer: … do you make other suggestions, like [having] policies in their data centre to help energy management or is it all technological kind of changes?
>
> Con_2: Yes, so I think we will encourage people to use the EU CoC which is a good stake in the ground from the point of view of encouraging you to put

> policies in place and or you know, encouraging you to select certain types of IT equipment and use the facilities in certain ways.

It appears that Con_2 recommend that data centres do have an energy management policy in place, and that they use the EU CoC to do that.

Limitations

During the discussion about the implementation of measures to increase efficiency, several of the participants mentioned certain limitations that they had faced. This is separate to the question about difficulties, the discussion for which follows in section 5.3.5. Most of the limitations mentioned at this point do cross over into the same themes that arose under difficulties, so the discussion here will be brief.

The main limitations that the participants had when trying to introduce efficiency measures were limitations with the building their data centre is situated in. Colo_2 run a data centre that is a '*below ground site, a basement site so pressure, fresh air cooling and so on was going to be rather difficult and it was retrofitted to an existing site as well, that wasn't going to work for us'*. Whilst free cooling and fresh air supply is advantageous from an energy use perspective, it is not necessarily available to all facilities. Building limitations also affected Colo_3 who said that they would '*like to use free air cooling in data centre one, they do have retrofit free air units which go on top of the air handling units but they won't fit in our building*'. This is because they don't have the room height to be able to accommodate them. Colo_4 has limitations with the building because it has Grade II listed status. Grade II status means that it is hard for them to carry out their plans to improve the insulation of the building as building fabric alterations are difficult to get permission for.

When talking about limitations, the owner operators mainly discussed costs. Budgets are difficult to stretch to cover further efficiency improvements and better controls. OO_2 said they would like more detailed power monitoring but the required PDUs are too expensive for example.

Increasing the inlet temperatures has been one of the measures employed by OO_1 however they found that there is a *'trade off'* as higher temperatures cause higher energy consumption by the server fans. They also found that they have more hot spots in the data centre when running at 27°C meaning they have to reduce it back down to 25°C. This limits their ability to reduce energy use by increasing inlet temperatures to the servers.

Colo_2 said that they have struggled to use containment as their data centre is a '*legacy site [so] it's a bit messy*'. After conducting some experiments using air curtains that have now had success.

5.3.4 Incentives and Reasons for Reducing Energy Consumption

After discussing energy efficiency measures with the participants, it was clear that most of them had made efforts where possible to decrease their energy use. It was also important to find out what the motives behind this were, especially in terms of incentives for them to make these efforts. Therefore, in order to both find out the reasons behind reducing energy consumption and to see if there are any incentives driving changes, the following question was asked;

> *For Colo and OO:* ***Q5:*** *What incentives are there for you to increase the energy efficiency of your data centres?*
>
> *For Con:* ***Q5:*** *Are there any incentives for you to encourage energy efficient design?*

Q5 was aimed at finding out which were the main reasons why companies might invest time and money in improving energy efficiency and energy use. It is important to find out the drivers so that they can be listed and focussed upon to encourage further savings.

Table 5.8 lists the number of occurrences of each sub-theme under the main theme of *incentives*. The discussions with each participant highlighted four main incentives for increasing their energy efficiency or encouraging their clients to do so. Commercial, costs, environment and marketing were all given as reasons for making the effort to increase energy efficiency. Commercial has been defined as a benefit to the participant as a business and marketing covers their desire to be seen as a good company to do business with, along with their wishes to be seen as eco-friendly. Citing the environment as a reason to be energy efficient also occurred as a couple of participants appeared to be aware of the ethical implications and felt a responsibility to be green beyond the marketing benefits. Costs were also a subject discussed under incentives, as it was recognised that in making an effort to be more energy efficient, the participants had noticed savings in running costs. From looking at Table 5.8 it is clear that the biggest incentives are commercial and marketing benefits to the company, a factor that was especially true for colocation providers.

Commercial and *marketing* incentives emerge as the biggest drivers to increasing energy efficiency, this is especially the case for *Colo* and *Con* participants, whereas *OO* participants tended to be more concerned with costs. Environmental reasons were mentioned, however it is clear that business benefits provide the biggest incentives for energy use reduction. Commercial drivers tended to be the fact that they may gain or have already gained more customers and business for their data centre. They also saw increasing their energy efficiency as beneficial to their marketing stance, and staying ahead of the competition. Colo_3 said that '*marketing perception*' was an incentive to them and being more energy efficient was '*how we want to be seen*' according to Colo-4.

Table 5.8 Sub-themes under the main theme of *'incentives'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
Commercial	2	2	0	4	3	2	0	5
Marketing†	4	1	0	5	4	1	0	5
Costs†	3	0	1	4	3	0	1	4
Environment	1	1	0	2	1	1	0	2

Colo_4 stated that they saw being '*eco-friendly'* as '*very important'*. Con_1 also explained that whilst they were a business and obviously needed to stay competitive, they acknowledge that '*saving the planet and the resources that it has'* is important. However they discuss the fact that the issue is not always understood by everyone, so it's important to them that they communicate that message to their clients.

Beyond commercial and marketing incentives, which can be seen also be seen as benefits, only one participant talked about an incentive in terms of financial help. OO_1 described how the university that their data centre belongs to will give money to offset any capital costs provided that it savings can be proven for new installations of equipment reducing energy use.

In summary of the incentives that are driving lower energy use in data centres, it appears that companies tend to cite lower costs and marketing benefits. No one mentioned any incentives given to them by external parties, i.e. financial incentives from government. Everything is driven by themselves and whether they have the desire to be seen as a 'greener' data centre to do business with.

5.3.5 Difficulties and Barriers

Whilst it is already clear that all of the participants have made some effort towards reducing energy use or encouraging energy efficiency, it is important to share difficulties with these efforts. By sharing any difficulties, it makes it easier for future changes and for data centres that have not yet made any changes to also improve their energy use. For this reason question 6 asked about successes and difficulties, also so that an initial assessment could be made of what barriers may be occurring preventing higher savings in energy. Discussion arising from this question serves to show which technologies and measures may be harder to implement, and the difficulties that the industry faces in terms of reducing energy consumption. Question 6 was worded as follows:

> *For Colo and OO:* ***Q6:*** *Thinking about the measures or policies you may have put into place, did you encounter any difficulties?*
>
> *For Con:* ***Q6:*** *Thinking about the technologies and energy efficient methods that you design for or suggest to clients, what kind of difficulties have you come across in terms of getting these accepted by the client, or do you not tend to have any difficulties?*

Initially, the wording of this question did not elicit detailed enough responses from the participants. Therefore it was reworded to directly mention barriers (such as money, training, knowledge and others) as the research progressed to gain better and more involved answers from the participants. The reasoning behind this has already been explained further in section 4.4.4.

Coding the transcribed interviews for themes following on from this question has highlighted several issues that participants found in trying to implement energy efficiency measures. These have been listed and ranked as sub-themes under the main theme of *difficulties and overcoming problems* in Table 5.9. Four themes came out on top in terms of difficulties that participants had faced in trying to increase energy efficiency; *costs, customer agreement, logistics* and *communication between teams*. Ranked fifth in Table 5.3 is *no difficulties*, after some participants noted that they had not really had any difficulties in trying to improve their efficiency in the data centre. It can be assumed that themes as they stand are classed as barriers to energy efficiency in data centres. As discussed in depth in section 3.11 barriers are difficulties that data centres face when trying to make improvements to their energy use. The following text examines the top three difficulties in more detail.

Table 5.9 Sub-themes under the main theme of *'difficulties and overcoming problems'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
Difficulties								
Costs†	3	1	2	6	3	2	4	9
Customer agreement	2	0	1	3	7	0	3	9
Logistics	2	0	2	4	3	0	3	6
Communication between teams	0	1	2	3	0	1	4	5
No difficulties	2	0	0	2	3	0	0	3
Justifying costs	1	0	0	1	2	0	0	2
Technology	0	1	0	1	0	1	0	1
Lack of knowledge	2	0	0	2	2	0	0	2
Overcoming Problems								
Policy	2	1	1	4	3	3	1	7
Education	2	1	1	4	2	3	1	6
Improvements over time	1	0	0	1	3	0	0	3
Financial help	1	0	1	2	1	0	1	2
SLA†	1	0	0	1	2	0	0	2
Balance resiliency and efficiency	1	0	0	1	1	0	0	1

Costs

Costs acting as a difficulty was predicted before the interviews took place, as costs are often a barrier especially as increasing energy efficiency usually means installing new equipment. Costs were something that was brought up by each group of participant. OO_2 said that costs were an issue in the past when they were trying to purchase new cooling equipment that would improve energy use. However, despite being able to '*prove a payback…of under a year*' it took them two or three years to have the funding for this approved. They noted that '*it seems crazy really because if the payback's under the year, why wouldn't you spend the money?*'. It would appear that even though benefits can be had in terms of energy savings, the funders for this data centre were not willing to pay out the capital cost. This was despite the fact that there was a short payback time. This could be to do with the fact that they are a university and may have other priorities such as investing money into research. OO_1 also worked within a university data centre and discussed the fact that costs are a difficulty;

> Interviewer: Does the university set aside much money to be able to do this kind of thing *(energy efficiency measures)* or do you struggle with budgets for your data centres?
>
> OO_1: I don't know what the university sets aside for budgets, but regards to budgets that I have to work within IT, it would be fair to say that it's an immense challenge to keep data centres going on the budgets that we have got allocated.

As both OO participants were data centre managers, it is likely that this is only one side of the story. However it is clear that as the people responsible for managing energy use, but with limited budgets decided and given to them by others, it is a difficult task.

The three Colo participants that mentioned cost as a difficulty talked about it in terms of the expense of investing in new equipment and the payback associated with this. Although Colo_1 mentioned that whilst cost in an issue, they had not needed to make '*any significant compromises*'.

Customer agreement

Customer agreement was identified as the next most common difficulty between the participants interviewed, although it was not applicable to the OO participants. Colo participants stated that they often found it difficult to persuade customers to agree to certain proposed measures to increase efficiency, this being the main difficulty after cost. Colo_2 stated that tuning the data centre was difficult because they had to '*get the customer to play along*'. They also had the same problem when thinking about air curtains saying that '*if we were to hang air curtains and they were to take them down again because they couldn't be bothered to live with them, then it wouldn't work anymore*'. They summarised by saying that '*we've got 90 contracts, 60 odd customers to persuade each one of them one by one to play along and that this is a good idea is a challenge'*. This explains the main difficulty that they face when trying to reduce energy consumption, persuading all of their customers to join in and agree. This is mainly likely due to the conflicting goals that each party has. Whilst the data centre owner may wish to become more energy efficient, the tenant still wants the best value for money and may not have their own concerns about energy use. Justifying the payback for each customer if the capital costs of improvements were to be included in their power formula can therefore be difficult. This creates a dilemma because it is hard for the colocation provider to justify spending the capital costs if they are not able to pass it on to their customers.

Carrying out efficiency methods such as increasing chilled water temperatures is especially difficult because it affects all tenants within the data centre. The chilled water temperatures are written into their SLA so if they were to increase them, they would need agreement from every single customer otherwise they would have problems with breaking the agreements of the SLAs in place.

Another difficulty that colocation providers face is with the IT equipment that their customers bring in and install in the data centre. Colo_4 said that '*unfortunately being a colocation facility, we have no immediate control over what kit customers bring in, so we're trying to keep a happy medium'*. They mentioned this when talking about increasing supply air temperatures. OO_2 stated that '*the problem with HPC* (High Performance Computing) *is that the funding certainly for the hardware side of things, doesn't come with any strings attached. So what they try to do is try to buy biggest bang for their buck and essentially go out and buy you know whatever they feel like, and because research is such a priority certainly in our institution, we find it really difficult to get ourselves in that process to prevent them from buying stuff that is very energy inefficient'*.

Having no control over what IT equipment was purchased to be place in the data centre is a significant problem for OO_2. This means that they could have no influence over the efficiency of the equipment being brought in and installed. Departments purchasing the equipment are not responsible for the running costs of them, hence they do not necessarily need to consider their energy use and efficiency, only their computing performance. They actually said that '*they could spend a bit more, get a little less grant and get more energy efficient, but they don't pay the bills [so] they don't care'*. This means that there is a disconnect between responsibilities for efficiencies, buying equipment and paying for the electricity[24]. Cases like this highlight the problem of communication between different users of the data centre, a point discussed further under the discussion of the sub-theme *communication between teams.*

---

[24] This issue was brought up in the review of literature in Chapter 3 as past research has found that often the person buying equipment is not the one that pays the bills.

Despite this difficulty of no control over the equipment being brought in, as owner operators it appears that they do not face problems on the scale of colocation providers. OOs provide the data centre services for themselves so they do not need to deal with persuading a multitude of customers to reduce energy consumption.

Logistics

Several participants mentioned logistics when discussing difficulties that they have had. The term logistics is used here to describe problems that may have arisen due to the physical situation that the facility is in (i.e. whether it is part of a larger building), or due to time and resource restrictions. OO_3 for example said that they struggle to implement the measures they want to because of constraints on their time, with energy use improvements being pushed down the priority list. This participant also said that they hit difficulties because of their facility being located in an old building which causes its own unique complications. They have had for example the capability to have chilled water delivered to the data centre from chillers that have been installed on the roof, but have no efficient means to distribute this down to the data centre rooms. One of the problems that Colo_6 say they face is that it is difficult for them to install certain technologies if their tenants will not be using them over a long time period, when the tenant moves on, they could be stuck with technology that future tenants do not want to take on.

Communication between teams

Communication between teams was mentioned by a couple of participants as a difficulty that they have when trying to implement new low energy use technologies or measures. In terms of consulting on a project for a client, Con_1 said that they can face difficulties when trying to put their ideas across. Depending on who they speak to in a company, they may get different responses regarding their suggestions. As Con_1 put it, *'in general in organisations there are differences between teams. They get stuck in their own sort of world … let's say for example you were engaged with the IT manager you would have different KPIs* (Key Performance Indicators) *to reflect on. The FM*[25] *manager will have different, it all depends on who's which and who's not really'*. This demonstrates that there are issues with differences of opinion and perhaps lack of communication between different managers within the data centre. An IT manager will have a plan and agenda for the IT equipment that is installed in the data centre, and the facilities

[25] Facilities Manager

manager will be more concerned with the infrastructure. There may then be a separate person who is responsible for the energy bill.

OO_2 also faced a communication problem within their data centre. As mentioned previously, they had issues with departments within the university purchasing IT equipment without any thought to the energy efficiency of them. Better communication between different managers and responsibilities within the data centre will help to coordinate a better focus on energy use. Everyone, whether IT or facilities need to consider the energy use of their constituent part, otherwise the people responsible for paying the energy bill struggle to keep the costs and energy use down.

5.3.6 Overcoming Difficulties

This question followed on from the discussion of difficulties and barriers that made it hard for participants to implement certain measures. Question 7 was worded as follows for all participants;

> ***Q7:*** *Thinking about these issues, what do you think would help to overcome them?*

After a few interviews, this question was further expanded to include the following statement;

> *Would it be an economic incentive for the client, a policy intervention that would require them to do more, or would it be something more simple like the client having more access to information about the issues?*

It was found this additional statement was needed to acquire more information from the participant. It is understood that the wording of this question will have given the participants ideas about what to answer, hence it was not completely unbiased. However it was found that a better response was received when a few ideas were given. Table 5.9 shows that in terms of overcoming problems, the participants said that *policy*, *education* and *improvements* over time would be the most useful things to do.

Education

Education links back to persuading clients and communications between teams in that educating people about the issues of energy use in data centres may help to encourage energy savings. This is especially true when it comes to colocation providers educating their customers and consultant engineers educating their clients on the issues. Colo_2 said that persuading and coaching people to '*play along with it*' (i.e. agreeing to efficiency measures suggested) would be important to get everyone to agree and be able to reduce energy use. However they mentioned that this would take a lot of effort.

Con_1 spoke about education in general in terms of '*what does energy efficiency mean to the market space'* and that this would need to '*start at a higher level, at a CEO* (Chief Executive Office) *CIO* (Chief Information Officer) *level and filter it down'*. They said that rather than providing incentives, people should be educated on the issues. This implies that the best methods would be to educate people on the issues and allow them to take these on board and drive themselves to make better energy choices.

Policy

Policy was brought up in terms of overcoming difficulties both for governmental policies and internal company policies, although the latter tended to be discussed more. Con_1 talked about having an in house policy for energy efficiency but unfortunately whilst there are companies that have these, '*the larger majority don't have energy efficiency policies'*. They recommended that internal policies are put into place for '*simple things like you know, you need to turn your lights off when you're not using it'.*

Both Colo_1 and Con_1 mention the CRC at this point in the discussion. Colo_1 commented that '*in terms of policy, obviously CRC is an enormous stick'*, however Con_1 was more negative about this policy. They said that '*the main issue with CRC is was it was complicated, it wasn't straight forward*'. Due to this, Con_1 said that we need policies that are clear over their reasoning, and it is highly important that everyone can understand the reasoning. They said the industry needs a legal policy put into place but it needs to be a policy that is '*actually clear to people and not some complicated thing that's never going to work for you'*.

OO_2 suggested that the solution to their problem of University departments installing inefficient IT equipment would be to have a policy in place. The policy would target where the funding for the equipment comes from, i.e. research councils. They suggested it should stipulate that only more efficient systems can be purchased and installed.

Improvements over time

Making improvements over time was mentioned as a solution by one Colo participant who mentioned it a few times during their interview. When discussing their approaches to improving their energy efficiency, they mention that they are making improvements over time. They consider it '*an evolution as we grow and as we build'* and have '*planned replacement programs and tuning'*. It is in this way that they can financially manage changes and plan for the future. When it came to talking about what solutions may help to encourage further energy savings and to overcome barriers, they mentioned that there is a '*time factor'*. By this they suggested that data centres should make the conscious effort to change over time as contracts are renewed and new buildings are built. This suggests that changes cannot be made quickly and so must be made deliberately and consciously over a period of time.

5.3.7 Use of Policies and Guidelines

The theme for the question about policies and guidelines was to find out which ones the participants used. The main focus was the EU CoC and its usefulness, the questions were worded as follows for all participants;

Q8a: *Are you aware of the EU Code of Conduct and do you follow/endorse it?*

Q8b: *Do you follow any other schemes such as BREEAM?*

Q8c: *Are you aware of any other governmental initiatives or legislation about energy efficiency in data centres?*

This question was aimed at finding out more about whether they used the EU CoC, the usefulness of the scheme to them and whether it brought them any benefits. The question also looked at whether Colo clients are interested in these kind of schemes, and whether that are used as a marketing tool.

Table 5.10 Sub-themes under the main theme of *'policy'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
EU Code of Conduct	5	2	3	10	6	5	4	15
Other schemes	3	1	1	4	4	2	1	7
BREEAM or similar	1	1	0	2	1	1	0	2
CRC	1	0	0	1	2	0	0	2

All participants were asked about their awareness and involvement in the EU CoC. The aim of this question was to find out what level of involvement they may have had with the scheme, whether it has benefitted them and also their opinion regarding it. The sub-theme of EU CoC could be further broken down into different categories, as represented in Figure 5.2.

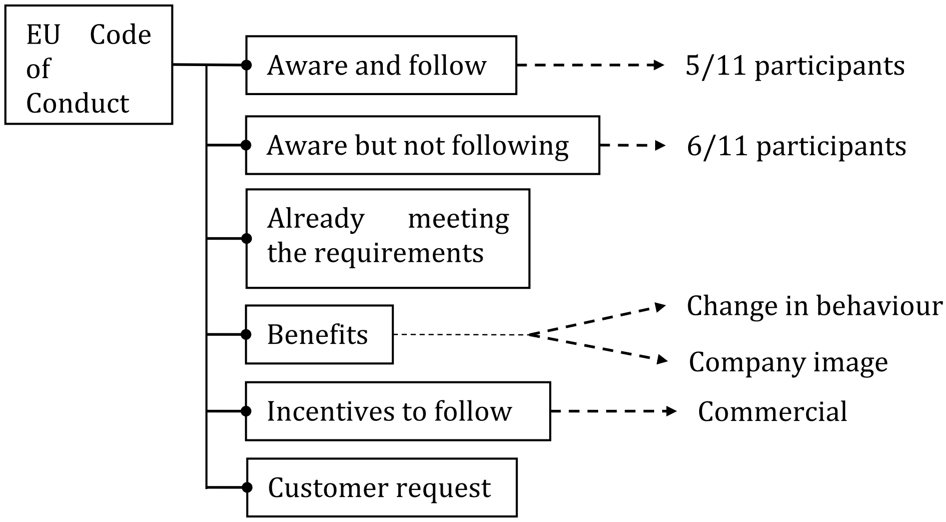


Figure 5.2 Illustration of the discussion about the EU CoC

From the 11 participants interviewed, there was roughly a 50% split between those who follow the EU CoC and those that don't. One of these participants also said that they were already meeting most of the requirements for the code before they signed up to it. Of the participants that said they did not follow the code, one colocation provider said that they could not give a reason for not following it as they were reluctant to do this on the recording. Another participant, another colocation provider, said that they did look through the code but once they passed it over to management it didn't get taken any further. They were unclear as to the reason for this. The third colocation provider that stated they did not follow it because '*it's not specified as part of any of our contracts*'. Because it is not a part of any of their contracts and agreements with clients, they do not follow it. The fourth participant that said they do not follow it was an owner operator who said that some parts of it were not possible to follow given their data centres location in old buildings;

> Interviewer: Are you aware of the EU Code of Conduct and does the university follow it at all?
>
> OO_2: Yes. Shall we say we are not signed up to it to my knowledge, we try to follow some aspects of it. There are certain aspects of the EU Code of Conduct that you cannot sign up to especially if you're looking at really old buildings.
>
> Interviewer: It's difficult for you
>
> OO_2: Very difficult. You can sign up to it and spend millions of pounds trying to meet every criteria of it. I think the problem of it it's not there as, it's there to

> shall we say it's like a framework. I treat it more as a framework you take bits from it and use it if you can. But unless you're a very modern data centre, you could be ending up spending millions of pounds just to meet some of the criteria.

OO_2 makes the point that unless the data centre is modern, it can be very expensive to meet the requirements (recommendations) of the code. Whilst the code does provide best practice guidelines, some changes that are required are out of the reach of a lot of data centres.

From the participants that said they follow the code, Colo_1 gave a detailed response to the reasoning behind their involvement with the main incentive for doing so being commercial. They gave the main benefits of being involved in the code as being *changes in behaviour* and *company image*. Changing their behaviour '*a little bit'* was mentioned as a benefit of the CoC. It encouraged them to do things like changing the air flow within their data centre and increase efficiencies by blocking off air flows properly. They also discussed the fact that the code brings the additional benefit of improving their image as an efficient data centre, describing it as *'a badge which helps us commercially*'.

As a consultant, Con_1 are involved as an endorser[26] of the EU CoC. They said that they are a '*thought leader*' and want to be able to help their clients to improve their energy efficiency, or to have high efficiency from the beginning. In the past they have already been following the guidelines in the code, however they said they have also learnt some new things from it. In order to find out client interest, the following question was asked;

> Interviewer: Is it something that clients are interested in, do they ask if you are an endorser?
>
> Con_1: Not at this point in time, but we have had some queries where they want to become participants within the EU code of Conduct and have asked if we can help them and review their applications. But again like I said before, there's no great appetite in this, it's just another thing you need to do to satisfy a tick box for green really.

It would appear that although the customers of Con_1 sometimes make enquiries about being a participant of the code, there is no real demand for it. When asked about customer demand, Con_2 said that half of their customers tend to ask about it. They

[26] A company that endorses the EU Code of Conduct by recommending and explaining its benefits to its customers

discuss the code with customers from the early stages of the project. Con_2 said that part of the reason they are involved in the code is for marketing purposes, but they also see it as a '*good tool*' for achieving best practice in design. They also mention that it is helpful when presenting clients with solutions and improvements as it '*certainly helps to reinforce what we're suggesting to them*'. They do note however that the changes suggested in the code, such as the use of blanking plates, are '*not something you can just jump in and do yourself necessarily. Having the skills to look at and understand the impacts are of the changes that you're making*'.

For new data centres, it is easier to comply with the code as it can be built in compliance with it, like the facility run by OO_2. OO_2 describes the code as '*a great resource for people to use to prove energy efficiency*'. In terms of the expense of carrying out the requirements, they said that '*there's a lot of low hanging fruit in there and inexpensive stuff that you don't have to spend vast amounts of money on*'. It must be noted that this participant is part of the technical committee for the code. They said that it has helped their reputation within the university sector, being involved in the code and making efficiency improvements.

When asked if they follow any other similar schemes, Colo_2 and Con_2 both said that they have explored BREEAM accreditation for new build data centres. Colo_2 said that whilst they have explored BREEAM and other similar schemes, there is a capital expenditure involved with achieving the requirements for them. They do ask customers if they are interested, as customers are not requesting it themselves. However they '*declared it's not something that would help them*' when asked if they would be happy with factoring in the capital costs into their rent. This shows the difficulty for colocation providers with these schemes, if their customers are not interested in paying the extra costs incurred from the improvements, then they are not likely to go through with it.

5.3.8 Use of and Receptivity to Liquid Cooling Technology

Question 9 aimed to find out the opinion of the industry when it comes to liquid cooling infrastructure. This question tended to vary a lot depending on the participant and the way that conversation had progressed. It also took different levels of explanation depending on the understanding of the participant. The basic wording of the question was as follows;

> *For Colo and OO:* ***Q9a:*** *Do you use, or have you considered using liquid cooling technologies in your data centre. This includes for example rear door cooling*

units, liquid submersion cooling and cold plate cooling, anything that brings liquid in close proximity to the server?

*For Con: **Q9a:** I would like to find out whether your company ever uses liquid cooling technologies in your projects, including for example rear door cooling, liquid submersion cooling, cold plate cooling, anything that brings liquids in close proximity to the servers. Have you done any designs that have involved technologies like this, if so, then did the client have any reservations about bringing water close to the servers?*

This question evolved a lot over the interview process as it was found it needed changing to get a response that included them discussing client response about the technology. Later interviews then went on to include a question about the reuse of heat, as liquid cooling infrastructure is one way that waste heat can be captured more easily than if air cooling was used. Therefore a second part to question was added:

***Q9b:** The reason I ask this question is because one way of reducing the impact of data centres is to reuse the heat from the servers. Using liquid based cooling is one way to recover the heat so that it can be used again.*

For Colo and OO: Have you considered the reuse of heat within your facilities?

For Con: Have you been involved in any designs where the reuse of heat was introduced?

The main idea behind this question was to find out whether they use any cooling infrastructure that brings liquid near to the servers. Liquid cooling is an efficient alternative to air cooling however there may be reservations as it could be seen as a higher risk solution.

The discussion arising from this question has not been coded the same as the rest, as it was a short discussion with participants that did not lead to many sub-themes beyond whether they used the technology or not. Therefore the analysis here is just based upon quotations from the interview transcripts.

A couple of participants said that they already use some form of liquid cooling. Colo_1 said that they use in-row chillers which require water to be brought into the server room. OO_1 also said that they use liquid cooling in the form of rear door coolers which involved bringing water into the server room. They say that they use this for the HPC equipment which has such a high heat density that they would not be able to achieve enough cooling with the existing air system that is in place. OO_1 '*for those systems* *water was definitely the most efficient means of cooling the HPC equipment. In fact I would say, it is the only method that they could use'*.

Con_2 mention that they have done designs that include liquid cooling to the servers before, but this has been for HPC equipment. They also say that in colocation facilities, '*the clients like to ensure that the IT spaces are for the IT guys only and where possible avoid the mechanical servicing within the space preventing access to the space for those sorts of maintainer*' (Con_2). Colocation providers would rather keep the IT and mechanical infrastructure separate; their customers like to know that limited people have access to their servers. Having mechanical maintenance teams being able to access the server rooms can be seen as a security risk from the customers' point of view. Having the mechanical infrastructure separate from the IT equipment gives them an additional level of security for their servers.

Of the participants that said they don't use liquid cooling, 3 out of 4 of them were colocation providers. Colo_2 said that they can't use that kind of technology because they are situated '*right up against [their] customers' activity'*. Meaning this kind of technology would '*need to be driven by customer requirement*'. They go into further detail by citing the fact that they are currently working on a chimney rack system with one of their customers. However '*the demarcation*[27] *point would move to include the rack*'. They mention this example because it is '*as close as [they've] got to something that influences what the customer can do, and limits flexibility*'. This means the customer is restricted as to the IT equipment that they can bring in. This example is cited because they clearly have difficulties persuading customers to do certain things, getting them to agree to something like liquid submersion cooling for example would be very difficult. Also because the client remains in control of the server but the colocation provider is in control of the cooling infrastructure. If those two parts of the data centre suddenly became one piece of equipment, then the contractual agreements between and responsibilities of the provider and customer would change. Colo_3 said that they did not have high enough density equipment to warrant using technologies like liquid submersion however they do use a chilled water system. Although this does not bring water into the server room. They have looked at using in-row cooling but have not yet installed any. Colo_4 and OO_2 simply said that they did not use such technologies.

[27] The point at which the responsibility changes from the colocation provider to the customer

> Colo_5:…if we decided to adopt that sort of cooling we would have to assess the risk and our stakeholders would therefore have to agree and say yes we accept that risk or we don't. So they might very well turn round and say no, it's water.

After a few interviews, further information was added to the question to ask participants about the reuse of heat. As liquid/water can hold significantly more heat than air, there is potential for better heat recovery and therefore reuse from the data centre. Colo_4 said that they are looking into exporting the heat from their facility but they were not able to give any further details on how that would work. Con_1 said that harnessing the waste heat from the data centre is a technology that '*is really in its infancy*'. Con_2 say that the biggest challenge with the reuse of heat is where it can be used afterwards as it is only of a low grade.

In summary of liquid cooling, some participants do use in-row or rear door coolers where water is brought into the facility. But again the colocation provider faces difficulties because of the setup of their business. They need to get client agreement generally to install new technologies like this. A lot of their customers like to keep mechanical infrastructure away from the IT equipment from a risk and security point of view. They also face the difficulty of contractual issues when combining the cooling infrastructure with the racks and servers.

5.3.9 Future of the Industry

In order to get opinions about where the industry is heading in the future, the following question was asked;

> *For Colo and OO:* ***Q10:*** *How do you see the issue of energy affecting this industry in the future?*
>
> *For Con:* ***Q10:*** *Thinking about the future of the data centre industry, do you think there will be a greater demand for more energy efficient designs and if that is the case, then what do you think may be the drivers towards this?*

A sub question was also asked regarding the focus of attention when it comes to efficiency measures;

> *Do you think the focus will shift towards the IT equipment rather than the cooling infrastructure and other systems?*

The aim of this question was to find out from the participants where they thought the industry was heading in the future, given that energy prices are rising. Most of the participants focussed on the cost of energy rising in the future and decreasing availability in regards to how this might affect their data centre operations. Colo_3 discussed the fact that any increases in energy price will be passed onto their customers which will be the case for all data centres, unless they are incorporating renewable energy into their facility. When asked about their future energy use;

> Interviewer: Do you think you're going to have to make more effort to reduce energy use?
>
> Colo_3: We couldn't do it here. I think you'd need a site with a larger roof space or larger open fields or something to make more use of it.

This brings it back to the point of older data centres having a tougher time when it comes to trying to lower energy use or increase efficiency.

Table 5.11 Sub-themes under the main theme of *'future of the industry'*

Sub-theme	No. of interviews				No. of occurrences			
	Colo	Con	OO	Total	Colo	Con	OO	Total
Energy	5	2	2	9	7	3	4	14
Reducing the '1'	5	2	1	8	6	4	2	12
Server technology	3	1	0	4	8	1	0	9
Costs†	2	1	0	3	4	1	0	5
Cooling	1	1	1	3	1	2	1	4
Cloud computing	0	1	1	2	0	1	3	4
Best practice	1	1	1	3	1	1	1	3
Rack density	1	1	0	2	1	2	0	3
Demand	1	0	1	2	1	0	1	2

A slightly different question was posed to Colo_5 to follow on from them saying that costs are going to be a big driver towards energy efficiency in the future;

> Interviewer: You just said that costs are going to be a driver. How do you think they (*the company*) are going to be able to bring energy use and costs down. Do you think it's going to be by bringing in more technology, or do you think it might be concentrating on other things like managerial practices or perhaps looking at IT rather than cooling? What do you think it's heading towards?

> Colo_5: I think there's a combination of everything in there. You're looking at your IT infrastructure to be more energy efficient … and that's where manufacturers are going. You try and buy a server now the things that they push on the website is how efficient that particular server is. On the other side the data centre technology obviously your cooling needs to become more energy efficient.

They also then go on to talk about management practices and that software allowing more effective management of the data centre will be important. This along with practices and processes will be another element in saving energy in the future. They actually mention as well that the power companies are competing and driving down their prices. So whilst the price of energy is increasing over time, the power suppliers will need to remain competitive.

Energy was also discussed by Con_2. They thought that availability of power will be an issue for data centres in the future. In order to avoid power supply issues, they suggest that in the future the typical data centre won't be '*necessarily sitting in an industrial estate in the middle of nowhere, they'll probably be located close to energy projects you know waste energy and so forth. Where we can, we can perhaps use heat from other processes within the data centre to provide cooling themselves and that's certainly something that we're looking at as well*'.

OO_1 also stated that the cost of energy is going to become a big problem in the future, although whilst their department does not pay the electricity bill, the participant still appeared to feel some level of responsibility to keep a lower energy bill for the university. When talking about lowering energy bills, they mentioned that the university is signed up to '*carbon treaties*' that require the university to cut its energy use by 20%. However the university needs to do this whilst '*meeting the research expansion commitments it needs to make to maintain and be a top university in the world…we have to compete with universities that are running these massive power intensive research systems*'. Competition to expand IT capabilities may hamper their efforts to reduce energy consumption. Whilst this is in a university setting, it may be applicable to other industries as well who would be required to keep up with increasing demand for access to internet and IT services, whilst also being encouraged to decrease energy use.

The second most common theme that arose from this question was *reducing the '1'*. This is concerned with focussing on reducing the 1 of the PUE figure, which requires a reduction in the energy consumption of the IT equipment. Nearly all of the participants discussed a shift in the focus from energy use by the facility infrastructure to the IT equipment, this was highlighted as an important point for the future. Con_1 said that '*I think from an M and E*[28] *side it has reached where it needs to be whereas the IT side is still lagging behind*'. Some participants used the phrase 'reducing the 1' when describing new server technologies designed to make them more energy efficient. The case study in Chapter 6 demonstrates this problem of the IT load being hidden 'within the 1'. Chapter 7 takes this theme forward by investigating a way of reducing energy consumption of the IT equipment and also presents the effect this has on the PUE metric.

Section 5.3.1 highlighted a difference of opinion between an IT consultant and an infrastructure consultant. The IT consultant (Con_1) mentioned that for their customers, energy efficiency was very much an after-thought however the infrastructure consultant (Con_2) said they push energy efficiency from the beginning of a project. This illustrates the disconnect that exists between the focus on energy use in the data centre infrastructure and the IT equipment that it supports. It is clear that if further savings are to be made throughout the data centre, then the focus needs to shift from the infrastructure (which has already made significant savings) to the IT equipment.

5.4 Summary

Eleven professionals within the data centre industry were interviewed with questions being asked about the efforts their company have made to increase energy efficiency in their data centre. Participants were asked what measures they had taken, whether these had been worth it and what difficulties they had in implementing them. They were also asked about the policies and assessment methods in place to encourage low energy use in data centres before asking their opinions about a particular low energy form of cooling. Participants represented three different types of businesses within the data centre industry namely colocation providers, owner operators and engineering consultants. This allowed a wide selection of opinions and thoughts to be collected. The following is a summary of the findings and the conclusions that have been made.

---

[28] Mechanical and electrical

5.4.1 Barriers to Saving Energy in the Data Centre

Once the results of the interviews had been considered, a more in depth review of the literature was carried out (following the principles of the methodology set out in Chapter 4). Recalling this post-research review of literature regarding barriers within the data centre industry in particular (section 3.11), several main barriers were identified. A study by Big Switch Projects found that risk aversion, standard practices, technology and a disconnect between costs and operation were the four main barriers (Big Switch Projects, 2004). The U.S. EPA study provided a longer list of barriers which included the same four as just mentioned but also lack of energy auditing and learning curves (U.S. Environmental Protection Agency, 2007). Whilst somewhat reinforcing these ideas, the research in this chapter also found additional barriers which did not turn up in the review of literature. The four main barriers as identified by the results of this research are; costs, customer agreement, logistics and communications. Whilst costs crosses over the findings from the literature, this study has demonstrated that efforts need also to be focussed towards solving problems with customer agreement, logistics and communication if energy efficiency barriers are to be reduced. These four main barriers are concluded in more detail as follows;

Costs

Costs are a major barrier to energy efficiency improvements in the data centre, and they affect different parties in different ways. The following points arose from the results of the interviews;

- Whilst colocation providers may want to reduce their energy use through efficiency (as this would lower the costs to their clients, and improve their environmental credentials), this is difficult to do as they do not see the cost benefits themselves. If investing in new technologies costs significant amounts of money (especially in a retrofit situation), then it is not a path they will follow if there is not enough interest from clients. Reducing costs can be aided by persuading their clients to take a share of the capital costs of improvements, but this is a difficult task.
- The owner operators have more incentives to introduce newer energy efficient technologies and measures as they will see a direct benefit in terms of a lower energy bill. However for operators like the ones interviewed in this study (university and research sector), it can be difficult for the IT department to

persuade the university to grant the funding for changes. This is despite the fact that research tends to be HPC heavy, which consumes more energy.

The results of the interviews have shown that the participants found it difficult to persuade funders to invest in energy saving technologies or measures for their data centre. The colocation providers would not necessarily directly benefit from improvements or may find it difficult to persuade clients, whereas the owner operators found it hard to persuade funding bodies of the benefits of investment. In the latter case the focus is on the speed of payback rather than on the longer term benefits. A better way of presenting the economic (and also the carbon saving) benefits may be to use a Marginal Abatement Cost (MAC) curve which would illustrate the costs of investment and the carbon emissions abated as a result of any investment. MAC curves can either be expert based or derived through the use of models. Either way, they illustrate the economics surrounding efforts to mitigate climate change and reduce emissions. The justification of investing money into a data centre to reduce energy use would benefit from the use of an expert based MAC. These types of MAC curve demonstrate the amount of emissions that could be abated by specific individual technologies and the costs associated with doing so. They must be used with care though as whilst they are easy to understand, they only show the maximum potential and do not take into account social aspects (Kesicki, 2011) such as the barriers discussed here. MAC curves would enable a data centre to demonstrate the emissions savings possible, which may become more important as regulations become stricter. A standard MAC generation tool would be useful to the industry. One which was pre-loaded with each available technology that a user could use as part of their business case for investing in energy efficient technologies would not only help them to see which ones offer the best economical and carbon payback, but would offer a clear visual representation for those granting the funding.

Customer agreement

Customer agreement and interest came up as a significant barrier, especially for colocation providers;

- Customer interest has a heavy influence over the changes that can be made within a colocation provider’s data centre. Unless it is a new build facility, the provider has to get agreement from all their existing tenants and write it into the SLA. This is a difficult process, as all tenants would need to agree.

- Costs and customer agreement came out as the most common difficulties that participants faced when trying to invest in more energy efficient technologies and practices. Customer agreement particularly refers to changes in colocation data centres and bringing IT equipment into those and owner operator data centres.

The issue of gaining agreement from all customers brings back the point made within section 2.3, that the data centre is a sub-system within the digital services sector and that other external sub-systems have an effect on how these facilities are designed and run. In this case, the goals of the tenants within the data centre are not aligned with the goals of the facility owner and this causes conflict when trying to achieve these goals.

Communication

Communication problems within the industry and on a smaller scale within businesses also emerged from the research as a barrier to energy efficiency. An organisation that owns and manages a facility for their own purposes would in theory have an easier time introducing energy saving features, due to the fact they do not have to hurdle over the problems of agreement from multiple customers. However, they are not without their own problems. A disconnect between the priorities of different departments, and also the communication between them was noted.

Logistics

Logistics of course also have a role in how many changes and alterations a data centre owner can make to their facility. It was found that some participants in particular were restricted to what measures they could do and hence their ability to reduce energy use is limited, mainly due to the building that their data centre was located in. The implications of this are that whilst new build data centres are able to achieve high efficiency levels, the existing stock of legacy data centres (and even some that are not that old) can struggle to make changes. What makes this a more important consideration, is that the existing stock of data centres is naturally much larger than modern facilities and those that are being built.

5.4.2 Summary of Other Findings from the Interviews

Attitudes towards energy use along with measures taken

It was found that amongst the participants of this research, energy efficiency did feature on their list of priorities as a business; however they had different levels of priorities and different reasons for considering energy use. For the OO participants, energy use was a key priority as they are responsible for the operational costs and are in the position to be able to make improvements. The Colo participants on the other hand have different drivers when it comes to reducing energy use. Their drivers for reducing energy use tend to arise from their need to remain competitive within the market and to attract new tenants to their facilities.

The other side of considering attitudes towards energy use is the views of customers or clients looking for space to rent in a data centre, or wishing to build their own facility. The results have shown that only half of the Colo participants have customers that are interested in energy efficiency, or discuss it as part of their initial inquiries. Perhaps the introduction of green SLAs as suggested by Klingert et al (2011) (discussed in Chapter 3) would help to educate customers and encourage them to consider energy efficiency. It is no surprise that costs have been found to play a major role in the consideration of energy efficiency by the participants. This is from the fact that costs can be saved by increasing energy efficiency, but high upfront costs must be dealt with first.

In terms of the measures that had been taken in order to reduce energy use and increase efficiency, a lot of discussion was focussed towards air management and cooling technology. Several different measures were mentioned such as free cooling, aisle containment and blanking, with some mentioning more efficient power distribution networks and UPS units. Management of the cooling and air distribution infrastructure is a key component of the guidelines given to EU CoC participants. A lot of the measures mentioned by participants were part of the compulsory guidelines, compulsory because of their relative ease of implementation relative to a beneficial savings in energy. Whilst most of the participants were not partaking in the EU CoC, it is a useful measure of whether they are placing their priorities in the relevant areas. Free cooling as a technology is only suggested in the code for new build data centres or those undergoing a retrofit. As most of the participants in the interviews mention they employed free cooling, it is clear that they have taken significant steps to curb their energy use. This is likely because free cooling has over recent years received a lot of attention and publicity as an energy saving technology, and helps to demonstrate to outside parties that the facility is going to be energy efficient and ‘greener’.

The EU CoC also places a heavy emphasis on IT equipment energy use. None of the participants mentioned considering the efficiency of the IT equipment, neither hardware nor software. It is noted that this is something which colocation providers would find hard to implement though as customers are able to install their own IT equipment in. It would be detrimental to the colocation providers business if they were to prescribe what equipment their customers could bring in. It is possible that some participants had made some steps towards improving their IT equipment, however it was not an option they mentioned. This may be reflected in the point made above, the technologies used to promote energy efficiency in data centres tend to be the cooling system employed, particularly when it comes to free cooling.

The use of metrics and measurement schemes

Many of the participants discussed the PUE metric and referred to it as not being a very useful metric to them. Whilst some do use it to present their facilities energy efficiency, others use it purely as a marketing device and for providing their customers with benchmark figures. A few noted that care needs to be taken when using the PUE and that a better alternative metric needs to be created, especially if it is to be used for schemes like the EU CoC and BREEAM. A positive however is that the metric has been cited as a useful educational tool that has encouraged participants to measure power use in more detail. In terms of measurement schemes, all of the participants were aware of the EU CoC but they had varying levels of involvement with it. Whilst some embraced the code and took on the guidelines provided, a couple had good reasons (although they were reluctant to share why) for not signing up. Following such schemes has brought benefits to those that have used them in terms of changing behaviour and improving their image as a business.

Reception and use of liquid cooling technologies

A couple of participants said that they do use liquid cooling technologies, but this had been used only for their HPC equipment. The colocation providers in the main said that it is difficult for them to implement a new technology in their data centre because of reasons mentioned above, agreement from all clients. Also these technologies tend to mean closer contact with the IT equipment, or specialist servers, something that is difficult for the colocation provider to provide. Their contracts normally allow clients to bring their own IT equipment in and to have control over their individual racks, something that would not necessarily be an option with liquid cooling technology.

The future of the industry

Energy of course played a key role in the participants thoughts about the future, mainly relating to how much costs may rise. Most participants said that this would lead them into considering their energy consumption and efficiency more closely. Whilst reducing energy consumption is still seen to be managed through focussing on the infrastructure, some participants mentioned a shift in focus towards the IT equipment.

This is going to be key heading into the future, it will become increasingly important to consider the IT equipment (both hardware and software), once infrastructure (cooling and electrical systems for example) has reached its peak efficiency levels.

5.4.3 Research Themes

Two themes that came through in the interviews are now carried forward for further in depth investigation, focussing on the operational and the technological level of the data centre. These themes are energy efficiency metrics and energy use of IT equipment. As demonstrated above these two themes are by no means the only two that came through from the interviews, they have been examined further in this thesis for the following reasons. Efficiency metrics are currently a key area of debate in the industry and whilst technology is advancing towards a point of peak efficiency, the methods used to assess and present this are falling behind. Therefore the examination of the most commonly used metric is important to be able to understand where metrics could be developed in the future.

Discussion in the interviews regarding the future of data centres was focussed on energy use in servers, and IT technology. Again, as mechanical and electrical infrastructure is reaching a plateau in energy efficiency, the server equipment is often left behind in the consideration of energy use. Being able to save energy within a standard existing server in a data centre may be a beneficial way of reducing overall energy consumption. Therefore Chapter 7 considers a way in which this could be achieved.

6 A Case Study and Critical Assessment of the Power Usage Effectiveness Metric

6.1 Introduction

The current chapter presents the results of a case study on the energy use effectiveness of a data centre located in the USA. Access to data for this study was made possible by a scheme known as the Open Compute Project which provides open source information and specifications for a data centre owned and operated by the social networking company Facebook.

Taking forward an emerging theme from the literature review and interview results, the case study is used as part of a critical analysis of one particular metric commonly used within the data centre industry to assess energy use, the PUE (introduced in Chapter 3). Measuring the PUE through using open source information also reflects on the difficulty of obtaining measurements to make accurate energy analysis in this industry.

The methodology used in this chapter can be applied to other data centres, particularly those that employ direct air cooling. Each system that consumes energy within the data centre is assessed and energy consumption is calculated wherever possible. A sensitivity analysis is conducted in order to find out how different parameters affect the PUE value. Further analysis of the sensitivity of the PUE value to server energy use is provided in Chapter 7.

6.2 Methodology

The aim of this section of the thesis is twofold, firstly to closely examine the PUE metric and secondly to highlight some of the difficulties that arise as a result of the data centre industry having a closely guarded approach to sharing data. It is generally very difficult to obtain enough data to carry out an energy analysis for a data centre, meaning that there is a lack of information sharing throughout the industry. The Open Compute project run by Facebook is a very unique project in that the company has provided and made available for general use a set of specifications and information files about one of its newest facilities. The aim of the Open Compute project is to share information with industry partners regarding the efforts that they have made to reduce their energy consumption in this particular facility, with the idea that sharing ideas is a positive approach to reducing overall industry energy use. Alongside the critical analysis of the

PUE metric, the Open Compute project will be indirectly reviewed in terms of how much useful information has actually been released.

In terms of using this open source information to find a PUE measurement, Figure 6.1 illustrates the methodology that will be used. The information within the open source documents will feed into each part of the PUE measurement including the energy consumption of the IT equipment, air distribution, cooling, and lighting systems. Losses through the electrical distribution network will also be estimated. All of this data will then feed into the PUE measurement which will be compared to the figure published by the Open Compute project.

A sensitivity analysis will then be performed on the data to investigate the sensitivity of the metric to changes in the values of certain parameters. The results from this alongside an analysis of the process by which the PUE was measured will be used to form a critical assessment of the metric.

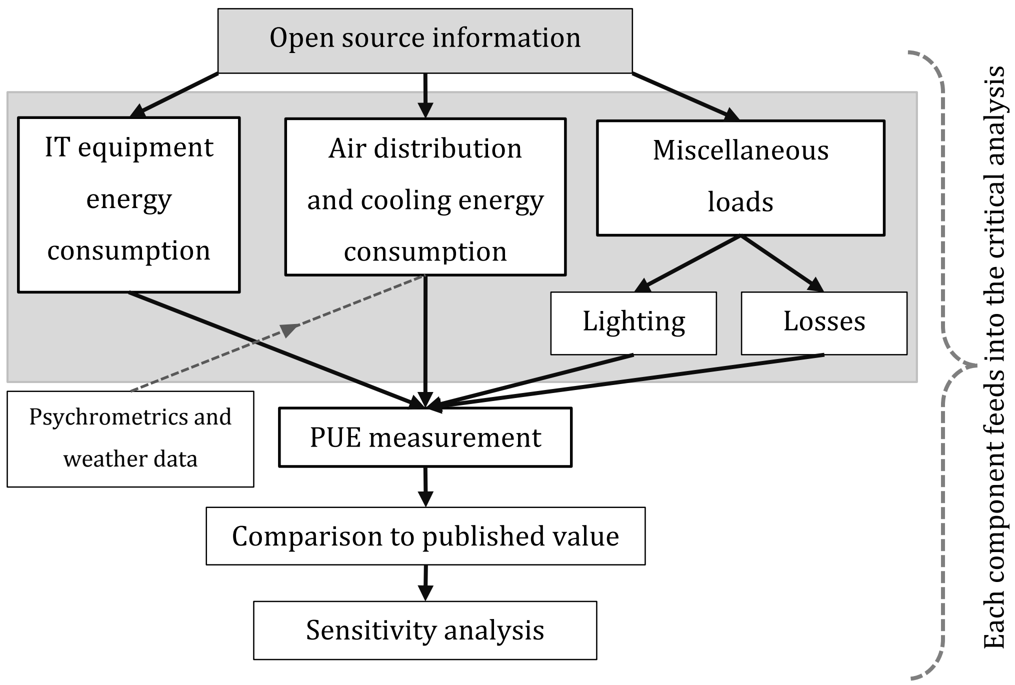


Figure 6.1 PUE measurement and analysis methodology using open source information

6.3 The Energy Consumption of the Servers

Measurement of the PUE requires the energy consumption of the servers to be accounted for. Whilst the Open Compute specifications do not indicate how much power is required to serve the total IT load in the data hall, the power supplied to each custom built server (illustrated in Figure 6.2, further details of this server can be found on the

Open Compute website (Open Compute Project, [no date])) is given as 450 W (this power will also supply the internal fans in the server units). It is also known that 30 servers are supported in each rack, the racks being arranged in groups of three, termed as '*triplet racks*' (Park, 2011). This leads to a maximum rating of 40.5 kW per triplet rack. However, there is no indication in the specifications as to how many of these triplet racks or servers there are in the entire facility, information which would make comparing energy use easier. Due to this, the energy consumption of the servers and the PUE will be estimated for one triplet rack. This does not account for other services for example switches or storage, and is the best estimate possible with this limited information.

A server's power draw varies over time and they rarely operate at 100% of their rated power. They will also rarely operate at 100% CPU utilization, sometimes operating at a value as low as 10% (Greenberg et al., 2009). Despite the variation though, an idle server (very low to zero CPU utilization) will still use a significant proportion of its rated power (Mohan Raj and Shriram, 2012). This proportion can be as high as 60% (Barroso and Hölzle, 2007). Without any usage data, a conservative assumption is made that the servers discussed above draw a continual 60% of the rated power, this gives an annual energy consumption of 212,868 kWh/annum per triplet rack. Using a peak value of 60% is a significant assumption to make (the sensitivity of this will be examined later in this chapter), given that the servers will not be using this level of power continuously throughout the year; however this assumption needed to be made given the limited amount of information provided in the Open Compute documents. Without direct access to the power use of the IT equipment, it is difficult to make an accurate prediction. This highlights a crucial inaccuracy that can arise from estimating the energy consumption or PUE for a data centre still in the design phase. The sensitivity of the PUE value to the assumed IT load is analysed in section 4.6.

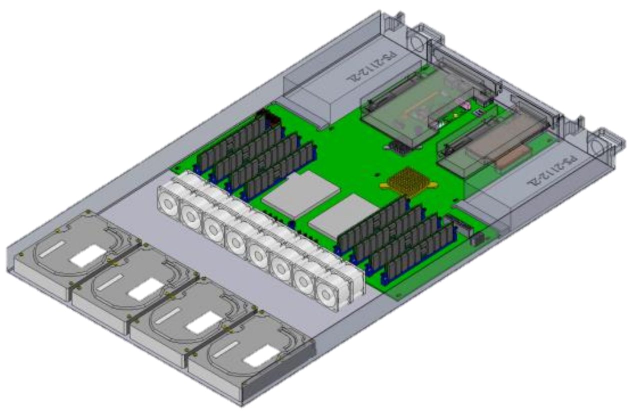


Figure 6.2 Drawing of an Open Compute server (Ogrey et al., 2013)

6.4 Energy Consumption of the Air Handling and Cooling Systems

Energy costs for cooling and distribution of the air to maintain integrity of the electrical components can be significant. Depending on site location, air conditioning can consume a significant fraction of the energy required to run the ancillary services. In the case of the Prineville data centre, cooling is achieved through air-side economisation, where filtered outside air is delivered directly to the servers, and a high pressure misting system provides evaporative cooling and humidification (Park, 2011). As discussed in Chapter 3, this is an efficient method of providing cool air as it negates the need for a compressor. The design of the cooling system for this data centre is unique though, in that it has a ductless air supply system. Illustrated in Figure 6.3, the cooling system consists of a series of rooms through which outside and return air flows. Outside air is drawn in through a series of filters before passing through a misting system which is used to increase humidity and to cool the air. After this, the air passes through a fan wall made up of large fans before being delivered into the data centre itself. Within the data centre, the servers are contained within hot aisles in order to separate air flows. Warm return air is drawn up into a ductless space above the server room before being either rejected to atmosphere or mixed with outside air. This is an efficient means of providing cool air to the servers, as the large spaces through which the air passes results in lower pressure drops.

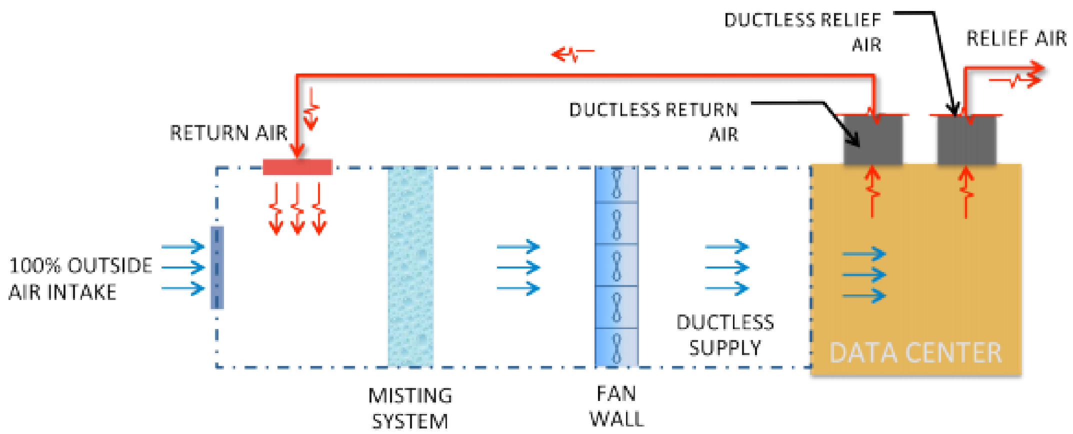


Figure 6.3 Diagram of air flow in the Prineville data centre (Park, 2011)

The cooling system has several different modes of operation which depend upon the properties of the outside air. As the local weather and hence air temperature and humidity change throughout the year, the cooling system needs to adapt to these constant changes. In the case of this data centre, there are seven modes of operation as summarised in Table 6.1. Which mode is in operation at any one time depends on the properties of the outside air.

Table 6.1 Description of each mode of operation for the Prineville data centre's cooling and air distribution system

Mode	*Operation of Cooling System*	*Outside Air Properties (OA)*
A	Evaporative cooling provides humidification, mixing of outdoor/return air to raise temperature	OA<11.11°C (T_{wb}) and <5.5°C(T_{dp})
B	100% outdoor air, evaporative cooling provides humidification	OA>11.11°C(T_{wb}) and <5.5°C (T_{dp})
C	100% outdoor air	OA>18.33°C (T_{db}), >5.5°C (T_{dp}), <26.67°C (T_{db}), <15°C (T_{dp}) and <65% (φ)
D	100% outdoor air, cooled by evaporative cooling	OA>26.67°C (T_{db}), >5.5°C (T_{dp}) and <18.76°C (T_{wb})
E	100% outdoor air, cooled by evaporative cooling	OA>26.67°C (T_{db}), >5.5°C (T_{dp}) and >18.76°C (T_{wb})
F	Mixing of outdoor and return air	OA>26.67°C (T_{db}), >5.5°C (T_{dp}) and >21.28°C (T_{wb})
G	Mixing of outdoor and return air	OA>18.33°C (T_{db}), <15°C (T_{dp}) and >65% (φ) or <65% (φ), >5.5°C (T_{dp}) and <15°C (T_{dp})

Note: T_{db} = dry bulb temperature, T_{wb} = wet bulb temperature, T_{dp} = dew point, φ = relative humidity

6.4.1 Air Distribution – The Power Required to Operate the Fans

The distribution of air through the data centre and across the servers consumes energy, as fans are required to move the supply air flow and overcome the pressure drops within the system. Air delivered to and from the servers will experience a series of pressure drops, for example due to filtration, distribution to the cold side through vents and down aisles, and similarly through the return path. This will depend heavily on the design of the data centre (e.g. flow paths, duct sizes, distribution tiles).

The other significant pressure drop is across the server itself. Manufacturers generally specify on-board fans to ensure a low net pressure drop from the front to the back; one of the costs of cooling high density electronics is the energy associated with accelerating air to high velocities to affect the necessary heat transfer from the surface of heat exchangers and components within the server. Since fans generally sit within the rack server, this implicitly includes part of the energy costs for air distribution within the IT load. A move away from the multitude of small (inefficient) on-board fans to a larger fan unit that pressurised the inlet side of the server could lead to a larger PUE, despite a total reduction in energy consumption. This will be addressed in Chapter 7.

In order to calculate or provide a design estimate of the power required to operate the fans, the pressure drop across the system needs to be quantified. The total volume flow rate of air required to deal with the heat load produced by the servers is also needed. Parameters such as these are necessary during the design stage of a project to provide an estimate of total operational energy use. The open source documents do not provide information regarding the fans in use or the power required to operate them, therefore pressure drops and air flow rate would be the only way to estimate power usage.

In terms of air delivery, the Open Compute documents indicate that each server would require a maximum 0.028 m^3/s of air flow equating to 2.55 m^3/s per triplet rack (Furuta, 2011). Along with pressure drop values for the whole system, this design air flow rate can be used to determine the total power required to operate the air distribution system. Open Compute state that the filters in use have a pressure drop of 67 Pa at an air flow of 2.54 m/s (Park, 2011). However, no further details on the pressure drops across the whole system have been provided in the open source documents.

Unfortunately, this means that it is therefore not possible to calculate the power required to provide air distribution within the data centre. Energy consumption for air distribution is therefore incorporated into the estimated energy use of the evaporative cooling system in the following section. By incorporating the energy required to operate the air distribution system in with the cooling system, a total energy consumption estimate can be arrived at. Discussion around the issues of doing this along with the problems of estimation whilst measuring the PUE can be found in section 6.8.2.

6.4.2 The Energy Consumption of the Evaporative Cooling System

Electronic components have increased failure rates when operating above or below allowable temperature and humidity levels (Ohring, 1998). The ASHRAE TC 9.9 guidelines recommend a delivery temperature range of 18°C-27°C and a humidity range of 5.5°C dew point to 60% relative humidity and 15°C dew point (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011), although the maximum allowable envelope within ASHRAE is actually wider than this, see Chapter 2 for more details. Data centres which employ air-side economisation take advantage of a recent expansion of the recommended ASHRAE operating envelope, as they allow ambient outside air to be delivered directly to the servers for a greater proportion of the year. Whilst there is an associated increase in expected failure rates associated with operating outside the allowable envelope, ASHRAE provide factors which show that this increase is actually very low, especially for regions such as North America (American Society of Heating Refrigeration and Air-Conditioning Engineers, 2011).

In order to calculate the energy consumption of the cooling system, weather data must be used for the local area. Data from the same weather station given in the Open Compute specifications was used for this case study (sourced from (National Renewable Energy Laboratory, 2008)). The data used is a Typical Meteorological Year 3 (TMY 3) data set for Redmond, Oregon, USA[29].

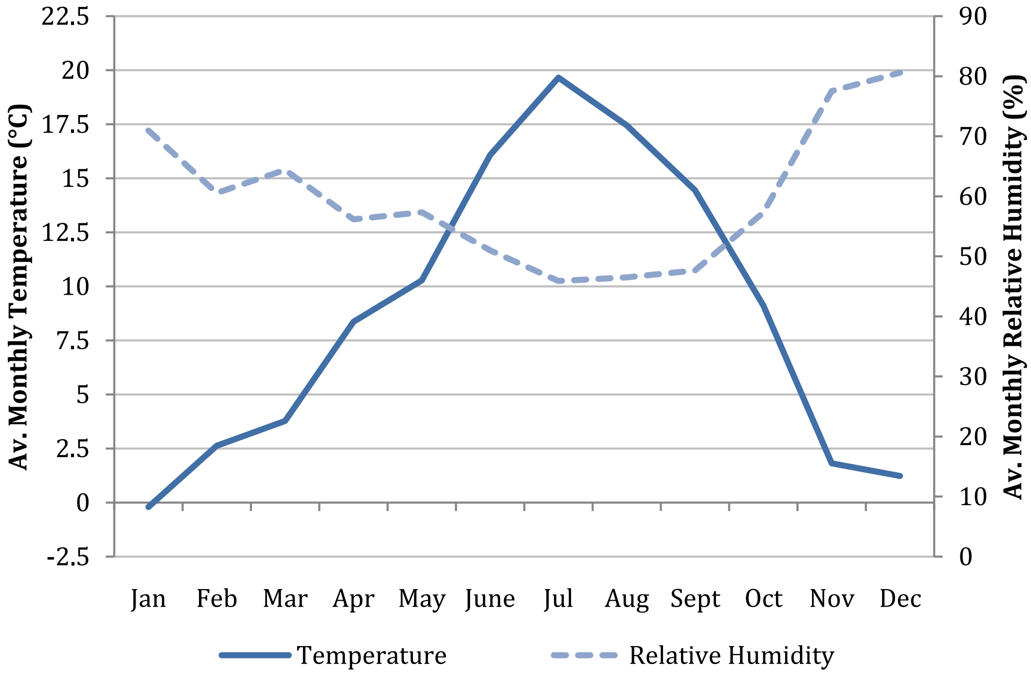


Figure 6.4 Average temperature and relative humidity for Redmond according to the TMY data set

The mechanical specifications for the Open Compute project provide a description of the modes of operation for the cooling system, and the psychrometric properties required for each one (Table 6.1). Figure 6.5 (plotted using a Matlab code (Ngai, 2011)) shows the weather data plotted on a psychrometric chart with the outside air properties for each hour of the year represented by a point on the graph. The chart has been used to calculate the proportion of the year that the misting system is required to provide evaporative cooling or humidification.

Also drawn on the graph are the areas within which the system is operating under certain modes. As a simplification, it is assumed that each mode of operation will be in

---

[29] A data set compiled from actual readings from representative years between 1978 and 2005.

use for an hour for every point on the graph which falls within it. Figure 6.6 shows the percentage of the year that each mode of operation would be used for based upon the TMY 3 data set.

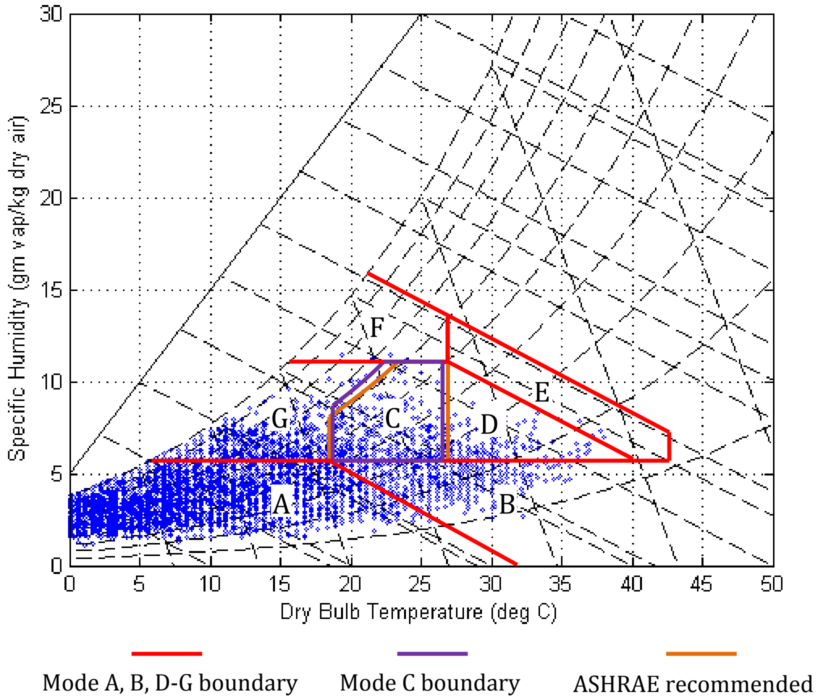
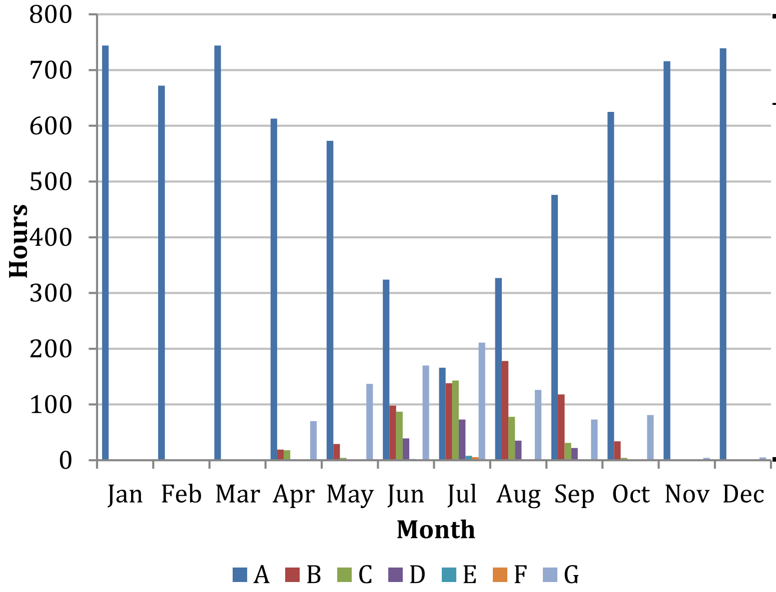


Figure 6.5 Psychrometric chart with weather data for Redmond, Oregon and the cooling system modes A to G for the data centre as per the Open Compute information. Also shown is the ASHRAE recommended envelope

What is immediately clear from Figure 6.5 is that for the vast majority of the year, the outside temperature is below that required in the data centre. However, a lot of hours fall within the 'A' boundary meaning that whilst the outside air can be mixed with warm return air, it will then require humidification to bring it to the correct relative humidity (see Figure 6.7). Therefore as boundary 'A' covers a large percentage of the year, humidification may be needed for a significant proportion too. This becomes more apparent when considering Figure 6.6, which illustrates the significant proportion of the year that mode of operation 'A' would be used, a mode which requires humidification according to Table 6.1.



Mode of operation	*Percentage of year (%)*
A	76.70
B	7.01
C	4.17
D	1.94
E	0.11
F	0.06
G	10.01

Figure 6.6 Number of hours per month and percentage of the year that each system mode (A-G) is in operation for, based upon TMY weather data

Figure 6.6 sets out the number of hours in each month that come under each mode of operation, it is clear that the warmer hours are in the summer, but are still accompanied by cooler hours overnight. The analysis shows through the results in the table within Figure 6.6 that the misting system will need to be in operation for 85.8% of the year (when either mode A, B, D or E is in use). Once this was known, it was possible to estimate the power required to operate the cooling system by considering the amount of water that needs to be pumped.

Estimating the power required to operate the cooling system meant finding out the power required to pump the water around the system and through the misting jets. In order to do this, it was necessary to calculate the amount of water that needs to be delivered to the supply air through the misting system.

The TMY 3 weather data was once again consulted to calculate the moisture content of the outdoor air relative to the required moisture content of the supply air to the data hall. Calculation of the moisture content (θ) was possible by using the psychrometric relationship between relative humidity (φ), air pressure (P), and saturation vapour pressure (P_s) as shown in equation 6.1. Detailed explanations of this equation can be found in (McQuiston, 2005, Singh et al., 2002);

$$\theta = \frac{(0.6219\varphi P_s)}{(P - \varphi P_s)} \qquad (6.1)$$

Given that;

$$For\ 0°C < T < 63.0°C \qquad P_s = 610.78e^{\left[\frac{17.269T}{237.3+T}\right]} \tag{6.2}$$

$$For\ T < 0°C \qquad P_s = 610.78e^{\left[\frac{21.874T}{265+0.9615T}\right]} \tag{6.3}$$

Equations 6.2 and 6.3 calculate the saturation vapour pressure (Singh et al., 2002), having been modified for different temperature ranges based upon original work by Dilley (Dilley, 1968). The above set of equations are used to calculate the quantity of water that needs to be added to the outside air per hour given the ambient conditions and the required conditions of the air supplied to the servers. The TMY 3 weather data was used to see how much cooling would be required for each hour of the year to bring the air to the minimum standards required by the design requirements for the data centre (boundary C in Figure 6.5).

For each hour within the year, the amount of water required was estimated using the above equation. Each hour was not necessarily brought into the middle of area 'C', but brought in to the outer limits of it. For example, the water requirements for hours that fell within boundary 'A' were estimated based on bringing the temperature up to 18.33°C and the relative humidity to 43%, as illustrated in Figure 6.7. A more accurate representation of the required amount of water could be found by considering every data point and hence every hour within the year. However, as this study is only looking for estimation, the above method is well suited.

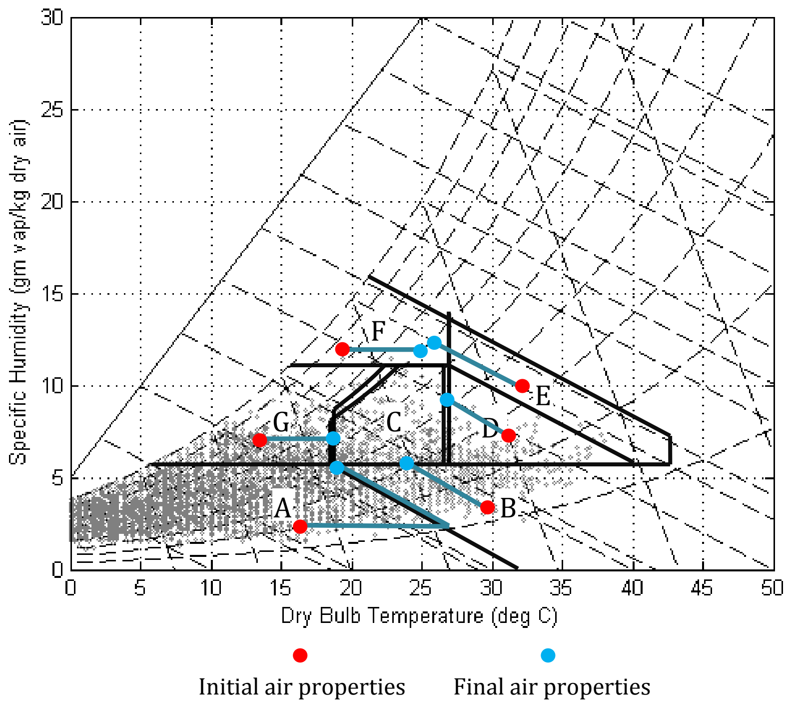


Figure 6.7 Psychrometric chart illustrating how each data point was treated during the analysis

Table 6.2 contains the results from the data analysis, presenting the required water flow rates in order to cool each rack. The figures displayed for maximum water flow and water required over the year are all only for one rack, as the total number of racks in the data centre is unknown. Figures are all based on the requirement of 2.55 m^3/s air flow rate per rack. Annual water consumption is also illustrated in Figure 6.8, showing the variation in the required water flow throughout the year.

Table 6.2 Maximum water flow and total water volume required for each mode of operation where evaporative cooling is used

		Mode of operation A	B	D	E
Temperature	(°C)	18.33	22.5	26.67	26.67
Relative humidity	(%)	43	34.5	37.5	55
Specific volume	(m^3/kg_{air})	0.83	0.85	0.86	0.87
Air flow rate per rack	(m^3/s)	2.55	2.55	2.55	2.55
Mass flow rate per rack	($kg_{air}/hour$)	11013.4	10857	10667.6	10594
Total Water Required at 15°C	($m^3/year$)	137.09	4.01	6.50	0.31
Max water flow at 15°C	**($m^3/hour$)**	**0.054**	**0.031**	**0.052**	**0.043**

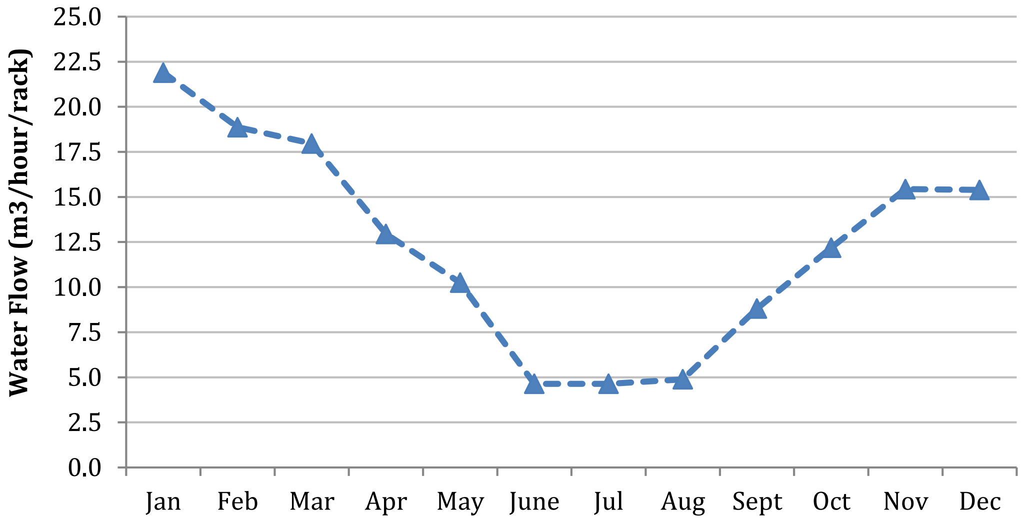


Figure 6.8 Water consumption by the cooling system over a full year

The total water consumption can be used to measure the WUE. Discussed in Chapter 3, this metric allows the water usage to be assessed in relation to the IT load. Using the figures above, and the IT load found previously, the WUE is measured as follows;

$$WUE = \frac{Total\ water\ usage\ (L)}{IT\ load\ (kWh)} \tag{6.4}$$

$$WUE = \frac{147.91 \times 10^3}{212{,}868} = 0.69 \tag{6.5}$$

The owners of the data centre, Facebook, stream their current PUE and WUE values online making them available for the public to view. According to their online figures, their peak WUE value between July 2013 and July 2014 was 0.68 l/kWh, with a trailing annual average for that period of 0.29 l/kWh (Facebook, 2014). The maximum figure given by Facebook is comparable to the figure calculated in this case study; however this is only a maximum. It is therefore likely that the figure measured here is an overestimation, which will be down to the assumptions that needed to be made during the analysis. It is therefore possible that the PUE measurement may also be an overestimate. Despite this, it is still a good close approximation to the published figures given the assumptions that were made.

In terms of calculating the energy consumption of the system, calculations conducted using the above equations estimate that the amount of water required during a typical year would be 148 m^3 per triplet rack (equivalent to 6.1 m^3/kW of cooling load given continuous 60% server power utilisation), with a maximum flow rate during the year of 0.054 m^3/hour per triplet rack (0.7 l/kWh). Given the water flow, it would be possible to then calculate the energy required to pump this through the misting system. However no details about the energy utilisation during pumping, spray generation or filtration are provided.

This means that it was necessary to use power consumption figures from another data centre in order to arrive at a suitable figure. A comparison was therefore made to a data centre in Bedford, UK which uses evaporative cooling and requires 4 kW of energy to run the air distribution fans and high pressure pumps when the misting system is in operation (Modern Building Services, 2010). Using a case study on this particular data centre presented in Modern Building Services, it was possible to estimate the energy requirements for a system like the one used at the Prineville facility. The misting system in this particular data centre serves a maximum capacity of 384 racks and delivers a total amount of water equal to 1.6 m^3/year/kW of cooling load. 4 kW for 384 racks implies that 10.42 W is required for each rack and given that Prineville uses triplet racks, it could be assumed that 31.25 W would be required for each triplet rack. However, the water flow rate in Prineville is 3.8 times that in Bedford and so it could be approximated that 119 W is required to operate the misting pumps and the air distribution fans in the Prineville data centre. Given that it was found out that the system would be required for 85.76% of the year, then the consumption will therefore be 893 kWh / annum per triplet rack.

However, since the data given in the case study incorporates air conditioning *and* air distribution (Modern Building Services, 2010), the figure above is increased to incorporate the 14.2% of the year when the misting system is not in operation, giving a consumption of 1,048 kWh/annum (the energy consumed for air distribution and air conditioning). This will be a slight over estimate in energy requirements, since the droplet misting is not required, however filtration has not been included in the calculation when it will in fact increase energy use.

This transposition of energy use measurements does pose a risk to the accuracy of results due to the difference between the two data centres involved. The Bedford data may have a different efficiency associated with its misting system compared to the Prineville one, it may for example has less or more efficient pumps and associated equipment. Distances between end use and plant rooms will be different and will inherently have an effect on the pumping power. Unfortunately, none of these factors are available to quantify and therefore they must remain a possible source of error. However, as one of the aims of this section of work is to examine the PUE metric in terms of what affect different parameters have on the results, the transposition provides a number that can be used for this purpose. Therefore it has been determined that for the purpose of this research, the transposition is an acceptable risk as it enables the examination of the PUE metric.

6.5 Miscellaneous and Other Loads on the Electrical System

6.5.1 Lighting

The data centre uses Light Emitting Diodes (LEDs) which are much more energy efficient than incandescent lamps, they also emit a much lower amount of heat. The facility also powers them in an unusual way, opting to not power them through a normal connection to the mains. Power is instead delivered to them through network cables rather than electrical wires, a method known as Power over Ethernet (PoE). Using this method of lighting in the data centre means that raw materials are saved through the reduced need for electrical wiring providing, thus providing an additional environmental benefit. It is therefore assumed for this study that the lighting load is included in the figures given earlier for the IT load.

Using PoE can have an impact on the final PUE value. Through using PoE it means that energy consumption for the lighting is part of the IT load within the data centre. Therefore it is included in the denominator of the PUE, which can lead to a reduction in its value when compared to more traditional lighting arrangements. Whilst there is a reduction in the energy consumption due to the use of LED lamps, there may be an additional decrease in the PUE by including the lighting in the denominator. This is one of the ways that the PUE metric can be used incorrectly.

6.5.2 Electrical Losses

The power distribution system in a typical data centre will incur losses of energy through inefficiencies in its equipment, mainly the transformers, PDU and the UPS units. The Prineville facility makes use of AC supply with a DC backup system to protect servers in the event of a power supply failure, which requires a continuous supply of DC power to the servers. With a 95% efficient power delivery network (Open Compute Project, 2011), this is another contributing factor towards high efficiency in the data centre. There is also a loss due to power transformation of 2% (Open Compute Project, 2011). The energy losses in the power network are calculated here based upon the IT load, air distribution and evaporative cooling systems, totalling a loss of 15,854 kWh/annum per triplet rack.

6.6 Energy Analysis Results for the Prineville Data Centre

Collating the values obtained from the analysis above, the PUE was then measured for one triplet rack within the Prineville data centre. Strictly speaking, this should be counted as a *partial* PUE or a pPUE as it does not count for the energy consumption of the whole facility. However the term PUE will be used in this case as one triplet rack is being treated as a representative proportion of the whole facility. If the figures used in this measurement were to be scaled up to include the whole facility (i.e. increased by a factor equal to the total number of racks), then the PUE measurement would be the same. Equation 6.6 shows the results of the PUE measurement using the figures presented earlier.

$$PUE = \frac{212{,}868 + 1{,}048 + 15{,}854}{212{,}868} = 1.08 \tag{6.6}$$

Facebook reports that the Prineville data centre achieved a PUE of 1.08 (for the end of quarter three in 2011) (Open Compute Project, 2011). The live PUE reporting tool provided for the data centre also indicates that the PUE has fluctuated between 1.11 and 1.07 over the year leading to June 2013 (Facebook, 2013). From the data analysis in this work, the PUE measurement results present a figure of 1.08. Therefore it would appear that the calculations above appear to enable this PUE figure to be reproduced accurately, even though loads such as plant room lighting and generator heaters have not been included. It is clear from the numbers in equation 6.6 that the energy usage of the cooling system is very small compared to the IT energy usage; this is due to the design of this particular data centre. Although assumptions were made during the calculation, this means that any inaccuracies in estimating the energy consumption of the cooling system are unlikely to cause large differences in the calculated PUE in this case. Differences in the cooling system energy usage are examined below in the sensitivity analysis.

6.7 Sensitivity Analysis

Despite the assumptions that needed to be made, the above PUE analysis provides a good estimate of the published value, concluding that the Open Compute value is a credible figure. In this section a brief sensitivity analysis of the effect of these assumptions on the calculated PUE is carried out. The PUE is defined as the total energy use divided by the energy consumed by the IT equipment (W_{IT}). Total energy consumption is equal to the total amount of energy used by the equipment and infrastructure in the facility (W_T) plus the energy losses due to inefficiencies in the power delivery network (W_L), hence;

$$PUE = \frac{W_T + W_L}{W_{IT}} \tag{6.7}$$

In a typical data centre the total energy consumption, W_T, includes the energy used by the IT equipment and supporting infrastructure. The following sensitivity analysis considers the effects that the electrical losses, IT load and the humidification system at Prineville have on the PUE. The analysis consisted of three scenarios to see which parameter causes the largest changes in PUE. The power requirement for the humidification system and the efficiency of the power distribution network were increased or decreased by 20% and 1% respectively. Then the IT equipment was investigated by decreasing the power draw of the servers to 20% and increasing it to 80%. Within each scenario, only one parameter was changed. The lower value of 20% was chosen as this would be representative of the server still drawing power whilst doing no useful work. It has been shown that servers can still draw a good proportion of their full power requirements when doing no work (SPEC, 2012). Table 6.3 shows that a server with the same name plate rated power uses around 25% of its total power rating when doing no useful work. It should be noted that of course server power use does fluctuate frequently, however in this case the sensitivity analysis is taking a snapshot view of the PUE and what factors affect its value. Figure 6.10 considers in great granularity the effects of IT power use on the PUE value. When it came to decreasing the utilisation of the IT equipment, the cooling and air distribution power requirements were not scaled down to account for the lower IT load (whilst the utilisation of IT equipment fluctuates, the supporting infrastructure does not always react to the changes (Malone and Belady, 2008)).

Figure 6.9 illustrates the percentage changes in the PUE value depending on which parameter has been altered. It demonstrates that an alteration in the energy use of a particular system causes a corresponding change in the PUE value. Changes in the energy consumption of the cooling system have minimal effect on the PUE value, whereas a much larger difference is noted when changes in the energy consumption of the IT load and efficiency of the power supply network are made. It is clear from the graph that the greatest impact on the PUE came from changing these latter values. For the power supply efficiency, the changes are even more important when considering that this was only altered by ±1%, leading to a corresponding ∓1% change in PUE. By comparison, the PUE changed minimally (±0.1%) when the power required for humidification was altered by ±20%. The significance of the power distribution efficiency (η_{pd}) and the IT load can be explained by the following expressions, given their relationship to the total energy use in the facility;

$$W_L = \frac{W_T}{\eta_{pd}} - W_T \tag{6.8}$$

Therefore;

$$PUE = \frac{W_T + \frac{W_T}{\eta_{pd}} - W_T}{W_{IT}} \tag{6.9}$$

$$PUE = \frac{W_T}{W_{IT}\eta_{pd}} \tag{6.10}$$

Equation 6.10 highlights the fact that whatever energy savings are made in the data centre, the power distribution efficiency and IT load will still have a significant effect on the overall PUE value. The effect of this is amplified in the case of the Prineville data centre due to the use of a direct air cooling system.

The sensitivity analysis also demonstrates that the PUE should be used with caution due to the effect of IT load changes. For example, if the IT load increases through improved utilisation, the PUE will reduce in spite of the increased overall power consumption in the data centre. This highlights the fact that a reduced PUE can predict that a data centre is operating more efficiently even though its overall energy consumption has increased, and vice versa. This is one of the PUE's main limitations as an energy efficiency metric.

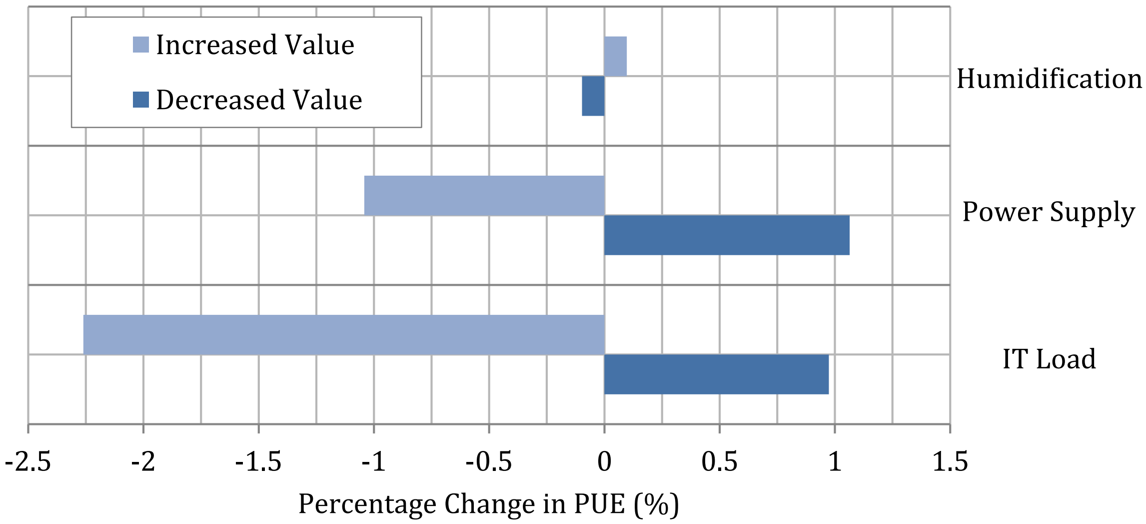


Figure 6.9 Percentage changes in PUE as a result of altering different parameters within the metric: (i) humidification by ±20%, (ii) η_E by ±1%, (iii) IT load by -40% and +20%

In order to carry out a closer examination of the effects of IT load on the PUE measurement, data from the SPEC has been used. SPEC provides an online repository of data regarding power use and efficiency within IT servers (SPEC, 2011). Using data provided by SPEC for a Dell Inc. PowerEdge R820 server[30] (with a maximum load power rating of 450 W) along with the results from the energy consumption analysis already carried out, the PUE is examined by considering how it changes with the IT load. Table 6.3 below contains the data for this particular server as provided through the SPEC data base. Alongside this, are estimated PUE figures based upon the analysis of the Prineville data centre. It must be noted that as the IT load is decreased, the cooling load is not altered. As mentioned previously, the cooling load is not always scaled back in real situations. Also, the aim is to see how the IT load alone can alter the PUE value.

[30] This server was chosen as it had a similar power rating to those used in the Prineville data centre, as stated in the Open Compute documents.

Table 6.3 Power use of a Dell Inc. PowerEdge R820 (Intel Xeon E5-4650L, 2.60 GHz) server (SPEC, 2012).

Performance		*Power*	*PUE*
Target Load (%)	Actual Load (%)	Average Active Power (W)	
100	99.70	446	1.0545
90	90.10	414	1.0547
80	79.80	358	1.0550
70	70.00	307	1.0554
60	60.00	268	1.0558
50	50.10	237	1.0562
40	39.90	218	1.0565
30	30.00	188	1.0571
20	20.00	167	1.0577
10	10.00	149	1.0583
0	0.00	110	1.0603

It can be seen in Table 6.3 and also Figure 6.10 that despite the significant increase in server power use, the PUE actually decreases. Whilst it does only decrease by a small fraction, it illustrates the fact that the metric is not representative of power use in a data centre and should not be regarded as an energy efficiency metric.

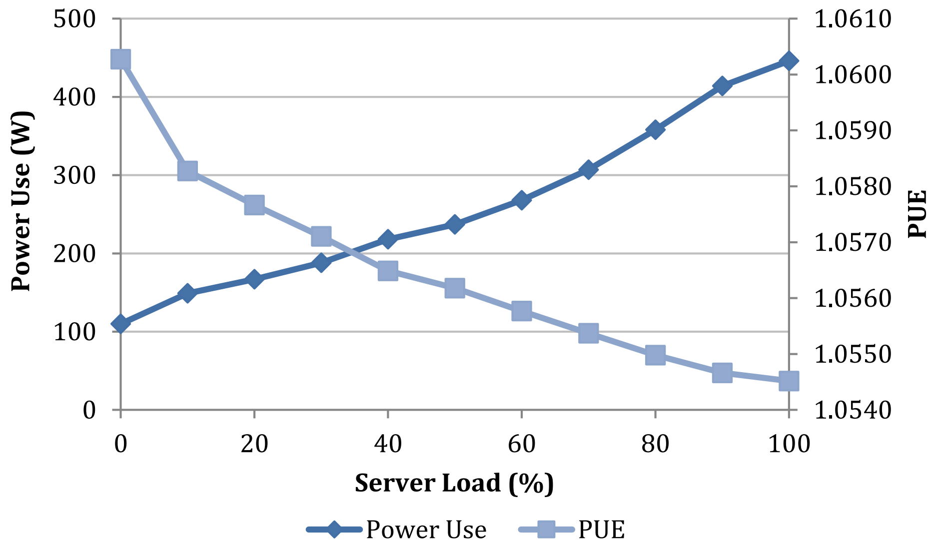


Figure 6.10 Changes in PUE with increasing server load

6.8 Critical Assessment of the PUE Metric

Open source data has enabled a PUE value to be measured for a particular data centre, despite the limited amount of information that was available. The main aim of this case study has been to assess the use of the PUE metric, whilst also drawing a parallel with the difficulties of accessing useful data regarding energy use within the data centre industry. The following text provides an in depth review of the PUE metric including its usefulness, potential issues and ability to assess energy use.

6.8.1 Benefits

Measuring the energy efficiency and consumption of a data centre is clearly very important if carbon emissions from the ICT sector are to be reduced. The PUE metric is a useful tool for the individual data centre that needs to assess overall energy consumption, as in the process of measuring it an energy audit needs to be carried out. A large PUE measurement for a data centre is a clear indicator that there is a significant overhead of power being used to support the IT equipment. The metric is also useful for recording annual variations in usage effectiveness as technically it should be monitored continuously over a year.

From the review of the literature in Chapter 3, it is also clear that the metric is commonly used within the industry to present energy use in the data centre. Therefore it has gone some way to creating competition between facilities and companies, driving up efficiencies, as advertised PUE values become lower. Companies will use the measurement to advertise their facility (as indicated in Chapter 5), hence introducing an element of competition regarding energy use. This has benefitted the industry and has encouraged reduced energy consumption within the data centre.

It is a simple metric providing the means to easily visualise how much energy is being used to support the IT equipment, however this is where its usefulness becomes limited. There are some issues with this metric as discovered through this case study, the following section explains some of these issues and how the case study has helped to highlight them.

6.8.2 Limitations

Through the case study measurement of the PUE, a few issues have become clear regarding the use of the metric. Measurement of the PUE involves accounting for the energy consumption of a number of different systems and processes within the data centre. It has been found that unless direct access to data and information is provided, detailed measurement of the PUE is a difficult task. Whilst the Open Compute project is open source, there was not sufficient information to conduct a detailed re-calculation of the PUE and hence the calculation could only be done based upon one triplet rack.

It is crucial that an accurate IT load is used for the PUE, and that it is not based upon the rated power use of the equipment (Itoh et al., 2010). Accuracy in the IT load is one of the major factors affecting the measurement of the PUE metric, as utilisation of the servers has an important effect on IT energy consumption and hence the overall PUE value. The sensitivity analysis showed that the PUE will tend to increase if the IT load decreases and the cooling system is not scaled back to account for this change. Careful use of virtualization where servers are progressively loaded to capacity could allow some servers to be switched off (Schulz, 2009). This will reduce energy consumption (although unless the cooling system power is reduced the PUE will increase).

The PUE does not accurately reflect the overall energy efficiency of a data centre as it does not include the efficiency of power use in the IT equipment (the IT load is hidden within the 1 of the PUE). A data centre with a low PUE but with low server utilisation can be less efficient overall than a facility with a higher PUE value and a higher server utilisation.

Problems can also arise with the reporting of the PUE metric. Ideally, the PUE should be reported along with information regarding what period it is measured over, and where the power measurements are taken. This would limit the problem that has arisen with the PUE value of different facilities being compared. This was never the original intention of the metric – differences in locality (and hence climate) will inherently affect the ultimate PUE achievable at a given location due to the requirements for cooling of the servers. The comparison of a facility where free cooling is available with one that is not, says little about the inherent efficiency of the equipment. A comparison of data centres that includes climate information and data analysis alongside the PUE would give a more comparable ranking of energy efficiency.

At the design stage of a data centre, an estimated PUE value will normally be used as a marketing tool to potential owners. However, once the data centre is operational it is normally only able to achieve this PUE value when it is at full IT capacity. This means that comparing a PUE value of data centres is somewhat meaningless unless it is known whether it is operating at full capacity or not. The PUE metric is more useful for data centres that are operating at full IT capacity, as this should mean that the supporting infrastructure is also operating at its designed capacity.

The reuse of waste heat either within the data centre or by an external consumer (such as heating in an office) is a valuable method of reducing overall energy consumption.

However, the PUE itself does not take into account any reuse of energy, for this purpose the Green Grid created the ERE or the Energy Reuse Effectiveness (The Green Grid, 2010a). As the PUE only assesses the actual consumption of energy within the data centre, it is not designed to incorporate any reuse of waste heat from the servers. Technology that enables the reuse of heat within the data centre itself such as the absorption chillers suggested by Haywood et al (Haywood et al., 2012) is not able to have its full benefits reflected in the PUE. Therefore in situations such as this, careful use of the correct metric is needed to enable the technologies full benefits to be realised, Haywood et al suggest the use of an alternative metric to the PUE.

6.9 Limitations of the case study

A number of significant assumptions were required to complete the PUE analysis using the Open Compute project, together with (at times) convoluted ways of estimating power requirements. A more realistic and accurate PUE could have been calculated if information about the number of racks and the mix of ancillary IT equipment, together with engineering specifications of the air handling units was available. Despite this, the estimated PUE agreed with that published through the Open Compute project.

Whilst the amount of information available within the open source documents was detailed compared to what is available as a whole from the industry, it did not go into enough detail to conduct a full PUE measurement. This is understandable as the Open Compute project has already gone further than other companies in the effort to make the industry more open. The limitations of this study have helped to show the problems that can arise from the lack of openness when it comes to energy usage within this sector.

6.10 Summary and Discussion

Metrics are essential if energy efficiency and energy consumption are to be assessed in a data centre; however they must be suitable and fit for purpose, whilst also being relatively simple to use. A case study has been conducted in order to analyse the PUE metric which has become a de-facto industry standard. This case study has demonstrated the type of detailed engineering data that is needed in order for meaningful PUE values to be calculated. Even when open source specifications are given around IT choice, this level of detail is not necessarily available. The assumptions made in this work enabled good agreement of the published PUE for a data centre to be completed. However the data centre in this study is a special case and has a unique design, leading to low energy consumption by the cooling system. Due to this, the energy use of the cooling system has little effect on the PUE value. The case study also illustrates the ease of simplifying the PUE calculation to achieve a good value, something which may occur during the design process.

The sensitivity analysis shows that once a data centre has made significant energy savings through reducing cooling system power requirements (as in the case of the data centre assessed in this work), the efficiency of the power supply network and the IT load will have the largest effect on the PUE value. The analysis also highlights problems with the relationship between changing IT loads and the PUE value, demonstrating that the PUE metric must be used with caution. This is due to the fact that energy saving measures such as virtualization can actually increase the PUE, falsely implying a less energy-efficient operation.

The PUE metric has over time become a marketing tool to present the overall energy efficiency of a facility. Despite the fact that PUE values cannot be directly compared, its use has helped to create an industry where data centres have become more competitive in their energy use and efficiency. The metric has helped to set benchmarks for energy consumption relative to the IT load in the data centre. However, the metric does not show a true representation of energy efficiency in a data centre, due to it not including the efficiency of the servers in their required operation. This is also reflected in the results of the interviews presented in Chapter 5, as some participants remarked that the PUE is not always useful and does not represent energy efficiency effectively. The data centre industry needs a metric which incorporates the energy efficiency of all equipment and infrastructure including the server units, but also a metric which is useful when comparing one facility to another. Incorporating IT equipment operational efficiencies and also climate or weather information may allow for a fuller picture to be painted about energy efficiency, allowing more direct comparisons between facilities.

There is also the aspect of aligning the goals of different external parties to the data centre, as discussed within Chapter 3 and highlighted in Chapter 5 when discussing the landlord – tenant relationship. A metric which could incorporate aspects wider than just direct equipment energy consumption may help to address wider goals than just reducing the energy use of the data centre. For example, if a business using data centre facilities as a tenant has a goal of reducing the environmental impact of their operations, then a metric which demonstrates how the use of IT contributes towards their business would help. The PUE metric allows the tenant or the user of the data centre see how efficient the facility is at providing a suitable environment for their data, however it doesn’t describe how suitable this environment is, how much energy is being used or how efficiently their data is being processed. For the energy provider, the PUE metric does not describe the fluctuations in energy use or when peak times, or periods of low energy use may occur.

Significant gains have been made in increasing the efficiency of data-centres through careful choice of location and ancillary equipment. PUE was never intended to bench-mark data-centres however the understandable desire to rank facilities has led to some distortion of the PUE calculation process. Little discussion has focussed on the efficiencies of the IT processes; the PUE metric does not necessarily drive such improvements. An analogy between a manufacturing environment and a data centre where raw materials (unprocessed data or requests for information) are processed (computed) before dispatch (through networks) would suggest that a careful examination of the unit cost of operation will lever efficiency gains. This will allow data-centre operators to balance responsiveness and resilience in a holistic sense.

7 Investigation of Server Air Flow and Fan Energy Use

7.1 Introduction

The previous chapter explored one of the metrics that is commonly used to assess energy use in the data centre, the PUE. One of the main outcomes of Chapter 6 was that the PUE metric places a heavy emphasis on the supporting infrastructure in the data centre (such as the cooling system) rather than the servers. Being hidden within the '1' means that the energy consumption of the servers is not always considered when assessing overall efficiency in the data centre. This point was also raised in Chapter 6 as part of the discussions with industry professionals, who also mentioned that energy use of IT equipment needs to be considered more in the future. Whilst efficiency within the supporting infrastructure has greatly increased over recent years, there are still meaningful savings that can be made within the server. The current chapter focuses on this point by performing an energy analysis on one set of internal server components, the fans. Whilst these fans are technically part of the air distribution system rather than being a computing component, they are the only part of the air distribution system to be controlled internally by the server and hence included in the '1' for the PUE.

As discussed in Chapter 2, within each server there are a series of small fans whose purpose it is to distribute air across the components, hence aiding in the convection of heat from these critical areas. The drawback with these fans however, is that they consume a significant amount of the total server power usage. As will be discussed in section 7.8.1, perversely, the larger this consumption the lower the PUE as energy associated with air movement is included within the server rather than the facility (or in terms of the PUE, the denominator rather than the numerator). Nevertheless, any energy savings that can be made through the fans will be translated into overall efficiency gains in the data centre. The current chapter presents the findings from experimental investigations into the energy use of these internal server fans, and one particular way to reduce it. Through the study and characterisation of a server and its fans using an air flow test bench, the work in this chapter has characterised the energy use of these fans and identified whether any savings can be made. The contributions from the investigations in this chapter will also allow for a more detailed comparison to be made between air and liquid based cooling. As was discussed in Chapter 3, liquid cooled servers offer the greatest savings in energy compared to traditional cooling methods (Kennedy, 2009). The study in this chapter helps to demonstrate the energy savings that can be made when internal server fans are not required.

7.2 Experimental Methodology

There are three main stages to the research in this chapter; fan curve production and initial testing, operating the server and finally, energy analysis. An illustration of the methodology used can be seen in Figure 7.1.

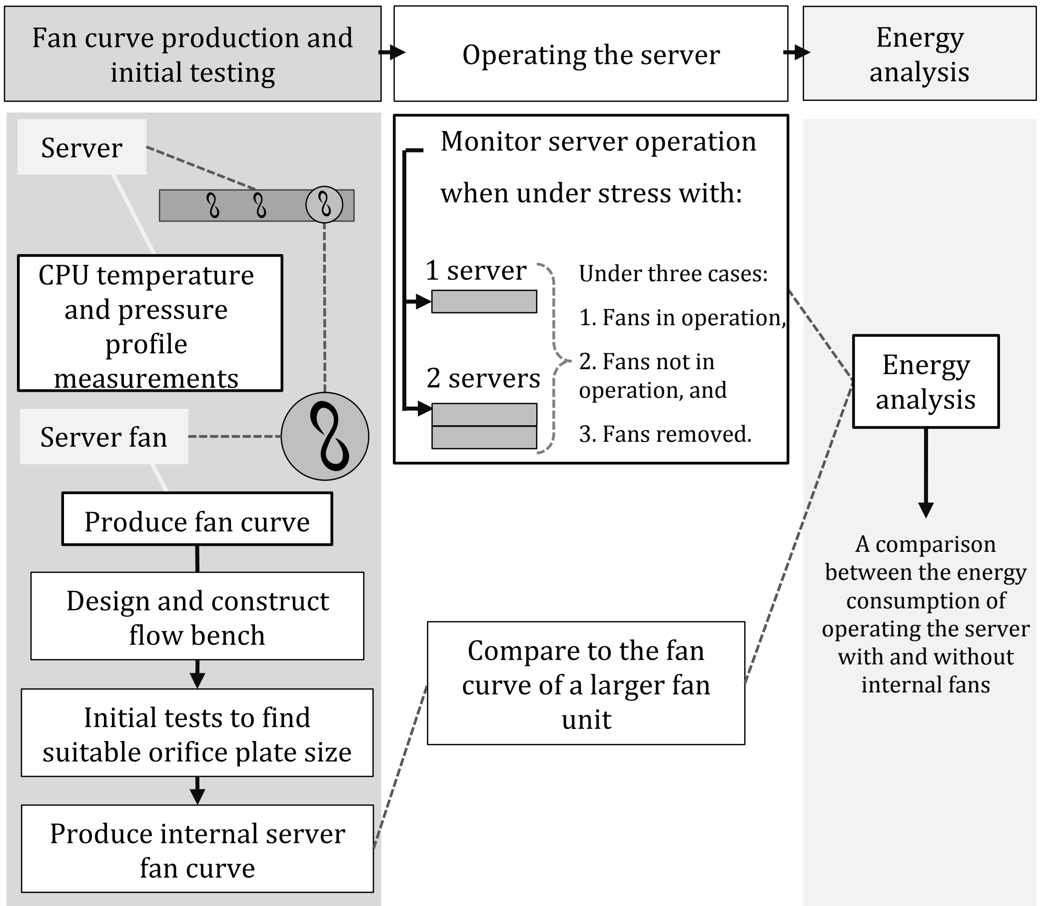


Figure 7.1 Visual representation of the methodology used for the experimental work within this chapter, showing the steps taken between the main three stages of the research

Initially, data regarding the operation of the server in terms of temperature and air flow was gathered in order to gain a better understanding of the performance of the server. Once this initial data was recorded, attention was paid to the internal server fans. In order to carry out the third stage (energy analysis), a fan curve was needed for an individual server fan. This process was involved and required the design and construction of an air flow bench (the theory behind which is detailed in Chapter 2).

During the construction of the bench and through initial testing, several iterations of orifice plate size were made before a suitable diameter was selected.

Once this integral part of the study was completed, a server was operated under three different cases whilst being given a virtual workload. First of all the server was run under normal conditions (i.e. internal fans in situ and turned on), it was then run with the fans in situ but disconnected, and then finally with the server fans removed all together. For these last two cases, air flow through the server was provided by a larger external fan. This process was then expanded to include the testing of two servers operating in unison. Finally, data collected from these operating cases were analysed to compare the energy consumption of each one.

7.3 System Description

The server used in the experimental analysis was a 1U Sun Fire V20z (see Figure 7.2) measuring 430 x 724 x 43 mm and weighing a total of 15.88 kg. The server has a power rating of 465 W and is capable of operating at 10 - 35°C and 10-90% RH at 27°C. In terms of environmental control, this particular server contains six 40 x 40 mm axial flow DC fans, rated at 0.5 A and 12 V (SanAce 40). They are capable of a speed of 13 x 10^3 RPM and can deliver an air flow of up to 0.68 m^3/min each. An additional two fans are mounted onto the power supply unit. Further details about the Sun Fire server can be found in Table 7.1.

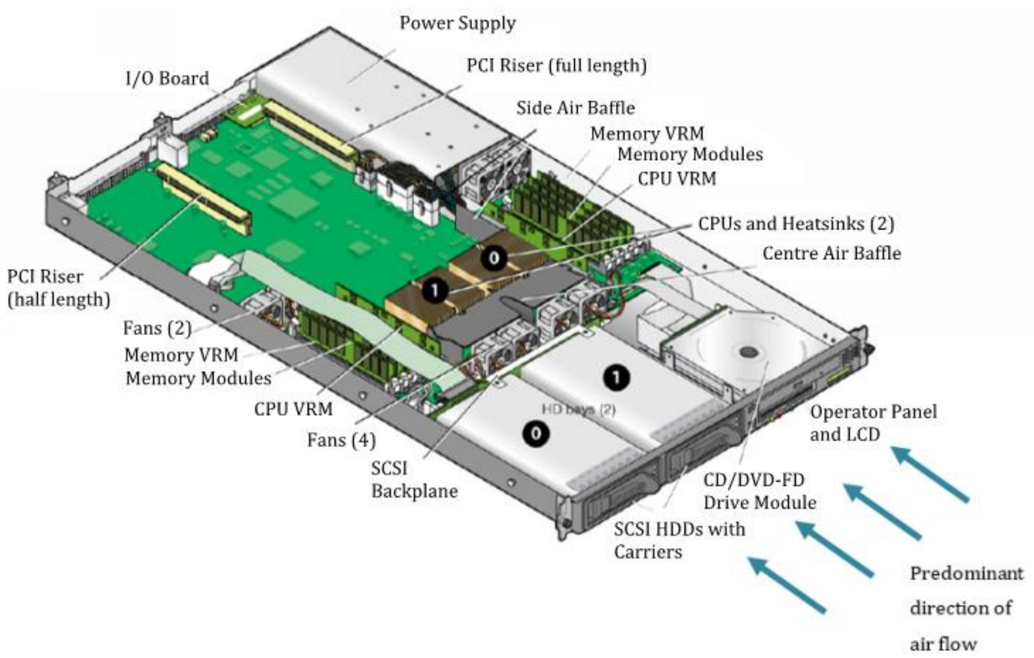


Figure 7.2 Isometric diagram of the Sun Fire V20z server (Sun Microsystems, 2008)

Table 7.1 Architecture of the Sun Fire V20z server, information from the manufacturers datasheet (Sun Microsystems)

Architecture	
Processor	One or two AMD Opteron™ 200 Series processors
Memory	1 GB to 16 GB DDR1/333 MHz ECC Registered DIMMs, 8 DIMM slots (4 per CPU)
Storage	One or two 36 GN or 73 GB Ultra 320 SCSI disk drives
Environmental	
Operating temperature	+10°C to +35°C
Operating altitude	3000 meters with maximum temperature of 30°C
Non-operating temperature	-40°C to +65°C
Power	
Power	100-240 V, 50/60 Hz, 465 W PFC supply

In order to monitor various readings from the inbuilt sensors in the servers, a laptop was networked to the server via the IP (Internet Protocol) address of the unit. Software already installed by the manufacturer on the server allowed monitoring to take place of parameters such as fan power use, CPU power use and various component temperatures. Whilst this software did not allow for logging of data, instantaneous readings could be made by entering certain commands depending on which parameter was required.

In order to test the server under operating conditions, it was necessary to impose a load on the computing components. Therefore, Stress Linux software was utilised in order to create a virtual load which would simulate the server under use. This software was written to a CD and loaded onto the server during power up. The Stress Linux software allowed the server to be stressed to different levels in order to vary the load on the CPU and hence total server power use. In order to use the Stress Linux software a visual display unit, with a keyboard, was connected to the server. This enabled commands to be sent to the server. Further detail about the Stress Linux and the commands used can be found in 7.4.3.

7.4 Experimental Setup

The following section describes the physical set up for this study including the air flow bench and how the server was operated.

7.4.1 Air Flow Test Bench

These investigations make use of an air flow test bench to measure air flow through the server fans. The theory behind this piece of equipment is explained in section 2.7.2. When constructing the flow bench used in this work, the ASME standards were consulted alongside information given in papers by Recktenwald. There are several standards available for the performance testing of fans and system curves. ASME 210-99 provides detailed requirements for an air flow bench and recommends the use of air flow nozzles for measurements. BS EN ISO 5167-1:2003 provide guidelines for the measurement of airflow within a circular cross-section, such as the air flow bench in this study. The air flow bench conforms to the ASME 210-99 standards, although it is fitted with a BS EN ISO 5167 orifice plate.

The flow bench (illustrated in Figure 7.3) used in this work consisted of a 290 mm inner diameter clear acrylic tube, with a wall thickness of 5 mm. Halfway along the length of the tube, an orifice plate was placed, which had the capability to be replaced by a larger or smaller diameter orifice. Either side of this central piece, flow straighteners in the form of wire mesh were placed in order to reduce turbulence within the tube.

On one end of the flow bench, an acrylic disc was fixed which had a hole in the centre of it that the device under test (DUT) could be attached to (this was either the server fan or the server itself, as in Figure 7.3). The other end of the flow bench was able to be opened or closed to a certain degree according to the requirements of the experiment. The degree of openness was achieved by utilising a plate within which a series of perforations had been created using a laser cutter. These are explained further later in this chapter.

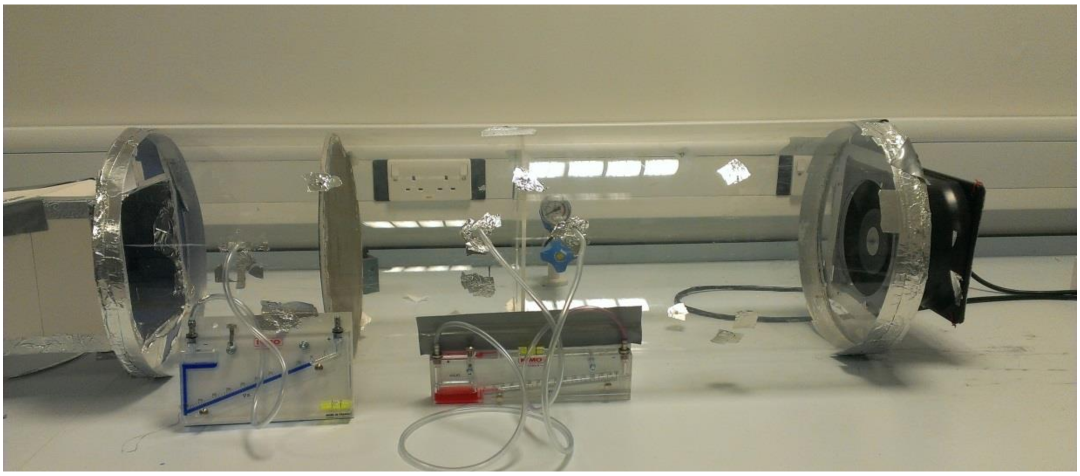


Figure 7.3 Air flow bench, pictured with manometers for measuring differential and static air pressure (note, the right hand air flow settling screen was not in place for this photo)

In terms of instrumentation, pressure was measured at two stages along the section of the tube by using manometers (which can be seen in Figure 7.3). A static pressure reading was taken in the front section of the tube, just after the DUT. Differential pressure was then taken across the orifice plate. A temperature probe was used to measure the temperature of the air before it passes through the orifice plate. K-type temperature probes were inserted into the tube and connected to a data logger. Both ambient temperature and the temperature within the flow bench were measured and recorded. The data logger was connected to the laptop to allow the temperature to be logged and recorded continuously, although the readings were taken manually during the fan testing.

Pressure was measured using liquid tube manometers, connected to the air flow bench using flexible plastic tubes. These were placed to ensure that the air flow was perpendicular to the tubes to ensure total, rather than dynamic pressure was measured. Tests were carried out in an air conditioned room, ensuring that the temperature rise of air passing through the server did not affect the inlet condition of the server inlet air.

7.4.2 Server Fan Setup

In order to find the fan curve for an internal server fan, one was removed from the server unit and wired into a motor controller (by connecting the live (red) and ground (black) wires to the controller). A Pololu Simple Motor Controller (Figure 7.4) was used which comes with the ability to connect to a laptop and be controlled via a piece of software. This software allowed the speed of the motor to be varied by altering the voltage supplied to the fan motor. The motor was increased in speed from 25%, to 50% then to

75% and finally 100% for the tests, this was in order to find fan curves at different fan speeds.

Figure 7.4 Pololu Simple Motor Controller 18v15, with a maximum voltage of 30 V and maximum continuous current of 15 A (Pololu Corporation, 2001)

7.4.3 Stress Linux

Stress Linux is a piece of software that creates a virtual load on the server by stressing it in various ways. Designed for testing a server by providing it with a high load and also used for monitoring certain parameters, the software is capable of stressing certain components of the server (such as the CPUs, memory and hard disk) for a set amount of time. For the investigations in this study, one type of stress test was used which stressed the CPUs within the server[31]. The particular stress test used was aimed at putting high stress on the CPUs within the server, doing this for a set period of time. There are many other tests that can be run including those that use the memory or storage components, however the focus here was on computing power or work done in the CPU as this creates the largest heat load within the server. Whilst these stress tests are aimed at simulating the loads that servers may operate under in real life, they do not accurately reflect the loads that they are subjected to in a data centre. Loads will vary depending on the purpose and use of the server; these load cases provide an approximation of real life operation.

When running the server during these tests, it took a few minutes for the Stress Linux software to load once the server is turned on. This leads to the long pause between the server main power being turned on and the test beginning.

[31] The command used for this test was as follows; stress --cpu 4 --timeout 10m

7.5 Server Characterisation

Prior to producing a fan curve for the internal server fans, the server itself was investigated in terms of the pressure drops and also the increase in temperature which can be noted when it is under operation.

The server was put under stress via Stress Linux, in order to find out when the CPUs would reach a steady state temperature. This load case (described in section 7.4.3) was applied for a duration of 10 minutes. It was found that both CPU 1 and CPU 2 reached a steady temperature after around 3 minutes into the test. Figure 7.5 illustrates this and shows that once the maximum temperature is reached, it remains steady for the rest of the test. This test was run under the default settings for the server.

This information informed the length of tests carried out later in this chapter; all tests were carried out for a minimum of 10 minutes to ensure that steady state temperatures had been reached.

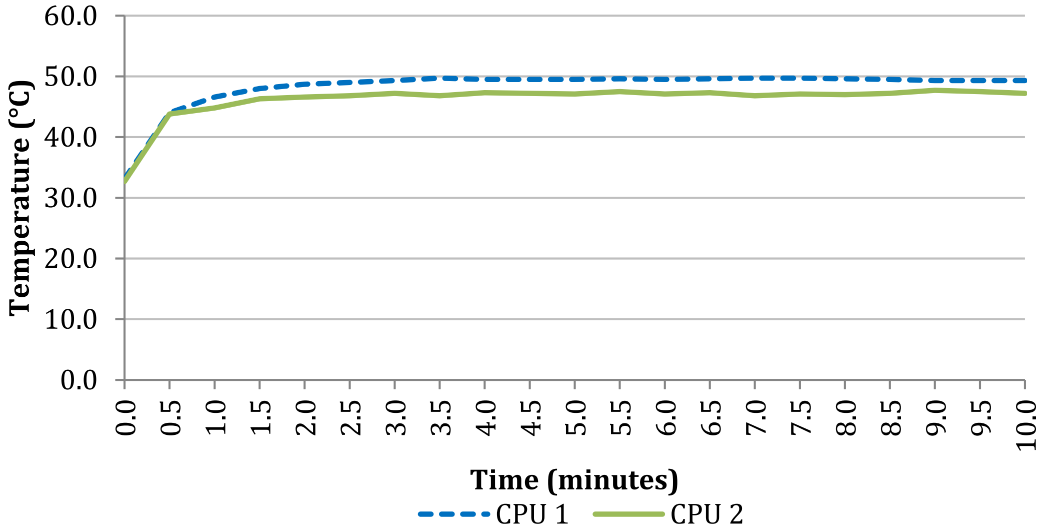


Figure 7.5 CPU temperature over time with server under stress

In order to produce a pressure loss profile for the server, several holes were drilled into the upper casing of the unit, the positioning of which are illustrated in Figure 7.6. These were located at strategic points within the server to enable key pressure drops to be accounted for. Each hole allowed for flexible plastic tubing to be inserted which was connected to an inclined liquid tube manometer. The outlet of the server was encapsulated in a card duct so that the air was channelled away from the server in a set direction. The differential pressure readings were taken between a certain point in the server and a set point at the outlet, within the card duct. Creating a pressure profile for the server is important in understanding the role of the fans within the unit, and where the largest pressure drops are located.

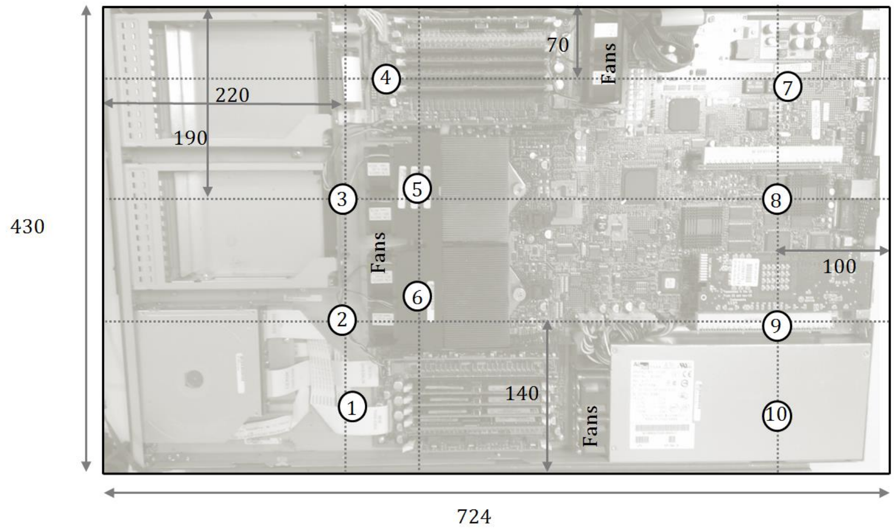


Figure 7.6 Layout of probe locations on the server casing relative to the internal fans, not to scale.

During this test, the servers' CPUs were stressed to simulate how it would operate in real life. The test was also carried out with two average fan speeds to simulate the pressure drops under different conditions and also without the fans in operation. However, within the server it is not possible to physically alter the fan speeds as they are programmed to react to the temperature of the CPUs (further demonstrating the lack of control a user may have over the energy consumption of the fans). Fan speeds were altered by changing a setting within the servers programming, to force the server into believing that the server was overheating. By setting the critical high[32] to 20,000 RPM, an average fan speed of 10,463 RPM was reached. An average of 14,033 RPM was achieved by setting the fan critical high to 9,000 RPM.

Figure 5.7 shows the layout of the server and the points at which the pressure was measured, along with the values that were recorded. The highest pressure drops can be noted across the heat sink which is situated over the CPU (for example a 54 Pa drop was noted from point 3 to 5 for a low fan speed), the server fans overcome this pressure loss by increasing the static pressure within the server. However, when the server fans are removed, this pressure drop dramatically reduces, even though the heat sinks are still in place. This would suggest that air velocity across the server is more uniform in these cases compared to when the internal fans are in operation.

[32] Forcing the server to believe that a fan speed of over 9,000 RPM meant that the server was overheating, hence driving the fans up to maximum speed.

Figure 7.7 Pressure difference (Pa) between a certain point and the server exit for four cases; (a) fan at 10,463 RPM, (b) fan at 14,033 RPM, (c) fans unplugged, and (d) fans removed. Cases (c) and (d) were operated by drawing air through the server using an external fan.

7.6 Fan Air Flow Rate Calculation

The theory behind the method for calculating air flow through the fan is explained in Chapter 2 (2.8.2), this theory leads to equation 7.1 (explained in Chapter 2) which describes the air flow rate through a flow bench based upon the pressure drop across an orifice.

$$\dot{V} = C_d \varepsilon A_o \sqrt{\frac{2\Delta P}{\rho(1-\beta^4)}} \tag{7.1}$$

Further explained here is how this relationship is used to calculate the flow rate through the fan for a given pressure drop, allowing a fan curve to be produced. The calculation method used here is set out by Recktenwald (Recktenwald, 2006), who explains the iterative process that must be used.

First of all, a value of ε must be found using equation 7.2 and values of α, β and γ (c_p and c_v are the specific heat capacities of air at constant pressure and volume respectively);

$$\varepsilon = \left[\frac{\gamma}{\gamma-1}\alpha^{2/\gamma}\frac{1-\alpha^{(\gamma-1)/\gamma}}{1-\alpha}\right]^{1/2}\left[\frac{1-\beta^4}{1-\beta^4\alpha^{2/\gamma}}\right]^{1/2} \tag{7.2}$$

$$\alpha = \frac{p - \Delta p}{p}, \beta = \frac{d_t}{D}, \gamma = \frac{c_p}{c_v}$$

This is then used to find a value for X, an intermediary figure used as part of the iteration process.

$$X = A_o \varepsilon \sqrt{\frac{2\Delta p}{\rho(1 - \beta^4)}} \quad (7.3)$$

An initial estimation of the coefficient of discharge, C_d, can then be used to calculate the flow rate. Throughout the analysis of the data in this chapter, an initial value of 0.98 was used.

$$\dot{V} = C_d X \quad (7.4)$$

$$Re = \frac{4\dot{V}}{\pi d_t \upsilon} \quad (7.5)$$

The initial value of the flow rate can then be used to calculate the Reynolds number for the flow, which in turn can be used to find a new and more accurate value for the coefficient of discharge using equation 7.6;

$$C_d = 0.9986 - \frac{7.006}{\sqrt{Re}} + \frac{134.6}{Re} \quad (7.6)$$

This new value for the coefficient of discharge is then used in equation 7.1 and the iteration is continued until the value of C_d converges to a suitable number of significant figures, in this case four.

7.7 Results and Discussion

Following in this section are results from the tests that have been conducted, the results are split into five main headings;

- the server system curve and loss coefficient
- fan curves;
- server internal fan efficiency and power use
- external fan efficiency and power use; and
- operating the server without using internal fans.

7.7.1 Server System Curve and Loss Coefficient

In order to further characterise the air flow through the server unit, a system curve was produced. The server became the DUT and was attached to the air flow bench, with a

180 x 180 mm fan used as a blower on the opposite end (rated at 70 W). During the system curve tests, mains power to the server was not turned on and so there was no load on the IT components.

A 90 mm orifice plate was used within the air flow bench and the speed of the blower was increased in order to increase the air flow rate through the server and bench. The fan speed was varied by increasing the power supplied to the fan in increments of 5 W from 0 W to 35 W. At each increment of power supply, the pressure drop across the orifice plate, the static pressure at the outlet of the server, and ambient and flow bench temperatures were recorded. The calculation method in section 7.6 was then used to calculate the air flow rate through the server. Following this, a system curve was then produced by plotting this against the pressure drop across the orifice plate, as illustrated in Figure 7.8.

This test was repeated three times to ensure repeatability of the results. Figure 7.8 shows that there was minimal variation in the results from one test to the next, demonstrating the reliability of the flow bench methodology. The set of tests were then run twice, once with the internal fans present in the server, and once when they had been removed. This was in order to demonstrate if there would be any change in the characteristics of the server once the server fans had been removed.

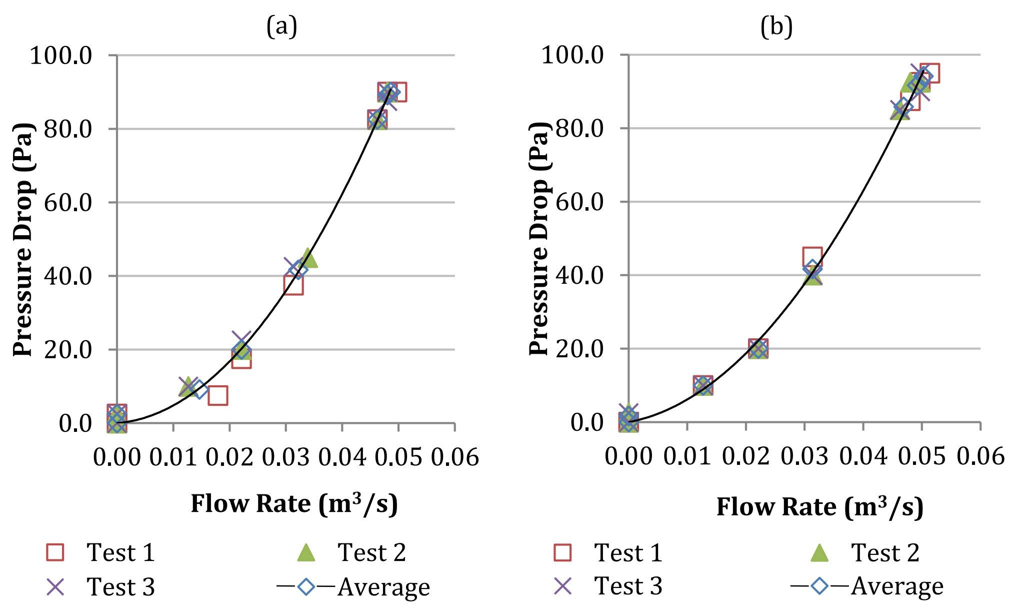


Figure 7.8 System curves for SunFire v20z server, repeated tests with internal fans (a) present and (b) not present

Figure 7.9 combines the server system curves for both fans present and fans not present. It is clear that there is not much difference between the two configurations, suggesting that the fans do not have much impact on the pressure losses through the server (at least when they are not in operation). This is due to the relative large open area of a fan and streamlined flow around the flow blades.

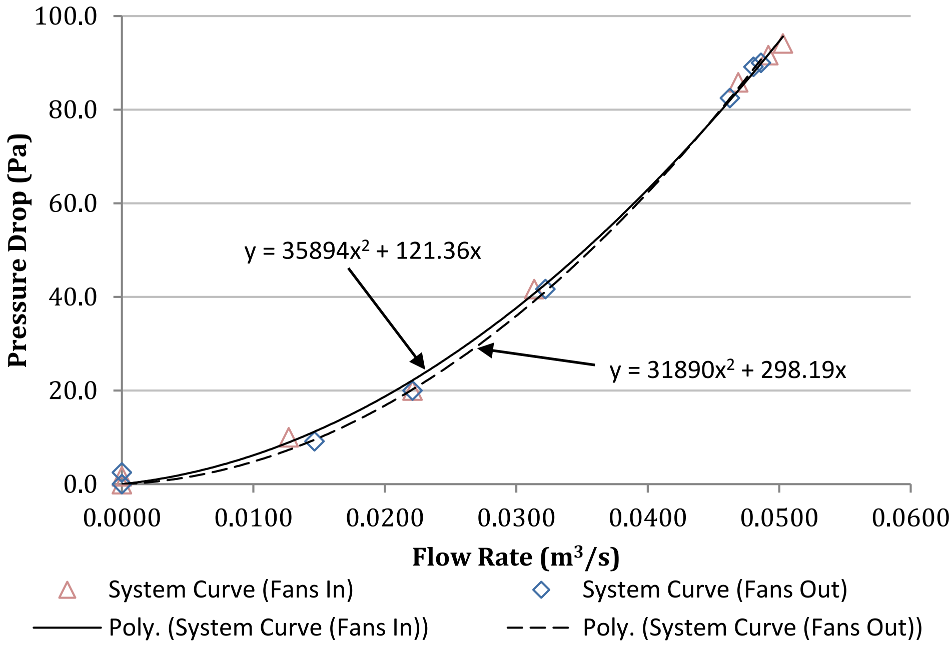


Figure 7.9 System curves with and without internal fans present for the SunFire V20z Server

The relationship between the pressure loss across the server and the air flow rate through it can therefore be written as follows (for fans present and not present respectively);

$$\Delta P = 35894\dot{V}^2 + 121.36\dot{V} \tag{7.7}$$

$$\Delta P = 31890\dot{V}^2 + 298.19\dot{V} \tag{7.8}$$

As described by Recktenwald, it is possible to then calculate a single value for the loss coefficient for the server. Given that for pipe systems, the head loss (h_L) is a function of the flow rate, area (A) and a loss coefficient (K), and that the flow rate is related to the change in pressure and some loss coefficient;

$$h_L = \frac{K\dot{V}^2}{2gA^2} \tag{7.9}$$

$$\Delta P = C\dot{V}^2 \tag{7.10}$$

It is therefore possible to calculate the loss coefficient;

$$C = \frac{\rho K}{2A^2} \quad (7.11)$$

Given that;

$$h_L = \frac{\Delta P}{\rho g} \quad (7.12)$$

A least squares principle can be used to obtain a single value of C, a detailed derivation of this can be found in work by Recktenwald (Recktenwald, 2006).

$$C = \frac{\sum_{i=1}^{m} \dot{V}_i^2 \, \Delta P_i}{\sum_{i=1}^{m} \dot{V}_i^4} \quad (7.13)$$

Equation 7.13 allows a single value loss coefficient to be calculated, which then describes the relationship between the pressure loss and air flow rate for the server. In this case, the following relationship was found (the calculations for this can be found in Table 7.2).

$$\Delta P = 15717.62\dot{V}^2 \quad (7.14)$$

Table 7.2 System loss coefficient data

Internal Fans Present				*Internal Fans Not Present*			
ΔP Pa	**$\dot{V}$** m^3/s	$\dot{V}^2\Delta P$	$\dot{V}^4 x10^{-6}$	**ΔP** Pa	**$\dot{V}$** m^3/s	$\dot{V}^2\Delta P$	$\dot{V}^4 x10^{-6}$
0	0.0000	0.0000	0.00	0	0	0	0
1.67	0.0000	0.0000	0.00	2.5	0	0	0
10	0.0127	0.0016	0.03	9.17	0.0146	0.0020	0.05
20	0.0221	0.0098	0.24	20	0.0221	0.0098	0.24
41.67	0.0314	0.0410	0.97	41.67	0.0322	0.0433	1.08
85.83	0.0469	0.1887	4.83	82.5	0.0462	0.1765	4.58
91.67	0.0492	0.2217	5.85	89.17	0.0480	0.2058	5.33
94.17	0.0503	0.2380	6.39	90	0.0486	0.2127	5.58
	Σ	0.2777	$18.30x10^{-6}$		Σ	0.6499	$16.85x10^{-6}$
		C	38,289.4			**C**	38,574.2

Given the loss coefficient values in Table 7.2, it is possible to plot the relationship between the air flow rate and pressure drop, as can be seen in Figure 7.10. Equation 7.14 has been plotted alongside the actual measured values, this curve will be used later in this chapter when considering the internal fan curves.

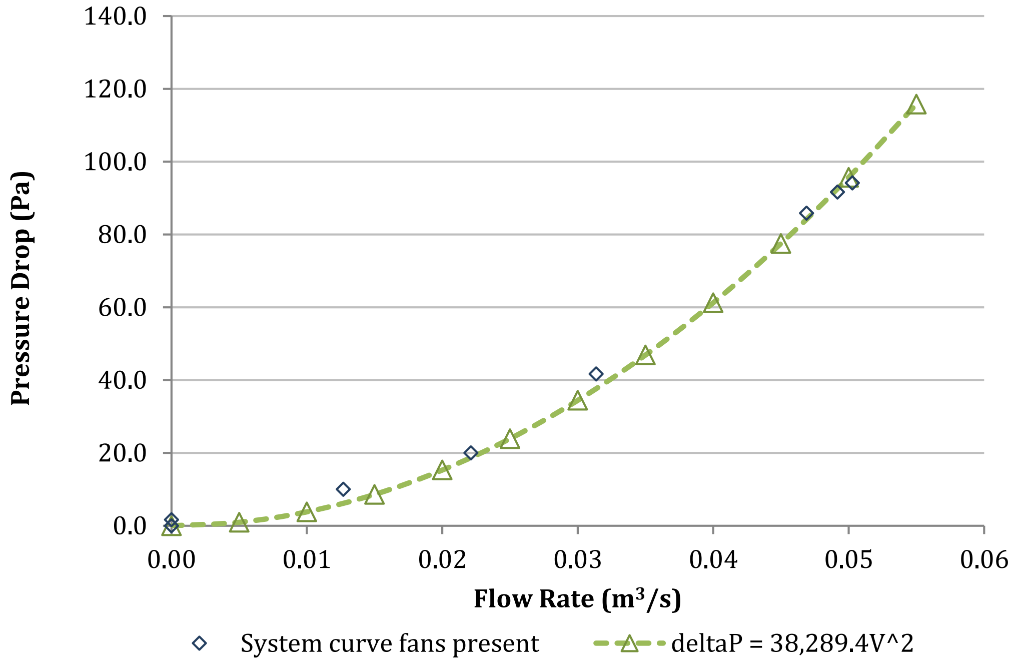


Figure 7.10 System curve from measured data (fans present in server) and calculated values from system loss coefficient

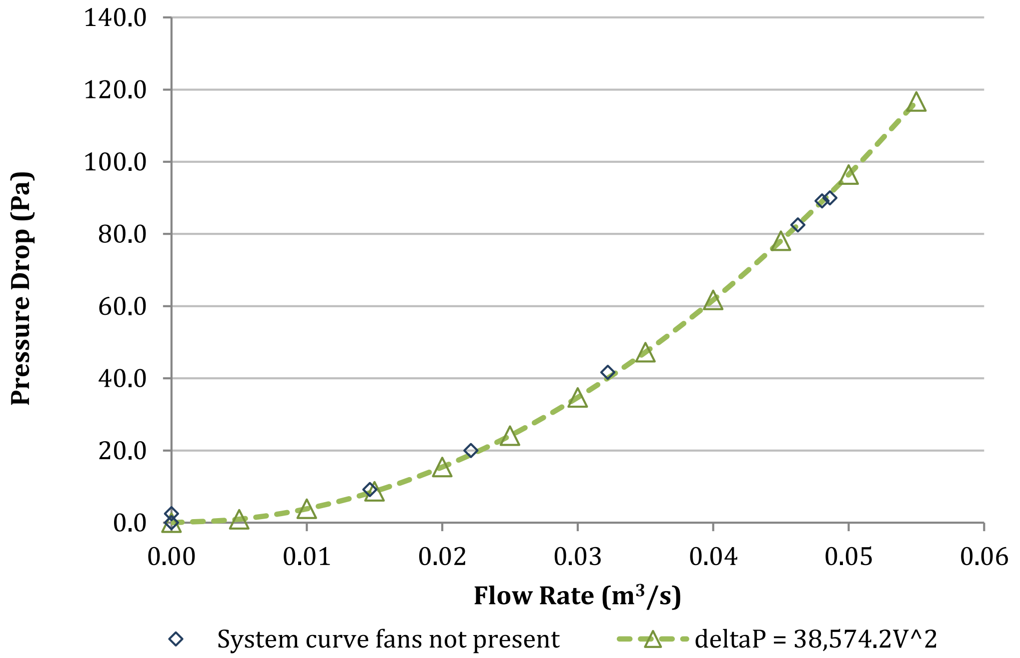


Figure 7.11 System curve from measured data (fans not present in server) and calculated values from system loss coefficient

7.7.2 Fan Curves

In order to find out the relationship between the air flow rate and static pressure produced by an internal server fan, one of these units from the server was attached to the flow bench as the DUT. The method used in this case to vary the air flow rate was a fixed plate on the opposite end of the flow bench with a fixed number of perforations of a certain diameter. This allowed portions of the perforations to be sealed and hence allow a range of flow rates from zero flow (maximum static pressure) to free flow conditions (zero static pressure). For each test, the fan was powered to 25%, 50%, 75% and then 100% in order to find a full range of fan curve relationships (this was achieved by using a motor controller as described in 7.4.2)

An initial test was done with an end plate that had 144 perforations of 1 mm diameter in it. These were uncovered in stages to allow for increased air flow within the flow bench. This initial stage was conducted with a 14.32 mm orifice plate. Table 7.3 and Table 7.4 show the results from the initial testing, it is clear to see that with the maximum number of holes open, the flow bench was still 99.83% closed. This range of openness did not allow the fan to achieve a full range of air flow rate.

Table 7.3 Results from 1 mm perforated gate and fan power at 25%

ΔP	P_{man}	P_f	T_1	T_{amb}	α	ε	C_d	$\dot{V}$	Re	Fan W_{out}	η_f
Pa	Pa	Pa	°C	°C				10^{-4} m^3/s		W	10^{-3}
0	137.2	36.2	20.2	20	1	0	0	0	0	0	0
1	135.5	34.5	20.5	20.4	0.993	0.996	0.9103	1.88	1067.24	0.006	3.2
2	133.9	32.9	20.6	19.9	0.985	0.992	0.9074	2.64	1498.34	0.009	4.3
3.5	133.1	32.1	20.5	20.1	0.974	0.986	0.9091	3.48	1973.51	0.011	5.6
5.5	133.6	32.6	20.5	19.9	0.959	0.978	0.9121	4.34	2461.66	0.014	7.1
7	133.7	32.7	20.7	20.7	0.948	0.972	0.9140	4.88	2765.41	0.016	8.0
9	132.1	31.1	20.7	20.1	0.932	0.963	0.9163	5.49	3115.31	0.017	8.5
10	132.4	31.4	20.7	20.5	0.924	0.959	0.9173	5.77	3273.32	0.018	9.1
12	132.5	31.5	20.8	20.2	0.909	0.950	0.9190	6.28	3560.77	0.020	9.9

*W_{in} at 25% power is 2W
Note: Results over a range of 100-99.83% closed

Table 7.4 Results from 1 mm perforated gate and fan power at 100%

ΔP	P_{man}	P_f	T_1	T_{amb}	α	ε	C_d	$\dot{V}$	Re	Fan W_{out}	η_f
Pa	Pa	Pa	°C	°C				10^{-4} m^3/s		W	10^{-3}
0	220	220.0	20.4	19.7	1	0	0	0	0	0	0
22.5	207	207.0	20.8	20.1	0.891	0.940	0.9258	8.57	4858.77	0.177	29.5
27	206	206.0	20.9	20.1	0.869	0.927	0.9276	9.28	5260.08	0.191	31.8
40	200	200.0	21.0	20.4	0.800	0.887	0.9311	10.84	6144.64	0.217	36.1
48	198.1	198.1	20.9	20.2	0.758	0.861	0.9326	11.54	6545.17	0.229	38.1
60	196.2	196.2	21.1	20.6	0.694	0.820	0.9340	12.32	6985.44	0.242	40.3
73	192.3	192.3	20.9	20.4	0.620	0.771	0.9349	12.78	7248.72	0.246	41.0
82.5	187.6	187.6	21.1	20.5	0.560	0.729	0.9350	12.84	7282.07	0.240	40.2
93.5	186.9	186.9	21.2	20.6	0.450	0.683	0.9349	12.82	7271.07	0.240	40.0

*W_{in} at 100% power is 6W
Note: Results over a range of 100-99.83% closed

As described in the methodology within section 7.2, several initial tests were required in order to find the correct orifice size and degree of openness within the flow bench. After running several iterations of these parameters, it was found that an orifice plate with a diameter of 55 mm was most suited, with 128 15 mm perforations within the end plate (allowing the flow bench to be up to 35 % open). Figure 7.12 shows the results of this increase in orifice size and open area. The results clearly show the inability of the fan to reach full flow rate under the conditions of the initial test. The stall region for the fan under each power rating is also clearly visible.

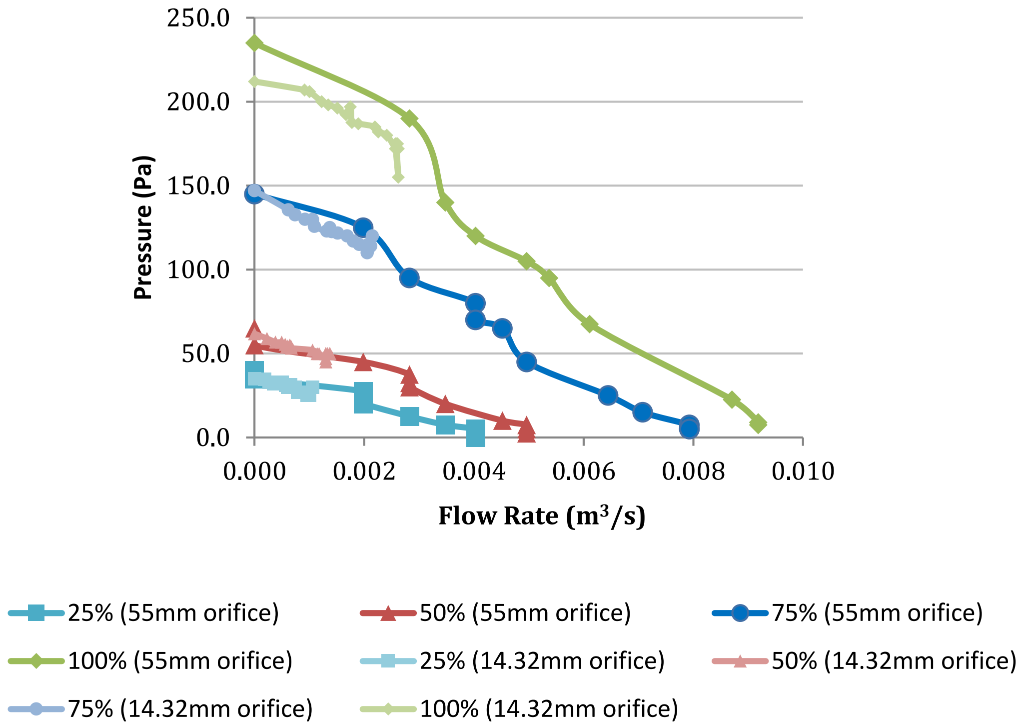


Figure 7.12 Fan curves comparison between orifice plate sizes, both using an end plate with 15 mm diameter perforations

Figure 7.13 shows a comparison between the fan curve as measured by the flow bench and the manufacturers stated fan curve. The manufacturers curve illustrates relationships for several different fans, the SH curve represents the fan that is used in the SunMicro V20z servers used in this analysis. It can be seen from the figure that there is a difference in the manufacturers curve and the measured one of 0.1 m^3/min (19% lower than the manufacturers curve) between the maximum air flow rate of each one. This is possibly due to loses within the flow bench and potential inaccuracies in the data readings. However, it is a good comparison and the two are in general agreement with one another.

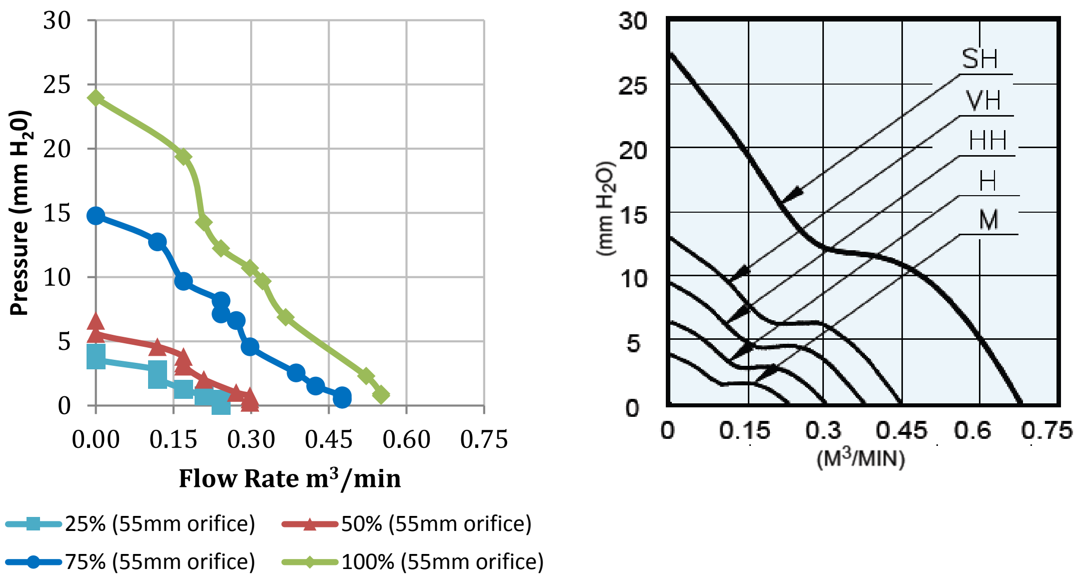


Figure 7.13 Comparison between measured and manufacturers fan curves

The results in Figure 7.13 show that the air flow bench is adequate for measuring air flow through the server fans.

7.7.3 Server Internal Fan Efficiency and Power Use

Fan efficiency is important as it varies depending on the static pressure and the air flow rate. Selecting a fan which has a high efficiency at the correct operating point is crucial to designing an efficient air distribution system. The operating point for the system is the point at which the system curve (in this case the server) and the fan curve cross. Figure 7.14 shows the fan curve for an internal server fan plotted alongside the server system curve. The server unit has 8 fans within it, and examination of the geometry suggests these are positioned in parallel. By treating them as being positioned in parallel, it was possible to keep the pressure the same but multiply the flow rate by 8, giving the second fan curve in Figure 7.14. In practice there will be losses associated with positioning of the fans (e.g. interaction of neighbouring units) meaning the capacity will be less than this, however it does provide a basis for analysis. The point at which the curve for 8 fans crosses the server system curve in Figure 7.14 is the point at which the fans will have been chosen for, this is also at a high efficiency level for these fans. Whilst 8% is quite a low efficiency, these fans only reach a maximum efficiency of less than 10% anyway.

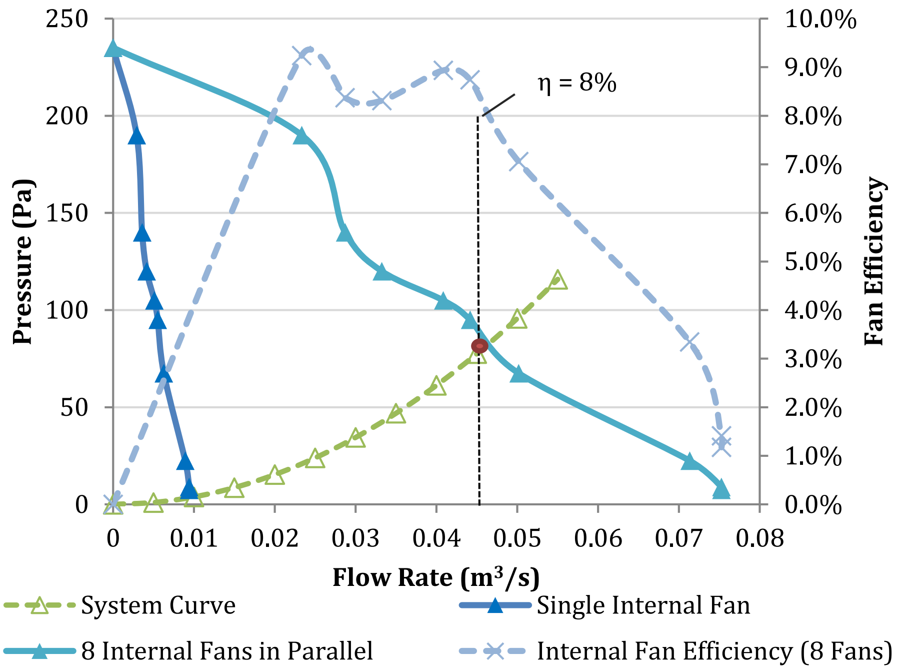


Figure 7.14 Fan curve for a single fan and six fans in parallel with the server system curve. The server system curve is measured here with the fans present and in operation.

7.7.4 External Fan Efficiency and Power Use

When considering whether it would be more efficient to remove the server fans, it is important to consult the graph of the system curve and fan curve. As it was not possible to use the air flow bench to measure the fan curve for the large external fan available, the manufacturers fan curve has been used as an estimate.

Overall, the external fan is more efficient, which is due to its size. Larger fans tend to be much more efficient than smaller ones (Holahan and Elison, 2007). When plotting the system curve against this larger fan curve, it is clear that the fan is suitable for delivering the correct pressure and flow rate, and at a higher efficiency.

In fact, it would be possible to use the large external fan to provide air flow through two servers at a much higher level of energy efficiency (18%). Figure 7.15 illustrates this clearly.

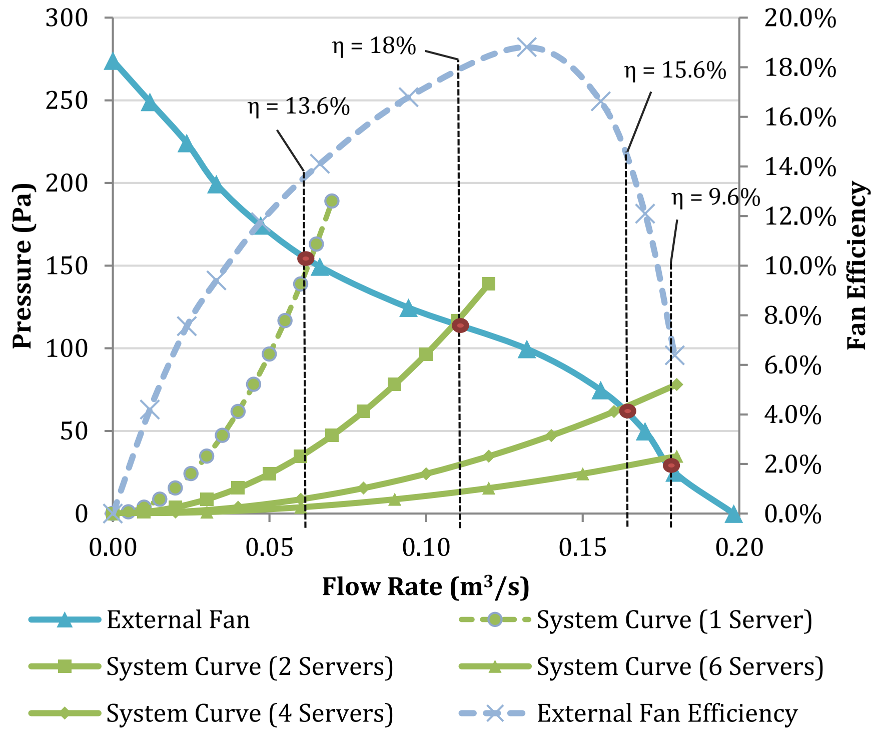


Figure 7.15 Fan curve and efficiency for the large external fan, with system curves for 1, 2, 4 and 6 servers. The system curve is plotted for when the internal server fans are not present.

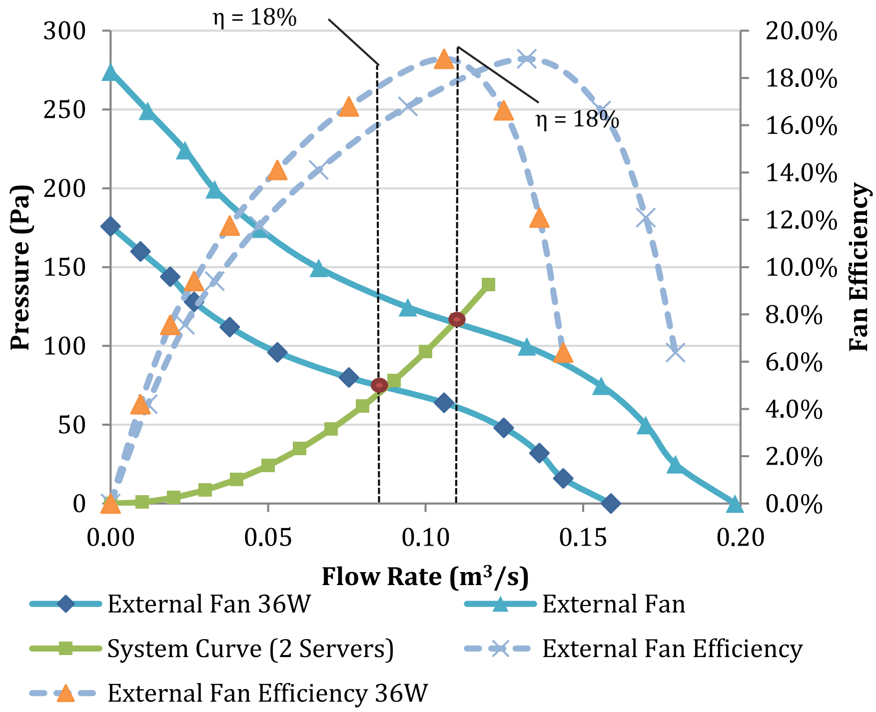


Figure 7.16 Fan curve and efficiency for the large external fan, operating at full rated and 36 W power

The next stage of the work in this chapter is to test the operation of the server when the internal fans are removed and the external fan is relied upon to provide the correct air distribution.

7.7.5 Operating the Server without Using Internal Fans

In order to test the operation of the server, the unit was connected to the air flow bench via a card duct that directed all the air flow through the server and bench. During each test, the seven internal server temperature probes were used to record temperature fluctuations[33], with the ambient temperature being recorded by a data logger with a K-type thermocouple. Air flow rate was monitored using the same method as before, by measuring the differential pressure across an orifice (in this case with a diameter of 90 mm).

The server was tested under three conditions, internal fans present and in operation, internal fans present but unplugged and internal fans not present. The external fan was in operation for the latter of these two. The external fan was fixed to the end of the air flow bench.

Each test was conducted in the same manner as follows;

- server mains power turned on;
- server power turned on (it then took a few minutes for the server to load up and the Stress Linux software to run);
- stress case 1 was run for 10 minutes;
- the server was then allowed to idle for 18 minutes;
- stress case 1 was then run again for a further 20 minutes; and
- the server was allowed to idle for 10 minutes before power and then mains power were shut off 2 minutes later.

[33] The internal temperature probes monitor the CPU die and memory temperatures (cpu.mem andcpu.die), (GigaBit on Broadcomm) gbeth, hard disk back plane (hdbp), service processor (sp)

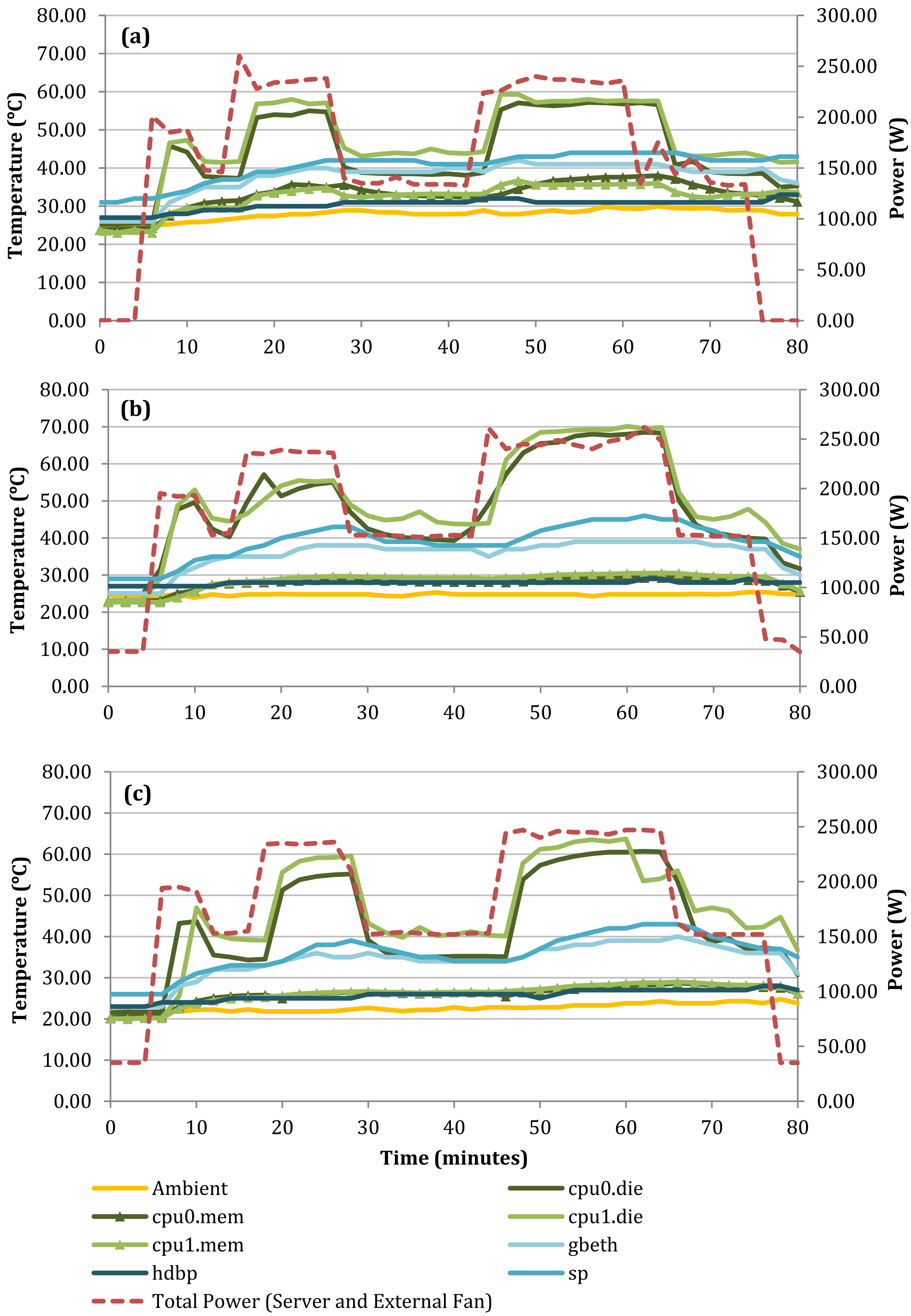


Figure 7.17 (a) Server operating with internal fans present and in operation, (b) server operating with internal fans present but disconnected, and (c) server operating with internal fans disconnected and not present

The changes in server power use and load can most easily be seen in the server power lines on the three graphs in Figure 7.17. A spike in power use can be noted when the power is turned on, with continuous high consumption throughout the stress case periods. It is also clear to see in Figure 7.17 the changes in CPU temperature as a result of the changes in load they are experiencing. There is a delay between the power consumption and the temperature increasing as it takes time for the components to heat up.

When operating using the external fan, the server operates almost the same as when under normal conditions. A slight difference is noted in that it appears from Figure 7.17 that when using the external fan, the temperature of the server components is reduced quicker than they were under normal operating conditions. This is probably because the external fan would actually be large enough to cope with distributing the air through more than one server.

The power line on each graph represents the total power use, so either server power alone (when internal fans are in operation) or server power and external fan power.

Making a closer comparison between the internal temperatures of each test case in Figure 7.18, there is a difference noted when using the external rather than internal fans. There was not a notable difference between each case during the short 10 minute stress scenario; however there was a difference when stressing the server for a second time for the longer period of 20 minutes. A higher temperature for the CPU can be seen when the internal fans were present but unplugged, suggesting that by remaining in place they are having an adverse effect on the air distribution. With the internal fans removed however, the external fan is able to maintain very similar conditions within the server.

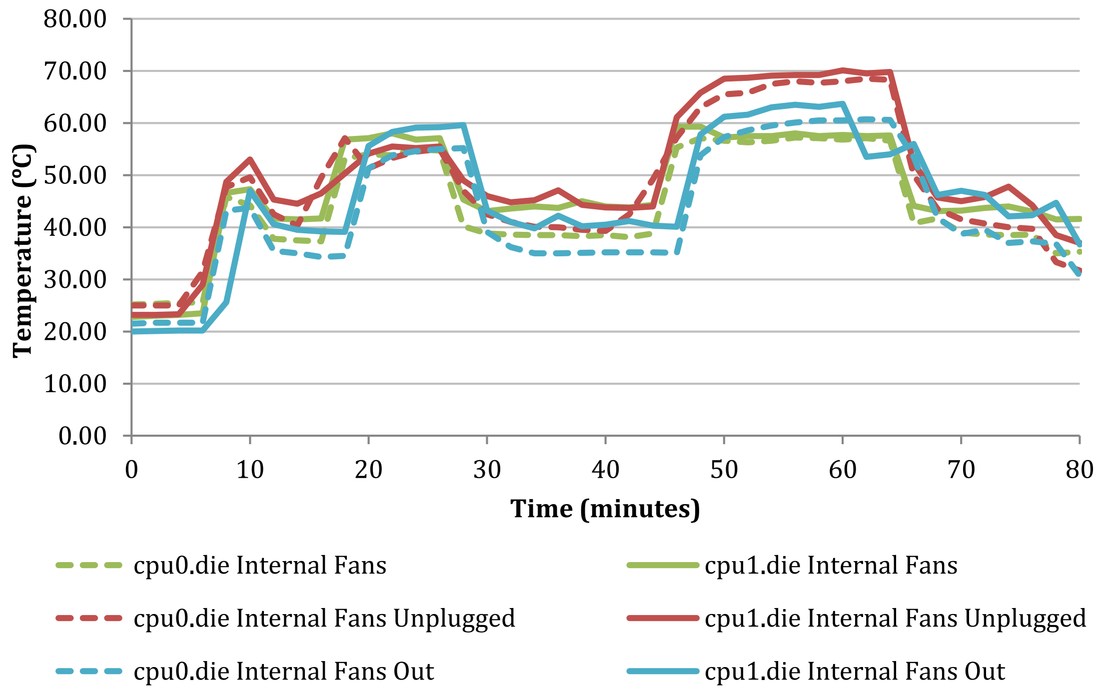


Figure 7.18 Comparison between CPU temperatures for different tests

The important part of this study is to investigate whether this produces any savings in power. Figure 7.19 shows the power use of the three scenarios. It is obvious from this graph that there is not much difference between the power used during each scenario.

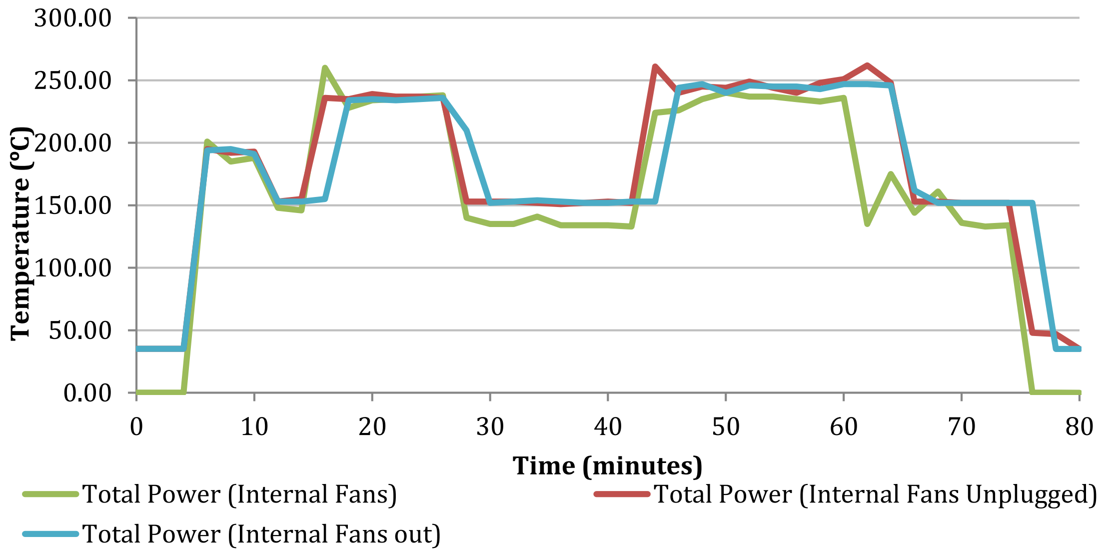


Figure 7.19 Total power use under different tests

The change in temperature across the server was recorded by taking temperature readings at the front and rear of the server unit. It is interesting to note that there appears to be a lower rise in temperature across the server when the internal fans are not used. This will be because when they are used, they produce energy in the form of heat themselves, contributing to the rise in temperature.

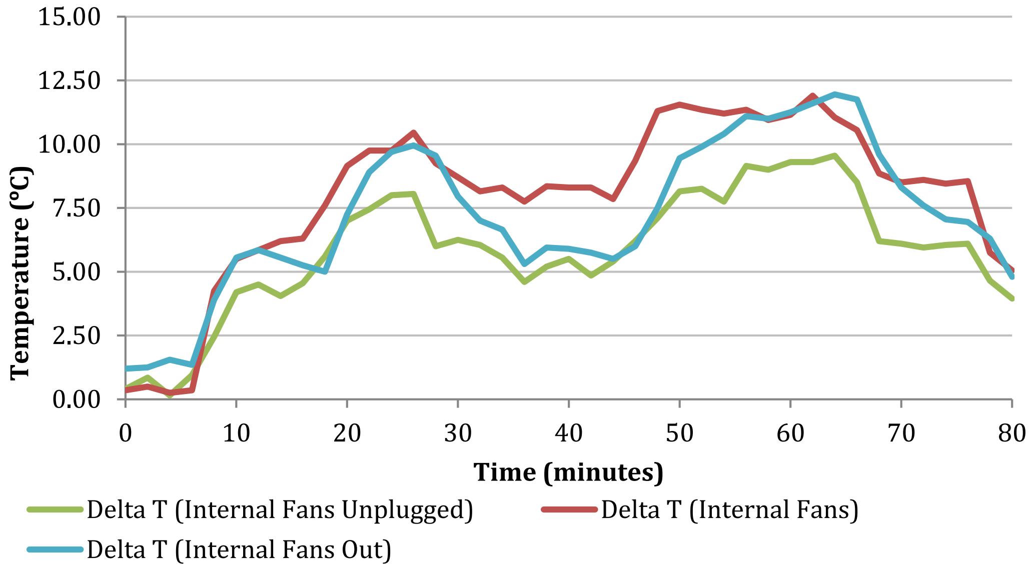


Figure 7.20 Change in temperature across the server under different tests

Once it was established that it was possible to run the server using the external fan only, tests were then conducted to find out whether it was possible to do the same but with two servers. Another identical server was placed on top of the first and the ducting around the end was enlarged so that they were both connected to the flow bench. The test were repeated as before, with the CPUs being stressed for 10 minutes first and then after for 20 minutes with idling time in between. The servers were both tested first with the fans present and operational, neither had any failures when running through the two stress tests. Figure 7.21 shows the two CPU temperature recordings for the two servers (A (bottom server) and B (top server)), compared with the temperature recordings from the one server alone. There is no obvious difference between the temperatures when comparing the single and double servers. However, problems were encountered when the internal fans were unplugged or removed from the two servers. During this stage of the testing, the CPUs within server B overheated and caused a failure leading to the server shutting down. This failure was likely due to the additional heat load imposed by heat transferring between the server units.

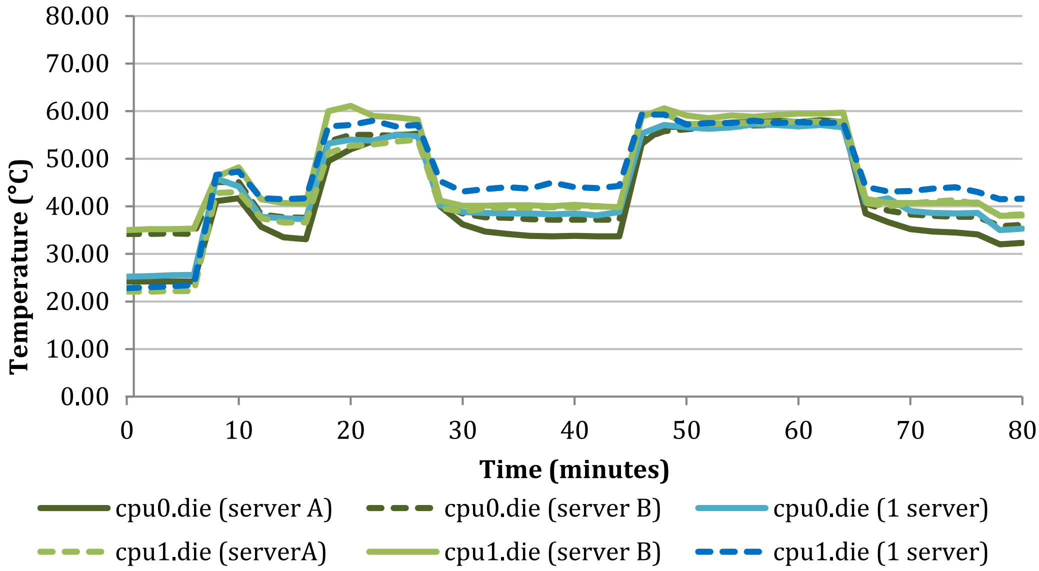


Figure 7.21 Variations in internal server temperatures across each stress test for either server A and B when in situ together, or one server in isolation

7.8 Summary of Results and Energy Analysis

As efforts made towards reducing energy use within the data centre are turning towards the server itself (a point discussed in Chapter 7), this hardware is coming under increasing scrutiny regarding energy consumption. Also, new technologies for removing the heat load from these units are emerging such as liquid rather than air distribution infrastructures. By examining the energy use of internal server fans and considering their efficiency during use, the work in this chapter has contributed to efforts made to reduce energy consumption by servers. The following is a summary of the results from this work.

The server was first characterised by experimentally producing system curves which illustrate the relationship between air flow rate and pressure drop across the server. A system curve was produced for both when the fans are present and not present within the server, with a minimal difference being observed between the two. This indicates that the fans have little impact on the air flow through the server when they are not in operation, implying that the fans could be turned off (unplugged) without much hindrance to the pressure drops in the system. The difference between the two scenarios is more noticeable however when considering the temperature rise across the server. With the server fans in place and in operation, there was a temperature rise of around 7.5°C when the server was idling. When the server fans were removed completely, this temperature rise dropped to 5°C. The efficiency of mechanical equipment such as CRAH units and heat rejection plant rely upon a higher return temperatures in order to deliver peak efficiencies. Therefore this decrease in server temperature rise may have an effect downstream on the energy consumption of the cooling system. This means that if this solution were to be employed in a data centre, then careful consideration would need to be given to the loss return air temperatures and the effect this would have on the energy use of the mechanical cooling plant.

Using an air flow bench is a standard method of measuring the air flow produced by a fan or a system. The air flow bench produced in this work provided consistent results across each test, although difficulties were noted when finding an appropriately sized orifice plate. In order to produce a full fan curve from the internal server fan, it was found a 55 mm orifice was most suited. The final fan curve produced did have a lower maximum velocity than the manufacturers curve, but this will be due to losses within the air flow bench itself. Once the fan curve had been characterised, it was plotted against the system curve to find the estimated efficiency of all the server fans acting in unison (8 fans in total). At an efficiency of 8% the fans were operating at almost peak efficiency, and whilst this is low it must be remembered that as small fans these units are not very efficient[34]. This is of course assuming that the fans are operating at 100% power. The graph was then reproduced for a large external fan (the fan curve for which was approximated from the manufacturers' data). It was found that a higher efficiency of 13.6% could be realised if there is one server per fan, or an even higher efficiency of 18% would be managed if there were two servers per fan (in addition, overall savings would be higher as the loss of efficiency from the internal fan due to it being situated within the server would not be a factor). In addition to this increase fan efficiency, in theory it would be possible to also reduce the total amount of power consumed by the server and fans.

Table 7.5 shows the results of an analysis done to find out how much energy could be saved by switching to a larger, more efficient, external fan. Data collected from testing one server in isolation was used to estimate the power use of two servers. When the internal fans were not being used, the power use of the external fan was taken into account. The data indicates that two servers under normal conditions and being stressed using Stress Linux for 10 minutes would use 477.33 W. Two servers operating without the use of the internal fans (and having had them removed) would operate at a lower power use of 434.60 W, inclusive of the 35 W being consumed by the external fan.

---

[34] This doesn't however take into account the loss of efficiency due to the fans being located within the server. As mentioned in Chapter 2, this is a difficult quantity to measure and values for this do not currently exist in the literature.

By doing this, it would be possible to reduce energy use by just under 9%. However, how much this saving would actually be under real life conditions would vary greatly depending upon the operation of the server. The above analysis is only based upon the power use of one particular server under a certain stress test.

Table 7.5 Theoretical energy savings if one external fan were used to provide air distribution over two servers, with fans either still present or removed

	Power Use (W)		
	Idle	**Stress 10min**	**Stress 20 min**
Internal fans in operation (no external fan)			
Server	145.82	238.67	233.67
Fan	N/A	N/A	N/A
Total for 1 server	145.82	238.67	233.67
Total for 2 servers	291.65	477.33	467.33
Internal fans unplugged (use of external fan)			
Server	124.39	201.67	213.36
Fan	35	35	35
Total for 1 server	159.39	236.67	248.36
Total for 2 servers	283.78	438.33	461.73
% saving	**2.70%**	**8.17%**	**1.20%**
Internal fans removed (use of external fan)			
Server	124.11	199.80	210.00
Fan	35	35	35
Total for 1 server	159.11	234.80	245.00
Total for 2 servers	283.21	434.60	455.00
% saving	**2.89%**	**8.95%**	**2.64%**

When attempting to run two servers under each test case however, problems arose that meant the above analysis for two servers could not verify this experimentally. When running the same testing procedure with two fans in unison, the top server failed due to overheating. This happened during several attempts, and could be due to the extra heat load imposed by transfer of heat between the two units. Also, the external fan was situated at the outlet of the server in this study, rather than the inlet. This may have contributed to the servers failing when under the operation of the external fan by reducing the efficiency with which the heat is convected away from the components. As discussed within the literature review, Grimes and Davies (2002) found that with the fan located at the outlet of an electronic device, cooling was restricted to the boundary layer (leading to higher temperatures). The situation in this study was different however, as due to the complex internal geometry of a server, mixing is more likely to occur as discussed in Chapter 3. Having said this, the results found in the work by Grimes and Davies may play a part here.

7.8.1 Effect on the PUE measurement

From the results presented above, this work has shown that replacing the internal server fans with a larger external unit offers the potential to save energy within the computing energy load. However, following on from discussion in Chapter 6 it is known that reducing the IT load can have an adverse effect on the results of a PUE measurement. Therefore, results from the experimental analysis are combined with the data from the case study in Chapter 6 in order to see how the PUE measurement might be affected. In Chapter 6, energy consumption values were calculated for the IT load, the cooling and air distribution loads along with the losses in the power delivery network. Here, the IT load is replaced with values obtained through the results in the current chapter, and the electrical losses are adjusted to reflect this. Table 7.6 contains the results of this analysis. Using values for each case (normal operation, fans unplugged and fans removed) and for both stress test durations, a change in the PUE value has been demonstrated.

Table 7.6 PUE measurements based upon data gathered in Chapter 6 and energy data from experimental work in Chapter 7

	IT Load			*PUE**
	W	*kWh/yr*[1]	*kWh/yr*[2]	
10 minute stress				
Normal operation	477.33	4,181.41	188,163.45	1.08
Fans unplugged	403.33	3,533.17	158,996.70	1.17
Fans removed	399.60	3,500.50	157,522.50	1.18
20 minute stress				
Normal operation	467.33	4,093.90	184,225.43	1.08
Fans unplugged	461.73	3,738.07	168,213.15	1.17
Fans removed	455.00	3,679.2	165,564.00	1.17

[1] kWh/year for two servers, based on results from Chapter 6

[2] kWh/year for 3 x 30 servers (equivalent to Open Compute triplet rack number of servers)

*PUE measurement based upon figures calculated in Chapter 6

The conclusion that the IT components are able to conceal high energy consumption within the '1' of the PUE measurement has been reinforced by these results. By removing the internal server fans and replacing them with an external one, the energy consumption is shifted to the supporting infrastructure rather than being included within the IT load. This causes an increase in the PUE to be noted. For example, the result of removing the internal fans completely caused an increase from 1.08 to 1.18 during the 10 minute stress test. This increase in PUE value is despite the fact that energy overall has been reduced, 9% in this case.

8 Conclusions and Suggestions for Further Research

8.1 Overview of Research

An in depth literature review considering energy use and efficiency within data centres has provided a summary of the technologies available and areas where energy use can be reduced. The review forming chapters 2 and 3 has demonstrated that there are numerous technological improvements, metrics and management strategies that can be employed to reduce the energy consumption of data centres. Despite this however, there are still both existing and new facilities that are not achieving their theoretical energy efficiency potential. It is for this reason that the first aim of this thesis was to find out whether there are any barriers or problems preventing further energy savings. Two aspects of research carried forward from this added more depth to the argument and provided further evidence of where industry choices regarding energy efficiency could be improved. These objectives employed interdisciplinary research skills to look into the issues surrounding further energy efficiency improvements in data centre facilities.

The following sections summarise the main conclusions from the research by asking what barriers there are to further energy savings, whether the PUE is a useful tool for driving better efficiency and lastly whether there is energy that can be saved within the server.

8.2 What barriers are there to further energy savings?

The qualitative research in Chapter 5 addressed the difficulties that participants face when trying to improve the energy use of their facility (or encouraging others to improve theirs). Agreement from customers, especially in the case of colocation providers, was one of the key barriers identified as preventing further energy savings in the data centre. Colocation providers need the agreement of their current tenants before they can make alterations, such as increasing the inlet temperature to the servers, as it may affect the terms of their SLA. As the results in Chapter 5 show, agreement from customers is one of the key barriers to improvements in colocation facilities. Demonstrating the problem of goal alignment, it shows the need for the goals of different parties to be similar if higher levels of energy savings are to be achieved. In the colocation providers' case, a mismatch between the energy saving goals of tenant and landlord leads to conflicting views and hence difficulties in reducing energy use.

The other key barriers presented in Chapter 5 were costs, communication, and logistics. Understanding the investment costs required and the potential energy or cost savings available through technology and infrastructure improvements is a key factor to obtaining the necessary funds. Most participants mentioned costs as a difficulty that they face, with some saying that energy is not a priority as it adds costs for their clients. One owner operator discussed the fact that they struggle to obtain funding even when they are able to demonstrate a short payback time. Whilst the scale of improvement costs participants were facing was not covered in the interviews, if was clear that most participants found it difficult to justify the costs involved or to source the required funding.

Whilst some may focus on obtaining funding for larger scale energy improvements, there are a number of enhancements that can be made in a data centre for lower cost, but which still offer energy saving benefits. Chapter 3 introduced the concept of MAC curves, a tool which can enable comparisons to be made between different technologies in terms of their carbon savings potential and investment required. Whilst measuring the carbon abatement potential may not appeal to data centre owners as a suitable measurement, a similar graph which ranks technologies and improvements based upon energy (or cost) saving would directly demonstrate the savings possible and which technologies to focus on first. This would enable the low cost and high impact changes to be targeted and also provide visual reference to how much energy could be saved. This kind of visual would enable better arguments for funding.

In conclusion, the results from Chapter 5 have shown that there is a need to focus on the management of data centres (i.e. how the facility is managed, communications between teams and parties, and education of staff and clients), the metrics that are used, the economics involved and also the logistics of how new technologies can be installed. Addressing these factors alongside focussing on which technologies to use would address what Backlund termed the extended energy efficiency gap, i.e. once significant savings are made through technology improvements, the management and operational side of a process needs to be considered (Backlund et al., 2012).

The interviews carried out in this work offer an alternative perspective to others currently in the literature in the sense that they found out the opinions of the data centre industry about energy efficiency and provided evidence of the problems that they face. These opinions were gathered from colocation providers, owner operators and also engineering consultants. This work has provided evidence of barriers to further energy savings in data centres. It is important that these are addressed if the industry is to make further energy savings and reduce its overall impact on the environment.

8.3 Is the PUE metric a useful tool for driving further energy savings?

Alongside barriers to improvements, the interviews also gathered opinions about various other topics on the subject of energy use and efficiency in data centres, including metrics and assessment methods. The results found that all participants measure or monitor energy use within their data centre, although to varying levels and through the use of different tools. When asked if they used any particular metrics to assess their data centre, almost all participants mentioned PUE. Several had actually already mentioned the PUE during the initial stage of the interview when describing their facility, indicating the use of it as a marketing tool.

Whilst the PUE was widely mentioned, there were stark contrasts in participant opinion regarding the usefulness of the metric. One participant, a consultant, said that they find the metric useful to give clients a benchmark to improve their facility on and a way to demonstrate improvements. On the other hand, a colocation participant said that using the metric made no sense to them, preferring to just judge improvements based upon a total power use measurement. Opinion is clearly divided on the usefulness of this metric as whilst it is a convenient tool to demonstrate improvements for a particular facility, it is not necessarily the best option for everyone. Despite these differences of opinion, PUE was the only metric mentioned by any of the participants demonstrating that the array of metrics available within the literature are not be successfully filtering down into the industry.

Several participants in the qualitative study mentioned that they use the PUE metric with some caution. The PUE metric is commonly used to assess energy use in the data centre, however the industry has also recently become more aware of the potential problems with using it. Some participants were of the opinion that the industry needed a new, more suitable metric for the purpose of energy efficiency measurement. This was in particular reference to the use of metrics within energy assessment schemes such as the EU CoC.

In order to further examine the potential pitfalls of the PUE metric, Chapter 6 conducted a case study measurement and sensitivity analysis. Open source information made available through the Open Compute Project provided the background for this case study. The Open Compute Project, operated by Facebook, aims to share information about the social media company's large new data centre in Prineville, USA so that other companies can learn about how they achieved high levels of energy efficiency. It was this desire of Facebook to share knowledge and information which allowed the PUE case study in this thesis to be carried out without direct access to a data centre. Detailed information regarding the infrastructure within and energy use of a data centre is very difficult to access (even the Open Compute Project did not provide a lot of detail). Nevertheless, having open source information meant that a PUE case study could be done whilst also highlighting issues surrounding access to data.

Whilst detailed information is needed for a truly accurate PUE measurement, it was possible to reproduce the published values for the Prineville data centre with a good level of agreement using the open source documents. Due to the limited information actually available through the open source documents and the amount of assumptions that had to be made to obtain a PUE measurement, this flagged up some issues with the use of the metric. Namely that it is easy to obtain a PUE value without detailed assessment of the data centre. The close agreement between the published and case study PUE values demonstrate the ease with which this metric can be calculated. However, it is this ease which also means that details can be negated in order to reach a believable estimation and results can (and are, in the industry) given without description of the measurement process. A description of the way in which the PUE was measured is key to understanding the answer that it gives.

Once the PUE measurement was reproduced, a sensitivity analysis was carried out on the inputs to see how they each affect the final result. It was found that once significant energy savings are made through improvements in the cooling infrastructure, the biggest effects on the PUE value come from the efficiency of power delivery and the IT load in the facility. In recent years there has been a strong focus on reducing the energy use of the cooling infrastructure, and significant steps have been made in achieving this. However the sensitivity analysis reinforces the idea that the PUE is only useful for considering the energy use of the supporting infrastructure. If further savings are to be made then the focus now needs to expand to include the IT systems and efficiency gains that can be made in the power distribution system.

Given the restrictions of the PUE this would suggest that a new metric is needed. One conclusion from the case study is that the PUE hides the amount of energy consumed by the IT equipment. The metric purely assesses the energy overheads required to run the IT equipment, without placing any scrutiny on the energy consumption of the hardware. LCA has clearly demonstrated that servers have the biggest impact on the carbon footprint of a data centre due to material make-up and energy consumption. However PUE focusses purely on the supporting infrastructure and places no emphasis on the servers or other IT equipment. Newer metrics in the literature such as ITUE (Patterson et al., 2013) and FVER (Newcombe et al., 2012) should help this problem, however they are not necessarily being used within the industry yet. Results from the qualitative study reinforce the idea that a new metric is needed that considers more than just the overheads of the infrastructure. Participants in the interviews said that heading into the future, IT equipment needs to be considered much more closely if further savings are to be made.

Whilst the industry is aware that the PUE metric is not entirely suited to its purpose, the work in this thesis has demonstrated what some of the issues are and why a new metric is needed. Nevertheless, as one participant in the interviews did mention, the PUE has served as a useful education tool that has acted as an encouragement for driving down energy consumption.

8.4 Is there energy that can be saved in a standard server?

Considering the fact that it is easy to conceal the energy consumption of the servers within the PUE measurement, one of the aims of Chapter 7 was to examine this further. Also, the case study in Chapter 6 and results from the qualitative study showed that there needs to be a stronger emphasis placed on the energy consumption of the IT equipment. Therefore, the work presented in Chapter 7 is also a technological contribution towards achieving this.

Chapter 7 presented the methodology behind and the results from an experimental approach investigating possible energy savings within a typical server. Internal server fans are technically part of the air distribution system but are included in the IT load section of the PUE measurement. They consume a significant proportion of overall server power and are inefficient, Chapter 7 shows that energy could theoretically be saved by removing them.

In order to more accurately plot the efficiency curve of an internal server fan, the manufacturers fan curve was reproduced under laboratory conditions. An air flow test bench was constructed in order to measure the flow rate and pressure delivered by the fan. The characteristics of the eight fans within the server were approximated by assuming that they act in perfect parallel nature and an efficiency curve was produced for the whole group of fans in the server. At only 8% efficiency, it was found that there is opportunity to increase the efficiency of the server as a whole. It should be noted that this low efficiency is one of the reasons that a move towards liquid, rather than air, cooling technologies would present significant savings in energy use.

The overall aim of Chapter 7 was to find whether it would be more energy efficient to remove the internal server fans and replace them with a single, larger external fan. It is noted that this is similar to the way that blade servers operate, however the key point here is that this method would allow normal servers to consume less energy rather than upgrading to blade servers. It has been shown experimentally that it may be possible to remove the internal fans from one server and replace them with a single external one. However in the study carried out and with the available equipment, energy will only be saved once there are at least two servers being supported by the external fan that was used. Of course another external fan could be sized and selected that would provide energy savings when supporting just one server (or likewise, more than two).

Unfortunately it was not possible to maintain two servers on one external fan without one of the servers overheating and failing, likely caused by the heat transfer between the two units. The conclusion is made that it would be possible to save energy, but the experimental set up would need further work in order for it to be applicable in practice. Further investigations would be needed into the air flow through the server under these conditions, Section 8.5 describes recommendations for further work.

Theoretically it has been shown that a saving of up to about 9% of energy consumption could be made by operating the servers with an external fan. When combining these results with data gathered during the case study in Chapter 6 it was found that this energy saving actually translates to an increase in the PUE value. The PUE analysis also highlighted the fact that within the '1' is a significant amount of energy that is not necessarily currently considered when making efforts to reduce energy use in the data centre. The conclusions from this unique study have also highlighted the energy benefits of making a switch to liquid based forms of cooling. By removing the internal server fans (and external if possible), significant amounts of energy could be saved.

8.5 Recommendations and Future Research Endeavours

Following on from the conclusions of this work, the recommendations in this section aim to address the problems highlighted. Suggestions for the future research required to achieve these are also made. Recalling Figure 1.3, Figure 8.1 shows that the results of chapters 6 and 7 could be used to influence the design of further qualitative study, consulting with members of the industry on the implications of the findings. As discussed in more detail below, this would involve surveying a larger number of data centre industry members on topics such as the use of metrics, new and emerging metrics, implications of using the PUE, IT equipment energy use and the barriers to altering the components in the server.

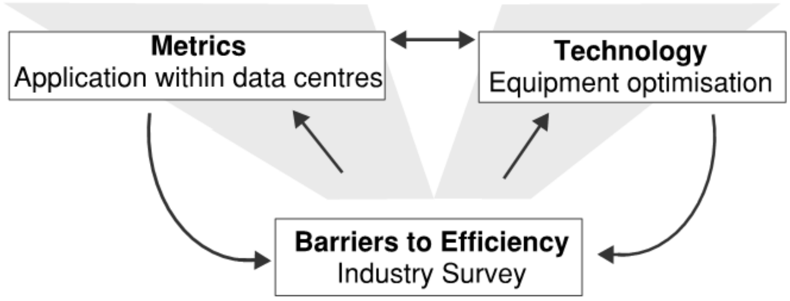


Figure 8.1 Feedback loop from the findings of chapters 6 and 7 into further industry surveys

- *Education, information and knowledge sharing*

Several approaches would enable the industry to overcome the problems discussed in this thesis, the first of which is education, information and knowledge sharing. Sharing of experiences and expertise is vital if knowledge is to be improved and better distributed. For a wider understanding of the problem of energy consumption in data centres, education of other sub-systems in the ICT sector (such as data centre users) would be a beneficial step. Departments and systems that use data centre provisions should be educated on the energy use problems of these facilities and the need to reduce energy. This would involve for example tenants of colocation facilities being given more information regarding the impacts of energy consumption and how they can better integrate the energy use of their IT service into the footprint of their business as whole. If tenants were more aware of their IT energy use, and had a responsibility to reduce it (through the use of internal or Governmental policy), then it would create a better driver for reduced energy use in data centres. Reducing the energy use of IT services is linked to financial benefits in the form of reduced costs and can also bring 'green image' benefits, these should both be drivers for Green SLAs to become more commonplace. Creating this drive for businesses to invest in 'greener' IT would more easily align the goals of the colocation provider and the tenant, a theme which has arisen throughout this thesis. Creating a greater demand for green IT would be hugely beneficial, a greater consumer demand for efficiency will help to negate the customer agreement barrier.

Engaging customers and users of data centres in the problem of energy use would be beneficial and help to drive a bottom up approach of improvements. The EU CoC, and likewise BREEAM, are useful tools to assess the efforts that a data centre has made to reduce consumption. However these schemes do not engage the tenant or the person using the data centres services. A scheme aimed at the users of data centres would enable the goals of these parties to be aligned. A scheme which would encourage (or enable) companies to focus on their IT energy consumption would more easily enable them to enter into a Green SLA. A better understanding of IT energy use on behalf of users and tenants would help to negate or reduce the barrier faced by colocation providers, whereby they find it difficult to persuade tenants to accept efficiency improvements.

Information and knowledge sharing in regards to the methods which a data centre may have employed to reduce energy use is also key. The Open Compute Project is leading the way in this area, sharing information about how Facebook reduces their energy use. Despite these efforts, it is still difficult to obtain information and data about this industry when it comes to energy use. The EU CoC publishes a list of participants and endorsers on its website, which is useful for companies to see who else involved. An improvement on this would be for these companies to provide a statement on which energy efficiency measures which they have used and the reduction in energy (or costs) that these interventions have achieved.

It will also be important to improve the communications between internal data centre teams. For facilities that are owned and operated for the use of a business, a lack of communications between teams can lead to inefficiencies. Again here is an opportunity for communication of the need to reduce energy use and how it can be done, which would hopefully translate into real savings. Improved internal communications would be introduced through the use of energy efficiency policies on behalf of the business, something which the EU CoC encourages.

For these recommendations to be developed, it is envisaged that the following future research tasks would be required. Firstly, a wider sample of participants would be needed to take part in an interview or survey study similar to the one in this thesis. Whilst the sample in this work was large enough for saturation of ideas, a larger study is needed to incorporate more representatives of the industry such as tenants and users. Surveying a larger representative section of the industry on their opinions about energy use would allow misconceptions and attitudes towards energy use to be brought into the open and addressed. Including tenants and users would enable a better understanding of their drivers and needs when it comes to renting data centre space. This would be a precursor to developing better incentives for companies seeking data centre services to choose more efficient facilities and aim to reduce the energy consumption of their ICT provisions. Such a survey would also help to identify the gaps in company's knowledge regarding energy consumption and the ICT sector. The landlord/tenant relationship would also benefit from further in depth investigation so that Green SLAs can be developed further and become more common place within the industry.

- *A better understanding of life cycle costing and which technologies to invest in*

The interview results provided evidence that data centre owners or colocation providers do not necessarily consider which technologies would offer the best energy savings relative to the investment cost required. Also, difficulty with obtaining funding for improvements would suggest a lack of understanding on the funder's behalf of the savings that could be achieved and the payback potential of such technologies.

The industry would benefit from the production of a MAC (or similar) graphic that ranks the energy saving potential of all available improvements (both technical and managerial) against the cost that may be required to install and maintain them. Such a graph could simply show relative cost, or could be adapted for different types and sizes of data centre to illustrate a closer estimation of cost. This is a tool that would be highly valuable as schemes such as EU CoC don't list which improvements are most beneficial or likewise which would be the lowest cost. As found from the results of the qualitative study, there appears to be a lack of life-cycle costing understanding from those responsible for obtaining funding for efficiency improvements. The integration of a MAC tool or certain economic indicators into data centre energy use policies or schemes such as the EU CoO or BREEAM would therefore be useful.

A simple version of a MAC graphic that is easy to access and distribute would also enable those facilities not involved in such schemes to consider what changes they could make, hence reaching a wider audience. A visual graphic showing the energy saving benefits of different improvements, both technological and managerial would also enable better communication of energy saving ideas. Better communication has already been suggested as an improvement that the industry could make, such a visual tool would be an integrated part of this.

In order to make this tool a reality, extensive research is required into the carbon savings that are possible for each measure and technological intervention that is available. Such research would require close collaboration with equipment manufacturers to find current efficiency ratings achievable with each technology available. This, along with current purchase and installation costs would provide better information for data centre owners, and a better case for investment.

- *Addressing the problem of using the PUE metric as a sole indicator of 'energy efficiency'*

Chapter 6 concluded that further development of metrics is needed if they are to drive additional energy savings in the data centre. This was reinforced by the qualitative results presented in Chapter 5, through divided participant opinion on the usefulness of

PUE and mention of the need for a new methodology. Designing a metric that can assess energy use (or more appropriately, useful work done) and allow figures to be compared from one data centre to the next however will always be a difficult task. There have been some attempts in the literature regarding the development of more inclusive metrics, yet the application of these within the industry still needs to be carefully considered. Repeating a case study and sensitivity analysis like the one in Chapter 6 is a useful exercise to consider any issues that there might be with new metrics.

Chapter 3 introduced the idea that each energy assessment metric or methodology is suited to either one particular section of the internal data centre system or an external party that has an influence on the facility. For instance, recirculation indexes are useful for owners to see inefficiencies in air delivery, PUE is useful for design engineers to demonstrate an efficient cooling system and FVER is useful for the tenant to assess the energy use relative to useful IT work done. Consequently a crucial part of any new metrics development is ensuring that the applicability is well thought out in terms of what the metric aims to do and which systems it is envisaged will use it.

In terms of recommendations and further research required, this thesis has shown the importance of not just considering the supporting infrastructure in energy use assessment. Metrics used within the industry need to move their focus to the actual work done and show a more holistic view of energy use and also sustainability. As energy use in the infrastructure has improved greatly over the last few years, driving the commonly cited PUE value ever closer to 1, energy use of IT equipment is being concealed making further savings difficult to justify. Moving forward from this, metrics now need to focus on overall energy consumption, actual work done (i.e. virtual services) and perhaps other factors such as life cycle emissions. The metrics required to achieve this already exist within the literature, but they are not successfully becoming commonplace for use in the industry. Only once other metrics or methodologies become as popular and easy to use as the PUE will they aid in driving overall energy use down.

In regards to the PUE, the applicability of this metric should be carefully considered. Looking back to the definition of energy efficiency in Chapter 3, it is clear that PUE does not represent the efficiency of a data centre. Indicators such as Energy Intensity (equation 3.2) offer better opportunities as they rate service outputs rather than energy outputs against energy inputs, which is more applicable to data centre services. Metrics such as FVER are based upon this principle.

The PUE is beneficial during the design stages of a data centre to assess the energy saving potential of different cooling options, and for the day to day running to monitor efficient operation, but should not be used to advertise a facility or market energy efficiency levels. When selecting a facility, the user of the data centre should have more information than a PUE value alone, they need access to a metric that encompasses more than just the energy use of the supporting infrastructure. PUE is a good initial indicator of energy use in the data centre, however the user of data centre services would also benefit from life cycle information to make a more rounded decision about which facility to place their IT equipment in. Couple this together with a better approach to IT equipment and software efficiencies, this would better enable businesses to assess the overall environmental impact of their use of IT.

Further research is required into the applicability of alternative metrics in the industry. An extended survey on a larger sample would need to be done in order to gather a wider base of opinion with regards to metrics in particular. Metrics such as FVER need to be tested within the industry for ease of use and applicability to those using it. It would be beneficial if the qualitative research in this thesis were extended to the users and tenants of data centres, to see which metrics would be best for their needs. Extensive research will be required into what impacts any regulation of businesses IT energy consumption would have, and whether it would be advantageous to drive further data centre energy savings with a demand led approach as suggested earlier.

- *Shifting some of the focus of energy saving towards the main component of the data centre, the server*

Whilst emphasis still needs to be placed on using low energy cooling and air distribution solutions, the server needs to receive more attention in terms of energy use. This will become more important in the future as infrastructure efficiency hits a peak and IT equipment also becomes denser in both computing power and physical footprint. The suggestions made above regarding new choices of metrics should help with this, as they will turn the focus towards actual work done rather than the energy required to support this work. Also, as suggested earlier, if companies had more responsibility for the consumption of their IT equipment (rather than just the efficiency with which they are supported), then the desire to decrease energy use of this major part of the data centre will improve.

Removing the internal server fans offers one opportunity to save energy, as demonstrated by the experimental work in Chapter 7, however further development of this investigation is required. Unfortunately, it was not possible to run an energy analysis on running two servers with one external fan because of problems encountered. Therefore further experimental work needs to be done to prove the theoretical savings as outlined in Chapter 7. The experimental procedures would be improved if they were conducted under better conditions, perhaps within a server rack with a mounted external fan. In order to develop the experiment, more complex ducting could be implemented either internal or external to the server. External ducting could proportion the amount of airflow that reaches each server and internal dividers would ensure correct distribution across the components that need the most heat removal.

If the energy savings were proved, then it would also be necessary to take a closer look at the effect that operating with an external fan would have on the components within the server. The practicalities of such a change also needs to be investigated in terms of whether it would be possible to do this in a working data centre, how much it would cost to implement, what risks would be involved, how it would affect warranties and also whether air would still be delivered to the internal components correctly. Given the results of the qualitative study showing that agreement from customers can be a significant barrier to changes in the data centre, it is also likely to be a barrier in this case too. A direct comparison to the energy used with liquid cooling technologies such as liquid submersion would also be valuable future work to demonstrate the benefit of removing fans from the servers. Chapter 3 touched on an idea by Meza et al who suggested the removal of all casing materials in order to reduce the overall carbon footprint (Meza et al., 2010). A next developmental stage of the experimental work in this thesis would be to investigate the carbon footprint impacts of removing the server fans, and the practicalities of the suggestion from Meza et al.

8.6 Final Summary

The data centre industry is vital to the running of todays' world, without the complex IT systems within these facilities, services that are crucial to everyday life would not be able to operate. In turn, these complex IT systems would not be able to function if they were not supported by an integrated network of infrastructure including cooling, air distribution, power delivery, lighting and security to name but a few. All of these systems and infrastructures together consume a vast amount of energy and are contributing to the worlds' emissions of GHGs. If the world is to continue expanding in its digital services, whilst also making efforts to reduce worldwide GHG emissions and the impacts of climate change, then the data centre industry must play a part in this. It is noted that whilst the services they provide help to reduce emissions in other sectors, energy consumption still needs to be reduced in the face of a rapid expansion in demand for services.

New technologies available for cooling the data centre are of course of high importance in the effort to reduce energy consumption, however they are not the only answer. Over the past several years, the mechanical aspect of data centres has been a real focus (and also target) for energy use reduction, partly as a result of the measurement tools and metrics available. However this has meant that the IT equipment has gone somewhat unchecked. Whilst there are schemes such as Energy Star that consider the energy efficiency of servers, energy analysis and commonly used metrics fail to take this into account. Also, once maximum savings are made through technological means, inefficiencies still arise because of issues with management, communications and education. Only once all of these are considered together will the data centre reach its theoretical energy efficiency potential.

This thesis has provided evidence of where the data centre industry is having problems reducing energy use, and suggested how further improvements may be made. Four main barriers were identified in the interviews with industry professionals, none of which involve the development or implementation of new technologies. The focus here was instead on overcoming the costs of such technologies, overcoming the communications barriers that exist, the difficulties that arise from a lack of customer agreement, and the logistical issues that are faced by older facilities.

The interviews in Chapter 5 along with the case study in Chapter 6 have shown that the industry needs a new approach to the way that metrics are applied. Overall, the industry needs to be using metrics available in the literature which relate actual work done to energy input, however they may need to be adapted to be easier and more accessible to use. If energy auditing schemes such as the EU CoC and BREEAM are to become more commonplace, then they should be based on a more applicable metric. Any new metrics that are suggested need to be tested thoroughly within the industry rather than an academic environment to ensure their viability. Better consideration needs to be given to the applicability of the metric being used. PUE is highly suited to assessing the energy use of the infrastructure, but in order to properly assess the environmental impacts of IT services a business needs more than just the results of this measurement. This thesis recommends that a more all-inclusive rating for a data centre is developed which incorporates not only the PUE but other indicators such as life cycle carbon footprint. The focus of schemes such as BREEAM is still on the performance of the building, rather than the overall impacts of the use of IT equipment, something that needs to be considered if the environmental impacts of data centres and the IT sector is to be properly realised.

As the industry progresses and demand for access to data increases, IT technology increases in density, and energy prices increase, it is going to become even more important that data centres focus on their energy consumption and overall environmental impacts. Data centres will come under closer scrutiny as they become more prevalent in the public eye. These facilities may become more integrated with society in the future with technologies such as waste heat recovery providing excess energy to industries that can use it (such as district heating).

This work has provided evidence of where the data centre industry is facing barriers to further energy savings and the interdisciplinary nature of the problem. It is an inescapable fact that demand for digital services will increase worldwide and that the role of the data centre is here to stay. It is this that makes it ever more important to consider the energy use of these facilities. The results of this research have shown areas within the industry that need to be improved. If these areas are all considered in close context with one another, then it is possible that the data centre can make further energy efficiency improvements beyond just improving the more obvious systems such as the mechanical infrastructure.

References

42U. 2009. The Data Center Temperature Debate [Online]. Available: http://www.42u.com/cooling/data-center-temperature.htm#comfort [Accessed 8th June 2012].

AGUSTI-TORRA, A., RASPALL, F., REMONDO, D., RINCON, D. & GIULIANI, G. 2015. On the feasibility of collaborative green data centre ecosystems. Ad Hoc Networks, 25, 565-580.

AHUJA, N. Datacenter power savings through high ambient datacenter operation: CFD modeling study. Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2012 28th Annual IEEE, 18-22 March 2012 2012. 104-107.

ALMOLI, A., THOMPSON, A., KAPUR, N., SUMMERS, J. THOMPSON, H. & HANNAH, G., Computational fluid dynamic investigation of liquid rack cooling in data centres. 2012. Applied Energy, 89, 150-155.

AMERICAN SOCIETY OF HEATING REFRIGERATION AND AIR-CONDITIONING ENGINEERS 2011. 2011 Thermal Guidelines for Data Processing Environments - Expanded Data Center Classes and Usage Guidance. ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment.

AMERICAN SOCIETY OF HEATING REFRIGERATION AND AIR-CONDITIONING ENGINEERS 2009. Design Considerations for Datacome Equipment Centres. Second Edition.

ANSI/AMCA 1999. Laboratory Methods of Testing Fans for Aerodynamic Performance Rating. Air Movement and Control Association International, Inc. and American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc.

ARIMURA, T. H., DARNALL, N. & KATAYAMA, H. 2011. Is ISO 14001 a gateway to more advanced voluntary action? The case of green supply chain management. Journal of Environmental Economics and Management, 61, 170-182.

AUVIL, M. 2013. Funding the Unfunded: Energy Savings Performance Contracts for Federal Data Center Consolidation [Online]. Available: http://blog.schneider-electric.com/datacenter/data-center-architecture/virtualization/2013/04/18/funding-the-unfunded-energy-savings-performance-contracts-for-federal-data-center-consolidation [Accessed August 2013].

AVASANT c2014. Big Data Analytics. California, USA.

BACKLUND, S., THOLLANDER, P., PALM, J. & OTTOSSON, M. 2012. Extending the energy efficiency gap. Energy Policy, 51, 392-396.

BANERJEE, A., MUKHERJEE, T., VARSAMOPOULOS, G. & GUPTA, S. K. S. 2011. Integrating cooling awareness with thermal aware workload placement for HPC data centers. Sustainable Computing: Informatics and Systems, 1, 134-150.

BARROSO, L. A. & HÖLZLE, U. 2007. The Case for Energy-Proportional Computing. IEEE Computer, 33-37.

BEAN, J. H. & MOSS, D. L. [no date]. Energy Impact of Increased Server Inlet Temperature. Dell.

BEATY, D. & LINTNER, W. Achieving Energy-Efficient Data Centers with New ASHRAE Thermal Guidelines. Federal Energy Management Program (FEMP) First Thursday Seminars.

BEITELMAL, M. H. & PATEL, C. D. 2006. Model-Based Approach for Optimizing a Data Center Centralized Cooling System. Hewlett-Packard Technical Report HPL-2006-67.

BELOGLAZOV, A., ABAWAJY, J. & BUYYA, R. 2012. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. Future Generation Computer Systems, 28, 755-768.

BIANZINO, A. P., CHAUDET, C., ROSSI, D. & ROUGIER, J.-L., 2012. A Survey of Green Networking Research. Communications Surveys and Tutorials. 3-20.

BIG SWITCH PROJECTS 2004. Data Centre Energy Efficiency Report.

BIRCHALL, S., 2011. An Appraisal of the Performance of a 'Green' Office Building. Ph.D Thesis. University of Leeds.

BLUMSTEIN, C., KRIEG, B., SCHIPPER, L. & YORK, C. 1980. Overcoming social and institutional barriers to energy conservation. Energy, 5, 355-371.

BOCCALETTI, G., LOFFLER, M. & OPPENHEIM, J. M. 2008. How IT can cut carbon emissions [Online]. Available: https://www.mckinseyquarterly.com/How_IT_can_cut_carbon_emissions_2221 [Accessed 4th July 2012].

BODENSTEIN, C., SCHRYEN, G. & NEUMANN, D. 2012. Energy-aware workload management models for opertion cost reduction in data centres. European Journal of Operational Reseearch, 222, 157-167.

BREEAM 2010. BREEAM Data Centres 2010. Scheme Document SD 5068.

BREEN, T. J., WALSH, E. J., PUNCH, J., SHAH, A. J. & BASH, C. E. From chip to cooling tower data center modeling: Part I Influence of server inlet temperature and temperature rise across cabinet. Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on, 2-5 June 2010 2010. 1-10.

BSI. 2014. BS EN 50600-2-1:2014, Information Technology – Data Centre Facilities and Infrasatructures. Part 2-1: Building Construction. London. BSI.

CAGNO, E., WORRELL, E., TRIANNI, A. & PUGLIESE, G. 2013. A novel approach for barriers to industrial energy efficiency. Renewable and Sustainable Energy Reviews, 19, 290-308.

CALIFORNIA ENERGY COMMISSION 2008. Uniterruptible Power Supplies: A Data Centre Efficiency Opportunity. California Energy Commision's Public Interest Energy Research Program.

CHAI, K.-H. & YEO, C. 2012. Overcoming energy efficiency barriers through systems approach—A conceptual framework. Energy Policy, 46, 460-472.

CHAN, M. A., YAP, C. R. & NG, K. C. 2009. Modeling and testing of an advanced compact two-phase cooler for electronics cooling. International Journal of Heat and Mass Transfer, 52, 3456-3463.

CHI, Y. Q., SUMMERS, J., HOPTON, P., DEAKIN, K., REAL, A., KAPUR, N. & THOMPSON, H. 2014. Case study of a data centre using enclosed, immersed, direct liquid-cooled servers. In Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM). 2014 30th Annual (pp. 164-173). IEEE.

CHO, J. & KIM, B. S. 2011. Evaluation of air management system's thermal performance for superior cooling efficiency in high-density data centers. Energy and Buildings, 43, 2145-2155.

CHO, J., LIM, T. & KIM, B. S. 2009. Measurements and predictions of the air distribution systems in high compute density (Internet) data centers. Energy and Buildings, 41, 1107-1115.

COLE, D. 2011. Data Center Energy Efficiency - Looking Beyond PUE. No Limits Software White Paper #4.

COOLING, G. R. The Carnot Jet System [Online]. Available: http://grcooling.com/GRC-Flyer-2011.pdf [Accessed 6th August 2012].

COPELAND, D. 2005. Review of Low Profile Cold Plate Technology for High Density Servers. Electronics Cooling.

COUNCIL, U. S. G. B. [no date]. LEED 2012 Development FAQ.

CRANE, J. 2013. Overcoming the 80/20 Barrier to Enable IT Innovation [Online]. Available: http://insights.wired.com/profiles/blogs/overcoming-the-80-20-barrier-to-enable-it-innovation#axzz2bIt3BXtU [Accessed August 2014].

DAIM, T., JUSTICE, J., KRAMPITS, M., LETTS, M. & SUBRAMANIAN, G. 2009. Data center metrics. Management of Environmental Quality, 20, 712-731.

DE LORENZO, D. S. 2002. Thermal design of a high-density server. Components and Packaging Technologies, IEEE Transactions on, 25, 635-640.

DEFRA, 2006,

DEPARTMENT OF ENERGY AND CLIMATE CHANGE. 2008. Climate Change Act 2008 [Online]. Available: http://www.decc.gov.uk/en/content/cms/legislation/cc_act_08/cc_act_08.aspx [Accessed 3rd June 2012].

DEPARTMENT OF ENERGY AND CLIMATE CHANGE 2011. Consultation on simplifying the Climate Change Agreements Scheme. London, UK.

DEPARTMENT OF ENERGY AND CLIMATE CHANGE 2013. 2012 UK Greenhouse Gas Emissions, Provisional Figures and 2011 UK Greenhouse Gas Emissions, Final Figures by Fuel Type and End-User.

DILLEY, A. C. 1968. On the Computer Calculation of Vapor Pressure and Specific Humidity Gradients from Psychrometric Data. Journal of Applied Meteorology, 7.

DOS SANTOS, A. H. C., FAGÁ, M. T. W. & DOS SANTOS, E. M. 2013. The risks of an energy efficiency policy for buildings based solely on the consumption evaluation of final energy. International Journal of Electrical Power & Energy Systems, 44, 70-77.

DUMITRU, I., FAGARASAN, I., ILIESCU, S., SAID, Y. H. & PLOIX, S. Increasing Energy Efficiency in Data Centers Using Energy Management. Green Computing and Communications (GreenCom), 2011 IEEE/ACM International Conference on, 4-5 Aug. 2011 2011. 159-165.

EBRAHIMI, K., JONES, G. F. & FLEISCHER, A. S. 2014. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. Renewable and Sustainable Energy Reviews, 31, 622-638.

ECKERT, J. P. 1946. Lecture 20: Reliability of Parts, MIT Press and Tomash Publishers.

den ELZEN, M., MEINSHAUSEN, M. 2005. Multi-Gas Emission Pathways for Meeting the EU 2°C. In: Schellnhuber, H. J., Cramer, W., Nakicenovic, N., Wigley, T., Yohe, G., eds. 2006. *Avoiding Dangerous Climate Change*. Cambridge University Press

EMERSON NETWORK POWER 2010. Precision versus Comfort Cooling Choosing a Cooling System to Support Business-Critival IT Environments. White Paper.

ENERGIZING TECH BLOG. 2010. Utility Sponsored Incentives for Data Centre Efficiency [Online]. Available: http://emidatacenterblog.wordpress.com/2010/03/18/utility-sponsored-incentives-for-data-center-efficiency/ [Accessed August 2014].

ENERGY INFORMATION ADMINISTRATION 1997. Annual Energy Outlook 1997 with Projections to 2015. Washington, DC: U.S. Department of Energy

ENERGY INFORMATION ADMINISTRATION. 2008. World Electricity Data [Online]. Available: http://www.eia.gov/iea/elec.html [Accessed 13th July 2011.

ENERGY STAR 2012. Understanding and Designing Energy-Efficiency Programs for Data Centres.

ENERGY STAR. 2014. Enterprise Servers for Consumers [Online]. Available: http://energy.gov/eere/femp/covered-product-category-enterprise-servers [Accessed September 2014].

ESOURCE. Evaporative Cooling [Online]. Available: http://www.esource.com/escrc/0013000000DP22YAAT/BEA1/PA/PA_Cooling/PA-42 [Accessed 6th July 2012].

EUROPEAN COMMISSION 2008. Code of Conduct on Data Centres Energy Efficiency Version 1.0. Institute for the Environment and Sustainability: Renewable Energies Unit.

EUROPEAN COMMISSION 2014. 2014 Best Practices: The EU Code of Conduct on Data Centres. Institute for Energy and Transport: Renewable Energy Unit.

EVANS, T. 2012. The Different Technologies for Cooling Data Centres. APC by Schneider Electric White Paper 59 revision 2.

FACEBOOK. 2013. Prineville, OR Data Center [Online]. Available: https://www.facebook.com/prinevilleDataCenter/app_399244020173259 [Accessed June 2013].

FACEBOOK. 2014. Prineville, OR Data Centre Dashboard: PUE & WUE [Online]. Available: https://www.fbpuewue.com/prineville [Accessed July 2014].

FAKHIM, B., BEHNIA, M., ARMFIELD, S. W. & SRINARAYANA, N. 2011. Cooling solutions in an operational data centre: A case study. Applied Thermal Engineering, 31, 2279-2291.

FANG, H., XIA, J. & JIANG, Y. 2015. Key issues and solutions in a district heating system uding low-grade industrial waste heat. Energy, 85, 589-602.

FLICK, U. 2009. An Introduction to Qualitative Research London, SAGE Publications Ltd.

FUKUE, T., ISHIZUKA, M., NAKAGAWA, S., HATAKEYAMA, T. & KOIZUMI, K. 2011a. Model for predicting performance of cooling fans for thermal design of electronic equipment (Modeling and evaluation of effects from electronic enclosure and inlet sizes). Heat Transfer—Asian Research, 40, 369-386.

FUKUE, T., KOIZUMI, K., ISHIZUKA, M. & NAKAGAWA, S. 2011b. Effects of Electronic Enclosure and Inlet Area on <I>P - Q</I> Curves of Installed Axial Cooling Fans. Journal of Environment and Engineering, 6, 650-664.

FURUTA, S. 2011. Open Compute Project: Server Chassis and Triplet Hardware v1.0.

GAO, Y., GUAN, H., QI, Z., SONG, T., HUAN, F. & LIU, L. 2014. Service level agreement based energy-efficient resource management in cloud data centers. Computers & Electrical Engineering, 40, 1621-1633.

GARG, S. K., YEO, C. S., ANANDASIVAM, A. & BUYYA, R. 2011. Environment-conscious scheduling of HPC applications on distributed Cloud-oriented data centers. Journal of Parallel and Distributed Computing, 71, 732-749.

GARTNER. Green IT: The new industry shock-wave. Symposium/ITXPO April 2007 2007 San Diego, CA.

GERMAIN-RENAUD, C., FURST, F., JOUVIN, M., KASSEL, G., NAUROY, J. & PHILIPPON, G. The Green Computing Observatory: A Data Curation Approach for Green IT. Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on, 12-14 Dec. 2011 2011. 798-799.

GLASER, B. G. & STRAUSS, A. L. 1967. The Discovery of Grounded Theory, London, UK, Weidenfeld and Nicolson.

GOLDEN, M. 2013. Stanford Expert Says Internet's Backbone Can Readily Be Made More Sustainable [Online]. Available: http://news.stanford.edu/news/2013/july/servers-energy-efficiency-071913.html [Accessed August 2014].

GREEN REVOLUTION COOLING. c2014. Gallery [Online]. Available: http://www.grcooling.com/ [Accessed 6th August 2012].

GREENBERG, A., HAMILTON, J., MALTZ, D. A. & PATEL, P. 2009. The Cost of a Cloud: Research Problems in Data Center Networks. ACM SIGCOMM Computer Communication Review, 39.

GREENBERG, S., MILLS, E., TSCHUDI, B., RUMSEY, P. & MYATT, B. 2006. Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers. Proceedings of the 2006 ACEEE Summer Study on Energy Efficiency in Buildings. Pacific Grove, CA.

GREENBLATT, D., AVRAHAM, T. & GOLAN, M. 2012. Computer fan performance enhancement via acoustic perturbations. International Journal of Heat and Fluid Flow, 34, 28-35.

GREENPEACE 2011. How Dirty Is Your Data? A Look at the Energy Choices That Power Cloud Computing. Greenpeace International.

GRIMES, R. & DAVIES, M. The effect of fan operating point and location on temperature distribution in electronic systems. Thermal and Thermomechanical Phenomena in Electronic Systems, 2002. ITHERM 2002. The Eighth Intersociety Conference on, 2002 2002. 677-684.

GUEST, G., BUNCE, A. & JOHNSON, L. 2006. How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. Field Methods, 18, 59-82.

GUZEK, M., PECERO, J. E., DORRONSORO, B. & BOUVRY, P. 2014. Multi-objective evolutionary algorithms for energy-aware scheduling on distributed computing systems. Applied Soft Computing, 24, 432-446.

HAHN, C. [no date]. Introduction to coding terminology [Online]. Available: http://qrtips.com/faq/FAQ--code%20terms.htm [Accessed 11th April 2014].

HAN, X. & JOSHI, Y. Energy Reduction in Server Cooling Via Real Time Thermal Control. 28th IEEE Semi-Therm Symposium, 18th-22nd March 2012 San Jose, CA, USA.

HAYWOOD, A., SHERBECK, J., PHELAN, P., VARSAMOPOULOS, G. & GUPTA, S. K. S. 2012. Thermodynamic feasibility of harvesting data center waste heat to drive an absorption chiller. Energy Conversion and Management, 58, 26-34.

Van HEDDEHGEM, W., LAMBERT, S., LANNOO, B., COLLE, D., PICKAVET, M. & DEMEESTER, P., 2014. Trends in worldwide ICT electricity consumption from 2007 to 2012. Computer Communications. 50, 64-76.

HERRLIN, M. K. 2005. Rack Cooling Effectiveness in Data Centers and Telecom Central Offices: The Rack Cooling Index (RCI). ASHRAE Transactions, 111, 725-731.

HERRLIN, M. K. Improved Data Center Energy Efficiency and Thermal Performance by Advanced Airflow Analysis. Digital Power Forum September 10-12, 2007 2007 San Francisco, CA.

HERRLIN, M. K. 2008. Airflow and Cooling Performance of Data Centers: Two Performance Metrics. ASHRAE Transactions, 114, 182-187.

HIRST, E. & BROWN, M. 1990. Closing the efficiency gap: barriers to the efficient use of energy. Resources, Conservation and Recycling, 3, 267-281.

HOLAHAN, M. F. & ELISON, B. P. 2007. Fan Laws for Rack Systems. ASME Conference Proceedings, 2007, 255-266.

HOPTON, P. & SUMMERS, J. 2013. Enclosed liquid natural convection as a means of transferring heat from microelectronics to cold plates. In Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2013 29th Annual IEEE (pp. 60-64). IEEE.

HUBER, P. & MILLS, M. 1999. Dig more coal - the PCs are coming. Forbes. 31st May.

IBM. [no date]. Bringing big data to the enterprise: What is big data? [Online]. Available: http://www-01.ibm.com/software/my/data/bigdata/ [Accessed October 2014].

IBRAHIM, M., SHRIVASTAVA, S., SAMMAKIA, B. & GHOSE, K. Thermal mass characterization of a server at different fan speeds. 2012 13th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 30 May-1 June 2012, 2012 Piscataway, NJ, USA. IEEE, 457-65.

ICEOTOPE. Iceotope Platform [Online]. Available: http://www.iceotope.com/assets/files/pdfs/iceotope-platform-spec.pdf [Accessed 6th August 2012].

IEA 2010. Energy Technology Perspectives 2010.

INTELLECT 2008. High Tech: Low Carbon, The role of technology in tackling climate change.

INTERNATIONAL ENERGY AGENCY 2013. World Energy Outlook 2013 Factsheet: How will global energy markets evolve to 2035?

IPCC 2007a. Changes in Atmospheric Constituents and in Radiative Forcing. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

IPCC 2007b. Historical Overview of Climate Change Science. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

IPCC 2007c. IPCC Fourth Asessment Report: Climate Change 2007, Cambridge University Press, Cambridge, UK and New York, NY, USA.

ISLAM, M. A., REN, S., PISSINOU, N., MAHMUD, A. H. & VASILAKOS, A. V. 2015. Distributed temperature-aware resource management in virtualized data center. Sustainable Computing: Informatics and Systems, 6, 3-16.

ITOH, S., KODAMA, Y., SHIMIZU, T., SEKIGUCHI, S., NAKAMURA, H. & MORI, N. Power consumption and efficiency of cooling in a data center. 2010 11th IEEE/ACM International Conference on Grid Computing (GRID 2010), 25-29 Oct. 2010, 2010 Los Alamitos, CA, USA. IEEE Computer Society, 305-12.

IYENGAR, M., DAVID, M., PARIDA, P., KAMATH, V., KOCHUPARAMBIL, B., GRAYBILL, D., SCHULTZ, M., GAYNES, M., SIMONS, R., SCHMIDT, R. & CHAINER, T. Server liquid cooling with chiller-less data center design to enable significant energy savings. 2012 IEEE/CPMT 28th Semiconductor Thermal Measurement &

Management Symposium (SEMI-THERM), 18-22 March 2012, 2012 Piscataway, NJ, USA. IEEE, 212-23.

JAFFE, A. B. & STAVINS, R. N. 1994. The energy-efficiency gap What does it mean? Energy Policy, 22, 804-810.

JILESEN, J., HARRISON, H., FUE-SANG, L. & MCCUMBER, D. Multiple High Performance Dual Redundant or DR FanTM Modules Optimized For 1U Server Processor Upgrade Program. Thermal and Thermomechanical Phenomena in Electronics Systems, 2006. ITHERM '06. The Tenth Intersociety Conference on, May 30 2006-June 2 2006 2006. 361-368.

JONES, P. 2013. UK Data Centers Get Climate Change Agreement [Online]. Datacenter Dynamics. Available: https://www.datacenterdynamics.com/focus/archive/2013/12/uk-data-centers-get-climate-change-agreement [Accessed January 2014].

JOUHARA, H. & MESKIMMON, R. 2014. Heat pipe based thermal management systems for energy-efficient data centres. Energy.

JUDGE, P. 2013. Ebay Shows On-site Generation Works [Online]. Available: http://www.greendatacenternews.org/articles/648718/ebay-shows-on-site-generation-works-by-peter-judge/ [Accessed August 2014].

KANT, K. 2009. Data center evolution: A tutorial on state of the art, issues, and challenges. Computer Networks, 53, 2939-2965.

KARLSSON, J. F. & MOSHFEGH, B. 2005. Investigation of indoor climate and power usage in a data center. Energy and Buildings, 37, 1075-1083.

KAWAMOTO, K., KOOMEY, J. G., NORDMAN, B., BROWN, R. E., PIETTE, M. A., TING, M. & MEIER, A. K. 2002. Electricity used by office equipment and network equipment in the US. Energy, 27, 255-269.

KEATS, D. 2000. Interviewing: A Practical Guide for Students and Professionals, Open University Press.

KEMFERT, C., KOHLHAAS, M., TRUONG, T. & PROTSENKO, A. 2006. The environmental and economic effects of Eurpoean emissions trading. Climate Policy, 6, 441-455.

KENNEDY, D. 2009. Understanding Data Center Cooling Energy Use & Reduction Methods. Rittal White Paper 507.

KESICKI, F., 2011. Marginal abatement cost curves for policy making – expert-based vs. model-derived curves. UCL Energy Institute, University College London.

KHANKARI, K. 2009. Rack Enclosures: A Crucial Link in Airflow Management in Data Centres. ASHRAE Journal, 48-51.

KIM, Y. J. & GONZALEZ, M. Exergy analysis of an ionic-liquid absorption refrigeration system utilizing waste-heat from datacenters. International Journal of Refrigeration.

KLIAZOVICH, D., ARZO, S. T., GRANELLI, F., BOUVRY, P. & KHAN, S. U. Accounting for load variation in energy-efficient data centers. Communications (ICC), 2013 IEEE International Conference on, 9-13 June 2013 2013. 2561-2566.

KLINGERT, S., SCHULZE, T. & BUNSE, C. GreenSLAs for the Energy-efficient Management of Data Centres. e-Energy, 31st May - 1st June 2011 2011 New York, NY, USA.

KOOMEY, J. G. 2007. Estimating Total Power Consumption by Servers in the U.S. and the World. Oakland, CA.

KOOMEY, J. G. 2008. Worldwide electricity used in data centers. Environmental Research Letters, 3.

KOOMEY, J. G. 2011. Growth in Data Center Electricity Use 2005 to 2010. Analytical Press.

KREITH, F. & GOSWAMI, D. Y. 2008. Energy Management and Conservation Handbook, Florida, USA, CRC Press.

KUMAR, P., SUNDARALINGAM, V. & JOSHI, Y. Effect of server load variation on rack air flow distribution in a raised floor data center. Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2011 27th Annual IEEE, 20-24 March 2011 2011. 90-96.

LBNL. 2001. U.S. Lab Study Says Computer Electricity Use Overestimated [Online]. Available: http://usinfo.org/wf-archive/2001/010206/epf205.htm [Accessed 08/07/2011.

LENT, R. 2015. Analysis of an energy proportional data center. Ad Hoc Networks, 25, 554-564.

LETTIERI, D. J., HANNEMANN, C. R., CAREY, V. P. & SHAH, A. J. 2009. Lifetime exergy consumption as a sustainability metric for information technologies. ISSST '09. IEEE International Symposium on Sustainable Systems and Technology. 18-20 May 2009. 1-6.

LIU, G. & LIU, M. 2012. Development of simplified in-situ fan curve measurement method using the manufacturers fan curve. Building and Environment, 48, 77-83.

LIU, S. H., HUANG, R. F. & LIN, C. A. 2010. Computational and experimental investigations of performance curve of an axial flow fan using downstream flow resistance method. Experimental Thermal and Fluid Science, 34, 827-837.

LOPER, J. & PARR, S. 2007. Energy Efficiency in Data Centers: A New Policy Frontier. Washington, USA: Alliance to Save Energy.

LU, T., LÜ, X., REMES, M. & VILJANEN, M. 2011. Investigation of air management and energy performance in a data center in Finland: Case study. Energy and Buildings, 43, 3360-3372.

LÜ, T., LÜ, X., REMES, M. & VILJANEN, M. 2011. Investigation of air management and energy performance in a data center in Finland: Case study. Energy and Buildings.

LUTZ, J., SCHLANGENOTTO, H., SCHEUERMANN, U. & de DONCKER, R. 2010. Semiconductor Power Devices. Springer Berlin Heidelberg.

MALMODIN, J., MOBERG, A. S., LUNDEN, D., FINNVEDEN, G. & LOVEHAGEN, N. 2010. Greenhouse gas emissions and operational electricity use in the ICT and entertainment media sectors. Journal of Industrial Ecology, 14, 770-790.

MALONE, C. & BELADY, C. Metrics to Characterize Data Center & IT Equipment Energy Use. 2006 Digital Power Forum, 2006 Richardson, TX, USA.

MALONE, C. & BELADY, C. L. 2008. Optimizing Data Centre TCO: Efficiency Metrics and an Infrastructure Cost Model. ASHRAE Transactions, v114 part 1, 44-50.

MARSHALL, M. N. 1996. Sampling for qualitative research. Family Practice, 13, 422-526.

MASANET, E., SHEHABI, A. & KOOMEY, J. G. 2013. Characteristics of low-carbon data centres. Nature Climate Change, 3, 627-630.

MASON, M. 2010. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. Forum: Qualitative Social Research 11.

MATHEW, P., GREENBERG, S., SARTOR, D., BRUSCHI, J., CHU, L., 2010. Self-benchmarking Guide for Data Center Infrastructure: Metrics, Benchmarks, Actions. Lawrence Berkeley National Laboratory.

MCALLISTER, S., CAREY, Van P., SHAH, A., BASH, C. & PATEL, C., 2008. Strategies for effective use of exergy-based modeling of data centre thermal management systems. Microelectronics Journal, 39, 1023-1029.

MCKENNA, R. C. 2009. Industrial energy efficiency: Interdisciplinary perspectives on the thermodynamic, technocal and economic constraints. Doctor of Philosophy, University of Bath.

MCQUISTON, F. C. 2005. Heating, ventilating, and air conditioning : analysis and design / Faye C. McQuiston, Jerald D. Parker, Jeffrey D. Spitler, Hoboken, N.J., John Wiley & Sons.

MEZA, J., SHIH, R., SHAH, A., RANGANATHAN, P., CHANG, J. & BASH, C. 2010. Lifecylce-based data centre design. Proceedings of the ASME 2010 International Mechanical Engineering Congress & Exposition, November 12-18, Vancouver.

MILLER, R. 2009. Google's Custom Web Server, Revealed [Online]. Available: http://www.datacenterknowledge.com/archives/2009/04/01/googles-custom-web-server-revealed/ [Accessed 6th August 2012].

MILLER, R. 2010. Data Centers That Recycle Waste Heat [Online]. Available: http://www.datacenterknowledge.com/data-centers-that-recycle-waste-heat/ [Accessed 6th August 2012].

MILLER, R. 2013a. The Efficiency Gap: Can Server Huggers Shift to the Cloud? [Online]. Available: http://www.datacenterknowledge.com/archives/2013/06/07/efficiency-and-the-cloud/ [Accessed 18th September 2013].

MILLER, R. 2013b. The Immersion Data Center: The New Frontier of High-Density Computing [Online]. Available: http://www.datacenterknowledge.com/archives/2013/07/01/the-immersion-data-center/ [Accessed August 2014].

MILLS, M. 1999. The Internet Begins With Coal: a preliminary exploration of the impact of the internet on electricity consumption (Executive Summary). A Green Policy Paper for the Greening Earth Society

MITCHELL-JACKSON, J., KOOMEY, J. G., BLAZEK, M. & NORDMAN, B. 2002. National and regional implications of internet data center growth in the US. Resources, Conservation and Recycling, 36, 175-185.

MITCHELL-JACKSON, J., KOOMEY, J. G., NORDMAN, B. & BLAZEK, M. 2003. Data center power requirements: measurements from Silicon Valley. Energy, 28, 837-850.

MODERN BUILDING SERVICES. 2010. 2.5 MW data centre is kept cool by evaporative cooling [Online]. Available: http://www.modbs.co.uk/news/fullstory.php/aid/8384/2.5_MW_data_centre_is_kept__cool_by_evaporative_cooling_.html [Accessed 11th November 2011].

MOHAN RAJ, V. K. & SHRIRAM, R. A study on server Sleep state transition to reduce power consumption in a virtualized server cluster environment. Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on, 3-7 Jan. 2012 2012. 1-6.

MOORE, G. E. 1965. Cramming more components onto integrated circuits. Electronics, 38.

MOORE, G. E. Progress in digital integrated electronics. Electron Devices Meeting, 1975 International, 1975 1975. 11-13.

MUROYA, K., KINOSHITA, T., TANAKA, H. & YOURO, M. Power reduction effect of higher room temperature operation in data centers. Network Operations and Management Symposium (NOMS), 2010 IEEE, 19-23 April 2010 2010. 661-673.

NATIONAL RENEWABLE ENERGY LABORATORY 2008. National Solar Radiation Data Base: Typical Meteorological Year 3. Data Set 726835 Redmond Roberts Field.

NEWCOMBE, L., LIMBUWALA, Z., LATHAM, P. & SMITH, V. 2012. Data Centre Fixed to Variable Energy Ration Metric DC-FVER. BCS The Chartered Institute for IT.

NGAI, T. 2011. Psychrometrics with Matlab [Online]. Available: http://www.tedngai.net/teaching/psychrometrics-with-matlab.html [Accessed 7th December 2011].

OFFICE OF ENERGY EFFICIENCY AND RENEWABLE ENERGY. 2014. Covered Product Category: Enterprise Servers [Online]. U.S. Department of Energy. Available: http://energy.gov/eere/femp/covered-product-category-enterprise-servers [Accessed September 2014].

OGREY, B., CHUNG, E. & JOE, O. 2013. AMD Open 3.0 Modular Server. Open Compute Project.

OHRING, M. 1998. Reliability and Failure of Electronic Materials and Devices, Boston, MA, Academic Press.

OPEN COMPUTE PROJECT. Specs & Designs [Online]. Available: http://opencompute.org/specs/ [Accessed 11th October 2011].

OPEN COMPUTE PROJECT. 2011. Energy Efficiency [Online]. Available: http://opencompute.org/about/energy-efficiency/ [Accessed 12th December 2011].

OPEN COMPUTE PROJECT. [no date]. Motherboard and Server Design [Online]. Available: http://www.opencompute.org/projects/motherboard-design/ [Accessed 21st June 2013].

PACHAURI, R. K. Statement of Dr R. K. Pachauri. Summit on Climate Change 2009, 22nd September 2009 2009 Copenhagen.

PARK, J. 2011. Open Compute Project: Data Center v1.0 [Online]. Available: www.opencompute.org.

PARLIAMENTARY OFFICE OF SCIENCE AND TECHNOLOGY 2008. ICT and CO2 Emissions. POST Note, 319.

PATTERSON, M. 2012. Energy Efficiency Metrics In: JOSHI, Y. & KUMAR, P. (eds.) Energy Efficient Thermal Management of Data Centers. Springer Verlag.

PATTERSON, M. G. 1996. What is energy efficiency?: Concepts, indicators and methodological issues. Energy Policy, 24, 377-390.

PATTERSON, M. K. The effect of data center temperature on energy efficiency. Thermal and Thermomechanical Phenomena in Electronic Systems, 2008. ITHERM 2008. 11th Intersociety Conference on, 28-31 May 2008 2008. 1167-1174.

PATTERSON, M. K., POOLE, S. W., CHUNG-HSING, H., MAXWELL, D., TSCHUDI, W., COLES, H., MARTINEZ, D. J. & BATES, N. TUE, a new energy-efficiency metric applied at ORNL's Jaguar. Supercomputing. 28th International Supercomputing Conference, ISC 2013, 16-20 June 2013, 2013 Berlin, Germany. Springer Verlag, 372-82.

PELLEY, S., MEISNER, D., WENISCH, T. F. & VANGILDER, J. W. Understanding and Abstracting Total Data Center Power. ISCA 2009: Workshop on Energy Efficient Design (WEED) in The 36th International Symposium on Computer Architecture, 20-24 June 2009 Austin, Texas, US.

PÉREZ-LOMBARD, L., ORTIZ, J. & VELÁZQUEZ, D. 2013. Revisiting energy efficiency fundamentals. Energy Efficiency, 6, 239-254.

POLOLU CORPORATION 2001. Pololu Simple Motor Controller User's Guide.

PONEMON INSTITUTE 2013. 2013 Study on Data Center Outages.

PRIYADUMKOL, J. & KITTICHAIKARN, C. 2014. Application of the combined air-conditioning systems for energy conservation in data center. Energy and Buildings, 68, Part A, 580-586.

RASMUSSEN, N. 2011. Air Distribution Architecture Options for Mission Critical Facilities. APC White Paper 55.

RECKTENWALD, G. 2006. A Flow Bench for Measuring Fan Curves and System Curves for Air-Cooled Electronic Equipment. Department of Mechanical Engineering, Portland State University.

REMSBURG, R. 2000. Introduction to Thermal Design of Electronic Equipment. Thermal Design of Electronic Equipment. CRC Press.

RUTH, S. 2011. Reducing ICT-related Carbon Emissions: An Exemplar for Global Energy Policy? IETE Technical Review, 28, 207-211.

SALDANA, J. 2009. An Introduction to Codes and Coding. The Coding Manual for Qualitative Researchers. SAGE Publications Ltd.

SANKAR, S., SHAW, M., VAID, K. & GURUMURTHI, S. 2012. Datacentre Scale Evaluation of the Impact of Temperature on Hard Disk Drive Failures. ACM Trans. on Storage.

SCHULZ, G. 2009. The Green and Virtual Data Center, Florida, USA, CRC Press.

SCHUTT, R. K. 2012. Qualitative Data Analysis. Investigating the Social World. SAGE Publications, Inc.

SHAH, A. J., CHEN, Y. & BASH, C. E. 2012. Sources of variability in data centre lifecyle assessment. 2012 IEEE Internationl Symposium on Sustainable Systems and Technology (ISSST), May 16-18, 1-6.

SHARMA, R. K., BASH, C. E., PATEL, C. D., 2002. Dimensionless Parameters for Evaluation of Thermal Design and Performance of Large-Scale Data Centers. American Institute of Aeronautics and Astronautics, AIAA-2002-3091.

SINGH, A. K., SINGH, H., SINGH, S. P. & SAWHNEY, R. L. 2002. Numerical calculation of psychrometric properties on a calculator. Building and Environment, 37, 415-419.

SLATERS, M. 2009. Data Centre Cooling: Benefits of Emerging New Technologies. Strategic Facilities.

SMITH, N. High efficiency electronic cooling fans. Semiconductor Thermal Measurement and Management Symposium, 2009. SEMI-THERM 2009. 25th Annual IEEE, 15-19 March 2009 2009. 92-97.

SODERMAN, M. Unknown. The Elephant in the Server Room: Bringing Server Rooms Into the 21st Century Sustainable World [Online]. 2 Degrees Network. Available: https://www.2degreesnetwork.com/groups/energy-carbon-management/resources/elephant-server-room-bringing-server-rooms-into-21st-century-sustainable-world/?goback=%2Egde_4973778_member_5828526705555308546#%21 [Accessed August 2014].

SORRELL, S., SCHLEICH, J., SCOTT, S., O'MALLEY, E., TRACE, F., BOEDE, U., OSTERTAG, K. & RADGEN, P. 2000. Barriers to Energy Efficiency in Public and Private Organisations. SPRU.

SPEC. 2011. SPECpower_ssj2008 [Online]. Available: http://www.spec.org/power_ssj2008/ [Accessed 8th May 2012].

SPEC. 2012. SPECpower_ssj2008 Dell Inc. PowerEdge R820 (Intel Xeon E5-4650L, 2.60 GHz) [Online]. Available: http://www.spec.org/power_ssj2008/results/res2012q2/power_ssj2008-20120515-00468.html [Accessed 2013 July].

STAFFORD, J., WALSH, E. & EGAN, V. 2010. Local heat transfer performance and exit flow characteristics of a miniature axial fan. International Journal of Heat and Fluid Flow, 31, 952-960.

STANSBERRY, M. Annual Data Center Industry Survey Results. Uptime Institute Symposium 2013, 2013 Santa Clara, CA.

STANSBERRY, M. & KUDRITZKI, J. 2012. Uptime Institute 2012 Data Centre Industry Survey. Uptime Institute.

STOVER, A., SACHS, H. & LOWENBERGER, A. 2013. Cryptic Barriers to Energy Efficiency. American Council for an Energy-Efficient Economy.

STRAUSS, A. 1987. Qualitative analysis for social scientists, Cambridge, UK, Cambridge University Press.

SUDHAKARA REDDY, B. 2013. Barriers and drivers to energy efficiency – A new taxonomical approach. Energy Conversion and Management, 74, 403-416.

SULLIVAN, A. Energy Star for Data Centers. The Green Grid Technical Forum, 2010.

SUN, H., STOLF, P., PIERSON, J.-M. & COSTA, G. D. Energy-Efficient and Thermal-Aware Resource Management for Heterogeneous Datacenters. Sustainable Computing: Informatics and Systems.

SUN, H. S. & LEE, S. E. 2006. Case study of data centers' energy performance. Energy and Buildings, 38, 522-533.

SUN MICROSYSTEMS Datasheet Sun FireTM V20z Server.

SUN MICROSYSTEMS 2008. Sun FireTM V20z and Sun Fire V40z Servers User Guide.

SVERDLIK, Y. 2013a. Ebay's Utah Data Center Offers A Glimpse Into The Future [Online]. Available: http://www.datacenterdynamics.com/focus/archive/2013/11/ebays-utah-data-center-offers-glimpse-future [Accessed August 2014].

SVERDLIK, Y. 2013b. Facebook Says Iowa Data Centre Will Be 100% Carbon Neutral [Online]. Available: http://www.datacenterdynamics.com/focus/archive/2013/11/facebook-says-iowa-data-center-will-be-100-carbon-neutral [Accessed August 2014].

THE BRITISH STANDARDS INSTITUTION 2012. Information Technology - Data Centre Facilities and Infrastructures. Part 1: General Concepts. BSI Standards Limited.

THE CLIMATE GROUP 2008. SMART 2020: Enabling the low carbon economy in the information age. Global eSustainability Initiative (GeSI).

THE GREEN GRID Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric. White Paper #32.

THE GREEN GRID Green Grid Data Center Power Efficiency Metrics: PUE and DCiE. White Paper #6.

THE GREEN GRID 2007. Green Grid Metrics: Describing Datacenter Power Efficiency: Technical Committee White Paper. White Paper #1.

THE GREEN GRID 2009. Usage and Public Reporting Guidelines for the Green Grid's Infrastructure Metrics (PUE/DCiE). White Paper #22.

THE GREEN GRID 2010a. ERE: A metric for measuring the benefit of reuse energy from a data center. White Paper #29.

THE GREEN GRID 2010b. The Green Grid Data Center Compute Efficiency Metric: DCcE. White Paper #34.

THE GREEN GRID 2011. Survey Results: Data Center Economizer Use. White Paper #41.

TRIANNI, A., CAGNO, E., WORRELL, E. & PUGLIESE, G. 2013. Empirical investigation of energy efficiency barriers in Italian manufacturing SMEs. Energy, 49, 444-458.

TRICKER, M. G. & LANGE, K.-D. The design and development of the Server Efficiency Rating Tool (SERT). 2nd Joint WOSP/SIPEW International Conference on Performance Engineering, ICPE 2011, March 14, 2011 - March 16, 2011, 2011 Karlsruhe, Germany. Association for Computing Machinery, 145-149.

TSUDA, K., TANO, S. & ICHINO, J. Lower data center power consumption through use of the climate characteristics of cold regions and inter-regional energy integration. 2010 1st IEEE International Conference on Progress in Informatics and Computing, PIC 2010, December 10, 2010 - December 12, 2010, 2010 Shanghai, China. IEEE Computer Society, 1298-1304.

TUMA, P. E. The merits of open bath immersion cooling of datacom equipment. Semiconductor Thermal Measurement and Management Symposium, 2010. SEMI-THERM 2010. 26th Annual IEEE, 21-25 Feb. 2010 2010. 123-131.

U.S. DEPARTMENT OF ENERGY & U.S. ENVRIONMENTAL PROTECTION AGENCY. Energy Efficiency in Data Centres: Recommendations for Government-Industry Coordination. National Data Centre Energy Efficiency Strategy Workshop 16th October 2008 2008.

U.S. ENVIRONMENTAL PROTECTION AGENCY 2007. Report to Congress on Server and Data Centre Energy Efficiency Public Law 109-431.

UK GOVERNMENT 2008. Climate Change Act.

UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE. Making those first steps count: An Introduction to the Kyoto Protocol [Online]. Available: http://unfccc.int/essential_background/kyoto_protocol/items/6034.php [Accessed 26th September 2013].

UPTIME INSTITUTE 2000. Heat Density Trends in Data Processing, Computer Systems, and Telecommunications Equipment.

VANGILDER, J. W., SHRIVASTAVA, S. K., 2007. Capture Index: An Airflow-Based Rack Cooling Performance Metric. ASHRAE Transactions, 113, 126-136.

VOGEL, M., CHEN, T., DOAN, S., HARRISON, H. & NAIR, R. New approach to system server air flow/thermal design development, validation and advancement in green fan performance. Semiconductor Thermal Measurement and

Management Symposium, 2010. SEMI-THERM 2010. 26th Annual IEEE, 21-25 Feb. 2010 2010. 136-142.

WAGNER, S. 2012. UPS Makes No Left Turns In Quest to Deliver Sustainability; Q&A [Online]. Bloomberg. Available: http://www.bloomberg.com/news/2012-09-20/ups-makes-no-left-turns-in-quest-to-deliver-sustainability-q-a.html [Accessed 24th September 2013].

WALKER, D. & MYRICK, F. 2006. Grounded Theory: An Exploration of Process and Procedure. Qualitative Health Research, 16, 547-559.

WALKER, W. 1985. Information technology and the use of energy. Energy Policy, 13, 458-476.

WALLACE, A. A. 2009. Reducing Carbon Emissions by Households: The Effects of Footprinting and Personal Allowances. Doctor of Philosophy, De Montfort University.

WEMHOFF, A. P. & ORTEGA, A. 2014. An exergy-based anaysis of the effects of rear door heat exchange systems on data centre energy efficiency. 14th IEEE ITHERM Conference, 1129-1136.

WHITEHEAD, B. The Screening Life Cycle Assessment of a Data Centre. 2nd LCA Conference, 6-7 November 2012 2012 Lille, France.

XU, D. & QU, M. 2013. Energy, environmental, and economic evaluation of a CCHP system for a data center based on operational data. Energy and Buildings, 67, 176-186.

ZACHARY WOODRUFF, J., BRENNER, P., BUCCELLATO, A. P. C. & GO, D. B. 2014. Environmentally opportunistic computing: A distributed waste heat reutilization approach to energy-efficient buildings and data centers. Energy and Buildings, 69, 41-50.

ZAFEIRIOU, G. 2000. Students' perceptions of issues arising from and factors influencing group interaction in computer conferencing: a grounded theory approach. Doctor of Philosophy, University of Sheffield.

ZHANG, H., SHAO, S., XU, H., ZOU, H. & TIAN, C. 2014. Free cooling of data centers: A review. Renewable and Sustainable Energy Reviews, 35, 171-182.

ZHENG, X. & Cai, Y. 2011. Energy-aware load dispatching in geographically located internet data centres. Computing; Informatics and Systems, 1, 275-285.

ZIMMERMANN, S., MEIJER, I., TIWARI, M. K., PAREDES, S., MICHEL, B. & POULIKAKOS, D. 2012. Aquasar: A hot water cooled data center with direct energy reuse. Energy, 43, 237-245.

Appendix

Appendix A: Interview Script

<table>
<tr><td><u>Introduction (begin recording)</u></td><td></td></tr>
<tr><td>Interview dated (date). Participant is (role) at a (engineering consultancy/company providing data centre space etc). Thank you for agreeing to take part in this research.</td><td></td></tr>
<tr><td><u>Opening Questions</u></td><td></td></tr>
<tr><td>Could you give me a brief description of what your company does and what role it plays in the data centre industry? For confidentiality purposes, you don’t need to mention the company name or any specific details about projects or facilities.</td><td>Data centre operators
Can you tell me a bit more about the data centres that you operate, i.e.;
Do you use the data centre(s) for your own purposes or do you rent the space out to others?
How many data centres do you operate?
What tier classifications are your data centres, if you use tier rating?
What kind of IT services does your data centre provide? i.e. colo, enterprise etc
Roughly how many servers are installed in your data centre(s)?
Do you know what the total power rating (in MW) of your data centre(s) is?
How many years has your data centre(s) been in operation for?
Are they all located in the UK?
Is the data centre(s) running at full capacity yet, or is it running out of space, power or cooling?
Consultant engineers
What scope do you normally have for working on a data centre?
How many years have you been working on data centre designs?</td></tr>
<tr><td>What is your role in the company?</td><td>Data centre operators
Do you have any direct responsibilities for the energy that is consumed, if not, who does?</td></tr>
<tr><td><u>Attitudes towards energy efficiency</u></td><td>Thank you, I’m now going to ask you some questions about energy efficiency.</td></tr>
</table>

Data centre operator	
What is the company approach and philosophy regarding energy efficiency within the data centre, how does it prioritise against other factors?	▪ So where would you rank energy efficiency in terms of priorities for example against costs, reliability, etc. For example would you rather have a cheaper design that is less energy efficient, or a more energy efficient design that is more expensive?
How do you ensure that your staff are up to date with current issues?	▪ Is it part of a training programme for all/some of the staff, or is that not really needed? ▪ How do you find out about new technologies that are available?
What measures/steps has the company taken to increase energy efficiency or reduce energy consumption in your data centres?	▪ Can you give me some specific examples? ▪ Could you tell me more about how successful these have been, and have they been worth undertaking? ▪ Would you recommend using these methods to others? ▪ Have these been technological changes or more managerial e.g. training, management of IT etc. ▪ Do you know what the return on investment has been like for these improvements, have they been financially worth it?
Are there any measures that you plan to implement in the future?	▪ Could you give me any details of these? ▪ Why have they not been implemented yet? ▪ How did you find out about these measures?
Are there any measures which you wish to use but are not able to for any reason?	▪ Are there any particular reasons for this?
What incentives are there for you to increase the energy efficiency of your data centres?	▪ For example, economics, policy or marketing.
Could you tell me about any monitoring of energy use that you may carry out?	▪ What metrics do you use? ▪ How often are the energy levels assessed?
For colo – do your clients request energy efficiency when considering your services?	▪ Do you use energy efficiency as part of your sales package? ▪ Are clients interested in energy efficiency issues?
Consulting engineers	

Does energy efficiency play a key role in the design of a new data centre or a retrofit?	▪ How important would you say that energy efficiency is for your projects?
Do your clients actively seek out energy efficient designs, if not, then do you encourage such designs?	▪ Have you found that your clients have become more aware of energy efficiency issues over the past few years? ▪ When putting together a bid for a project, is energy efficiency a key issue?
Are there any incentives for you to encourage energy efficient design? E.g. more likely to win a project bid, company marketing, economics etc.	
How do you find out about new technologies that are available?	▪ How do you ensure that you remain up to date on the latest energy efficient technologies and designs.
Barriers to achieving higher energy efficiency	
Data centre operators	
Thinking about the measures or policies you may have put into place, did you encounter any difficulties?	▪ Given this, what would you say are the main barriers that have either prevented you from achieving higher energy efficiency, or you have had to overcome.
What have been the biggest issues in trying to consume less energy?	
Thinking about these issues, what do you think would help to overcome them?	▪ Would economic or policy incentives work?
Consulting engineers	
Have you encountered or do you encounter any difficulties in encouraging energy efficient design?	
Thinking about these issues, what do you think would help to overcome them?	
Awareness	
Data centre operators	

Are you aware of the EU code of conduct, and do you follow it?	▪ Have you noticed any improvements in energy usage as a result of following the CoC? ▪ What were the reasons for you to follow the CoC and have you noticed any other benefits, apart from any direct energy savings?
Do you follow any other schemes such as BREEAM?	▪ Again, have these had any other benefits to you apart from direct energy savings?
Are you aware of any other government initiatives or legislation about energy efficiency in data centres?	
Are you aware of the recent climate change agreement for the data centre industry?	
Consulting engineers	
Are you aware of the EU Code of Conduct and do you endorse it?	▪ Have you/your clients noticed any improvements in energy usage as a result of the CoC? ▪ Does endorsing the CoC have any direct benefits to you? ▪ Are your clients aware of the scheme and do you have many use it?
Do you actively follow the BREEAM scheme; is it something that client's request?	
Are you aware of any other government initiatives or legislation about energy efficiency in data centres?	
Are you aware of the recent climate change agreement for the data centre industry?	
<u>Liquid Cooling</u>	Thank you, I would now like to ask you a few questions on your opinions about the use specific technologies in the data centre. I wanted to find out your opinions about liquid cooling technologies,

	including for example rear door cooling units, liquid submersion cooling, cold plate cooling.
Data centre operators	
What do you know about these technologies?	▪ Do you use them? ▪ Do you know the benefits of using these technologies?
Consulting engineers	
What do you know about these technologies?	▪ Do you use them? ▪ Do you know the benefits of using these technologies?
Future	
Thinking about future and current demands, are there increasing requests for more HPC in the data centre?	
How do you see the issue of energy affecting this industry in the future?	
Is there anything else that you would like to comment on whilst we are still recording?	

End

(*end recording*)	Thank you again for taking part in this study, any further information about it then please ask. Your recording will now be taken back to the University of Leeds where it will be inscribed in order to be analysed along with the other participant's data.

Appendix B: Invitation Letter

UNIVERSITY OF LEEDS

Energy Research Institute

Faculty of Engineering

University of Leeds

LS2 9JT

Date

Dear ___________________,

You *(or company name)* are invited to participate in a research project being carried out at the University of Leeds entitled **'Investigating the Barriers to Energy Efficiency in Data Centres'**. This research is being conducted as part of a wider investigation into energy efficiency in data centres and is seeking opinions of and discussions with professionals within the data centre industry.

Please take time to carefully read this letter as it explains the research in more detail and describes what will be required should you decide to take part. Please contact us on the details provided at the end of this letter should you require any further details. We will be in touch shortly to enquire as to whether you have decided to participate, or you can contact us if you wish. **If you decide to participate then we would like to arrange a meeting with you so that we can carry out an interview**.

What is the purpose of this project and why have I been chosen?

As part of a wider PhD research project centred on the subject of energy efficiency in data centres, this study is focussed on gaining insight into the barriers to energy efficiency within the data centre industry. It is also concerned with finding out opinions about certain energy saving technologies that are available. The research will be carried out by conducting face-to-face or telephone interviews with professionals within the data centre industry. The project will be completed in October 2014, although interviews need to be completed by April 2014.

As a professional/company *(delete as appropriate)* within the data centre industry, you have been selected to take part in this research. We would be interested in meeting with you to have an interview to integrate your views into this research project.

Do I have to take part?

As a professional/company *(delete as appropriate)* within the data centre industry you/*company name (delete as appropriate)* were selected based upon the subject area of this research. Taking part in this research is entirely voluntary, and it is your decision whether or not to take part. If you decide to take part, then we would like to arrange an interview with you at which you will be requested to sign a consent form confirming your agreement to participate.

What will I be required to do as part of this research?

Should you be willing to take part, the research will be conducted through the means of a single semi-structured interview which could be done face-to-face either at your place of work or at the University of Leeds, or alternatively over the telephone. The interview will last approximately **45 minutes**, and will involve you answering pre-written questions by the researcher on the subject detailed above. There will be opportunities for in depth discussion. The interview will be recorded for the purposes of the research.

Will my taking part in this project be kept confidential and what will happen to the results of the research project?

Strict **confidentiality and anonymity will be maintained during the research**. All data collected from the interview will be kept in the strictest confidence; data will be stored securely on the University of Leeds system and only accessible by the researcher. Interviews will be kept anonymous so that you or your company name will not be able to be traced back to your interview responses.

The results of this research will be used in a PhD thesis, which is publishable work. Hopefully, there will be paper published in a scientific journal paper as a result of this research. Any published work which does arise from this research will not identify you in any way, all the data will be kept anonymous.

What are the possible benefits of taking part and why is the collection of this information relevant for achieving the research project's objectives?

Your participation will be highly valuable to the overall objectives of the research, as it involves engaging in discussion and gathering opinions from professionals within the data centre industry. Whilst there are no immediate benefits for you participating in this research, it will be a chance for you to **engage in university research and to open discussion about the issues behind energy efficiency in data centres**.

Who is organising and funding the research?

This research project is funded by the Engineering and Physical Sciences Research Council (EPSRC) and is part of a PhD project being undertaken in the Doctoral Training Centre for Low Carbon Technologies at the University of Leeds.

Contact for further information

You can contact the research team for further information or to inform us of your decision to participate on the following email address.

Email: cen5gab@leeds.ac.uk (Gemma Brady, PhD student)

Telephone: 0113 3432152 (Prof Nik Kapur)

Address: ERi, Energy Building, School of Process Environmental and Materials Engineering, Faculty of Engineering, University of Leeds, Leeds, LS2 9JT

Kind Regards,

Gemma Brady (PhD student) Prof Nik Kapur (Supervisor)

Appendix C: Example Interview Transcript

Time	
0:00.0	**START OF RECORDING**
- 1:32.8	***Interviewer:*** *If you could just give me a brief description of what the company does and the role it plays in the data centre industry.* **Participant:** Ok, *** is an SME company, we turn over about one and half million pounds at the moment and that turn over comes from two sources. We are primarily a data centre, a small bespoke data centre that is expanding. We also carry out some technology strategy board and European union funded research and development work. All of that research and development work is focussed on the idea that it will expand our data centre capacity. So we are very much a data centre company. We also participate in the data centre alliance so we are, we're interested in the way in which the industry moves forward and as a spin out company from the University of Liverpool we're quite focussed on the area of research and development and competitiveness through research and development. We currently employing I think 17 people on an FT basis, and we're also nurturing and sheltering a PhD, a PhD student who's working with us on a couple of projects for those next 3 years while they complete their PhD.
1:32.8 - 2:49.6	***I:*** *So could you tell me a bit about the data centre that you operate, is it something that you rent out to other companies?* **P**: we operate two small... ***I****: ...it's not something you use...* **P**: Two small data centres on two separate, in two separate facilities on a single large site. They're distinct and separate to give us a degree of resilience to the site. One is a very small facility with a capacity of around a quarter of a megawatt of capacity. The other one will eventually have a 1 megawatt IT load capacity and that is in a separate building downstairs. Both are modular data centres that can be expanded on. We are unlikely to expand the smaller one, the larger one is specifically designed to be built on a modular building block system. The nature of the business is that we are primarily a real and virtual hosting provider. So colocation whether that be by using by clients buying virtual machines, virtual space from us or by mounting their own IT with us.
2:49.6	***I****: could you give me an idea of kind of how many servers or the load that's in*

- 3:33.3 *there at the moment? if you're happy to do that.*

P: The load in the load in the smaller centre is about, the IT load fluctuates obviously but it's around, it averages out somewhere in the region of abut 90kw. The other one, the other unit is growing and is probably around the 60kw range at the moment, with a maximum capacity of 250kw in its present form.

I*: so that's still not at full capacity then, it's still expanding.*

P: Still not at full capacity.

3:33.3 ***I****: How long have they been in operation for, the two data centres?*

- 4:06.5 **P**: The smaller data centre has been in operation since the end of 2007, January 2008 realistically when it came on stream and the other one came fully on stream in December 2011.

I*: Ok.*

P: In fact we were still carrying out testing in January 2012, but you know that (laughter).

4:06.5 ***I****: so I just wanted to ask quickly what your role is here, just to get an idea.*

- 4:31.9 **P**: I'm a project director but I also manage operations so my role is both new project development in terms of the building and construction side and making sure the facilities run. But also I manage the research projects that we have recently taken on.

4:32.7 ***I****: So that was just kind of an introduction to what the company does. I was now*
- 6:20.3 *wanting to ask some questions about energy efficiency in your facilities. So what is the company approach and kind of philosophy towards energy efficiency in your facilities, how does that prioritise against other factors such as economics or security and stuff.*

P: Because we are a colocation facility, we are almost invariably driven by cost. the however is, we believe that one way of driving costs down, and a very effective way is make sure that we are energy efficient. So energy efficiency is very tied to the economics the way we run the business. at the time that we built the first data centre, it was built to a specification which probably made it one of the most efficient data centres in the UK at that point in 2007 and it's still being continuously tested, We're running it at about a true PUE, a lazy PUE if you like of 1.56. PUE may be a crude measurement of energy efficiency but if you do it as crudely as we do, it works as a comparator. The other unit has been set up to run a PUE of 1.4 from a very early stage, and it has been designed to do that. IT isn't as yet running at that low level because no matter what kind of efficiencies you can put into a data centre, there is a hump to get over of the parasitic loads and we're only just getting over that parasitic load.

6:20.3 - 7:41.3 ***I**: So what kind of measures and steps have you taken to try and increase your energy efficiency or reduce energy consumption?*

P: We specifically designed, in both cases, the centres were specifically designed to have highly efficient cooling systems, managed air flows the older system runs as you would probably gather older chillers whereas in the new one we've gone for much more control over things like pumps, fans speeds, diverting flows and we've gone for free cooling chillers as well. But we manage the air flow, the in room air flow very carefully as well in order to optimize the, in order to optimise the pickup. we carried out a piece of research in January 2012 to try to find out whether or not it was possible to optimize those units by taking early trigger indicators including current increase. But unfortunately that proved to be no more valuable than simply taking the temperature of the cooling medium which in this case is air.

7:41.2 - 8:01.2 ***I**: So you'd say that some of your efforts have been successful in reducing your energy consumption but..*

P: Some of them have been considerably successful, *[inaudible]* very successful and we have to remember the time when we built this at 1.56 the average across the industry was considerably over 2.

8:00.8 - 9:02.3 ***I**: So would you say that the efforts you've made to reduce your energy efficiency have been worth doing, worth doing in the long run?*

P: In the long run they have, but it obviously increases the cost in the short term we've tried to come to a compromise where we have done as much as is , as is competitively practical. There are other things that we could do but the costs would outweigh the benefits.

***I**: So...*

P: Sorry, if i can just illustrate that one, I'll give you a really careful illustration. We have gone for free cooling chillers, we could have increased the efficiency by going for adiabatic free cooling chillers. In our market in the north west quite a number of our customers and potential customers are public sector and there is a real concern about stored and released water, no matter what we say. so commercially its becomes a more difficult sell although it would give us an uplift in energy efficiency so that’s where we've tried to come to a compromise, so does that make sense.

***I**: yeh that makes sense*

9:02.2 - 10:12.4 ***I**: So have you have all these changes and things you've tried to do, are they all technological or have you introduced some managerial efforts as well like monitoring energy, measuring in, the kind of managing the data centre type side?*

P: Our efforts have been both technological and managerial. We do measure all of the energy at various points now. Including right down to server level in the individual racks. So we have a very clear idea about what's happening. We also do use a data centre information management system. We use an APC Structureware system and we monitor, monitor temperature. we monitor air pressure, we monitor air flow, we monitor, we monitor pretty much everything that goes on, we monitor humidity and so that in all those cases what we're trying to do it strike a balance to make sure that we're operating in the most economic and efficient way.

10:11.9 - 10:40.6 ***I**: So you said that it has been mostly worth doing it, have you seen a benefit to the business, have you seen a return on your investment?*

P: We're still not quite over the, on data centre one yes. In data centre two we're still not quite over the parasitic load hump, and therefore, we're just there now so we expect to see it...

***I**: You expect to see it ok,...*

P: In fact our figures, we've just been doing some graphs, we think that should be in, very quickly.

10:40.3 - 11:22.8 ***I**: Have you got any other measure that you plan to implement in the future as it expands?*

P: One of the things that annoys us as a research orientated organisation is that we don't like energy waste and no matter how efficient a data centre is, it wastes a lot of energy in current form. So we are, with the data centre alliance Leeds university and the University of East London, trying to persuade the EU to move in the direction of energy use optimization. In other words what can we do with the waste heat that data centres produce.

***I**: So is that something you thing you'll be able to do in the future? or is it just…*

P: …we plan to be part of it yes.

11:22.8 - 11:43.8 ***I**: Is there anything that you wish you could do here but you're not able to for any reason?*

P: That's one of our main thrusts, is that we'd like to find out an economic, an economic and an efficient way of optimizing the energy that comes into the data centre.

11:43.8 - 12:54.4 ***I**: So are there, for increasing your energy efficiency, are there any incentives for you as a company to do that?*

P: There's a really strong commercial driver, because our clients pay for the facility space and then pay the energy on top of that. The lower cost we can pass on for energy the better it is.

***I**: So when you have clients approaching you, to rent your spaces, do they request energy efficiency, is that something that they mention?*

P: They mention it, it's almost invariable second to cost.

***I**: So they don't really mind as much.*

P: I wouldn't say they don't really mind, but it's not the number one priority. Almost invariably, no matter what anyone else says, there are a range of factors including energy efficiency and including things such as security, business continuity. But the hard reality is that when all those things are taken into account, the priority will in fact be cost.

12:54.3 - 13:35.7 ***I**: So when you're talking about new ways to increase energy efficiency, how do you keep up to date with the current methods of reducing energy use, and how do you find out about these technologies?*

P: We participate in a range. We visit exhibitions and seminars in the UK and outside of the UK. We participate in research calls and research programmes and partners as I mentioned before, and we do try keep a reasonable contact with some of the leading figures in the industry which the data centre alliance helps us to do.

13:35.6 - 15:13.7 ***I**: so i just wanted to go back to the technologies that you've already used or measures that you've put in place. Do you find any difficulties in trying to use those or implement those at all?*

P: No.

***I**: So...*

P: I think there is a reason for that. We're trying to be if you like be at the cutting edge of technologies rather than the leading edge. There are other technologies which may on the surface look more efficient and then maybe more energy efficient by the time they're installed. But at the moment, because we're not playing with our own data, and we enter service level agreements, we're always trying to balance resilience against energy efficiency.

***I**: Ok so you wouldn't say that there is any kind of barriers in the way of you increasing energy efficiency in way, I'm thinking of economic ways here, costs…*

P: …cost is always a barrier! If we consider energy optimization rather than straight energy efficiency, there are barriers. There is a general belief in the industry that we can't do anything with low grade, with what's described as low grade heat. Although interestingly there is a ground source heat business which uses much lower grade heat and considers that there are uses for it. So our challenges are, what can we do with the water temperature. with water at the temperature that we have it or air, it doesn't matter which, but whichever medium we choose .

15:13.6 - 16:43.7 ***I****: so when you try to research this technology what kind of things do you thing would help you to overcome the barriers that are in the way, costs or policy?*

P: lottery win (laughter)

I*: Are there things you think would help, would it be a policy intervention that would come in to help you to use these technologies quicker or do you think it would be a financial incentive?*

P: I think the industry, we're perhaps odd so I'm to a certain extent the wrong person to ask that question to because the opposite side of the coin is if you spend next to nothing on energy efficiency, you can actually run inefficiently and still be competitive because you can use very cheap equipment and you can live in very cheap premises and and and (etc) and can be cost competitive in that way and some operators still do. I believe that in general that approach is, is dying through the industry and it so in terms of policy obviously CRC is an enormous stick. We're not a the level that it impacts on us yet, however at the point at which it does it will bite quite hard and we've already had to put plans in place in terms of how that particular policy will affect us.

16:43.7 - 18:16.8 ***I****: ok so I wanted to ask a couple of questions about the policies that are in place in the industry, so are you aware of the EU code of conduct?*

P: Yes broadly we follow it yes, when I say broadly we follow it. We've just submitted so we'll find out whether or not we've been accepted. But yes we're aware of it. We've actually been a little bit part of *[inaudible]*.

I*: Ok and why the reasons for you to sign up to it, was there incentive behind it to sign up or...*

P: The incentive is again commercial because it helps, it is I don't, how shall I put this, there are two sides, two benefits to sign up to any code of conduct or standard. The first is that you change your behaviour to meet that and to a certain extent we have changed our behaviour a little bit, we've actually changed our air flow even more than we had done before. So for example we've ensured that all of the blocking off is properly done to ensure that the air flow is efficient as it can be through the actual data hall. the other side of it is of course that if you're seen to be doing that then when I said before people ask for other things, they do ask for energy efficiency how you are treating this, what are your green policies. and again with the public sector it is a badge which helps us commercially.

18:16.8 - 18:44.1 ***I**: So you mentioned before though the CRC doesn't affect you yet, but it might do soon. Are you aware of the recent climate change agreement that the data centre industry has been...*

P: …yes, in fact we participated in both the survey...

***I**: …Oh did you...*

P: …and submitted views to it

***I**: So that might be something that affects you in the future then?*

P: Yes

18:44.1 - 21:12.3 ***I**: As a last thing i just wanted to ask what you thought about certain technologies, so one of the more energy efficient technologies is to use liquid cooling so bringing water into the data centres and the different methods you could use to do that so there's rear door cooling, there's liquid submersion and there's cold plate cooling as well, do you know much about these and do you use them at the minute?*

P: In our existing data centre we use in row chillers so we certainly bring liquid into, literally into the data centre. The pros and cons to doing that. As far as the cold plate idea is concerned, rear door technologies, there is very little difference in some respect between the rear door technology and the in row technology, shoving one on the side and one on the back so no we don't know anything about it at all *[laughter]*. We have, because of the nature of our business we have a pretty clear idea of the interest in the technologies out there. In fact we're sending our technical staff down to data centre world in February for a refresher.

***I**: So I know you mentioned before that people sometimes have an issue with that kind of system, where you bring lots of water in coz it's, they might see it as a risk. Have you found that before?*

P: We have to carefully explain how we manage, how the water is managed. And we haven't gone for that in the new data centre, we've gone for a straight although, I mean there is water involved but its' not in the data hall, it's kept separate from it. As far as the liquid cooling of componentry, submerging the circuit boards etc. in a benign liquid, the 3m solution. It would be interesting to see how that goes from a manufacturer's point of view. At some point we will still be rejecting, it's a more efficient way of carrying the heat. It's not necessarily a more efficient way of getting rid of it. It's in a way moving it from one place to another, you've still got to get rid of it to atmosphere. And it will be interesting to see what the costs, the eventual costs at manufacturer level of servers, racks etc. which can take advantage of that technology.

21:12.2 - 21:46.0 ***I**: Do you find that clients who come to you, are they interested the specific system that you use, do they want to know whether you're using any kind of liquid cooling, does that bother them?*

P: No, not in the least, as long as it is at a temperature, in fact one of the problems we do have is that in order to gain efficiency, we run our data centre, we're currently running our data centre at 26.5 degrees in the data centre. Quite deliberately, the complaint is that it's probably too warm whereas of course it's well within the ASHRAE limits.

21:45.9 - 22:10.9 ***I**: Yep. So i just wanted to think about the future now. So thinking about future and current demands, have you seen any increase in requests for high performance computing? Or is that not something that you…*

P: No we've seen exactly the same number of requests for high performance computing on a month by month basis.

22:10.8 - 25:38.7 ***I**: Ok so how do you think the issue of energy efficiency or energy usage is going to effect the industry in the future, do you think it's going to get more prevalent or not something that's going to increase in problem?*

P: That's a big question, a short question with a long answer. Ok I think we'll chop that into various different bits. One, energy is only going to get scarcer. It almost never, and certainly never for long periods gets cheaper. It has been cheaper in periods but by and large it becomes more expensive. Particularly in Europe when we started applying tariffs and green tariffs and climate change policies. So we confidently expect that energy will become more expensive so how does that affect data centres. Well we will also see i think a number of different technologies emerging around the server itself to make the server more efficient and emit less heat. so we're already seeing experimentation in that area from companies like arm bit abortive at the moment, we've been part of the trials of that the hard reality is that when we test it in order to get the same compute power, we're using about the same amount of energy *[laughter]* so what looks like an attractive future solution so far isn't there. There's also of course bio computers where we're going to change or move away from silicon based chips to some other system and *** announced that they are particularly interested in that but again we're in early stages of development. All of those things will eventually reduce the 1 if you like in the PUE it will reduce the prime requirement for energy, that of the it load. Because if they have less heat they use less energy in terms of fans which is bigger areas, they'll also use less in terms of the processors themselves so we can see that emerging. We also see higher density chips. Now whether Jevons law takes over and we discover that everything we do means that we just use more and use the same amount of energy and therefore data centres don't need to. Is something that's kind of impossible to predict. One thing that hasn't happened to the extent that we thought it would is we believe that the density, the rack density is going to rise over the last 5 years. We have next to no evidence that that's happening. In fact the opposite is pretty much staying stable, that's not fair, it is creeping up but it is creeping up much more slowly than i think both manufacturers and the industry expected. so if you've attended symposiums some 5 years ago we were talking about blades, we were talking about increased rack density and that was the way to efficiency we were moving up so that a 4kw rack was going to become very much old hat. And we were moving towards 8, 10, 12 up to 20 kw racks, the hard reality is not many people are asking for that at all.

I*: so it' not going quite the way that people predicted?*

P: certainly not in the colo market.

25:38.7 - 25:50.7 ***I****: So I think that's about it from my questions, unless you have anything else you want to add?*

P: No.

I*: Nope ok, thank you*

END OF RECORDING