Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Modelling and Control of Cooling Systems for Data Center Applications

Ph.D. candidate
Michele Lionello

Advisor
prof. Alessandro Beghi

Director & Coordinator
prof. Andrea Neviani

Ph.D. School in
Information Engineering

Department of
Information Engineering
University of Padova
2019

University of Padova
Department of Information Engineering

Università
degli Studi
di Padova

Ph.D course in: Information Engineering
Curriculum: Information Science and Technology
Cycle: XXXII

Modelling and Control of Cooling Systems for Data Center Applications

Thesis written with the financial contribution of Vertiv

Director: Prof. Andrea Neviani
Advisor: Prof. Alessandro Beghi

Ph.D. candidate: Michele Lionello

Year 2019

Abstract

Nowadays, the Data Center industry is playing a leading role in the world economic development and it is growing rapidly and constantly. Beside this, it has become more concerned with energy consumption and the associated environmental effects. Since about half of the total energy consumption in a typical Data Center is devoted to cooling the IT equipment, energy efficiency must be the primary focus in the design and management of the cooling infrastructure.

In this Thesis, we consider the problem of optimizing the operation of cooling systems in Data Centers. The main objective is that of maximizing the energy efficiency of the systems, while provisioning the required cooling demand. For this purpose, we propose a two-layer hierarchical control approach, where a supervisory high-level layer determines the optimal set-points for the local low-level controllers. The supervisory layer exploits an Extremum Seeking model-free optimization algorithm, which ensures flexibility and robustness against changes in the operating conditions. In particular, a Newton-like Phasor-based Extremum Seeking scheme is presented to improve the convergence properties and the robustness of the algorithm.

The proposed control architecture is tested in silico in optimizing the operation of an Indirect Evaporative Cooling system and a Liquid Immersion Cooling unit. Simulations are performed by exploiting First-Principle Data-Driven models of the considered systems and the results demonstrate the effectiveness of the proposed approach.

Sommario

Al giorno d'oggi, l'industria dei Data Center ricopre un ruolo fondamentale nello sviluppo dell'economia mondiale e sta crescendo in maniera rapida e costante. Peraltro, la crescente sensibilità verso tematiche di risparmio energetico, dei cambiamenti climatici, e di sviluppo sostenibile, coinvolge anche le nuove politiche industriali. Ad esempio, se si tiene conto che circa la metà dell'energia tipicamente consumata in un Data Center viene utilizzata per il raffreddamento delle apparecchiature informatiche, l'efficienza energetica deve essere considerata l'obiettivo primario nelle fasi di progettazione e di gestione dell'infrastruttura di raffreddamento.

In questa Tesi, si considera il problema di ottimizzare il funzionamento dei sistemi di raffreddamento nei Data Center. L'obiettivo principale è quello di massimizzare l'efficienza energetica di tali sistemi, fornendo al contempo la capacità frigorifera necessaria a raffreddare le apparecchiature e i server. Specificatamente, viene proposto un approccio di controllo gerarchico basato su due livelli, in cui un supervisore determina i set-point ottimali per i controllori locali di basso livello. Il livello di supervisione sfrutta un algoritmo di ottimizzazione Extremum Seeking non basato su modello, che garantisce flessibilità e robustezza al variare delle condizioni operative. In particolare, viene presentato uno schema Newton-like Phasor-based che permette di migliorare la convergenza e la robustezza dell'algoritmo.

Le prestazioni dell'architettura sono state verificate in silico per ottimizzare il funzionamento di un sistema di raffreddamento evaporativo indiretto e di un'unità di raffreddamento a immersione in liquido. Le simulazioni vengono eseguite sfruttando modelli First-Principle Data-Driven dei sistemi considerati e i risultati confermano l'efficacia dell'approccio proposto.

Acknowledgments

First, I want to express my deepest gratitude to my advisor Prof. Alessandro Beghi for the continuous support, encouragement, and excellent advice during my PhD studies.

I am particularly indebted to my co-advisor Mirco Rampazzo for being a constant presence and an endless source of support and advice. I could not have imaged having a better mentor throughout my PhD.

My sincere thanks go to my Scholarship Sponsor Vertiv S.r.l. for financing my studies. In particular, I thank Giuseppe Dalla Mana, Tommaso Venturelli, and Gabriele Menegazzo.

I want to thank Prof. Damiano Varagnolo for hosting me in his group at the Luleå University of Technology and for his encouragement and support. I would like to thank also Riccardo Lucchese and Khalid Atta for all the long and stimulating discussions and the ideas shared.

I would like to express my gratitude to all the staff at RISE SICS North for giving me the access to their laboratories and for their time and technical expertise.

I want to thank my PhD colleagues for making the Department such a pleasant working environment. Special thanks go also to all my friends in Montegrotto for all the unforgettable moments spent together during these years.

A special thank goes to Alessandra for constantly encouraging and supporting me during the writing of this Thesis. Thank you for always being by my side.

Last but not least, I want to thank my family for their unconditional love and for all the support throughout my years of study.

Contents

Introduction

Data Centers are computing infrastructure facilities that host large amounts of Information Technology (IT) equipment used to process, store, and transmit digital information. In recent years, the crucial role of the telecommunications industry in almost every sector of society is leading to a significant growth in Data Centers number and power densities. As a consequence, Data Centers are becoming more and more energy intensive and their energy efficiency is attaining fundamental importance. The power consumption of Data Center enterprises on a global scale is estimated around 120GW, which is approximately 2% of the world energy consumption, ABB (2013).

Most of the electrical power consumed by the IT equipment in Data Centers ultimately transforms into heat, which must be removed by the cooling systems to prevent malfunction and breakage of hardware. As a result, nearly 50% of the power supplied to a typical Data Center is spent on cooling, Dayarathna et al. (2015). Hence, the improvement of cooling systems efficiency represents a great opportunity for energy and cost savings in Data Centers. Adopting the best cooling method according to the size and location of the Data Center is a key opportunity to reduce the amount of energy consumed for cooling. Currently, most of existing Data Centers rely on cooling systems which use air as a cooling medium to remove the heat produced by the servers. In typical setups servers are arranged in racks so that cooling air first enters the room into a so-called cold aisle placed in front of the racks, then passes through the servers, and exits on their back into a so called hot aisle. Here, hot air rises and moves to a Computer Room Air Conditioning (CRAC) unit, where air is cooled and recirculated. Typically, CRAC units are based on the vapour compression cycle, where compressors are extremely energy-hungry. However, if the environmental conditions are favourable, the integration of air free cooling offers opportunities to drastically reduce the electrical power consumed by the cooling systems. Moreover, the free cooling technology can enhance even more its cooling capabilities by leveraging on the evaporative and adiabatic cooling principles. A prominent alternative to air-based cooling methods is given by liquid cooling solutions, since liquids used in cooling have generally higher heat transport capacity than air. Along

with the significant thermal advantage, there are some other benefits of liquid-cooled solutions. Among the others, with the high heat capacity of liquids, liquid-cooled Data Centers are able to harvest the waste heat more efficiently facilitating the transformation of Data Centers into active players in the local heat network.

Adopting an appropriate cooling architecture is not enough to establish an efficient cooling system without effective cooling controls. Traditional control approaches consider static policies tuned against nominal conditions or based on knowledge or intuition of Data Center operators. However, cooling systems exhibit a wide variability of operating conditions, numerous technological constraints, and an intrinsically complex nature of their underlying physics, which make traditional control approaches not effective and require the design of more advanced control systems. Within this context, modelling and simulation tools provide invaluable support for the analysis of the cooling system and the design and test of its control. The use of Computer Aided Control Systems Design (CACSD) software tools is generally cheaper and safer than conducting real-world experiments. Simulations can be conducted faster than real time and they can be even more realistic than traditional experiments, as they allow the free configuration of environment parameters found in the operational application field of the final product.

In the latest literature various works focus on Data Centers modelling and control. For example, in Dayarathna et al. (2015) a comprehensive taxonomy of Data Centers power modelling and prediction techniques are presented. In Parolini et al. (2010a) a control-oriented model of Data Centers, including the coupling in the dynamics between the computational and cooling resources, is depicted. The research presented in Parolini et al. (2010b) shows a control strategy for realizing best trade-off between satisfying user requests and energy consumption. In Ogawa et al. (2015) a Model Predictive Control (MPC) approach is used to develop a cooling control system for Data Centers which utilizes indirect fresh air. In Simonazzi et al. (2018) a control-oriented model is proposed for the identification of different air mixing regimes at the computer room level in general room configuration scenarios. In Lucchese and Johansson (2019), a network oriented formalism to develop models of Data Center air-cooled equipment is used and an airflow provisioning controller that balance the cost of the server fan power absorption with that of temperature-dependent leakage effects in the electronics is proposed.

In this Thesis[1] a hierarchical control approach is considered, with the overall objective of optimizing the energy efficiency of Data Center cooling systems. More specifically, we consider that cooling systems are managed by means of a hierarchical two-layers control

[1]This Thesis work is carried out within an industrial partnership with Vertiv S.r.l., Piove di Sacco (PD), Italy, a global leader in designing, building and servicing critical infrastructure that enables vital applications for Data Centers, communication networks, and commercial and industrial facilities.

scheme, which consists in:

- a supervisor (high layer), which specifies the set-points for some of the main variables of the targeted process (e.g., temperatures and flow rates);
- local controllers (low layer), which operate the system controllable devices (e.g., fans, pumps, compressors) in order maintain the set-points specified by the supervisor.

The supervisor is the focus of this work; the aim is to determine the set-points for the local controllers that optimize the efficiency of the system, while provisioning the needed amount of cooling power. To cope with this problem, we exploit Extremum Seeking (ES) optimization algorithms, which target the problem of maximizing (or minimizing) the steady-state value of a performance (or cost) index via real-time adaptation of one or more process inputs. The main benefit of this class of methods is that the optimization is based solely on feedback from online measurements, which ensures flexibility and robustness against changes in the operating conditions. More precisely, the ES method exploits an estimation of first-order information obtained by filtering the online direct or indirect measurements of the targeted index in response to the perturbation of one or more control inputs. In contrast to model-based optimization schemes, ES do not base its decisions according to predictions provided by mathematical models which reproduce the behaviour of the plant. Thus, ES method can be applied also in complex processes where deriving accurate models may be even more difficult than the control system design. For example, when some system parameters are not completely known or the system structure or disturbances change over time. To improve even more the convergence properties and the robustness of the ES method, we propose a Newton-like Phasor-based ES scheme, which is able to estimate and exploit also second order information.

By exploiting the flexibility of the considered control architecture, we applied it to optimize the operation of different cooling units and we tested its effectiveness by means of simulations. For example, we considered the problem of minimizing the electrical power consumed by an Indirect Evaporative Cooling (IEC) unit during its operation. Moreover, we targeted the problem of optimizing the heat-recovery capabilities of a Liquid Immersion Cooling (LIC) system within heat-reuse scenarios. In both cases, the effectiveness of the control approach has been tested in silico by exploiting simulation environments based on First-Principle Data-Driven (FPDD) mathematical models of the considered systems.

Main Contributions

As regards the problem of designing advanced control systems for optimizing the operation of the Data Centers cooling units, the contributions are twofold. First, a flexible hierarchical control architecture with a supervisory layer based on the Extremum Seeking model-free optimization technique is developed. Also, the performance of the considered optimization algorithm have been improved by proposing a Newton-like Phasor Extremum Seeking method. Second, we provide simulation environments based on First-Principle Data-Driven models of different cooling units and we assessed the effectiveness of the proposed control strategy by means of simulations. This work achieved the following papers:

- M. Rampazzo, M. Lionello, A. Beghi, E. Sisti, and L. Cecchinato. *A static moving boundary modelling approach for simulation of indirect evaporative free cooling systems*. Applied Energy, 250, 1719-1728. 2019.

- M. Lionello, M. Rampazzo, A. Beghi, D. Varagnolo, and M. Vesterlund. *Energy-based modelling and simulation of liquid immersion cooling systems*. International Conference on Computational Heat, Mass, and Momentum Transfer (ICCHMT). 2019.

- M. Lionello, M. Rampazzo, A. Beghi, D. Varagnolo, and M. Vesterlund. *Lumped-parameters Control-oriented Gray-box Modelling of Liquid Immersion Cooling Systems*. 18th IEEE European Control Conference (ECC). 2019.

- A. Beghi, M. Lionello, and M. Rampazzo, *Efficient Operation of Indirect Evaporative Data Center Cooling Systems via Newton-like Extremum-Seeking Control*. IEEE Conference on Control Technology and Applications (CCTA). 2019.

- R. Lucchese, M. Lionello, M. Rampazzo, M. Guay, and K. Atta. *Gradient-free optimization of data center cooling using Extremum Seeking*. Nordic Process Control Workshop. 2019.

- R. Lucchese, M. Lionello, M. Rampazzo, M. Guay, and K. Atta. *Newton-like phasor extremum seeking control with application to cooling data centers*. 11th IFAC Symposium on Nonlinear Control Systems (NOLCOS). 2019.

- Rampazzo M., Lionello M., Carpignani Panebianco F., and Beghi A. *Model-Free Control of Data Center Compressor-Based Cooling Systems*. 4th IEEE International Forum on Research and Technology for Society and Industry (RTSI). 2018.

- Beghi A., Dalla Mana G., Lionello M., and Rampazzo M. *Data-Driven Supervisory Control of Indirect Adiabatic Cooling Systems*. IEEE Conference on Control Technology and Applications (CCTA). 2018.
- Beghi A., Cecchinato L., Dalla Mana G., Lionello M., Rampazzo M., and Sisti E. *Modelling and control of a free cooling system for data centers*. Energy Procedia, 140, 447-457. 2017.
- Beghi A., Dalla Mana G., Lionello M., Rampazzo M., and Sisti E. *Energy-efficient operation of an indirect adiabatic cooling system for data centers*. IEEE American Control Conference (ACC). 2017.

Thesis Outline

In Chapter 1, a brief overview of Data Centers is provided. In particular, the typical structure and the main components of a Data Center are described and an analysis of the electrical power consumption within a typical Data Center is reported.

In Chapter 2, the main challenges and the state of the art of the thermally efficient design and management of Data Centers are discussed.

In Chapter 3, the problem of designing a hierarchical control architecture for optimizing the operation of Data Centers cooling sytems is faced by means of a Computer Aided Control System Design (CACSD) workflow.

Chapter 4 is dedicated to the Extremum Seeking Control (ESC) optimization method, and the novel Newton-like ESC scheme is presented.

In Chapter 5, the problem of optimizing the cooling efficiency of an Indirect Evaporative Cooling unit is considered. A simulation environment based on a First-Principle Data-Driven model of the system is presented and the effectiveness of the proposed hierarchical control architecture have been tested by means of simulations.

In Chapter 6, we consider the problem of optimizing the heat-recovery capabilities of a Liquid Immersion Cooling unit. A First-Principle Data-Driven model of the system is presented and a simulation environment has been designed accordingly. The effectiveness of the proposed hierarchical control architecture has been tested by means of simulations.

1 Data Centers

Data Centers are the supporting backbone of the telecommunications industry's infrastructure. The number and size of Data Centers have grown significantly in the past years and so they are becoming more and more energy-intensive. This Chapter provides a brief overview of Data Centers. In particular, next sections describe the typical structure and the main components of a Data Center and report an analysis of the electrical power consumption within a typical Data Center.

1.1 Introduction

Data Centers are facilities hosting a large number of servers dedicated to process, store, and transmit digital information. The crucial role of the telecommunications industry in almost every sector of society is leading to significant growth in the number and size of Data Centers in recent years. This trend is driven, for example, by electronic commerce, cloud computing, and other Internet-based applications such as social media platforms. Moreover, the Internet of Things (IoT), which interconnects remote assets and pushes massive amounts of data to centralized sites, will further increase the need for large-scale Data Centers.

According to Cisco, the annual global IP traffic first exceeded that of one zettabytes (10^{21} bytes, or ZB) in 2016, starting the Zettabyte Era, Barnett Jr (2016). Moreover, it

is expected to grow to 396 exabytes (EB) per month by 2022, up from 122 EB per month in 2017, Fig. 1.1. This huge amount of data needs to be collected and processed by Data

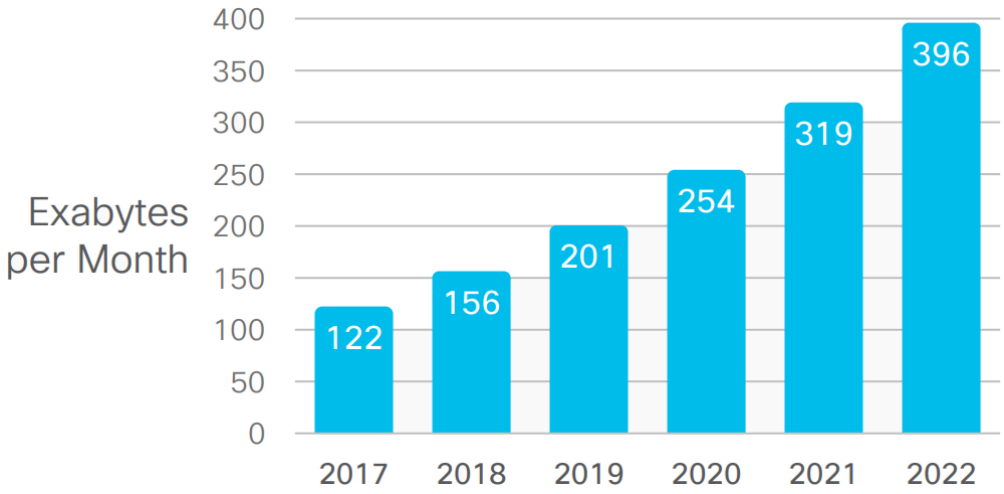


Figure 1.1: Forecast of the annual global IP traffic from 2017 to 2022, Cisco (2019).

Center infrastructures. Such processing requires highly scalable, high-density servers that are optimized for running massive, parallel, computationally-intensive workloads and algorithms. This necessitates a huge and ever increasing demand of energy, an obvious resource in any computation. As a result, in addition to dimensions and processes attaining close to atomic scale limits, the energy efficiency of computing has become one of the most important challenges. As the performance improvement is dictated by the Moore's law, Moore et al. (1965), the trend of the energy efficiency of computing hardware is described by the Koomey's law. According to Koomey, Koomey et al. (2010), the computations per kW h doubled every 1.57 years until 2009, but today it takes almost 3 years for peak-output efficiency to double. One theoretical bound for the energy efficiency of computing systems is dictated by the Landauer's principle, Landauer (1961), which states that the lower limit for energy consumed in erasing one bit of information can be expressed in the following manner:

$$L_l = kT \ln 2, \tag{1.1}$$

where k is the Boltzmann constant and T is the temperature of the system in Kelvin. However, Landauer's limit should not disturb Koomey's law before the 2060s, as shown in Lääkkölä et al. (2015). As most of the energy consumed by processors ultimately transforms into heat, a significant amount of power needs to be used for cooling the Information Technology (IT) devices in Data Centers to prevent malfunction and breakage

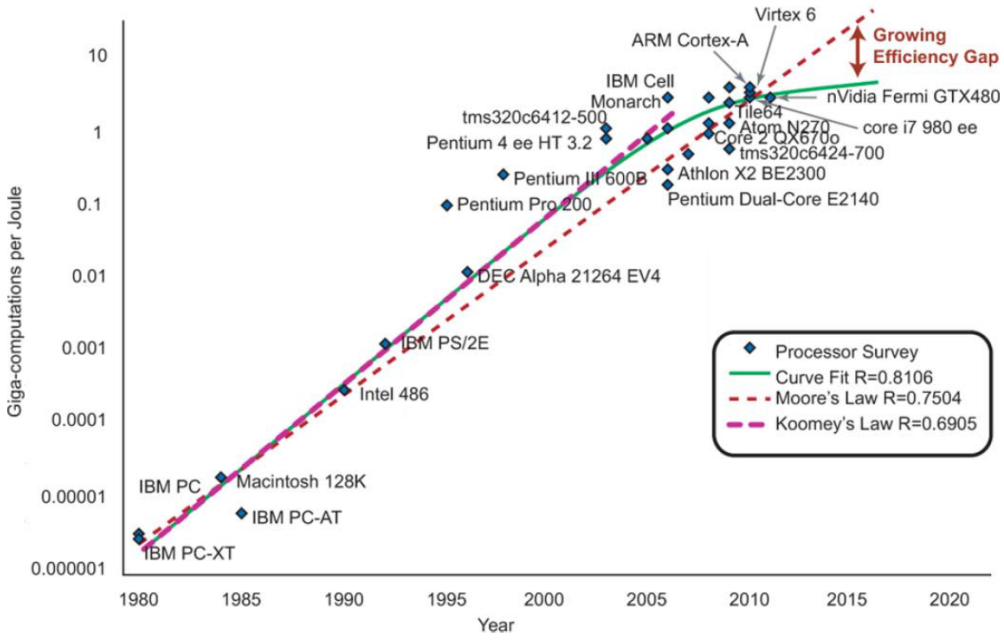


Figure 1.2: Energy efficiency versus year for commercial digital processors. Note that Koomey's law, which assumes operations per energy increases at a faster rate than the Moore's law rate, no longer holds as of about 2009 and operations per unit energy starts to fall off creating a larger gap between practice and theory.

of hardware. Thus, the improvement of the energy efficiency of the IT devices must be combined with an energy efficient management of the cooling infrastructure, which is the main focus of this Thesis.

In this first Chapter, we report an overview of the structure and an analysis of the energy consumption of a typical Data Center. Fig. 1.3 shows the structure of a generic Data Center: it includes the IT systems and a supporting infrastructure. The IT systems are the main elements within a Data Center since they supply the main functions that should be provided by the Data Center itself. They include servers, storage devices, and network hardware. The supporting infrastructure consists of all the ancillary equipment needed to guarantee the safety operation of the IT devices, for example:

- the power distribution system, that must ensure continuity and stability of the electrical power supplied to the Data Center;
- the cooling infrastructure, that must remove all the heat produced by the IT equipment.

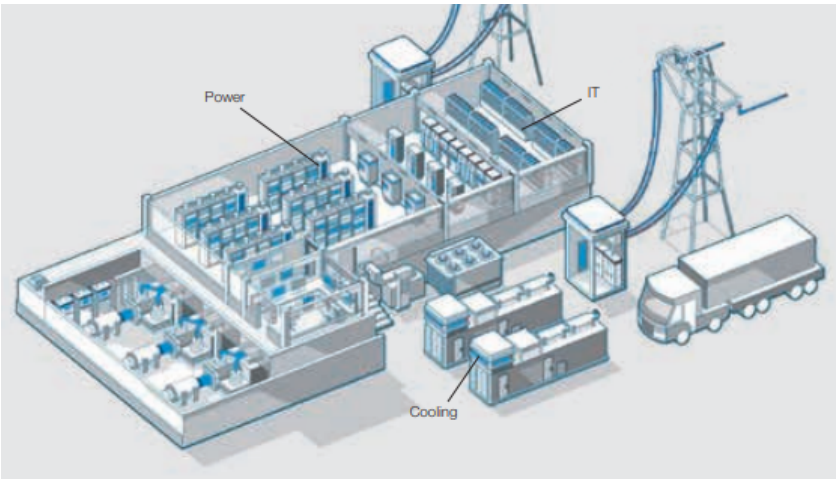


Figure 1.3: Physical layout of a generic Data Center, ABB (2013). The IT equipment includes servers, storage devices, and network hardware; the power system provides continuous and stable electrical power; the cooling infrastructure removes the heat produced by the IT devices.

1.2 The IT Equipment

The IT equipment is the heart of a Data Center since it performs the main tasks that should be provided by the Data Center itself. The IT infrastructure mainly comprises:

- servers, which manage the digital information;
- memory devices, that are used for storing data;
- network hardware, which is devoted to transmitting the data to the network.

To reduce the risk of data losses and communication faults as much as possible, the redundancy of all these components is crucial. It is also worth pointing out that the choice of the IT devices should not take into account only the performance (in terms of computational speed), but also energy efficiency. For example, efficient servers produce less heat that must be removed by the cooling infrastructure and require less electrical power.

1.3 The Power Distribution System

The power distribution system consists of a hierarchy of electrical devices that transmit electrical power from a utility feed to the Data Center equipment, Yeh and Manjrekar (2007). One or more feeds arrive from the electrical utility before or after their voltages are transformed to usable levels. Among other purposes, switchgear provides a disconnect

point from the utility feed when needed. Typically, the utility supplies a Medium Voltage (MV) to the Data Center, where it is stepped down to Low Voltage (LV) by a transformer. Then, the electrical distribution equipment distributes the LV power to the different electrical loads such as IT devices, cooling systems, and lighting.

The heart of the power distribution system is the Uninterruptible Power Supply (UPS), which guarantees protection from disturbances and power discontinuities. The UPS system is a backup source of power, installed between the central power supply and the critical end loads to provide two functions. First, it provides a secure power source when the main Alternating Current (AC) power supply fails. Second, it provides a clean, stable, and regulated power supply when the mains supply is present. An ideal UPS should be able to deliver uninterrupted power and, at the same time, to provide the necessary power conditioning for critical loads. UPS systems can provide power for up to several minutes. However, a standby power system is needed for more prolonged outages, Nasiri et al. (2008). Fig. 1.4 shows an example of a UPS unit produced by Vertiv.

Figure 1.4: An example of UPS unit produced by Vertiv.

Even a momentarily loss of power can cause IT systems and sensitive equipment to crash and lose data in Data Centers. A UPS maintains power by switching instantaneously to batteries in the event of utility power failure. They also condition the power supply to reduce unwanted spikes and harmonics. Different types of UPS have been designed to correct a variety of problems, such as:

- power supply failure: the total loss of power;
- sags: transient under-voltage;
- brownouts: under-voltage for a more extended period;
- spikes: very brief but high energy burst which may cause data loss and may damage sensitive solid-state components;
- surges: relatively short duration high voltage power surges;
- line noise: distortions that are caused by electromagnetic interference;
- frequency variation: deviation from the nominal frequency that causes motors to increase or decrease speed;
- switching transient: instantaneous under voltage, which may result in data loss and components stress;
- harmonic distortion: multiples of power frequency superimposed on the power waveform, which causes excess heating in wiring and fuses.

UPS technology comes in two types: static or rotary UPS. However, each of those classifications contains several types of technology and significant differences in what the technology provides. The most common UPS systems deployed in Data Centers are either double-conversion static UPS systems or Flywheel Rotary UPS systems. Further details of the two types of UPS can be found, for example, in Milad (2017).

1.4 The Cooling Infrastructure

The role of the cooling infrastructure is to transfer heat from the Data Center indoor environment to the outside, to avoid overheating that may lead to IT systems shutdown or permanent failure. Traditionally, there are two primary ways of cooling a Data Center: air-based cooling and liquid-based cooling, Dai et al. (2016).

1.4.1 Air-based Cooling

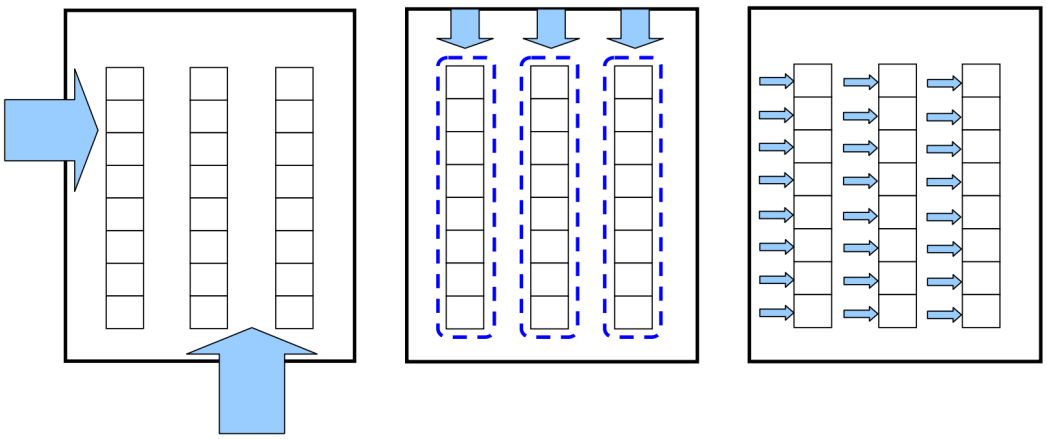
Currently, most of existing Data Centers use air-based cooling systems to maintain the desired thermo-hygrometric conditions in the indoor environment. The IT devices are organized in rows of racks located within a computer room. A cooling unit is used to supply cold air, which is distributed to the IT devices. The cooling air absorbs heat from

the IT equipment, and then it is recirculated back to the cooling unit. Successful thermal management of air-cooled Data Centers lies in the provision of an appropriate airflow rate at an appropriate temperature for all the IT equipment and under all operating conditions. Poor thermal management of the Data Centers cooling infrastructure may have several adverse effects, which result in a substantial increase in the operating cost, Khalaj and Halgamuge (2017). Examples are the early failure of servers due to the poor distribution of the airflow, increased downtime, and unstable reliability. In designing an air-based cooling architecture, the following three aspects are of fundamental importance:

- the location of the cooling unit that supplies cold air to the IT devices;
- the design of the air distribution system that must ensure the correct air circulation;
- the choice of the heat removal method that is exploited by the cooling unit to supply air at an appropriate temperature and humidity.

1.4.1.1 Room, row, and rack-based cooling architectures

Depending on the power density within the computer room, three different approaches can be considered for locating the cooling unit: room, row, and rack-based architectures. In the three approaches, the cooling units are associated with the room, rows, and individual racks, respectively. Fig. 1.5 depicts the three architectures. Black squares



(a) Room-based. **(b)** Row-based. **(c)** Rack-based.

Figure 1.5: The room, row, and rack based architectures: black squares represent racks arranged in rows, and the arrows represent the logical association of the cooling units to the loads in the IT racks.

represent racks arranged in rows, and the arrows denote the logical association of the

cooling units to the loads in the IT racks. The room-based configuration is usually considered in the case of power density up to 3-4 kW/rack. Conversely, the other two architectures are typically used if the power density exceeds 5-6 kW/rack. Further details can be found, for example, in Dunlap and Rasmussen (2006).

1.4.1.2 Airflow distribution

In the room-oriented configuration, an appropriate distribution of the air within the computer room is crucial. There are three basic approaches to distribute the air in the computer room: flooded, targeted, and contained. In a flooded air distribution system, the only constraints to the airflow are the walls, ceiling, and floor of the room. This approach leads to heavy mixing of the hot and cold airflows. In a targeted air distribution system, a mechanism (e.g., duct, perforated tile, cooling unit placed within IT rows) directs the supply and return airflows close to the IT equipment intake and exhaust. In a contained air distribution system, the IT equipment inlet and outlet airflows are entirely enclosed to eliminate air mixing between the supply and the return air streams. Each of the three approaches can be used in either the supply air path or the return air path. This results in 9 possible combinations, or types, of air distribution. All of these types have been used in various circumstances and, occasionally, different types are mixed in the same Data Center. Rasmussen (2012) provides a detailed analysis of the advantages and disadvantages of all these possible configurations. Typically, the targeted approach is used for both the air sides. In this configuration, the rows of racks are partitioned by aisles that are either cold air intake or hot air exhaust. The chilled supply air provided by the cooling unit enters through an under-floor plenum into the cold aisles in front of each rack. The chilled air blows into the racks and absorbs the heat generated by the IT devices. Then, the resulting warm air exits on the back of the racks to the hot aisles, it rises to the ceiling, and then it is gathered and recirculated to the cooling unit. Fig. 1.6 reports a schematic representation of the typical hot/cold aisles configuration. The primary cooling inefficiency occurring for this raised floor configuration is the hot air recirculation from hot aisle into the cold aisle. A possible solution to this inefficiency is to consider a contained return airflow, that, on the other hand, would be more difficult to be designed.

1.4.1.3 Principal air-based cooling methods

There exist several different heat removal methods that can be used by the cooling units to cool the IT environment and transport the unwanted thermal energy to the outdoor. Examples of different cooling technologies are discussed in Evans (2012). The purpose of

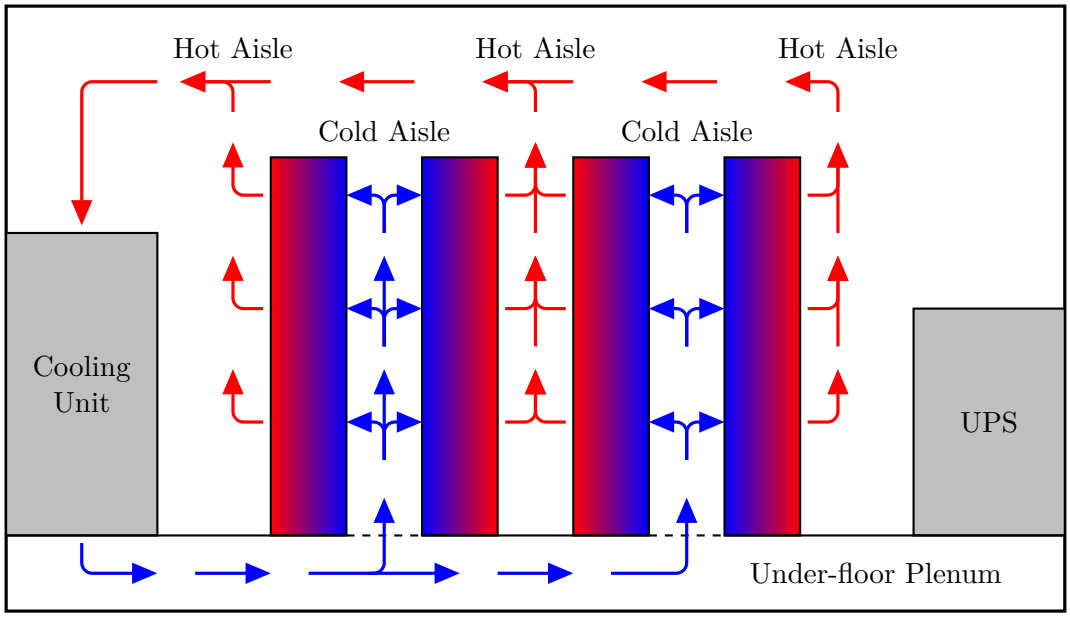


Figure 1.6: The hot/cold aisles configuration: the chilled supply air enters into the cold aisles through an under-floor plenum and absorbs the heat generated by the IT devices, and then it exits on the back of the racks into the hot aisles where it rises to the ceiling and is recirculated to the cooling unit.

all these cooling methods is to transfer heat from an indoor point (i.e., the computer room) to an outdoor point. This energy transport is generally accomplished by using an intermediate fluid (liquid or gas) that carries the thermal energy between the two points. Typically, a heat exchanger is used to transfer heat from the computer room air to the intermediate fluid.

A common method consists in combining a Computer Room Air Handler (CRAH) with a chiller. This combination is generally known as a chilled water system. This solution uses chilled water as an intermediate medium, which is pumped in pipes from the chiller to the CRAH units located in the computer room. They draw warm air from the IT environment through chilled water coils filled with the circulating cooling water. The heat removed from the computer room flows out with the chilled water exiting the CRAH units and returning to the chiller. Then, this removes the heat from the warmer chilled water and transfers it to another stream of circulating water called condenser water. This stream flows through a device known as a cooling tower, which rejects the heat from the condenser water to the outdoor environment. More specifically, the warm condenser water is sprayed onto sponge-like material (called fill) at the top of the tower. The water spreads out, and some of it evaporates away as it drips and flows to the bottom of the cooling tower. The small amount of water that evaporates from the cooling tower

serves to lower the temperature of the remaining water. The cooler water at the bottom of the tower is collected and sent back into the condenser water loop via a pump package. Typically, the cooling tower is located on the top of the Data Center building. Fig. 1.7 reports a schematic representation of a typical chilled water system. To prevent the

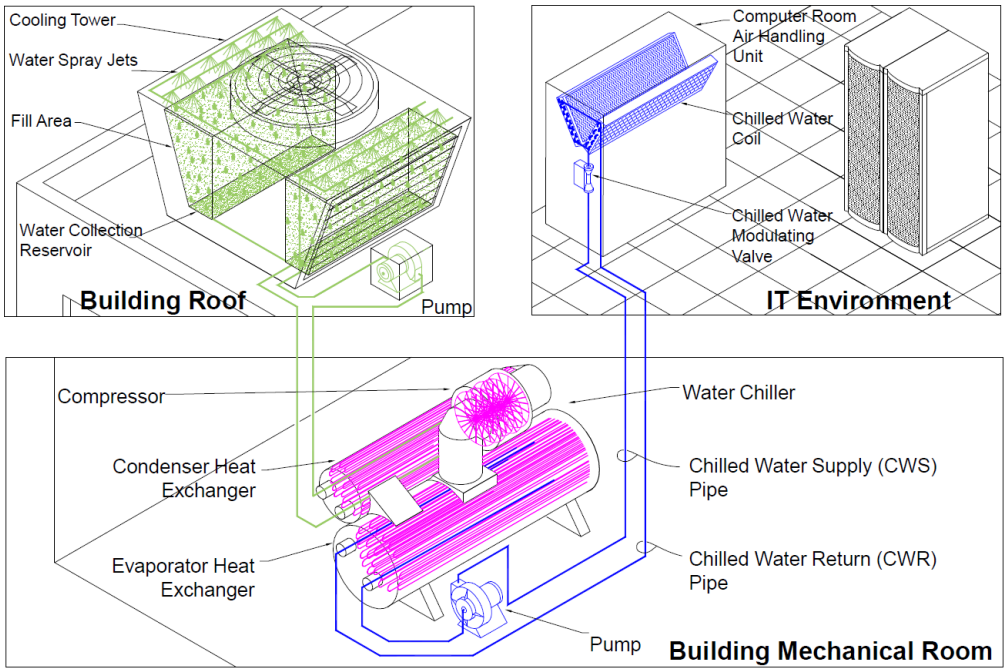


Figure 1.7: Schematic representation of a chilled water system, Evans (2012).

risk of damage to the IT equipment due to water leaks, the cooling method described above can also include a pumped refrigerant heat exchanger. Typically these systems use a refrigerant[1] or other non-conductive fluids. Chilled water is pumped in pipes from the chiller to a heat exchanger which transfer heat from the pumped refrigerant to the cooling water. The colder refrigerant then returns to the cooling unit to absorb more heat and returns again to the heat exchanger.

Another widely used technology is obtained by joining an air-cooled Computer Room Air Conditioning (CRAC) together with a condenser. This combination is generally known as an air-cooled CRAC Direct Expansion (DX) system[2]. In this solution, half of the components of the refrigeration cycle are in the CRAC unit located in the computer room, and the rest are in the air-cooled condenser located outdoor. Refrigerant circulates between the indoor and outdoor components in pipes called refrigerant lines. The heat

[1]The two main refrigerants used in Data Center cooling systems today are R-410A, used in small to medium-sized Data Centers, and R-134a, used in larger systems.

[2]Any system that uses a refrigerant and an evaporator coil can be called a DX system.

from the IT environment is transferred to the outdoor using this circulating flow of refrigerant. Fig. 1.8 shows a schematic representation of an air-cooled CRAC DX system.

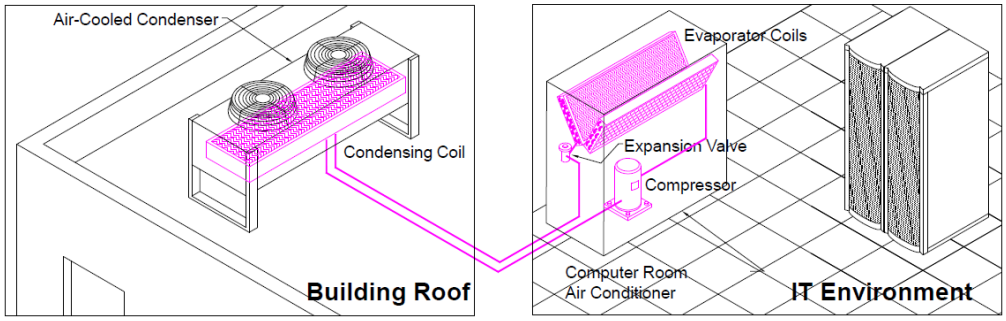


Figure 1.8: Schematic representation of an air-cooled CRAC DX system, Evans (2012).

In many climates, the costs of cooling Data Centers can be reduced by taking advantage of external environmental conditions to cool the IT equipment. Under the right conditions, these technologies can save so much in electrical costs that we often call it Free Cooling. This favourable conditions may occur when the outside air is at a lower temperature than the air inside the computer room. Turning off cooling units leads to a considerable energy saving since the cooling of Data Centers must be ensured both in summer and winter and for all hours of the day. Already existing conventional cooling units can be used as backup elements to cover load peaks. Furthermore, the integration of already existing units and Free Cooling systems is not very expensive. Fig. 1.9 shows a schematic representation of a typical free cooling system. The simplest implementation of the Free

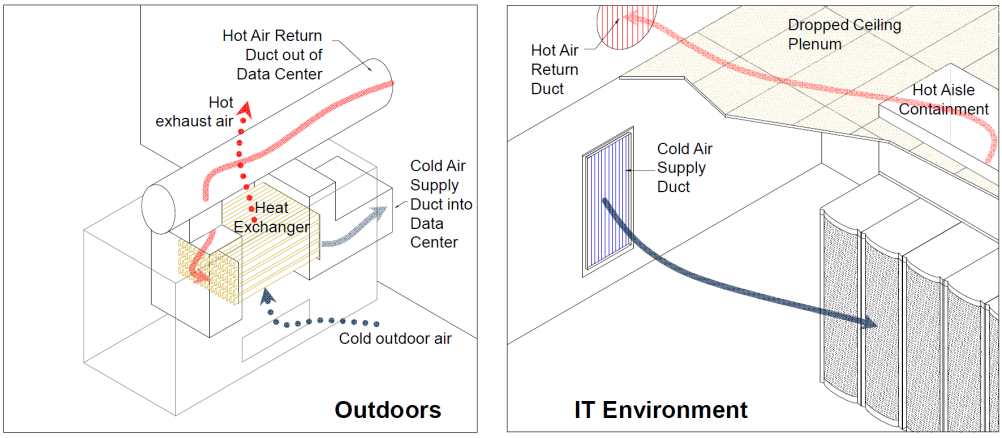


Figure 1.9: Schematic representation of a typical free cooling system, Evans (2012).

Cooling technology is direct Free Cooling. The low-temperature outdoor air is filtered and

then directly forced within the computer room. An alternative implementation consists in indirect Free Cooling. The fresh outdoor air is used to indirectly cool the recirculated warm computer room air through a compact air-to-air heat exchanger. Since the internal air never mixes with the outside air, there is a reduced possibility of external pollutants contamination. Also, the water vapour content of the internal air is not affected by the external humidity.

The Free Cooling technology can enhance even more its cooling capabilities by leveraging on the evaporative cooling principle. Evaporative cooling was used in ancient Egypt, Rome, and India by hanging straw mats in the windows that were wet with water. The evaporation of water cooled the air blowing through the window. Wind towers and flowing water ducts were some other methods used for cooling buildings, Dixit et al. (2017). This process exploits the fact that water absorbs a relatively large amount of heat in order to evaporate. As a result, the temperature of the dry air can be dropped significantly through the phase transition of liquid water to water vapour. The growing diffusion of these systems is due to the structural simplicity and to the ability to exploit natural energy resources, without using refrigerants. These devices ensure a performance significantly greater than that guaranteed by conventional vapour compression systems. In fact, the electrical power absorbed by pumps used for vaporizing water is much lower than that consumed by typical chillers. On average, the energy consumption can be 4 times lower than that of a conventional device with the same cooling power, but can go down to 10 times lower in a hot and dry climate Jaber and Ajib (2011). The overall operation cost is approximately 20 times lower than that of a standard vapour compression system, Lazzarin (2015). The main limit of exploiting water evaporation is the high dependence of its effectiveness on external climatic conditions. In fact, this cooling method strongly depends on the difference between the dry-bulb temperature of the primary inlet air and the wet-bulb temperature of the outdoor air. Therefore, considering a system installed in regions with a wet or mild climate, the temperature difference mentioned above may be excessively reduced, resulting in a very small cooling capacity. There exists two different implementations of this cooling method: Adiabatic Cooling and Evaporative Cooling. In Adiabatic Cooling systems, the outdoor air is forced through a humidifier, where some nozzles are located. The vaporization of water adiabatically reduces the temperature of the air, while increasing its humidity. The resulting cold air can be then directly forced into the computer room (Direct Adiabatic Cooling), but this affects the humidity of the internal air. The Adiabatic Cooling system can be combined with an indirect Free Cooling scheme to avoid problems related to the primary air humidification. An alternative to Indirect Adiabatic Cooling is the Indirect

Evaporative Cooling method. The outdoor air is humidified directly in the channels of the compact air-to-air heat exchanger, which are wetted by a film of vaporized water, as shown in Fig. 1.10. Therefore, both mass and heat transfers occur at the same time in the heat exchanger, making the analysis of the system more complicated. It is worth noting

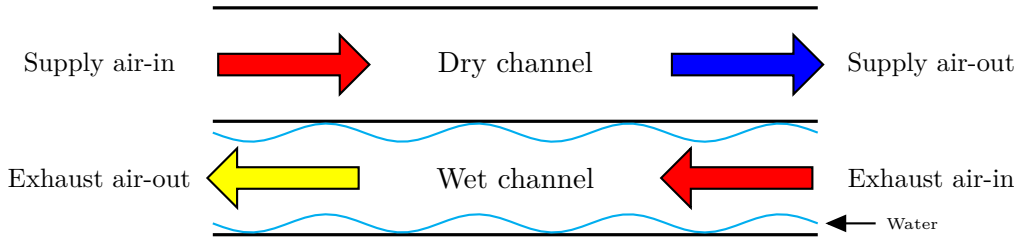


Figure 1.10: Functioning of an IEC system.

that the Indirect Evaporative Cooling technology can be straightforwardly included in an already existing indirect Free Cooling unit since it does not require the installation of an adiabatic humidifier. An example of an indirect evaporative free cooling system is shown in Fig. 1.11. In both Adiabatic and Evaporative cooling, water consumption can

Figure 1.11: An example of an indirect evaporative free cooling system: the EFC300 unit (Vertiv).

be a concern, particularly in arid regions. To mitigate this problem, the water that does not evaporate can be recovered and recirculated back to the pumps.

1.4.2 Liquid-based Cooling

Air-based cooling methods are the most well established for the current Data Centers infrastructure. However, the air is fundamentally an inefficient cooling method due to its low density and low heat removal capacity, Khalaj and Halgamuge (2017). This limitation will be more tangible with the emergence of high-performance microprocessors in servers, which will dissipate more than $100\,\mathrm{W\,cm^{-2}}$ heat fluxes inside Data Centers, where the maximum capacity of air is around $37\,\mathrm{W\,cm^{-2}}$. Therefore, if the size of Data Centers is due to expansion with high-performance IT equipment, other cooling solutions are required. Liquid cooling is one of the most prominent and efficient methods that can be directly or indirectly implemented.

1.4.2.1 Indirect Liquid Cooling

Indirect liquid cooling is a heat removal method without direct contact between the heat source and the liquid coolant. In a typical indirect liquid-cooled Data Center, a Coolant Distribution Unit (CDU) supplies close controlled chilled liquid from an external cooling source to the internal closed coolant loop attached to the electronic devices. Inside the IT devices, coolant is delivered to the racks, servers, and chips, depending on the cooling mechanism. Since processors are the primary source of heat flux, coolant typically goes there, while the rest of the electrical components are air-cooled. Standard air-cooled heat sinks need to be replaced with appropriate liquid-cooled heat sinks to implement this method. The cold plates and water-blocks are among the traditional indirect liquid cooling methods. However, recent research in this area focuses on microchannel heat sinks which have superior heat transfer characteristics. The utilization of coolants with considerably superior thermophysical properties than air can effectively extend thermal barriers and reduce Data Center energy consumption by allowing higher coolant temperatures and reduced airflow rates. The trade-off, on the other hand, is versatility reduction due to the installation of sealed enclosures and rack and server level piping system. As a result, the cooling system of a Data Center must be tailored to the facility layout and specific server. Moreover, fluidic connections constrain the easy hot-swap access to server equipment that commonly exists in the air-cooled servers. Quick-disconnect fluidic connectors can fix this problems. However, they represent further reliability concerns due to increasing the probability of leakage. An example of server with micro-channel liquid coolers is shown in Fig. 1.12.

Figure 1.12: Water-cooled IBM BladeCenter HS22 with two high performance micro-channel liquid coolers that remove heat from the microprocessors and various further heat sinks that collect heat from other components, Sridhar et al. (2011).

1.4.2.2 Liquid Immersion Cooling

The presence of a solid–solid contact resistance limits the thermal efficacy of indirect liquid cooling techniques. In fact, they involve conductive heat removal at the chip or chip package followed by convection to a liquid. Attention was directed to the enormous promise of direct liquid cooling to overcome this limitation. In particular, a promising alternative is given by liquid immersion cooling, which consists in the complete immersion of electronic components in a dielectric liquid. By doing so, the liquid captures all the heat generated by the IT equipment. Liquid submersion is a well-established method for cooling large power distribution components such as transformers. Suitable dielectric liquids can absorb much more thermal energy than air with the same volumes and temperatures. As a consequence, running the IT devices in higher environmental temperatures still allows the IT components to operate well below the maximum component temperatures. The main advantages of liquid immersion cooling systems over air-based ones are examined, for example, in Prucnal (2013). Today, suitable dielectric liquids are, for example, oil and chemically engineered liquids. A comparison of the effectiveness of different liquids for cooling servers is provided, for example, in Matsuoka et al. (2017). The study in Shah et al. (2016) investigated the effect of mineral oil on cables, printed circuit boards, packages, passive components, and oil properties over time. The circulation and the rejection of heat from the primary liquid is an important aspect of liquid immersion cooling. Traditional oil approaches use infrastructures with pumps and piping for circulating oil from a basin,

through a facility, to a cooling tower and back into the immersion bath. This same circulation pushes the oil through the IT chassis and over the micro-electronics, thus removing their generated heat. An alternative approach is to circulate the oil by natural convection, without making it leaving the immersion system. The oil circulates by the heat generated by the IT components and the presence of immersed cooling plates where cooling water circulates. Thus, the thermal expansion of the oil and gravity completely drive the primary circulation. Fig. 1.13 reports a schematic representation of the natural convection immersion cooling. The only requirement for heat rejection is any common

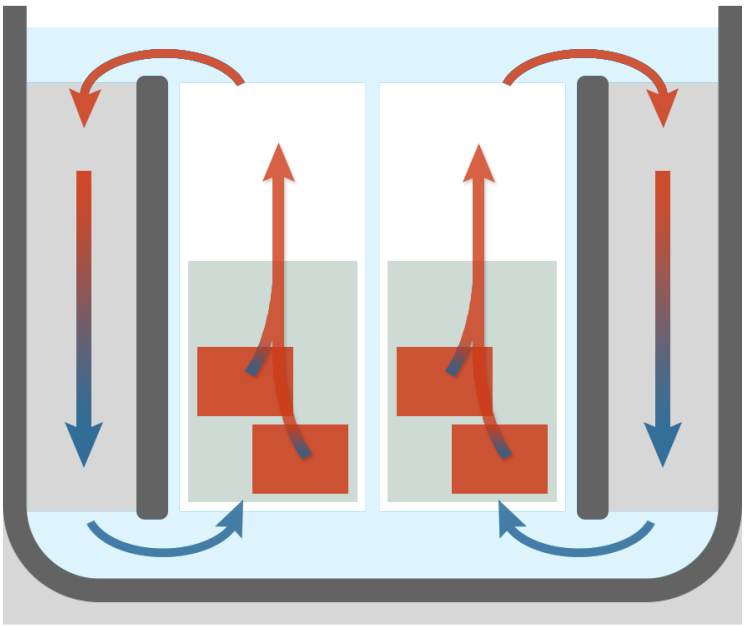


Figure 1.13: Functioning of a liquid immersion cooling system based on natural convection.

water infrastructure. Since there are no moving parts and everything in touch with the oil has been tested before insertion, the oil does not degrade in immersed cooling with natural convection. No air is ever mixed into the oil in the system, and the lid is closed by default since a closed lid improves the effectiveness of the natural circulation. This prevents oxygen from reacting with the oil. A double hull can be sealed to the vessel to protect the environment of the system from any leakage or additional heat. The double hull is insulated and provides an optimal safeguard against leakage. The hull also prevents heat from escaping from the liquid. The only place for the heat to go is into the water circuit. Ideally, this means that the result of liquid immersion cooling is water with

the same amount of thermal energy as the IT consumption. Fig. 1.14 shows an example of a liquid immersion cooling system which exploits natural convection. Any server can

Figure 1.14: Snap shot of a liquid immersion cooling system with natural convection, Matsuoka et al. (2017).

be inserted into the system, although it does not maintain the traditional server shapes as we know them in air-cooled racks. After all, traditional servers are not designed to allow the oil to flow through the chassis by itself. Liquid immersion-cooled servers consist of IT components like mainboards with components attached, power supplies, and storage. Limiting factors are related to the use of liquid. Thermal paste dissolves in any liquid, so this must be replaced with a non-dissolvable material (i.e., aluminium foils). Solid State Drive (SSD) can be directly used, while traditional mechanical hard drives must be certified for immersion by the manufacturer. The natural convection approach and the containment of liquid in the system allow shifting focus from cooling to compute. The self-contained approach allows for a further level of integration. Everything required to facilitate the IT is integrated in the system. A single system requires only redundant power, access to a water loop, and data connectivity. This limited requirement enables high flexibility when determining appropriate deployment sites and scenarios. A single water loop can be shared across multiple modules, and modules can be placed back-to-back and side-to-side. Since there is no air required to operate, large rows of

interconnected systems can be placed in relatively small spaces. The reduced floor space is one of the obvious benefits of liquid immersion cooling over an average air-based cooled Data Center. Indeed, no raised floors and aisles separation schemes are required. Since there is no air involved with immersion cooling, there are no airflows to separate. Although raised floors are fully supported, they are not a requirement for liquid immersion cooling. Another important benefit provided by liquid immersion cooling is the reduction of energy consumption. Any air-cooled IT equipment requires air circulation. In servers, this consumes between 10% and 45% of the total energy footprint. 1U[3] servers with average processor power with good utilization often end up with the highest fan overhead. Larger servers (2U-5U) have larger fans which consume less energy, but these servers take up a lot more space in a rack, which makes it less effective. Liquid immersion cooling eliminates this fan overhead and space limitation. This power consumption reduction has a significant impact also on the UPS systems. They need to be sized to allow all the IT equipment to function during power outages, as well as the cooling infrastructure to facilitate the IT. The power systems can be downsized or, when already present, facilitate more IT devices. Another benefit provided by liquid immersion cooling systems from an energy efficiency point of view is the high energy reuse potentiality. In fact, almost all of the IT equipment energy is captured in the form of heat inside water. Warm water can easily be transported or even stored for energy reuse scenarios.

Two-phase immersion cooling solutions can be considered to augment even more the heat transfer rate. Two-phase cooling is a latent heating process of circulating coolant where the liquid-vapor phase change occurs. There are various dielectric fluids with low boiling temperatures and refrigerants that can be utilized as a proper coolant in this process. The added advantages of this method compared to the single-phase cooling are the reduced temperature gradient across the heated surfaces and the augmented heat transfer rates, both of which are associated with the latent heat of vaporization as well as isothermal nucleation processes exist in the two-phase flow. The main parameters impacting on the performance and thermo-flow behaviour of two-phase cooling are typically similar to the single-phase cooling. The main drawbacks of this method are associated with flow instabilities. They cause flow reversal, temperature and pressure fluctuations, and other thermo-flow behaviours that may finally cause super-heating and burnout of the heated surfaces. Therefore, this cooling method has been scarcely developed in Data Centers industry. The working principle of the two-phase immersion cooling method is schematized in Fig. 1.15.

[3] A rack unit U is a unit of measure frequently used as a measurement of the overall height of rack frames, as well as the height of equipment that mounts in these frames. For example, a typical full-size rack cage is 42U high, while equipment is typically from 1U to 5U high.

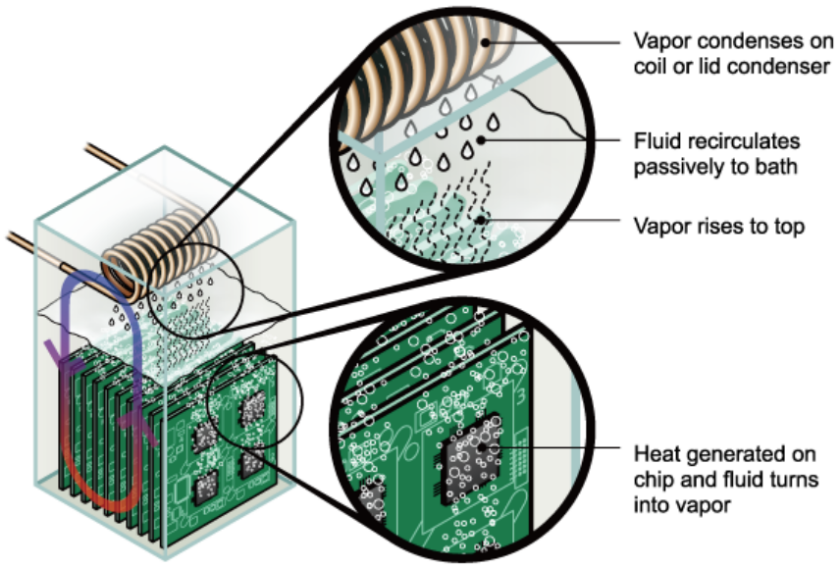


Figure 1.15: The working principle of two-phase immersion cooling solutions, Dai et al. (2016).

1.5 Design Considerations

Main aspects of the design of a Data Center are Bramucci et al. (2010):

- reliability, which can be guaranteed by the redundancy of devices;
- scalability, ensured using modular devices and avoiding useless oversize of the system;
- choice of high-efficiency devices.

All these factors must be considered in the choice of IT devices and components of cooling and power supply systems. Also, they must be considered in the choice of the geographical location and the structure of the building. In particular, the most important aspects to consider in the realization of the system are:

- the position of the computer room within the building;
- ease of connection to the electricity network, without limits in the demandable power in prevision of possible expansions in the future;
- the proximity of the structure to water basins for heat dissipation;
- the proximity to network bands with redundant paths;
- the proximity to an urban center to facilitate access to the structure;

- the evaluation of building realization costs;
- the height of the room to ensure the smooth passage of cables and pipes;
- the presence of a floating floor or a false ceiling for the passage of the cooling air;
- easy access for maintenance operations in case of malfunctions, guaranteed by proper location of wiring and piping;
- the protection of all components from possible external disturbances;
- the possibility of using a Free Cooling system according to the geographical location.

The Clumeq Data Center realized in Canada is an example of Data Center that has been designed to facilitate the use of Free-Cooling. Its shape is cylindrical and servers are placed along the circular perimeter of the building, so the produced heat is concentrated in the central zone of the room and it can be easily recovered, Fig. 1.16. Another

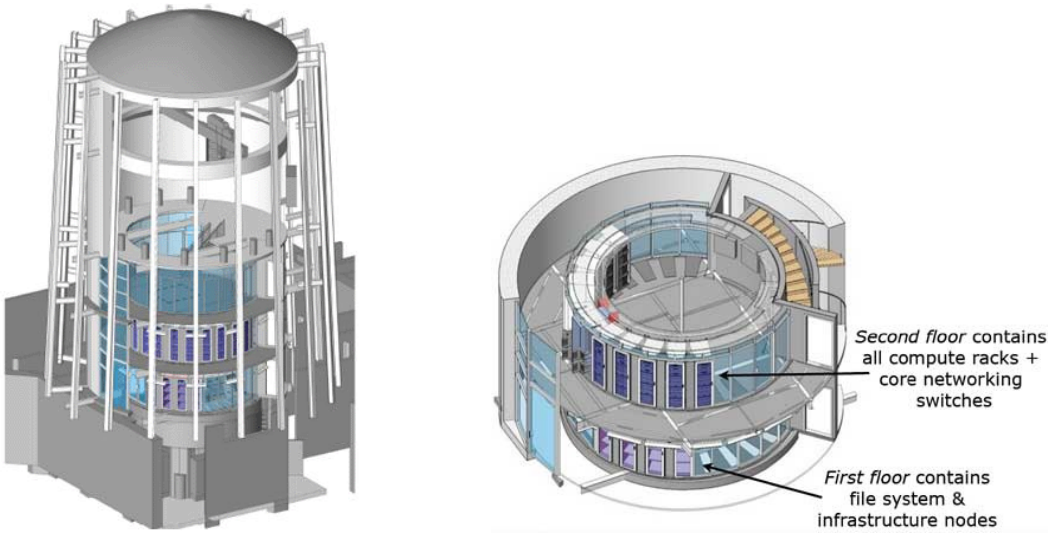


Figure 1.16: The Clumeq Data Center in Québec: its 65 feet high and 36 feet wide cylindrical layout has been designed to facilitate the use of Free-Cooling. The racks are arranged in a ring on each floor and the floors are comprised of grates to facilitate airflow through the facility.

example of peculiar Data Center concept is given by the Microsoft Project Natick, Simon (2018). On 5th June 2018, Microsoft announced the launch of self-sufficient underwater Data Centers, being submerged in a shipping-container-size capsule, 100 ft below the surface of the North Sea near the UK's Orkney islands. The world's oceans at depth are consistently cold, offering ready and free access to cooling. The project adapted a heat-exchange process commonly used for cooling submarines to the underwater Data

Center: a cooling system pipes seawater directly through the radiators on the back of each of the server racks and back out into the ocean. The underwater Data Center is powered by renewable energy provided by tidal turbines and wave energy converters, which generate electricity from the movement of the sea. The Microsoft underwater Data Center is shown in Fig. 1.17.

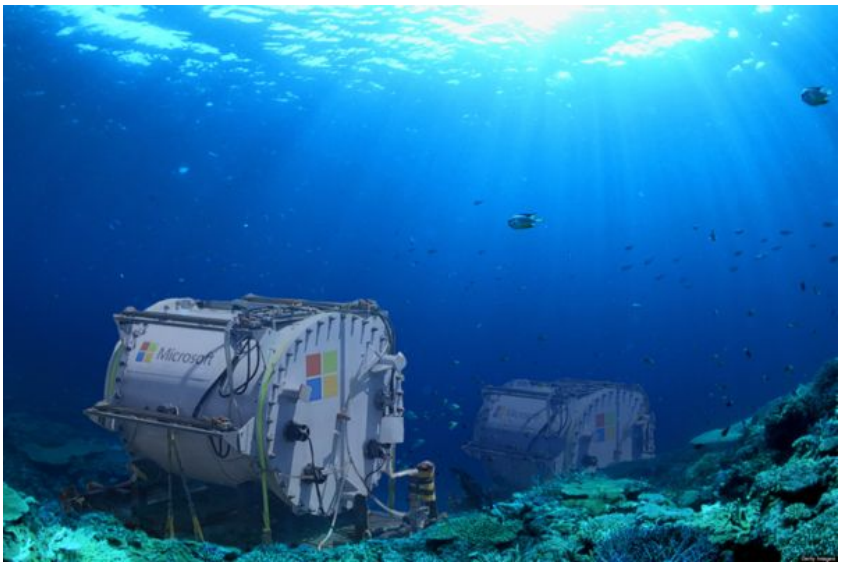


Figure 1.17: The Microsoft underwater Data Center, submerged 100 ft below the surface of the North Sea: cold seawater is circulated directly through radiators placed on the back of the racks and then it is pumped back into the ocean.

The specific application is also an essential factor that must be taken into account in the design of a Data Center. In particular, it determines the size of a Data Center, which can mainly be categorized in:

- small size Data Centers: these systems employ a hundred racks or less, and are typically used for smaller businesses, like experimentation facilities (e.g., the small 10-racks Data Center module hosted by Swedish Institute of Computer Science ICE (SICS) in Luleå, ICE, or modular server containers, e.g., the Sun Microsystems Modular Datacenter shown in Fig. 1.18);
- mid size Data Centers: these systems employ between a hundred to a thousand racks, and are generally used for medium-sized businesses;
- large size Data Centers: these systems host more than a thousand racks, and are typically used by huge corporations like Microsoft, Google, and Facebook. These premises may be very wide (e.g., the Facebook Data Center in Luleå, Sweden, is shown in Fig. 1.19).

Figure 1.18: The Sun Microsystems Modular Data Center, a portable Data Center built into a standard 20-foot long (6.1 meters) intermodal container.

Figure 1.19: Aerial view of the 300 meters wide by 100 meters long Facebook Data Center in Luleå, Sweden.

1.6 Analysis of the Energy Consumption

The energy efficiency of Data Centers has attained fundamental importance in recent years due to its high economic, environmental, and performance impact. Data Center spaces may consume up to 100 to 200 times as much electricity as standard office space, Poess and Nambiar (2008). According to energyliteracy.com, nowadays Data Centers in the US consume a total of 0.24 quads[4], with 0.133 quads wasted and 0.106 quads effectively consumed, Fig. 1.20. Moreover, the energy costs of powering a typical Data Center are continuously increasing, doubling every five years. Data Center energy usage also creates several environmental problems, Whitehead et al. (2014). For example, in 2005, the total Data Center power consumption was 1% of the total US power consumption and created as many emissions as a mid-sized nation like Argentina, Corcoran and Andrae (2013). Due to these reasons, Data Center energy efficiency is now considered chief concerned for Data Center operators, ahead of the traditional considerations of availability and security.

The energy consumed by a Data Center can be broadly categorized into two parts: energy used by IT equipment and energy used by the supporting infrastructure facilities (e.g., cooling and power supply systems). The amount of energy consumed by these two subcomponents depends on the Data Center design as well as the equipment efficiency. For example, according to the statistics published by the Infotech group and shown in Fig. 1.21, the more significant energy consumer in a typical Data Center is the cooling infrastructure (50%), Group et al. (2007). Server and storage devices rank second in the energy consumption hierarchy. Note that these values might differ from Data Center to Data Center.

1.6.1 Performance Metrics

The wide variety of different components which are present in a Data Center makes the evaluation of its energy efficiency a complicated task. Different metrics have been defined for quantifying and comparing Data Center efficiencies. The most common ones are the Power Usage Effectiveness (PUE) index and its inverse, the Data Center Infrastructure Efficiency (DCIE) index:

$$PUE = \frac{1}{\text{DCIE}} = \frac{\text{IT Power}}{\text{Total Power}}, \tag{1.2}$$

[4]quadrillion BTUs (British Thermal Units). BTU is defined as the amount of heat required to raise the temperature of one pound of water by one degree Fahrenheit. The metric unit equivalent is the joule, where one BTU is about 1055 joules.

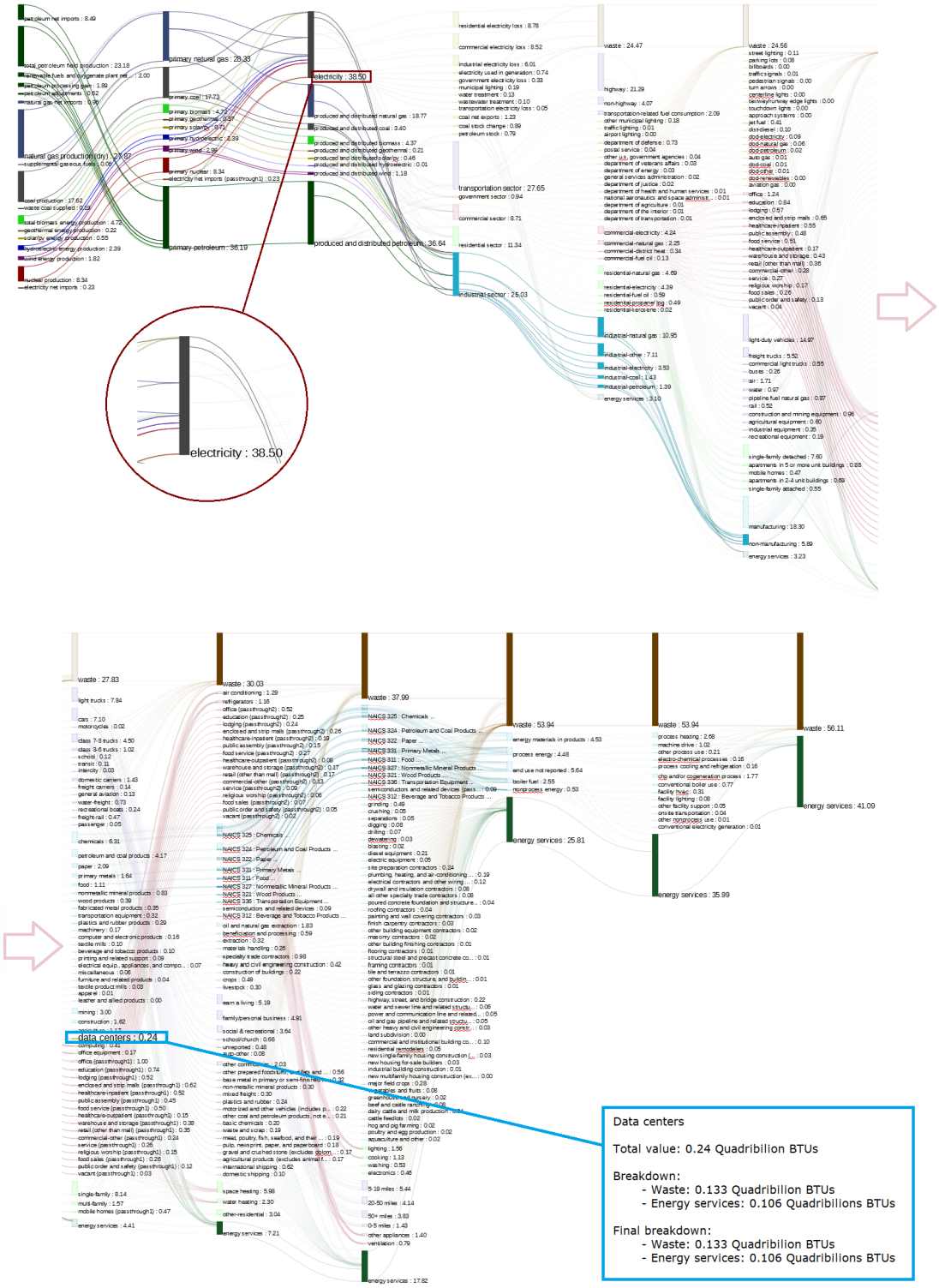


Figure 1.20: Graph of all energy production to power consumption in the US (http://energyliteracy.com/ produces an interactive graph: by clicking through the chart, you can see exactly how much energy is used for each activity). The comparison of production to waste gives us a good look at the net efficiency of the energy systems in the US. Data Centers consume a total of 0.24 quads, with 0.133 quads wasted and 0.106 quads effectively consumed.

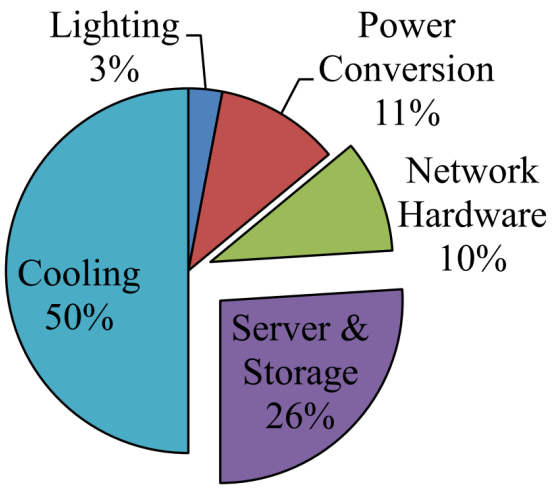


Figure 1.21: Analysis of the energy consumption of the different components within a Data Center.

where:

- IT Power is the total electrical power consumed by the IT equipment;
- Total Power is the total input power provided to the Data Center, including also the supporting infrastructure facilities.

The ideal value of PUE is 1, but values closed to 1.5 are also acceptable, considering that the average value for existing Data Centers is close to 2.5. On January 13, 2010, various industry and government organizations met to discuss energy metrics and measurements. Among the other agreements, they recommended four categories for the measurement of the PUE:

- PUE_0: ratio of the peak power during a 12-month measurement period to the IT power measured at the output of the UPS units;
- PUE_1: ratio of the total annual consumed energy to the one absorbed by IT devices, measured at the output of UPS units;
- PUE_2: ratio of the total annual consumed energy to the one absorbed by IT devices, measured at the output of PDUs;
- PUE_3: ratio of the total annual consumed energy to the one absorbed by IT devices, measured at their outputs.

It is worth noting that all the metrics described above do not consider how the electricity provided to the Data Center is produced. For example, it may come from renewable

sources or co-generation. Also, the PUE is affected by the eventual heat recovery only when the heat recovered is directly reused within the Data Center, for example for supplying cooling units.

To improve the energy efficiency of a Data Center and reduce its PUE index, operators can act on all the main contributors to the overall electrical power consumption. For example, using more efficient servers, exploiting virtualization, and optimizing the allocation of the tasks among the servers can improve the energy efficiency of the overall IT infrastructure. Also, the use of new UPS technologies can increase the efficiency of the power supply systems. Often, actual efficiencies of the power supply equipment are lower than their rated efficiencies, since they seldom operate at the same loads under which the rated efficiencies are calculated. However, the most significant opportunity to reduce the overall energy consumption in Data Centers is given by the cooling systems. In fact, they account for roughly half of the total energy absorbed by a typical Data Center, Fig. 1.21. The next Chapter of this Thesis discusses the problem of efficiently design and manage the cooling infrastructure in Data Centers.

2

Thermally Efficient Data Centers

This Chapter focuses on the challenges of the thermally efficient design and management of Data Centers. Provision of sufficient and reliable cooling is a central aspect of the operation of Data Centers. Moreover, as discussed in Section 1.6, a large amount of the electrical power provided to a Data Center is typically consumed by the cooling infrastructure. Nowadays, one can exploit both innovative cooling systems hardware design and advanced control systems which can actually improve both system efficiency and performance by mostly acting on software components. The following sections provide an overview of the state of the art in Data Center thermal design and management.

2.1 Introduction

Data Centers house a large number of IT systems that are used to process, store, and transmit digital information. Most of the electrical energy consumed by the IT equipment in a Data Center ultimately transforms into heat, which must be removed continuously in order to prevent overheating that may lead to malfunction and breakage. Provision of sufficient and reliable cooling is thus a central aspect of the Data Center operation. As discussed in Section 1.6, cooling systems absorb a large amount of the total electrical energy consumed by a typical Data Center. Thus, energy efficiency must be the primary focus in the design and management of the cooling infrastructure.

Designing a new Data Center facility or changing an existing one to maximize the cooling efficiency can be a challenging task. Adopting the best cooling method according to the size and location of the Data Center is a key opportunity to reduce the amount of energy consumed for cooling. As discussed in Section 1.4, there exists several different cooling methods that can be adopted in Data Centers. Using liquid cooling solutions provide a means to efficiently cool densely packed high-end servers, rather than to ease the implementation of heat-recovery technologies. However, most of currently existing Data Centers still use air as cooling medium to remove the heat produced by the IT devices. If the environmental conditions are favourable, the integration of air free cooling offers opportunities to reduce the electrical power consumed by the cooling systems. In the typical hot/cold aisles configuration, the airflow distribution has a big impact on the thermal environment in the Data Center and eliminating the cooling inefficiencies related to a maldistribution of the air allows to achieve significant energy savings without the need of large investments. Any number of different cooling system design strategies or floor layout variations can affect the efficiency, creating hotspots, or altering the amount of infrastructure required for the design. Computational Fluid Dynamics (CFD) tools offer a method of evaluating new designs or alterations to existing designs before they are implemented.

Choosing and designing an appropriate cooling architecture is not enough to establish an efficient cooling system without effective cooling control. Data Center cooling is full of challenges due to load dynamics and cooling system dynamics. The limitations or drawbacks of traditional control approaches make the situation worse. Designing a cooling system with an effective control system is a best practice to solve these challenges. Traditionally, the design of the cooling system and the synthesis of its control architecture are two separated activities. The conventional strategy, also known as sequential approach, requires that first the process is designed, in most cases at steady state, and then the control strategy is developed. However, since the dynamic performance of a process strongly depends on its design, the decision-making domains of process and control engineers overlap. Thus, a simultaneous approach to the design and control would lead to significant economic benefits and improved dynamic performance during plant operation.

In this Thesis, we focus on the problem of designing advanced control systems for optimizing the operation of Data Center cooling systems. For completeness, in this Chapter we discuss both design and control aspects and how they influence each other.

2.2 Thermally Efficient Design

The first step in designing an efficient cooling system for new or already existing Data Centers is choosing an appropriate cooling architecture based on the key Data Center attributes like location, design capacity, average power density, and Data Center preferences and constraints. There are several different heat removal methods for cooling Data Centers and transporting the thermal energy from the IT environment to the outside. Different cooling techniques have been discussed in Section 1.4.1. Multiple methods can be also combined to increase the efficiency and to provide redundancy. Currently, most commonly used cooling methods are air-based chilled water systems and CRAC DX systems, which are both based on a refrigeration cycle. The hot air or the warm water returning from the computer room are cooled by passing through a cold evaporator coil. As a result, the refrigerant inside the evaporator is evaporated. The refrigerant is then compressed at constant entropy by a compressor, which increases the temperature of the vapour. The heat is then rejected to the outside as the superheated vapour travels through a condenser. The refrigerant is then back in liquid form and flows through an expansion valve back to the evaporator coil, closing the refrigeration cycle. The compressors, pumps, and fans are driven by AC motors, which are the energy hungry parts of the chiller. The chiller system is typically designed to provide the maximum cooling demand during the hottest summer times, Hartman (2001), which, however, take place only for a short duration. For majority of the time, the chiller system does not run at the full-load, rather in substantially partial loading condition, especially during the winter. Thus, variable-speed chillers, taking into account load variations, offer opportunities to reduce the energy demand significantly compared to the constant-speed ones, Oliveira and Ukil (2019). In geographic areas with cold climate conditions, the energy consumption can be reduced drastically by combining the use of traditional chillers with Free Cooling units. As stated by the second law of thermodynamics, heat can not flow spontaneously from a colder location to a hotter one and, consequently, work is needed to obtain cooling. However, for the vast majority of regions, the outdoor temperature is lower than that of the Data Center in a long period of the year, Zhang et al. (2014). When the outdoor temperature is sufficiently below that of the Data Center, the heat will naturally flow to the outside without the need of compressors and vapour-compression refrigeration systems. Also, the use of Free Cooling can be extended to warmer and dry climates by exploiting the adiabatic and evaporative cooling processes. Lee and Chen (2013) analyze the potential energy savings of air-side Free Cooling in 17 climate zones by employing a dynamic building energy simulation program.

In air-based cooling architectures, the airflow distribution has a big impact on the

thermal environment in the Data Center. In most air-cooled Data Centers, an air supply plenum below a raised-floor is used to provide cooling air directly to the IT equipment. The air enters the room in front of individual racks through perforated floor tiles. The area below the raised floor is typically used also for other ancillary functions, e.g, cabling. With this design approach, the hot air either recirculates freely to the air conditioning units in the room or is channeled via vents in a dropped ceiling to reduce the intermixing with ambient air. An alternative less popular design approach provides cooling air via a dropped ceiling from top to bottom. In many Data Centers, severe airflow inefficiencies can be observed. Most common ones include:

- hot air recirculation from the rear of the rack directly to the server air inlets because of insufficient cold air supply or missing blanking panels in the racks;
- intermixing of hot and cold air streams, which raises server inlet air temperature and lowers the return air temperature of the cooling units;
- leakage of the cooling air into the room, for example at cable cutouts;
- underfloor obstructions impeding the air flow in the air supply plenum.

Eliminating this cooling inefficiencies allows to achieve significant energy savings without the need of large investments. Hot and cold aisle containment can improve the energy efficiency of air-cooled Data Centers. The basic benefits are the same for both containment methods: by physically separating the hot and cold air streams, hot spots at server air inlets can effectively be eliminated. Consequently, the cold air supply temperature can be raised, which saves energy and increases the cooling capacity.

To achieve an energy-efficient cooling performance, next generation high-density Data Center design needs to be based on a solution that eliminates the requirement for air heat transfer and to be replaced by liquid-cooling. The main advantage of a liquid-cooled solution is the higher heat transfer capacity per unit. In general, air has a very low density (about 1/1000 of water) and to carry a given quantity of heat (between two fixed temperatures) requires a large volume of air (i.e., high speed) rather than small volume of liquid (i.e., low speed), Chi et al. (2014). In some cases, such as density driven natural convection, the fluid velocity is very low and therefore liquid would be a better choice with a larger specific heat (per unit volume). Along with the significant thermal advantage, there are some other benefits of liquid-cooled solutions. A liquid cooled Data Center usually requires fewer fans and rotating components, therefore expending less energy and providing increased reliability due to fewer moving parts. In a fully immersed liquid-cooled solution the primary coolant is contained in a static (sealed) condition. A

typical 500 kW Data Center with fully immersed liquid-cooled systems would require only 32 pumps instead of more than 1000 spinning fans in an equivalent air-cooled solution, and the noise level of a fully liquid-cooled approach is negligible. Further benefits of a direct liquid-cooled Data Center are that all the computer nodes will be inside a sealed container and in a fully controlled environment, which would have minimum level of dust and vibration. With the high heat capacity of liquids, liquid-cooled Data Centers will be able to harvest the produced heat efficiently and reuse it in a multiplicity of scenarios. Being electrical energy consumers as well as heat producers, the Data Center can be successfully integrated into both electrical and thermal energy grids and the waste heat generated by the IT components can be effectively re-used either internally for space heating and/or domestic or district heating network operators, Antal et al. (2018).

CFD simulation tools are being increasingly used to numerically investigate the temperature fields and flow distributions in designing new and analyzing existing Data Centers. They were introduced to the Data Center community in early 2000s through the works Kang et al. (2001), Schmidt et al. (2001), and Karki et al. (2003), among others. In the design stage, CFD tools are used to validate and optimize different design parameters. For existing Data Centers, they are used to identify cooling issues and to explore various cooling improvements without risking the operation of the Data Center. Once the best practical and cost-effective remedies are identified, they can be implemented in the real Data Center. For example, Sundaralingam et al. (2015) experimentally investigated the performance of the cooling system in different setups with either an open, partially contained or fully contained cold aisle. Nada et al. (2016) numerically studied partial cold aisle containment and also compared different arrangements of cooling units and server racks with respect to airflow management. Wibron et al. (2019) use the CFD tool ANSYS CFX together with the Reynolds stress turbulence model to numerically investigate the effects of using partial aisle containment between the server racks for hard and raised floor configurations. Simonazzi et al. (2018) consider the specific flow modelling problem connected to the identification of different air mixing regimes at the computer room level in general room configuration scenarios. In Matsuoka et al. (2017) an interesting evaluation of the cooling performance of liquid immersion cooling technology with natural convection is conducted by means of both CFD simulations and actual experiments. In particular, authors show that the change in any CPU task and any server removal gives a limited cooling effect on the other servers and other CPUs. Also, it is shown the benefit of attaching a heat sink to the CPUs by means of CFD simulations. In Akhremenkov and Tsirlin (2016) authors propose a detailed numerical investigation of natural convection phenomena that occur in an immersion cooling tank where the heat

source and the cooling source are separated by a vertical partition, in order to facilitate the convective flow. The analysis in Almaneea et al. (2014) combines empirical curve fits and flow analysis with CFD simulations of liquid-immersed servers placed in parallel in a rack of a Data Center. The potential heat reuse offered by liquid-cooled Data Centers is evaluated in Oró et al. (2018), where authors propose a TRNSYS model of a liquid cooled Data Center coupled with the heat demand application, that in that case is an indoor swimming pool.

2.3 Thermally Efficient Control and Management

A thermally efficient design of the cooling architecture is not enough to establish an efficient cooling system without effective cooling controls. An effective cooling control system should automatically shift between different modes of operations and operate the different cooling devices in order to match the cooling requirement dynamically and efficiently. The designed control system should also be robust to variations of the IT load and outdoor air conditions and flexible to change certain setting based on customer requirements.

However, traditional control approaches limit how well the cooling system adapts to changes in the Data Center environment. Cooling devices are often adjusted manually by Data Center operators who change the set-points or turn the devices on and off based on their knowledge or intuition. But, sometimes the correct response is counter-intuitive. A large portion of traditional cooling control approaches mostly focus on monitoring the operation of the cooling system and do not perform control functions like adjusting the speed of compressors, fans, or pumps to optimize the performance. Some cooling devices do not have Variable Frequency Drives (VFDs) to change the speed. Moreover, adjacent cooling devices in the IT environment often work independently, which leads to demand fighting among these devices and wastes a lot of energy. Each cooling device in a traditional system is designed to optimize its own performance, regardless of the impact on the total cooling system energy consumption.

The first step in designing a control system is identifying the controlled variables and the control objectives. The number of controlled variables must be less or equal to the Degree Of Freedom (DOF), which represents the maximum number of variables in the control system that can be manipulated by it. The DOF strongly influences the potential effectiveness of the control system and it is determined by the design of the cooling system architecture. The design processes of the cooling architecture and control system should thus be integrated. For example, VFD fans, pumps, and compressors are typically less

economic than fixed-speed ones, but they allow to manipulate their speed augmenting the DOF of the control system. The primary objective of the cooling control system must be to reject all the heat load from the IT equipment in order to prevent overheating. This can be translated in satisfying certain constraints on the thermal status of the cooling medium, i.e., temperature and humidity of the air or temperature of the cooling liquid. Traditional air-based cooling approaches considered the return air temperature as main control parameter. Typically, the cooling units were operated to guarantee a return air temperature between 21 °C and 26 °C and, consequently, supply air temperatures to the computer room between 10 °C and 14 °C to be considered standard. Today, the best practices are very different. In a modern Data Center, the supply air temperature is the main control parameter. The Recommended Envelope suggested by American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) changed in ASHRAE (2008) and ASHRAE (2011), expanding the use of chiller-less Data Centers. The ranges for temperature and relative humidity of the supply air recommended by ASHRAE are reported in Tab. 2.1 and Fig. 2.1. In ASHRAE (2011), ASHRAE defined also five classes

	ASHRAE Recommended Ranges
T_{min}	18 °C
T_{max}	27 °C
RH_{min}	Dew point temperature 5.5 °C
RH_{max}	60% and dew point temperature 5.5 °C

Table 2.1: Temperature and relative humidity ranges provided by ASHRAE.

of temperature ranges for indirect liquid cooling solutions, where water or a cooling liquid are circulated in the proximity of the IT equipment heat sources. The five ASHRAE classes for liquid cooling are reported in Tab. 2.2. Beside the thermal constraints, an

Class	Water supply temperature
W1	2 °C to 17 °C
W2	2 °C to 27 °C
W3	2 °C to 32 °C
W4	2 °C to 45 °C
W5	>45 °C

Table 2.2: Temperature ranges provided by ASHRAE for indirect liquid cooling solutions.

.

effective cooling control system must take into account also the energy efficiency of the cooling process. This control objective can be typically formulated as minimizing the

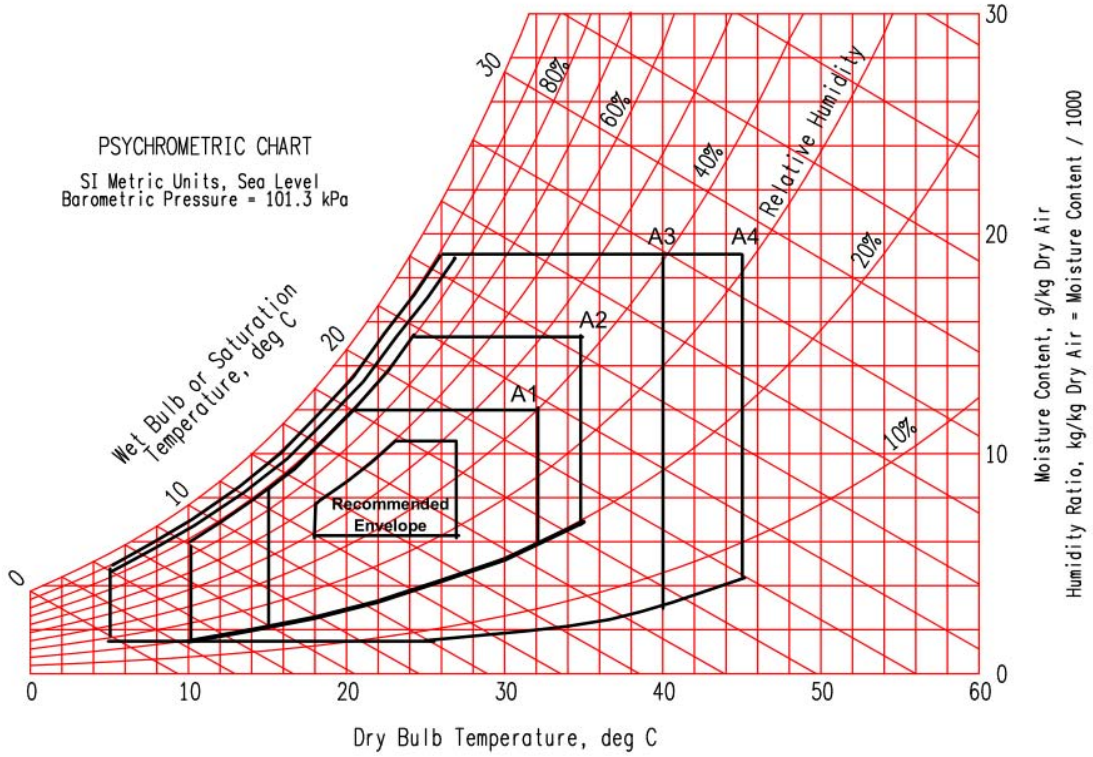


Figure 2.1: Temperature and relative humidity ranges recommended by ASHRAE (2011). The Recommended Envelope specifies a long-term operating environment that can result in greater IT equipment reliability. The other envelopes (A1, A2, A3, A4) should be used only for short-term operation or in the case of a cooling malfunction. The allowable envelope limits define how far the environment can stray from the Recommended Envelope and still expect the IT equipment to remain functional.

amount of energy used for cooling. However, especially in new liquid-cooled Data Centers, this can be extended in order to include the maximization of the amount of reusable thermal energy stored in the coolant. Satisfying the thermal constraints and optimizing the energy efficiency are typically conflicting objectives: requiring more cooling capacity means requiring more effort to the cooling units, which translates in more electrical energy need. It comes as no surprise that much of cooling (and more in general Heating, Ventilating, and Air Conditiong (HVAC)) control is about compromise: a balance that usually results in reasonable temperatures at minimum net energy use (Fig. 2.2) and financial costs.

Cooling systems and more in general HVAC systems represent one of the most complex challenges for control and optimization due the intrinsically complex nature of their underlying physics and the numerous technological constraints. The main complexities that characterize this kind of systems are, for example:

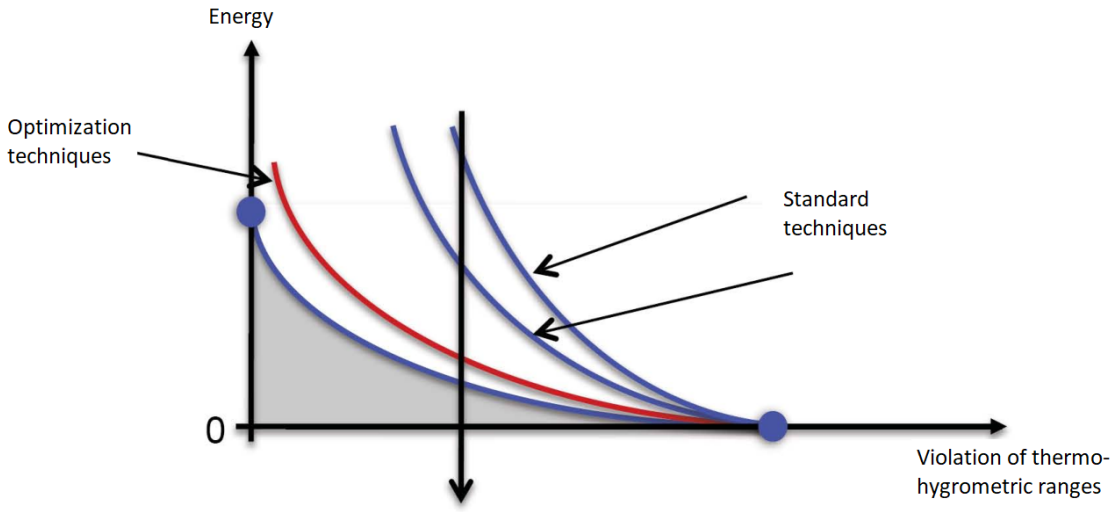


Figure 2.2: Energy-temperature range: much of cooling control is about compromise between reasonable temperatures at minimum net energy use.

- Data Centers involve physical quantities which belong to different physical domains, for example electrical power provided to the servers, temperature and flow rate of the air, and rotational speed of fans, pumps, and compressors;
- Data Centers are an example of Cyber-Physical Systems (CPS): the energy management depends critically upon the management of both computational (cyber) resources and cooling (physical) resources, Parolini et al. (2011);
- Data Centers are dynamic environments where the equipment and layout change over time and the heat load also changes constantly in response to IT computing traffic;
- the dynamics of the controlled processes span over multiple time scales: for example, servers power state can be changed as frequent as milliseconds, the actuators operate on a time scale of seconds, and temperatures in the computer room evolve in the order of minutes;
- cooling system capacity is always oversized, i.e., cooling capacity is larger than the actual IT load, but Data Centers typically operate under 50 % load;
- cooling systems typically exhibit non-linear and coupling behaviours;
- the efficiency of the cooling systems varies with Data Center load, outdoor air conditions, and cooling settings;

- the cooling systems are normally comprised of devices from different vendors, thus compatibility and coordination between these devices is a big challenge.

In this context, the design of the control system and the evaluation of its performance under different operating conditions are non-trivial tasks. Modelling and simulation tools provide invaluable support for this purpose. Simulation environments based on accurate models of the plant can be exploited to better understand the dynamic characteristics of the targeted process and to evaluate control objectives and constraints as functions of the controllable variables. Also, they can be used to derive Computer Aided Control Systems Design (CACSD) software tools, which facilitates designing, tuning, and testing in silico different control strategies in a generally cheaper and safer way than conducting real-world experiments. The mathematical models are required to capture the most relevant dynamics of the cooling process and the main cause/effect relationships, without the need of describing in details the spatially distribution of temperatures and flows within the system. For this reason, lumped-element physical models or surrogate data-driven models can be used instead of more conventional CFD ones, which would be too computational burden and they may lead to simulations taking many minutes, hours, or even days to complete. The next Chapter of this Thesis focuses on the problem of designing advanced control strategies for Data Center cooling systems by exploiting the potentialities offered by CACSD tools.

3

Computer Aided Control System Design

This chapter focuses on the challenging problem of designing advanced control strategies to optimize the operation of cooling systems in Data Centers. Within this context, modelling and simulation tools are the main elements of a typical CACSD framework, which provides invaluable support in analyzing the system and designing effective control strategies. In Section 3.1 we give a brief introduction to CACSD tools. Then, in Sections 3.2 and 3.3 we provide some considerations on the modelling and simulation steps and finally in Section 3.4 we propose a hierarchical control approach to cope with the typical complexities of cooling systems.

3.1 Introduction

CACSD encompasses a broad range of methods, tools, and technologies for process modelling, control system synthesis and tuning, dynamic system analysis and simulation, as well as validation and verification, Malinowski et al. (1994). The development of CACSD tools for system analysis and synthesis started around 1960, immediately after general purpose digital computers and new programming languages became available for research and development purposes. The domain of CACSD enlarged progressively over decades from simple collections of algorithms and programs for control system analysis and synthesis, to comprehensive tool sets and user-friendly environments supporting all

aspects of developing and deploying advanced control systems in various application fields. In the context of this Thesis, these tools offer a great support in designing advanced control strategies for cooling systems, which represent one of the most complex challenges for control due the intrinsically complex nature of their underlying physics and the numerous technological constraints. In the next sections, we described the main steps of a typical CACSD workflow, which consists in the following interlaced activities, Varga: modelling, design of simulation tools, system analysis and design of the controller, and performance assessment.

3.2 Modelling

Modelling is a necessary first step consisting in developing suitable mathematical models to accurately describe the plant behaviour and the main relationships between the input and output variables of the process. The main models classification concerns the information provided by the model about the time evolution of the considered process:

- dynamic models, which account for time-dependent changes in the state of the system;
- static (or steady-state) models, that compute the system states in stationary conditions.

Dynamic models are usually defined by a set of Ordinary Differential Equations (ODEs), differential-algebraic equation (DAEs), or Partial Differential Equations (PDEs), while static models are typically defined by algebraic equations. Deciding which components to be described by dynamic models and which ones by static models is particularly crucial in complex systems where different time-scale dynamics are involved.

The initial phase in modelling of a technical system is collecting and systematic treatment of the available knowledge. The a priori knowledge about a given phenomenon comes from the analysis, comprising of finding all possible connections to other phenomena and physical laws, preceding the modelling. The a priori knowledge is of key importance in modelling although its availability is always limited by the complexity of the physical system. Even if the governing physical principia are known, it is usually difficult to formulate the specific relationships and obtain particular values of the parameters. Availability of the a priori knowledge and the modelling purpose determine the following:

- the final type of the model;
- the accuracy requirements;

- the type of specific modelling procedure;
- the complexity of the model;
- the method and the cost of its realization.

According to the degree to which the a priori knowledge is available, then either a First-Principle (FP) or a Data-Driven (DD) model, or a combination of both, can be applied, Czop et al. (2011). FP models use understanding of the system underlying physics to derive its mathematical representation. FP models are expensive in development since expertise in the area of knowledge at the advanced level is required to derive equations from physical laws, while DD models use system test data to derive its mathematical representation. The advantage of the former approach is the depth of the insight into the behaviour of the system and thus ability to predict the performance, while the advantage of the latter is the speed in which an accurate model can be constructed and confidence gained thanks to the use of the data obtained from the actual system. The difficulty of the former approach lies in the determination of the phenomenological parameters like dampings or heat transfer coefficients. FP models are frequently adjusted by trial-and-error, which can lead to non-optimal results. On the other hand, the disadvantage of DD models is the need to handle multiple data sets in order to cover the range of system operation. A good compromise is given by considering a combined first-Principle Data-Driven (FPDD) model: the equations of the model are derived according to the physical laws, and some of their most relevant parameters are calibrated according to the operational data. This approach allows to obtain models which provide a satisfying compromise between accuracy and simplicity, Fig. 3.1.

To derive the physical mathematical representation of the technical system, three important aspects have to be considered, i.e., the system isolation, the inputs-outputs selection, and the model economy. System isolation allows various types of interactions existing between the system and its surrounding environment to be discerned. System isolation enables only the most relevant and important interactions to be selected and represented in the form of inputs and outputs. Using principles of isolation and selection, a model is always simplified in accordance to the purpose of modeling. The number of inputs and outputs of a model is limited due to economical-type constraints (model economy). The principle of economy implies the simplicity of a structure and that a minimal number of parameters and state variables are considered. Once derived the model equations, the calibration of their most relevant parameters consists of the selection of representative data sets, including a sufficiently broad operating range that the model proved could work correctly using a data fit measure and the adjustment of model

parameters to fit a model response to data.

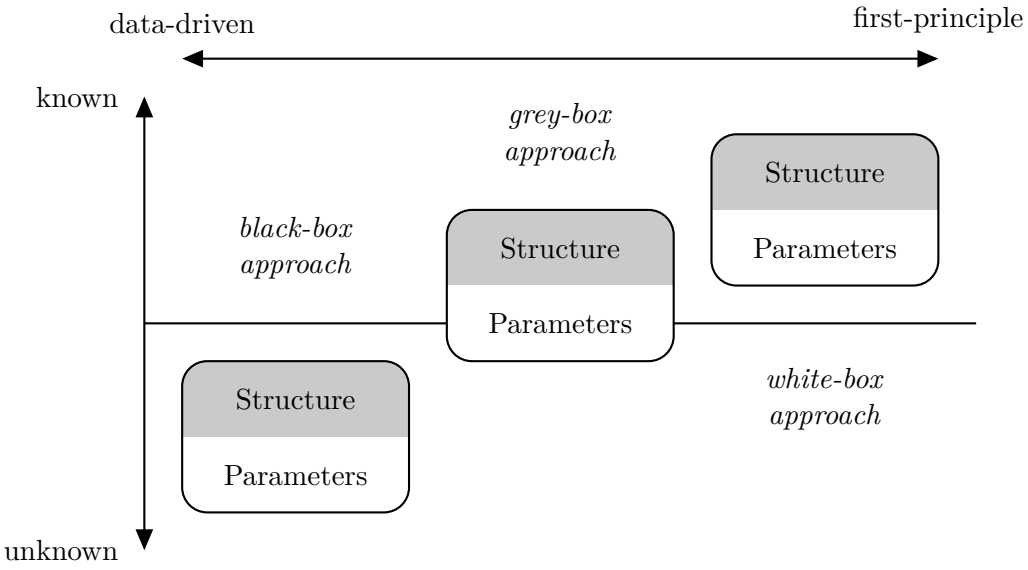


Figure 3.1: From the left to the right: Data-Driven (DD) Model, First-Principle Data-Driven (FPDD) Model, First-Principle (FP) Model. FPDD models provide a good compromise between simplicity and accuracy.

3.3 Design of Simulation Tools

Simulation of systems is an activity closely related to modelling and it is concerned with performing virtual experiments on a given plant model to analyse and predict the dynamic behaviour of a physical plant. Simulation is also a powerful tool for the validation of mathematical models and their approximations. In the context of CACSD, simulation is frequently used as a control system tuning-aid. Simulations are also useful for evaluating the performance of the control systems. For a reliable assessment, simulation-based analysis can be conducted at several different operative conditions in order to prove the performance robustness of the synthesized control system in the presence of parametric uncertainties, disturbances, and variabilities. By means of simulations, the control system can be tuned and tested without direct implementation in the real system, which can be time and cost consuming and may lead to unsafe operation of the hardware.

When analysing and comparing different simulation tools, the basic distinction is given by the term causality, Zupancic et al. (2008). This term can explain the evolution which was in the past from causal block-oriented tools to acausal object-oriented ones. Block-oriented tools, e.g. Matlab/Simulink, work with hierarchically interconnected blocks, where the connections between blocks conduct signals that transmit the values of

individual variables from the output of a block to the inputs of other blocks. The blocks process input information into output information. The hierarchically arranged block-oriented description clearly shows how the values of individual variables are calculated in the model – i.e., what the calculation algorithm is. The interconnection of blocks in Simulink thus reflects the calculation procedure rather than the actual structure of the modelled reality. This is called causal modelling. On the other hand, acausal object-oriented tools give the possibility to describe the individual parts of a model directly as a system of equations rather than an algorithm for solving the equations. The notation of models is declarative (the structure and mathematical relations are described, not the calculation algorithm) – thus the notation is acausal. Acausal modelling tools work with interconnected components that are instances of classes in which equations are directly defined. The components (i.e., instances of classes with equations) can be interconnected by means of precisely defined interfaces – connectors; this defines a system of equations. The latest versions of Simulink provides certain options for using acausal tools as well. Mathworks, the producer of the Matlab/Simulink simulation tools, responded to the new trends by creating a special acausal Simulink library – Simscape – and related domain libraries such as SimElectronics, SimHydraulics, SimMechanics, etc. A modern simulation language that is built directly on the acausal notation of models is Modelica, Fritzson (2010). It is clear that the object-oriented modelling is obviously much easier and more intuitive than block-based programs if the sub-models already do exist. But if the underlying equations are more complex, just building up these sub-models can become difficult, Richert et al. (2004).

3.4 Design of the Control System

In recent years, the main problem considered in Data Centers management is the minimization of energy consumptions and of management costs. The implementation of automatic control algorithms allows the achievement of these results. Generally, they play a central role in environment comfort satisfaction, efficient management, and interfacing of different systems. Certain aspects, such as the considerable number of subsystems composing Data Centers, and the heterogeneity of their operation in terms of physics domain and time scale, have favoured the integration of hierarchical control systems which coordinate the various parts of the system to meet Data Center energy performance specifications. A multi-layer hierarchical controller consists of a number of layers, each of which is constructed so as to perform functions providing specified services for an upper level and to determine, at a given time-scale, instructions required for operation of a

lower level. In a hierarchical control structure, CACSD tools assist the designer in the development of all the control layers, which may operate at different time-scales and may involve different physical domains. Usually, in a typical building energy management scenario, control systems are organized in three different levels, as depicted in Fig. 3.2.

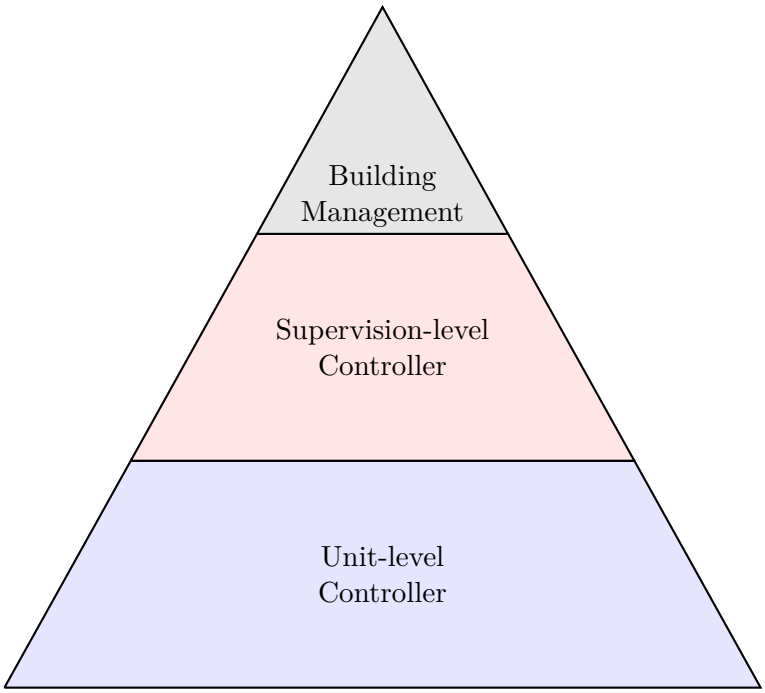


Figure 3.2: The three-layers hierarchical control architecture in a typical building energy management scenario.

The lower level is the Unit-Level Control, which has to maintain the set-point of each component of the system using local controllers, such as ON/OFF and standard regulators. The Supervision-Level Control generates the best combination of set-points for unit-level controllers in order to optimize the performance and/or the efficiency of the plant, e.g., comfort requirements on close control satisfaction and consumption minimization, ensuring a coordinated and smart control of individual components of the overall system. These goals can be carried out either using a dedicated hardware to which are connected all the unit controllers, or exploiting the individual units controllers hardware, as happens in simplest systems. Typically, complex HVAC systems are equipped with the two lower levels of Fig. 3.2, as shown in Fig. 3.3. However, it is more and more frequently required the ability to monitor the whole building, to integrate the cooling unit with the other subsystems and to have a central system able to provide monitoring and diagnostic functionalities. These tasks are carried out by the Building Management System, which represents the top of the controller.

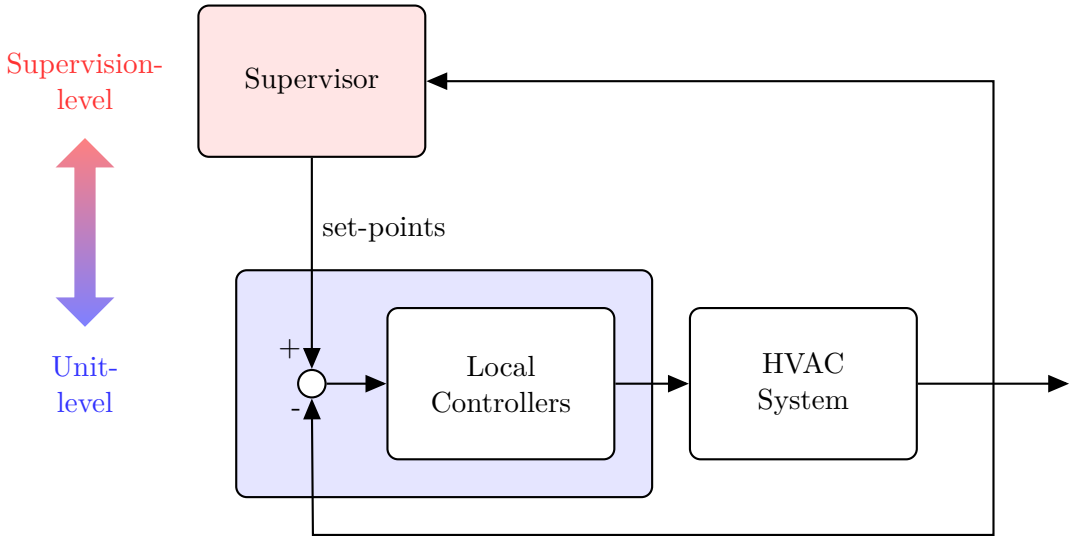


Figure 3.3: The two control layers for HVAC applications: the supervision-level provides the optimal set-points to the local controllers of the unit-level.

The main focus of this Thesis is the Supervision-Level Controller, which has the role to ensure coordinated and smart control of individual system components in order to guarantee the required performance and efficiency of the overall system. More specifically, the supervisory layer has to determine the values of the set-points for the low-level controllers that are optimal under a certain criterion, i.e., the set-points are computed as the solution of an optimization problem. In the context of this Thesis, the optimization problem consists in optimizing the energy efficiency of the cooling system, while provisioning the required amount of cooling power. For the solution of this optimization problem, the supervision layer can exploit model-based or model-free optimization methods. In all model-based optimization algorithms the first step is modelling the system, and then designing the controller based on the plant model with the faith that the model represents the true system, Hou and Wang (2013). These controller design methods would lose their utility if the model is inaccurate but, on the other hand, the more accurate the model is, the more effort or cost must be spent on the design of the model-based control system. When considering systems that are intrinsically complex due to their underlying physics and that exhibits, for example, non-linear, coupling, and time-varying behaviour, accurate modelling can be more difficult than control system design. Even if an accurate model of the system is available, its complexity makes it not suitable to use in practice and reduction of model or controller order must be performed. Modelling an accurate high-order model to target high performance for a control system design, then having to perform a controller order reduction or model simplification

for a low order controller, seems paradoxical. An alternative appealing solution is to design model-free data-driven optimization schemes, in which the controller optimizes the operation of the system by using only on-line or off-line I/O data or knowledge from the data processing without using explicit or implicit information of the mathematical model of the controlled process. Among all the model-free data-driven control techniques, Extremum Seeking Control (ESC) has drawn significant attention for optimizing the operation of HVAC systems due to its flexibility and robustness properties and by being model-free. In this Thesis, we consider the two-layer control architecture depicted in Fig. 3.3, where the supervisor finds the optimal set-points for the low-level controllers by means of an ESC algorithm, to whom the next Chapter is dedicated.

4

The Extremum Seeking Control Method

In this chapter, we consider the model-free ESC optimization algorithm. First, in Section 4.1 we introduce the main features and the advantages of the algorithm and in Section 4.2 we provide a brief historical survey of the method. Then, in Section 4.3 we provide to the reader an overview of the classic perturbation-based ESC scheme, describing its structure and some of its main properties. Finally, in Section 4.4 we propose a Newton-like Phasor-based ESC scheme, a variation of the classic pertubation-based ESC method that allows to improve the roboustness and the convergence properties of the optimization algorithm.

4.1 Introduction

The Extremum Seeking Control (ESC) denotes a class of model-free optimization methods which target the problem of maximizing (or minimizing) the steady-state value of a process output via real-time adaptation of one or more process inputs. The main benefits of all the ESC methods over other optimization techniques are that the optimization is based on feedback from online measurements[1] and that they do not require a model of the process for the optimization. By being model-free, ESC methods can be applied

[1]The output measurements can be directly used in the feedback or they can be filtered or mapped through some auxiliary functions that we like to optimize.

also in cases where some fundamental difficulties such as too limited available modelling resources and the presence of uncertainty prevent accurate modelling. Even if a model that is accurate enough can be obtained, modelling can be more difficult than control system design. Moreover, there is no well-recognized means of addressing certain types of complexity, such as that observed in plants whose parameters vary quickly or whose structures change over time, Hou and Wang (2013). In addition to the advantages of being model-free, ESC also inherits some of the flexibility and robustness against changes in the operating conditions which is associated with feedback-based methods. In particular, the feedback structure of ESC generally enables the methods to not only find an optimum, but also to track it if it moves over time, e.g. due to disturbances or process changes. ESC hence is an appropriate solution to most process control problems, where systems are typically very complex to be modelled and several disturbances are present.

ESC may be considered as a subfield of adaptive control. In adaptive control, the model used by the controller is updated online, using the information available in the measurements. A typical adaptive controller consists of two separate loops: one is a normal feedback loop, the other is a parameter-adjusting loop. In ESC, the model, which is adapted, is essentially the local gradient of the equilibrium map. A feedback loop is then applied to drive the system to a point, where the gradient is zero. Many methods in ESC are based on the gradient information of the performance function. In these methods, it is assumed that the cost function of the target system is continuous and has continuous derivatives. The optimal operating point for the system, which minimizes or maximizes the performance function, is the point with zero gradient. Thus, these methods can be regarded as an approximation of gradient descent. The most common ESC method is the perturbation-based ESC, where the gradient information is extracted from the response of the system due to the perturbation of the system inputs. Several variations of the classic algorithm have been presented in the literature, with the main goal of improving the convergence properties of the method. Among the others, the Phasor-based ESC presented in Atta et al. (2015) aims at improving the performance of the algorithm in the case of plants characterized by larger phase shifts. In this Thesis, we propose a Newton-like scheme to improve the convergence of the Phasor-based ESC by taking into account estimates of also the second order derivatives. In the following sections, after giving a brief historical overview of the ESC, we describe the classic perturbation-based ESC scheme and then we present the Newton-like Phasor-based ESC.

4.2 A Brief Historical Survey

As stated in Tan et al. (2010), the origin of ESC may be dated back to 1922, when Leblanc proposed a control mechanism to maintain the desirable maximum power transfer using what is essentially an extremum seeking solution. In his invention disclosure Leblanc (1922), Leblanc does not provide any analysis, nor a practical evaluation, of the control scheme. The first literature paper that describes in details and provide an evaluation of its performance may be the 1951 paper Draper and Li (1951), where authors consider the problem of optimizing an internal combustion engine. After that, ESC became a popular research topic in the 1950s and 1960s, like other forms of adaptive control. Citing Åström in his paper Astrom (1996), during those years there was very little theoretical analysis between the ideas and the real tests of adaptive controllers. Over the next three decades, research on ESC continued but clearly the mainstream adaptive control research emphasis had shifted to studying other forms of adaptive control that address the more demanding problem of system stability with performance control. The ESC started to receive significant interest from the control community again not until the turn of the century, largely due to the result presented in Wang and Krstic (2000). In that work, Krstic and Wang proved that the classic perturbation-based ESC method possesses a stable stationary solution in a local neighborhood of the steady-state optimum when applied to a wide class of nonlinear dynamic systems. This study renewed the interest in the field of ESC as is evident in the large number of publications following the breakthrough. The first method for multi-parameter extremum seeking was presented in Rotea (2000) and Walsh (2000), and the extremum seeking controller with a dynamic compensator from Krstić (2000) was extended to the multi-parameter case in Ariyur and Krstic (2002), in which a rigorous stability analysis was also provided. Beside the classic ESC scheme, other approaches have been considered in the decades. In the 1990s, a stochastic random signal was used as a perturbation instead of the classic periodic excitation, Spall (1998). Later, in Manzie and Krstic (2007) and Manzie and Krstic (2009), the stochastic approach was extended by adopting Gaussian distributed perturbation signals, and the convergence toward the extremum of a static map was demonstrated in addition to the case with dynamic actuators. Another extremum seeking approach is called sliding mode extremum seeking and it was first presented in Korovin and Utkin (1974). Later, this approach became popular and witnessed some additional publications, for example Drakunov et al. (1995) and Haskara et al. (2000), and subsequently, a stability analysis was presented in Pan et al. (2003). In Tan et al. (2013), the problem of using the ESC method for solving optimization problems with inputs constraints is considered. Due to the presence of an integrator in the classic perturbation-based ESC to drive the parameters that are being

adapted, it is possible that these parameters wander outside their physically relevant domain as the underlying adaptation technique is blind to this constraint. As these constraints may represent realistic operational limits, it is important to design ESC to respect them. Two such ESC schemes are proposed. One is based on a constrained optimization approach in which some penalty function is used to adapt the search so as not to violate the constraints. The other technique uses an anti-windup scheme, a widely used mechanism in engineering to prevent windup of integral action in a controller. In the paper, it is also demonstrated that both methods are essentially equivalent. After 2009, a new ESC approach based on the Newton optimization method became popular. This approach is called Newton-like ESC. In this method, the second derivative is required for optimization of one parameter, and the Hessian matrix is required for the multi-parameter case. The Newton-like approach for the one parameter case was presented in Moase et al. (2009) and Moase et al. (2010). The derivatives were calculated using an estimator, which provides the first and second derivatives of the objective with respect to the manipulated variable. A stability analysis was also presented. A multi-parameter Newton-like method was presented in Ghaffari et al. (2012) and Liu and Krstic (2014). In this work, the gradient was estimated using the classical band filtering approach (HPF, multiplier, and LPF). The signal for multiplication is a weighted form of the original perturbation signal as in the classic case, whereas the Hessian matrix is calculated using band pass filtering and the multiplication signal is carefully selected to eliminate other unwanted derivatives. The inverse of the Hessian matrix was calculated using a Riccati equation. This work also included a stability analysis. In Atta et al. (2015), an ESC algorithm that is based on estimating the phasor of the output instead of the gradient is presented. This method allows to deal with plants characterized by larger phase shifts, thus relaxing the assumption that the plant is a static map or that the perturbation frequency is very slow. The estimator is based on a variable gain observer which is derived from a continuous time Kalman filter. By means of simulations, it is also shown that the algorithm can be preferred in the case of systems with variable phase.

4.3 Classic Perturbation-based ESC

In this Section, we consider the classic perturbation-based ESC method which is illustrated in Fig. 4.1. Despite the origins of this scheme date back to 1922, nowadays this method is still considered the first choice in ESC, since it is applicable to a wide range of applications under relatively mild assumptions on the system structure and it is easy to implement.

In order to understand the operation of the loop scheme, let first consider its applica-

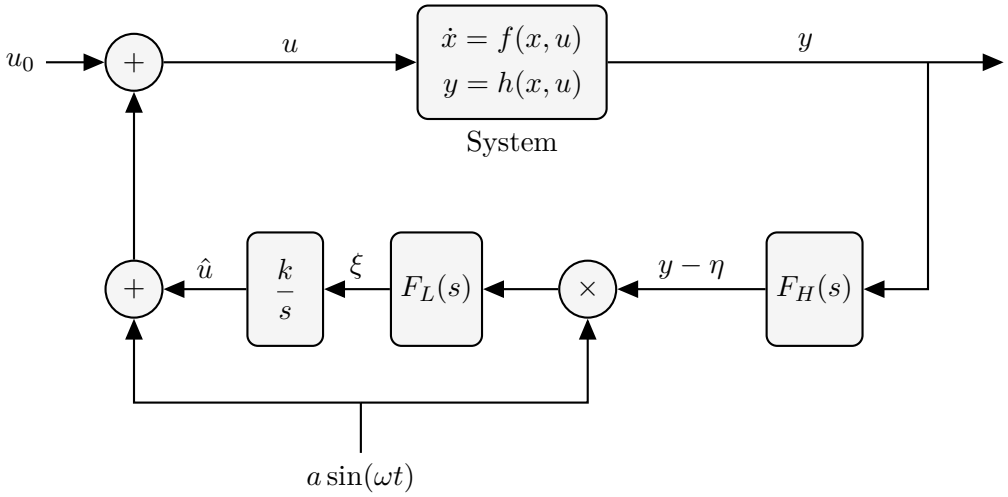


Figure 4.1: Sinusoidal perturbation based ESC scheme.

tion to a static plant. Assume first that the loop is opened at the integrator. In order to obtain an estimate of the gradient, the nominal input to the plant u_0 is perturbed by a sinusoidal signal. This will result in a plant output y given by a DC component corresponding to the current operating point and a sinusoidal component due to the perturbation. The amplitude of the oscillation in the output will approximately be scaled by the local gradient at the point of operation, i.e., the amplitude is modulated by the local gradient, see Fig. 4.2. In order to extract the gradient information from the output, the DC component is first removed by using a high-pass filter before the output is demodulated by multiplying it with the original perturbation signal. The demodulated signal then consists of a DC component proportional to the local gradient and some higher order harmonics. The latter are then attenuated by a low-pass filter, so the output ξ of the low-pass filter is essentially a scaled estimate of the local gradient. By closing now the loop with the integrator with a low gain, the operating point $\hat{u}$ will start to move in the direction of the gradient until a point of zero gradient is reached.

The choice of the perturbation signal $a\sin(\omega t)$ is important in the perturbation-based ESC scheme. A large value of the amplitude a is needed to get a good Signal-to-Noise Ratio (SNR) in order to overcome the noise in the output measurements and get a good estimation of the gradient. Moreover, as we will see later, the gradient estimate is proportional to the amplitude a, so it will affect the convergence rate and choosing a larger a is preferred. However, this choice implies large fluctuations in both the input and the output, which turn into stressing the plant and increasing the effect of nonlinearities. Thus, a trade off should be made in order to strike a balance between these two considerations: the value of a should be chosen properly to get a good estimation

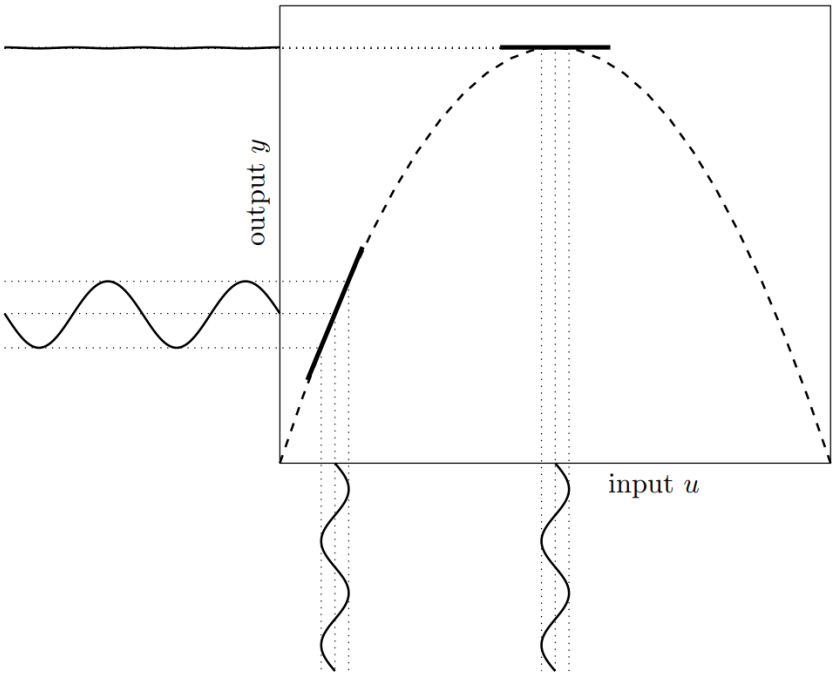


Figure 4.2: Principle for estimating the gradient of a static map.

of the gradient while maintaining the fluctuations as small as possible. Also the choice of the cut-off frequencies of the two filters is crucial in order to guarantee the correct operation of the scheme. The cut-off frequency of the high-pass filter should be set to a value lower than the frequency of the perturbation signal, so only the fluctuating component in the output signal which contains the gradient information passes through the filter. Also the cut-off frequency of the low-pass filter should be chosen lower than the perturbation frequency ω, in order to eliminate the high frequency component of the signal after the demodulation. Finally, also the gain k of the integrator may influence the whole control system. The gain essentially decides the bandwidth of the controller, which should normally be lower than that of the gradient estimator, since the updated gradient information should be provided to the controller. If the speed of the controller is higher than that of the estimator, the controller may make the whole system unstable. Since the bandwidth of the gradient estimator is decided by the low-pass filter frequency, which in turns depends on the choice of the perturbation frequency, the value of the integrator gain k should be chosen according to the value of the perturbation frequency.

Once the main elements of the scheme are settled, the perturbation-based ESC loop is able to find the optimal operating point automatically. Consider a general static

non-linear system, for which the performance function can be represented as:

$$y = J(\theta), \tag{4.1}$$

where $y \in \mathbb{R}$ is the output, $\theta \in \mathbb{R}$ is the input, and J is the static performance function that is assumed to be a C^2 function. Assume that the output of the system presents a maximum value and that the integrator gain k is positive. Also, the optimal input is denoted by θ^* while the current estimate of the optimal input is denoted by $\hat{\theta}$. The input error $\tilde{\theta}$ can thus be defined as:

$$\tilde{\theta} = \theta^* - \hat{\theta}. \tag{4.2}$$

By taking the second-order Taylor expansion about $\theta = \theta^*$ and dropping the higher order terms, it gives:

$$J(\theta) \approx J(\theta^*) + \frac{J''(\theta^*)}{2}(\theta - \theta^*)^2, \tag{4.3}$$

where $J''(\theta) > 0$ is the second-order derivative of $J(\theta)$ and the linear part of the Taylor expansion is missing because $J(\theta)$ has a maximum value at $\theta = \theta^*$. In the perturbation-based scheme, the input θ is equal to the sum of the current estimate $\hat{\theta}$ and the perturbation signal $a\sin(\omega t)$. By substituting θ in Eq. (4.3), expanding the square, and using the trigonometric identity $2\sin^2(\omega t) = 1 - \cos(2\omega t)$, it follows that:

$$y \approx J(\theta^*) + \frac{J''(\theta^*)}{4}a^2 + \frac{J''(\theta^*)}{2}\tilde{\theta}^2 - aJ''(\theta^*)\tilde{\theta}\sin(\omega t) - \frac{J''(\theta^*)}{4}a^2\cos(2\omega t). \tag{4.4}$$

The output signal y is passed through the high-pass filter, which removes the constant terms in (4.4):

$$y - \eta \approx \frac{J''(\theta^*)}{2}\tilde{\theta}^2 - aJ''(\theta^*)\tilde{\theta}\sin(\omega t) - \frac{J''(\theta^*)}{4}a^2\cos(2\omega t). \tag{4.5}$$

The output of the high-pass filter is demodulated by multiplication with the original perturbation, which gives:

$$a\sin(\omega t)(y - \eta) \approx \frac{aJ''(\theta^*)}{2}\tilde{\theta}^2\sin(\omega t) - a^2J''(\theta^*)\tilde{\theta}\sin^2(\omega t) - \frac{aJ''(\theta^*)}{4}a^2\cos(2\omega t)\sin(\omega t). \tag{4.6}$$

By using both the previous trigonometric identity and $2\cos(\omega t)\sin(\omega t) = \sin(3\omega t) -$

$\sin(\omega t)$, Eq. (4.6) becomes:

$$a\sin(\omega t)(y-\eta) \approx -\frac{a^2 J''(\theta^*)}{2}\tilde{\theta}^2 + \frac{a^2 J''(\theta^*)}{2}\tilde{\theta}\cos(2\omega t) + \frac{aJ''(\theta^*)}{2}\tilde{\theta}^2\sin(\omega t) + \\ + \frac{a^3 J''(\theta^*)}{8}(\sin(\omega t) - \sin(3\omega t)). \tag{4.7}$$

Assuming that the low-pass filter is perfect and completely removes any components with frequency above the break-off frequency, its output ξ can then be calculated as follows:

$$\xi \approx -\frac{a^2\tilde{\theta}J''(\theta^*)}{2}. \tag{4.8}$$

The differentiating of both sides of Eq. (4.2) gives $\dot{\tilde{\theta}} = -\dot{\hat{\theta}}$ and thus the evolution of the input error can be represented as:

$$\dot{\tilde{\theta}} = -\dot{\hat{\theta}} = -k\xi \approx \frac{ka^2 J''(\theta^*)}{2}\tilde{\theta}. \tag{4.9}$$

Since it is assumed that $k > 0$ while $J''(\theta) < 0$ since $J(\theta)$ has a maximum, it follows that:

$$\dot{\tilde{\theta}} \approx \frac{ka^2 J''(\theta^*)}{2}\tilde{\theta} < 0. \tag{4.10}$$

Thus, the designed scheme is stable and the input θ asymptotically converges to a neighbourhood of the optimal value θ^*.

It is worth stressing out that the analysis above is valid only for static systems. A difference between static systems and dynamic ones is that the latter will introduce a phase lag. Perturbation-based ESC works also for dynamic systems, under the hypothesis of very slow estimation and control. The average theory for nonlinear non-autonomous systems can be used to examine the stability and convergence of the extremum-seeking systems. In a more qualitative sense, the main idea behind averaging method is that a time-varying periodic system can be approximated by the time-invariant system derived from integrating the original system over a single period. Consequently, if the resulting time invariant system is stable around an equilibrium point, the time-varying periodic system converges to a periodic orbit around this equilibrium point. For more details on the average theory and its application for stability and convergence of nonlinear systems, one can refer to Chapter 9 in Khalil (2002). A detailed proof for convergence and stability of this scheme is given in Krstic and Wang (2000). The solutions of the closed-loop system converge to a small neighborhood of the extremum of the equilibrium map. The size of this neighborhood is proportional to the adaptation gain, the amplitude and

frequency of the periodic signal used in the design. Nešić (2009) explains the trade-off existing between the convergence rate and the size of the domain of attraction. Namely, if the domain of attraction is larger, the slower the convergence of the algorithm will be. In Chioua et al. (2007) it was shown that the neighborhood of convergence is proportional to the square of the dither frequency. The implication of this is that even when the amplitude of excitation goes to zero, the size of this neighborhood will not go to zero. Another aspect that must be taken into account into the design of perturbation-based ESC methods is that they require a very slow perturbation in such a way that the system will appear as a static map. This ensures convergence to the optimal solution, but this also slows down the system response. Increasing the perturbation frequency ω would allow to increase the integrator gain k, which would lead to a faster response but it may lead to a sub-optimal solution Trollberg et al. (2014) since it can no longer be considered as a static map. The problem can be mitigated by adding a phase compensation, but this may instead lead to instability of the overall system especially in the case of time varying phase lag. Another appealing solution to this problem is given by the ESC based on phasor estimation presented in Atta et al. (2015). This approach allows to deal with plants which exhibit larger phase shifts, relaxing the assumption that the plant is a static map or that the perturbation frequency is very slow. In the next Section, a Newton-like phasor-based ESC is presented, where the derivative estimator underlying the phasor approach is augmented in order to capture also the Hessian of the plant index, and then these estimates are used to steer the system along a Newton-like direction.

4.4 Newton-like Phasor-based ESC

Among the characterizing properties of any optimization strategy, the speed of convergence to an optimal solution plays a crucial role in applications Beiranvand et al. (2017). Within the context of process control, a faster convergence to the economic point of operation means a more efficient usage of the resources, and a potential reduction of the environmental impact. Here we consider on-line optimization problems for a multi-variable plant with unknown nonlinear dynamics and scalar measurable index. Toward faster optimizing controllers, a Newton-like acceleration of the phasor-based ESC Atta et al. (2015) is proposed. The scheme is inspired by the work in Moase et al. (2010), in which the steepest descent strategy of the classical perturbation based ESC is substituted by a Newton-like descent. This study proposes and evaluates a phasor-based strategy in which the first and second-order features of the plant's performance index are filtered from a truncated Fourier expansion of the output.

We address nonlinear control systems of the form

$$\Sigma \triangleq \begin{cases} \dot{\mathbf{x}} = f(\mathbf{x}, \boldsymbol{\theta}) \\ y^{\mathrm{o}} = h(\mathbf{x}) \end{cases} \tag{4.11}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state, $\boldsymbol{\theta} \in \mathbb{R}^m$ is the control value, $y^{\mathrm{o}} \in \mathbb{R}$ is the scalar measured output, and the vector fields f, h are sufficiently smooth. In this context, the plant index y^{o} has the meaning of either a cost or a utility. Without loss of generality, we consider only minimization problems in the remainder of this Chapter.

Assumption 4.4.1. There exists a sufficiently smooth function $\ell : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x} = \ell(\boldsymbol{\theta}). \tag{4.12}$$

Assumption 4.4.2. For each $\boldsymbol{\theta} \in \mathbb{R}^m$, the equilibrium point $\ell(\boldsymbol{\theta})$ is locally exponentially stable for the autonomous dynamics $\dot{\mathbf{x}} = f(\mathbf{x}, \boldsymbol{\theta})$ with constants uniform in $\boldsymbol{\theta}$.

Let $Q : \mathbb{R}^m \to \mathbb{R}$, $Q \triangleq h \circ \ell$, denote the steady state value of the plant index given the parameter. To ensure global convergence, we require the following characterization.

Assumption 4.4.3. The index Q is strongly convex (with everywhere positive definite Hessian).

We stress that f, h in (4.11), ℓ in (4.12), and the steady state index Q are otherwise unknown. The control task is to optimize the plant operation by steering Σ along the profitable directions inferred from the continuous measurement of y^{o}. This entails to learn the optimal parameters online,

$$\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \; Q(\boldsymbol{\theta}), \tag{4.13}$$

and then to bring the system operation near this optimum. Here, we take an ESC controller to be a new dynamical system of the form

$$\begin{cases} \dot{\boldsymbol{\zeta}}(t) = f_{\mathrm{esc}}\left(t, \boldsymbol{\zeta}(t), y^{\mathrm{o}}(t)\right) & \text{(4.14a)} \\ \dot{\widehat{\boldsymbol{\theta}}}(t) = \mathrm{diag}(\mathbf{k}) h_{\mathrm{esc}}\left(\boldsymbol{\zeta}(t), y^{\mathrm{o}}(t)\right) & \text{(4.14b)} \\ \boldsymbol{\theta}(t) = \widehat{\boldsymbol{\theta}}(t) + \mathrm{diag}(\mathbf{a})\mathbf{s}(t) & \text{(4.14c)} \end{cases}$$

where: f_{esc} is the dynamics of the direction finder and h_{esc} a mapping of the internal state to the estimated descent direction; $\widehat{\boldsymbol{\theta}}$ is the estimated minimizer of (4.13) and $\mathbf{k} \in \mathbb{R}^m_{>0}$ a

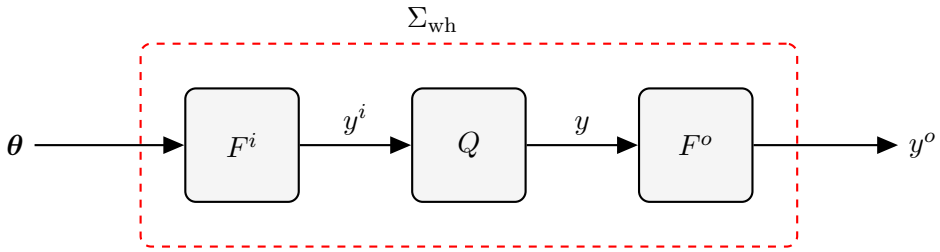


Figure 4.3: For the purpose of deriving the formal structure of the derivative estimator, the plant index is locally approximated by a Wiener-Hammerstein model.

vector of integration gains affecting the convergence speed; $\mathbf{s} = \mathbf{s}(t) \in \mathbb{R}^m$ is a vector of additive sine waves

$$\mathbf{s}_k(t) = \sin(\omega_k t), \quad k = 1, \ldots, m \tag{4.15}$$

with $\omega_1 < \omega_2 < \ldots < \omega_m$, while $\mathbf{a} \in \mathbb{R}^m_{>0}$ is a vector of perturbation amplitudes.

Remark 4.4.4. Similarly to Ghaffari et al. (2012) and Atta et al. (2017), we handle multi-variable problems by injecting independent sinusoidal perturbations to each input channel. The corresponding angular frequencies must be chosen such that for any choice of distinct indices $k, v, z = 1, \ldots, m$ there holds: *i)* $\omega_k \neq 2\omega_j$, *ii)* $\omega_k \pm \omega_v \neq \omega_z$, and *iii)* $\omega_k \pm \omega_v \neq 2\omega_z$.

For the purpose of deriving the estimator, we focus the following class of Wiener-Hammerstein models (see Fig. 4.3):

$$\Sigma_{\text{wh}} \triangleq (F^{\text{i}}, Q, F^{\text{o}}) \tag{4.16}$$

where F^{i} is an $m \times m$ input transfer function, $Q : \mathbb{R}^m \to \mathbb{R}$ is a sufficiently smooth nonlinearity[2], and F^{o} is a scalar output transfer function. For the sake of a more compact notation, as in Rotea (2000), we restrict our attention to diagonal F^{i}s and thus address problems in which the input cross-couplings are negligible. We moreover assume[3] $|F^{\text{i}}(0)| = |F^{\text{o}}(0)| = 1$. We stress that this simplified setting is a means to derive the functional structure of the derivative estimator. This block can then be applied to the more generic setting (4.11) as shown later. We approximate the input/output mapping of (4.16) by exploiting the underlying model structure. In particular, we consider the

[2]We warn the reader of the notational abuse: throughout this Section Q denotes the static nonlinearity of (4.16) and not the steady state output of (4.11) as everywhere else.
[3]Since non-unit gains can otherwise be moved into Q.

component-wise perturbations (cf. (4.14))

$$\boldsymbol{\theta}(t) = \overline{\boldsymbol{\theta}} + \operatorname{diag}(\mathbf{a})\mathbf{s}(t) \tag{4.17}$$

where $\overline{\boldsymbol{\theta}}$ is a constant parameter value in $\mathbb{R}^m$ and write the output of the linear input block as the sum of an average and perturbed contributions $\mathbf{y}^{\mathrm{i}} \approx \overline{\mathbf{y}}^{\mathrm{i}} + \widetilde{\mathbf{y}}^{\mathrm{i}}$ using

$$\overline{\mathbf{y}}^{\mathrm{i}}_k(t) \triangleq \sum_{v=1}^{m} |F^{\mathrm{i}}_{kv}(0)|\overline{\boldsymbol{\theta}}_v = |F^{\mathrm{i}}_{kk}(0)|\overline{\boldsymbol{\theta}}_k = \overline{\boldsymbol{\theta}}_k \tag{4.18}$$

$$\begin{aligned} \widetilde{\mathbf{y}}^{\mathrm{i}}_k(t) &\triangleq \sum_{v=1}^{m} |F^{\mathrm{i}}_{kv}(j\omega_v)|\mathbf{a}_v \sin\Big(\omega_v t + \arg F^{\mathrm{i}}_{kv}(j\omega_v)\Big) \\ &= |F^{\mathrm{i}}_{kk}(j\omega_k)|\mathbf{a}_k \sin\Big(\omega_k t + \arg F^{\mathrm{i}}_{kk}(j\omega_k)\Big) \end{aligned} \tag{4.19}$$

for $k = 1, \ldots, m$. We propagate the harmonics (4.19) through the static nonlinearity Q by approximating the latter with its truncated Taylor expansion (about the value $\overline{\mathbf{y}}^{\mathrm{i}}$ and along the direction $\widetilde{\mathbf{y}}^{\mathrm{i}}$) as[4]

$$y \approx Q(\overline{\mathbf{y}}^{\mathrm{i}}) + (\nabla Q)^{\top}\widetilde{\mathbf{y}}^{\mathrm{i}} + \frac{1}{2}(\widetilde{\mathbf{y}}^{\mathrm{i}})^{\top}(\mathbf{H}Q)\widetilde{\mathbf{y}}^{\mathrm{i}}. \tag{4.20}$$

After applying trigonometric identities, the latter response is rewritten as a linear combination of sines and cosines at angular frequencies $\{\omega_k, \omega_k - \omega_v, \omega_k + \omega_v\}_{k,v=1}^{m}$. Finally, the measured variable y^{o} is captured as the output of the linear transfer function F^{o} with (4.20) as input, resulting in a truncated Fourier series[5]. Define the shorthands

$$\varrho^{i}_{\omega_k} \triangleq |F^{\mathrm{i}}_{kk}(j\omega_k)|, \quad \varphi^{i}_{\omega_k} \triangleq \arg F^{\mathrm{i}}_{kk}(j\omega_k), \tag{4.21}$$

$$\varrho^{o}_{\omega_k} \triangleq |F^{\mathrm{o}}(j\omega_k)|, \quad \varphi^{o}_{\omega_k} \triangleq \arg F^{\mathrm{o}}(j\omega_k). \tag{4.22}$$

[4] We let $\cdot^{\top}$ denote transposition and adopt the compact notations $\nabla Q = \nabla Q(\overline{\mathbf{y}}^{\mathrm{i}})$ and $\mathbf{H}Q = \mathbf{H}Q(\overline{\mathbf{y}}^{\mathrm{i}})$.
[5] We adopt this terminology to aid in the exposition, even if the basis vectors are not normalized.

Then

$$
\begin{aligned}
y^{\mathrm{o}}(t) \approx\; & Q(\bar{\mathbf{y}}^{\mathrm{i}}) + \frac{1}{4}\sum_{k=1}^{m}(\mathbf{H}Q)_{kk}(\varrho^{o}_{\omega_k}\varrho^{i}_{\omega_k}\mathbf{a}_k)^2 \\
& + \sum_{k=1}^{m}(\nabla Q)_k \varrho^{o}_{\omega_k}\varrho^{i}_{\omega_k}\mathbf{a}_k \cos\left(\varphi^{i}_{\omega_k} + \varphi^{o}_{\omega_k}\right)\sin\left(\omega_k t\right) \\
& + \sum_{k=1}^{m}(\nabla Q)_k \varrho^{o}_{\omega_k}\varrho^{i}_{\omega_k}\mathbf{a}_k \sin\left(\varphi^{i}_{\omega_k} + \varphi^{o}_{\omega_k}\right)\cos\left(\omega_k t\right) \\
& + \frac{1}{4}\sum_{k=1}^{m}(\mathbf{H}Q)_{kk}(\varrho^{o}_{\omega_k}\varrho^{i}_{\omega_k}\mathbf{a}_k)^2 \\
& \quad \cdot \Big(-\cos\left(2\varphi^{i}_{\omega_k} + \varphi^{o}_{2\omega_k}\right)\cos\left(2\omega_k t\right) \\
& \qquad + \sin\left(2\varphi^{i}_{\omega_k} + \varphi^{o}_{2\omega_k}\right)\sin\left(2\omega_k t\right)\Big) \\
& + \frac{1}{2}\sum_{\substack{k,v=1,\dots,m \\ v<k}}(\mathbf{H}Q)_{kv}\varrho^{o}_{\omega_k+\omega_v}\varrho^{i}_{\omega_k}\mathbf{a}_k\varrho^{i}_{\omega_v}\mathbf{a}_v \\
& \quad \cdot \Big(-\cos\left(\varphi^{i}_{\omega_k} + \varphi^{i}_{\omega_v} + \varphi^{o}_{\omega_k+\omega_v}\right)\cos\left(\omega_k t + \omega_v t\right) \\
& \qquad + \sin\left(\varphi^{i}_{\omega_k} + \varphi^{i}_{\omega_v} + \varphi^{o}_{\omega_k+\omega_v}\right)\sin\left(\omega_k t + \omega_v t\right)\Big) \\
& + \frac{1}{2}\sum_{\substack{k,v=1,\dots,m \\ v<k}}(\mathbf{H}Q)_{kv}\varrho^{o}_{\omega_k-\omega_v}\varrho^{i}_{\omega_k}\mathbf{a}_k\varrho^{i}_{\omega_v}\mathbf{a}_v \\
& \quad \cdot \Big(\cos\left(\varphi^{i}_{\omega_k} - \varphi^{i}_{\omega_v} + \varphi^{o}_{\omega_k-\omega_v}\right)\cos\left(\omega_k t - \omega_v t\right) \\
& \qquad - \sin\left(\varphi^{i}_{\omega_k} - \varphi^{i}_{\omega_v} + \varphi^{o}_{\omega_k-\omega_v}\right)\sin\left(\omega_k t - \omega_v t\right)\Big).
\end{aligned}
\tag{4.23}
$$

In what follows, for the sake of a more compact notation, we let $\boldsymbol{\zeta}_0$ indicate the constant term in (4.23) and $\boldsymbol{\zeta}^{\mathrm{c}}_k, \boldsymbol{\zeta}^{\mathrm{s}}_k$ capture the coefficients corresponding to terms $\cos(\omega_k t)$ and $\sin(\omega_k t)$, respectively. We moreover let $\boldsymbol{\zeta}^{\mathrm{c}\pm}_{k,v}$ be the coefficient of $\cos(\omega_k t \pm \omega_v t)$ and extend this notation to the sine components using $\boldsymbol{\zeta}^{\mathrm{s}\pm}_{k,v}$.

The unknown coefficients in (4.23) include as factors all the entries of ∇Q and $\mathbf{H}Q$, and thus encode first and second order local information about the nonlinearity. Following Atta et al. (2015), the latter information is retrieved using a periodic gain observer. Formally, (4.23) is first rewritten as the inner product

$$
y^{\mathrm{o}}(t) \approx \boldsymbol{\chi}(t)^{\top}\boldsymbol{\zeta}, \tag{4.24}
$$

where $\boldsymbol{\chi}(t)$ is obtained by stacking the sine and cosine basis functions. By assimilating

$\boldsymbol{\zeta}(t)$ to a vector of state variables, we introduce the Luenberger observer

$$\dot{\widehat{\boldsymbol{\zeta}}}(t) = L(t)\left(y^{\mathrm{o}}(t) - \boldsymbol{\chi}^{\top}(t)\widehat{\boldsymbol{\zeta}}(t)\right), \tag{4.25}$$

where the entries of the gain are, in general, amplified and phase-shifted versions of the entries of $\boldsymbol{\chi}(t)$. For example, in the case $m = 1$, letting

$$\boldsymbol{\chi}(t) \triangleq \begin{bmatrix} 1 & \sin(\omega t) & \cos(\omega t) & \sin(2\omega t) & \cos(2\omega t) \end{bmatrix}^{\top}, \tag{4.26}$$

one has the general form

$$L(t) = \begin{bmatrix} l_1 & l_2 \sin\left(\omega t + \varphi_1^L\right) & l_2 \cos\left(\omega t + \varphi_1^L\right) & \cdots & l_3 \sin\left(2\omega t + \varphi_2^L\right) & l_3 \cos\left(2\omega t + \varphi_2^L\right) \end{bmatrix}. \tag{4.27}$$

Remark 4.4.5. As in Atta et al. (2015), one may alternatively consider the dynamics

$$\begin{cases} \dot{\boldsymbol{\zeta}}(t) = \boldsymbol{\varepsilon}_1(t) \\ y^{\mathrm{o}}(t) = \boldsymbol{\chi}(t)^{\top}\boldsymbol{\zeta}(t) + \varepsilon_2(t) \end{cases} \tag{4.28}$$

where $\boldsymbol{\varepsilon}_1, \varepsilon_2$ are zero mean Gaussian random processes with covariances $V \triangleq \mathrm{var}(\boldsymbol{\varepsilon}_1)$ and $r \triangleq \mathrm{var}(\varepsilon_2)$. A Kalman-Bucy observer gain structured similarly to (4.27) can then be evaluated by solving a periodic Riccati Differential Equation (RDE):

$$\dot{P}(t) = V - rL(t)L^{\top}(t), \quad L(t) = r^{-1}P(t)\chi(t). \tag{4.29}$$

The state (4.25) captures the derivatives' estimates only up to unknown multiplicative factors. For the purpose of driving the ESC we adopt

$$\begin{aligned} (\widehat{\nabla Q})_k &\triangleq \frac{\widehat{\boldsymbol{\zeta}}_k^{\mathrm{s}}}{\mathbf{a}_k} \approx (\nabla Q)_k \varrho^o_{\omega_k} \varrho^i_{\omega_k} \cos\left(\varphi^i_{\omega_k} + \varphi^o_{\omega_k}\right), \\ (\widehat{\mathbf{H}Q})_{kk} &\triangleq -4\frac{\widehat{\boldsymbol{\zeta}}_{kk}^{\mathrm{c}+}}{\mathbf{a}_k^2} \\ &\approx (\mathbf{H}Q)_{kk} (\varrho^o_{\omega_k} \varrho^i_{\omega_k})^2 \cos\left(2\varphi^i_{\omega_k} + \varphi^o_{2\omega_k}\right), \\ (\widehat{\mathbf{H}Q})_{kv} &\triangleq -2\frac{\widehat{\boldsymbol{\zeta}}_{kv}^{\mathrm{c}+}}{\mathbf{a}_k \mathbf{a}_v} \\ &\approx (\mathbf{H}Q)_{kv} \varrho^o_{\omega_k+\omega_v} \varrho^i_{\omega_k} \varrho^i_{\omega_v} \cos\left(\varphi^i_{\omega_k} + \varphi^i_{\omega_v} + \varphi^o_{\omega_k+\omega_v}\right). \end{aligned} \tag{4.30}$$

The amplitude distortion is induced by the frequency dependent amplitudes and phase

shifts in (4.21)-(4.22). To reduce the effect of the unknown gains $\varrho_\star^i$, $\varrho_\star^o$ it is possible to tune the design parameters $\omega_1, \ldots, \omega_m$, trading off between faster convergence and smaller estimation errors.

Finally, computing the Newton direction requires the computation of the inverse of the estimated Hessian. As in Ghaffari et al. (2012), we consider the RDE

$$\dot{\mathbf{\Gamma}} = \omega_\Gamma \left(\mathbf{\Gamma} - \mathbf{\Gamma} \widehat{\mathbf{H}Q} \mathbf{\Gamma} \right), \tag{4.31}$$

where $\omega_\Gamma \in \mathbb{R}_{>0}$ is a tuning gain, and the state $\mathbf{\Gamma}$ is an estimate of $(\widehat{\mathbf{H}Q})^{-1}$. We design a Newton-like acceleration for the phasor ESC by specializing the general scheme in (4.14). Specifically, we redefine:

- f_{esc} as the state-update law in (4.25), augmented by the Hessian inversion dynamics in (4.31), that is,

$$f_{\text{esc}} \left(t, \begin{bmatrix} \widehat{\boldsymbol{\zeta}} \\ \mathbf{\Gamma} \end{bmatrix}, y^{\text{o}} \right) \triangleq \begin{bmatrix} L(t) \left(y^{\text{o}}(t) - \boldsymbol{\chi}^\top(t) \widehat{\boldsymbol{\zeta}}(t) \right) \\ \omega_\Gamma \left(\mathbf{\Gamma}(t) - \mathbf{\Gamma}(t) \widehat{\mathbf{H}Q} \mathbf{\Gamma}(t) \right) \end{bmatrix} \tag{4.32}$$

  thus endowing the ESC with a phasor-based derivative estimator;

- h_{esc} as the Newton-like direction

$$h_{\text{esc}}(\widehat{\boldsymbol{\zeta}}) \triangleq -\mathbf{\Gamma} \widehat{\nabla Q}. \tag{4.33}$$

It can be shown that the proposed Newton-like ESC enjoys stability properties similar to those of the classical perturbation-based multi-variable ESC. In the following, we discuss the single-variable results, using the estimators (4.30), and reserve the lengthier treatment of the multi-variable case to a future work.

Proposition 1. Consider $m = 1$, $\varphi_i^{\text{L}} \in (-\pi/2, \pi/2)$ and $l_1, \ldots, l_3 > 0$ in (4.27), f_{esc} and h_{esc} as in (4.32)-(4.33). For a sufficiently small, the Newton-like phasor ESC renders an asymptotically stable equilibrium point for the averaged dynamics of the closed-loop

system

$$\begin{cases} \dot{\mathbf{x}}(t) = f\left(\mathbf{x}(t), \theta(t)\right) \\ y^{\mathrm{o}}(t) = h(\mathbf{x}(t)) \\ \dot{\widehat{\boldsymbol{\zeta}}}(t) = L(t)\left(y^{\mathrm{o}}(t) - \boldsymbol{\chi}^{\top}(t)\widehat{\boldsymbol{\zeta}}(t)\right) \\ \dot{\boldsymbol{\Gamma}}(t) = \omega_{\Gamma}\left(\boldsymbol{\Gamma}(t) - \boldsymbol{\Gamma}(t)\widehat{\mathbf{H}Q}(t)\boldsymbol{\Gamma}(t)\right) \\ \dot{\widehat{\theta}}(t) = -k\boldsymbol{\Gamma}(t)\widehat{\nabla Q}(t) \\ \theta(t) = \widehat{\theta}(t) + a\sin(\omega t) \end{cases} \tag{4.34}$$

and such that the equilibrium parameter $\widehat{\overline{\theta}}$ lies in a neighbourhood of the optimum θ^* of (4.13). In particular, there holds

$$\widehat{\overline{\theta}} = \theta^* - \frac{a^2}{8}\frac{\partial^3 Q(\theta^*)}{\partial\theta^3}\left(\mathbf{H}Q(\theta^*)\right)^{-1} + O(a^3). \tag{4.35}$$

A proof of Proposition 1 can be developed along the lines of the classical ESC stability proof Krstic and Wang (2000).

Proof. The complete stability proof has three main steps: 1) first, the slow controller dynamics and the fast plant dynamics are segregated; 2) then, the reduced and averaged system accounting for the slow dynamics alone is shown to admit an asymptotically stable equilibrium point; 3) finally, a singular perturbation argument is invoked to prove the stability of the closed loop. Here, we treat in some detail only the study of the averaged dynamics (parts 1 and 2). Part 3 follows by applying well known arguments from Krstic and Wang (2000).

We start by scaling time using $\tau = \omega t$ and derive the closed-loop dynamics

$$\begin{cases} \omega\dfrac{d\mathbf{x}}{d\tau} = f\left(\mathbf{x}, \widehat{\theta} + a\sin(\tau)\right) \\ \dfrac{d\widehat{\boldsymbol{\zeta}}}{d\tau} = \delta k' L\left(y^{\mathrm{o}} - \boldsymbol{\chi}^{\top}\widehat{\boldsymbol{\zeta}}\right) \\ \dfrac{d\boldsymbol{\Gamma}}{d\tau} = \delta k''\left(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\widehat{\mathbf{H}Q}\boldsymbol{\Gamma}\right) \\ \dfrac{d\widehat{\theta}}{d\tau} = -\delta k'\boldsymbol{\Gamma}\widehat{\nabla Q} \end{cases} \tag{4.36}$$

where the state variables, inputs, outputs, and the observer gain are implicitly evaluated at time τ/ω. We notice that applying also part 3 of the framework in Krstic and Wang (2000) requires $\delta > 0$ and $k' \triangleq 1/(\omega\delta)$, $k'' \triangleq \omega_{\Gamma}/(\omega\delta) > 0$ to be sufficiently small scalars, with the effect of limiting the perturbation frequency ω.

We then set to study the averaged dynamics (see Khalil (2002), Chapter 10) of the slow subsystem about the optimal parameter θ^*. Neglecting the plant's transients, we set $\mathbf{x} = \ell\big(\widehat{\theta} + a\sin(\tau)\big)$, $y^{\mathrm{o}} = Q\big(\widehat{\theta} + a\sin(\tau)\big)$, and consider

$$\begin{cases} \dfrac{d\widehat{\boldsymbol{\zeta}}^{\mathrm{a}}}{d\tau} = \delta k' \left(-A\widehat{\boldsymbol{\zeta}}^{\mathrm{a}} + B\Psi\right) & (4.37) \\ \dfrac{d\boldsymbol{\Gamma}^{\mathrm{a}}}{d\tau} = \delta k'' \left(\boldsymbol{\Gamma}^{\mathrm{a}} - \boldsymbol{\Gamma}^{\mathrm{a}}\widehat{\mathbf{H}Q}^{\mathrm{a}}\boldsymbol{\Gamma}^{\mathrm{a}}\right) & (4.38) \\ \dfrac{d\widehat{\theta}^{\mathrm{a}}}{d\tau} = -\delta k' \boldsymbol{\Gamma}^{\mathrm{a}}\widehat{\nabla Q}^{\mathrm{a}} & (4.39) \end{cases}$$

where

$$A \triangleq \operatorname{diag}\left(l_1, \frac{l_2}{2}\Phi_1, \frac{l_3}{2}\Phi_2\right),$$

$$B \triangleq \operatorname{diag}\left(l_1, l_2\,\Phi_1, l_3\,\Phi_2\right), \quad \Phi_j \triangleq \begin{bmatrix} \cos(\varphi_j^L) & \sin(\varphi_j^L) \\ -\sin(\varphi_j^L) & \cos(\varphi_j^L) \end{bmatrix},$$

φ_1^L, φ_2^L are the phase-shifts in (4.27), and

$$\Psi \triangleq \frac{1}{2\pi}\int_0^{2\pi} \left(Q(\widehat{\theta}^{\mathrm{a}} + a\sin\tau)\cdot\begin{bmatrix}1 & \sin\tau & \cos\tau & \sin 2\tau & \cos 2\tau\end{bmatrix}^\top\right)d\tau.$$

Given that A is invertible, (4.37) admits the equilibrium

$$\widehat{\overline{\boldsymbol{\zeta}}}^{\mathrm{a}} = A^{-1}B\Psi = \operatorname{diag}(1,2,2,2,2)\,\Psi. \tag{4.40}$$

Notice that $\widehat{\overline{\boldsymbol{\zeta}}}^{\mathrm{a}}$ is parametrized by $\widehat{\overline{\theta}}^{\mathrm{a}}$ through the dependence on Ψ. We proceed to study the $\widehat{\overline{\boldsymbol{\zeta}}}^{\mathrm{a}}$ that induces $\widehat{\nabla Q}^{\mathrm{a}} = 0$. If the estimate of the gradient vanishes then

$$\widehat{\overline{\boldsymbol{\zeta}}}_1^{\mathrm{a,s}} = 0. \tag{4.41}$$

Substituting (4.41) in (4.40), we deduce in particular

$$\int_0^{2\pi} Q\left(\widehat{\overline{\theta}}^{\mathrm{a}} + a\sin\tau\right)\sin\tau\, d\tau = 0.$$

Taking a Taylor expansion of Q (about $\theta^\star$) to develop the above integral reveals

$$\widehat{\overline{\theta}}^{\mathrm{a}} - \theta^* = -\frac{a^2}{8}\frac{\partial^3 Q(\theta^*)}{\partial\theta^3}\left(\mathbf{H}Q(\theta^*)\right)^{-1} + O(a^3), \tag{4.42}$$

namely the estimate in (4.35). Developing the remaining integrals in (4.40), taking into

account that $\widehat{\overline{\theta}}^{\mathrm{a}} - \theta^* = O(a^2)$, gives

$$\begin{aligned}\widehat{\overline{\boldsymbol{\zeta}}}^{\mathrm{a}} = \Big[& Q(\theta^*) + \frac{a^2}{4}\mathbf{H}Q(\theta^*) + O(a^3), \\ & 0, \;\; 0, \;\; 0, \;\; -\frac{a^2}{4}\mathbf{H}Q(\theta^*) + O(a^3)\Big]^\top.\end{aligned} \tag{4.43}$$

We consider then the following equilibrium state[6] for (4.38)

$$\widehat{\overline{\mathbf{\Gamma}}}^{\mathrm{a}} = \left(\widehat{\mathbf{H}Q}^{\mathrm{a}}\right)^{-1}, \tag{4.44}$$

namely the inverse of the Hessian estimate evaluated at (4.43). Notice that $\widehat{\overline{\mathbf{\Gamma}}}^{\mathrm{a}}$ is strictly positive for sufficiently small a. The Jacobian of (4.37)-(4.39) at the point given by (4.42), (4.43) and (4.44) reads

$$J^{\mathrm{a}} = \delta \begin{bmatrix} -k'l_1 & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & 0 & k'l_1C_2 \\ \mathbf{0}_{2\times 1} & -\frac{k'l_2}{2}\Phi_1 & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} & \frac{k'l_2}{2}\Phi_1 C_3 \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} & -\frac{k'l_3}{2}\Phi_2 & \mathbf{0}_{2\times 1} & * \\ 0 & \mathbf{0}_{1\times 2} & * & -k'' & 0 \\ 0 & C_1 & \mathbf{0}_{1\times 2} & 0 & 0 \end{bmatrix},$$

$$C_1 = \begin{bmatrix} -k' \left(\mathbf{H}Q(\theta^*) + O(a)\right)^{-1} & 0 \end{bmatrix},$$

$$C_2 = \frac{1}{2\pi}\int_0^{2\pi} \nabla Q(\widehat{\overline{\theta}}^{\mathrm{a}} + a\sin\tau)d\tau,$$

$$C_3 = \frac{1}{2\pi}\int_0^{2\pi} \nabla Q(\widehat{\overline{\theta}}^{\mathrm{a}} + a\sin\tau)\begin{bmatrix}\sin\tau & \cos\tau\end{bmatrix}^\top d\tau.$$

Applying a full rank transformation T to rearrange the rows and T^{-1} to rearrange the columns, the Jacobian is recast in the block lower-triangular form

$$\widetilde{J}^{\mathrm{a}} = \delta \left[\begin{array}{ccccc} [ccc|c|c] - k'l_1 & \mathbf{0}_{1\times 2} & k'l_1C_2 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{2\times 1} & -\frac{k'l_2}{2}\Phi_1 & \frac{k'l_2}{2}\Phi_1 C_3 & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ 0 & C_1 & 0 & \mathbf{0}_{1\times 2} & 0 \\ \hline \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} & * & -\frac{k'l_3}{2}\Phi_2 & \mathbf{0}_{2\times 1} \\ \hline 0 & \mathbf{0}_{1\times 2} & 0 & * & -k'' \end{array}\right].$$

The constants k', k'', l_1, l_2, l_3 are positive and the matrices Φ_i have eigenvalues in the open

[6] We disregard the other possibility $\widehat{\overline{\mathbf{\Gamma}}}^{\mathrm{a}} = 0$ which does not correspond to a plausible estimate of $\mathbf{H}Q(\theta^*)$ under our standing assumptions.

right half plane as long as $\varphi_i^{\mathrm{L}} \in (-\pi/2, \pi/2)$. It follows immediately that the second and third diagonal blocks are asymptotically stable. As for the spectrum of the first block, we refer to the analysis in Atta et al. (2015) in which the same structure appears. Collecting all these considerations, $\widetilde{J}^{\mathrm{a}}$ is asymptotically stable for sufficiently small amplitudes a. Finally, since J^{a} and $\widetilde{J}^{\mathrm{a}}$ are similar, they share the same spectrum. ■

5

Energy-efficient Operation of an Indirect Evaporative Cooling System

This chapter considers the problem of designing a control system to efficiently operate an Indirect Evaporative Cooling (IEC) unit. Following the workflow described in Chapter 3, we first derive in Section 5.2 a First-Principle Data-Driven mathematical model of the considered cooling unit and in Section 5.3 we design a simulation environment based on Matlab/Simulink accordingly. Then, a two-level control architecture with a supervision layer based on the ESC algorithm is designed in Section 5.4. By means of simulations, the effectiveness and the robustness of the proposed control system in optimizing the operation of the Indirect Evaporative Cooling unit are tested and the results are reported in Section 5.5.

5.1 Introduction

The evaporative cooling technique takes the advantage of water evaporation to achieve cooling effect. An IEC system is able to produce the cool air by exploiting the evaporative process, but without moist change. In a typical IEC unit, the primary air (or supply air) and the secondary air (or exhaust air) flow in separate passages of a heat exchanger. The secondary-side channels are wetted by a film of water, so the exhaust air is cooled down

and humidified by the evaporative cooling process. As a result, the temperature difference between the two airflows is increased and more heat can flow between them through the walls of the channels of the heat exchanger. Since the two airflows are not in direct contact, there is not moist change in the primary air. An example of transformation of the thermal conditions of both supply and exhaust air during an IEC process is depicted in the simplified psychrometric chart reported in Fig. 5.1. The transformation of the exhaust air is the result of the combination of two effects: the evaporative cooling process causes a decrease of its temperature and an increase of its relative humidity, while the convective heat exchanged with the supply air causes simultaneously an increase of its temperature. As a result of the transformation, the supply air temperature is decreased, while its humidity is not affected by the cooling process.

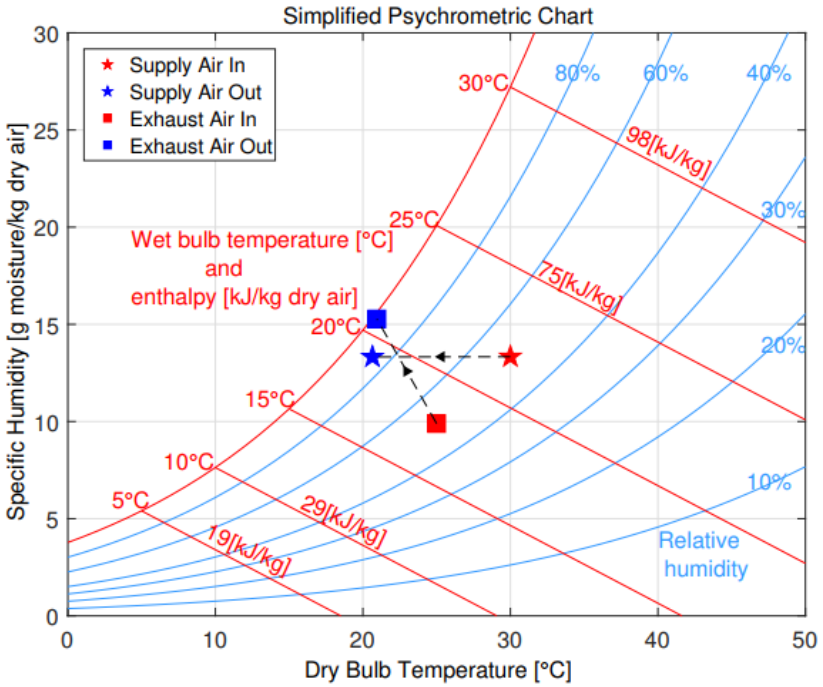


Figure 5.1: Example of an IEC process shown in a simplified psychrometric chart. The squares depicts the exhaust air conditions, while the stars depicts the supply air ones. The red colour refers to the air conditions in the inlet sections of the heat exchanger, while the blue colour refers to the outlet conditions.

In this chapter, we consider the problem of optimizing the operation of a specific IEC unit suggested by the industrial partner, i.e., the Vertiv EFC unit shown in Fig. 5.2. The operating principle of the targeted cooling system is schematized in Fig. 5.3. As it can be observed, the warm air heated by the IT equipment is recirculated by a group of fans from the computer room to the channels of a cross-flow air-to-air heat

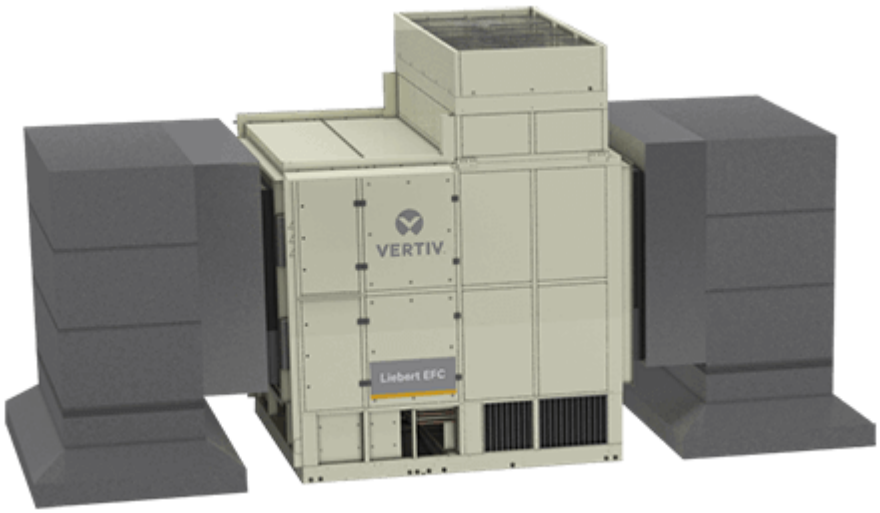


Figure 5.2: The Vertiv EFC indirect evaporative cooling unit.

exchanger (Supply air-in). Through the channels of the other side of the heat exchanger, air taken from the outdoor environment is forced by another group of fans (Exhaust air-in). Sprinklers placed above the heat exchanger draw water from a reservoir and spray it into the process air channels. Part of the water evaporates into the exhaust air flowing through the heat exchanger, while the remaining falls down into the reservoir. The process air absorbs heat from the primary air and then it is exhausted back in the outdoor environment (Exhaust air-out), while the resulting cool supply air is circulated back into the computer room (Supply air-out).

The first goal within this work is to derive a mathematical model representing the considered IEC system. In order to obtain a good compromise between simplicity and accuracy, we derive a lumped-parameters First-Principle Data-Driven model, which is described in Section 5.2. Then, in Section 5.3 we propose a simulation environment based on the mathematical model of the system and developed in Matlab and Simulink. By exploiting the simulation environment, we finally design, tune, and test a two-layer supervisory controller based on an ESC algorithm to optimize the operation of the considered cooling unit. The control system design phase is described in Section 5.4 and examples of simulations are reported in Section 5.5.

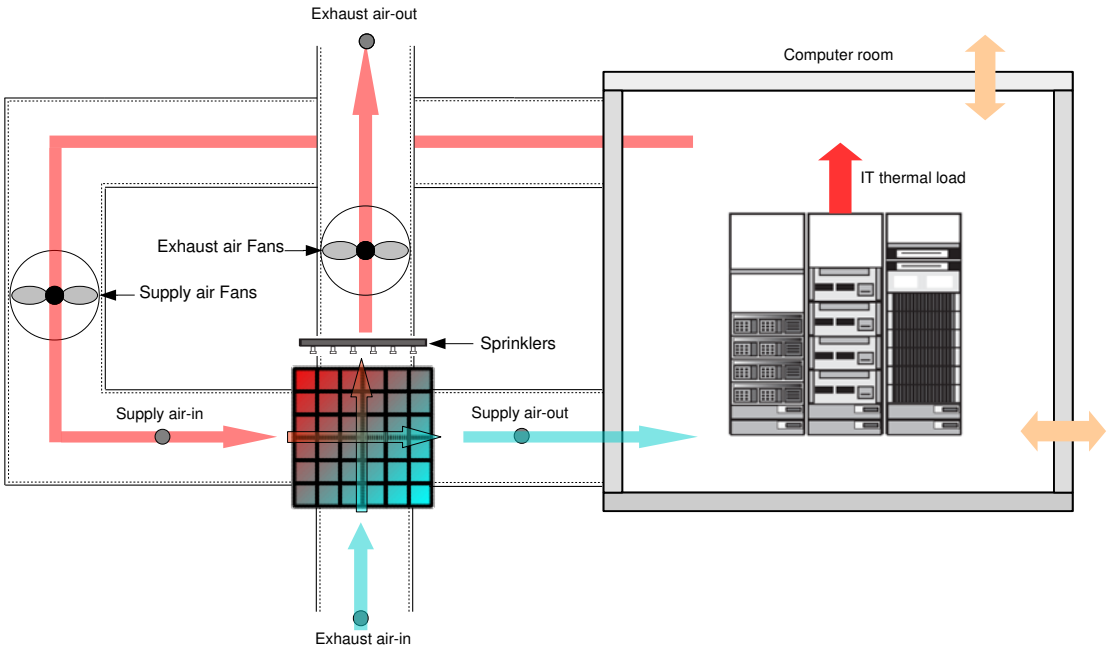


Figure 5.3: Schematic representation of a typical IEC system. The main components are: a compact cross-flow air-to-air heat exchanger, fans circulating the supply and the exhaust air, and sprinklers spraying pressurized water into the process air channels of the heat exchanger.

5.2 Modelling the IEC System

In this Section, we propose a lumped-parameters First-Principle Data-Driven model representing the considered IEC system. The overall model considers all the main components of the system, i.e., the indirect evaporative heat exchanger, the fans, the water pump, and the computer room. Due to its higher thermal inertia, the dynamics of the computer room are much slower than those of all the other components. For this reason, the computer room is described by a dynamic model, while all the other components are considered as static elements, in order to obtain simpler models. In the following, the models describing each of the components of the system are described. Particular focus is given to the model of the indirect evaporative heat exchanger model, which is the most particular and complex component of the system.

5.2.1 Indirect Evaporative Heat Exchanger Model

In literature several IEC models for various purposes and with different degree of complexity are presented. In Alonso et al. (1998), heat and mass transfer model based on basic principles has been developed for thermal calculations of an indirect evaporative cooler performance. The model can be used to analyze different indirect evaporative

cooler designs and conditions. In Kiran and Rajput (2011), fuzzy inference system (FIS), artificial neural networks (ANN), and adaptive neuro fuzzy inference (ANFIS) approaches are used to predict the performance of IEC systems. The results shows the usefulness of an intelligent way to predict the performance of indirect evaporative cooler. In Hasan (2012) an analytical model is developed for finding the thermal performance of indirect evaporative coolers based on modifications of the effectiveness-Number of Transfer Units (NTU) method. In Liu et al. (2013), a simplified model using a modified effectiveness-NTU method for thermal performance simulation of indirect evaporative heat exchangers is presented and then validated utilizing experimental data from the literature. The model facilitates efficient computer simulations of indirect evaporative cooling or hybrid indirect evaporative cooling/direct expansion vapor compression systems, both of which include an indirect evaporative heat exchanger as a core component. Anisimov and Pandelidis (2015) presents a numerical study of heat and mass transfer in indirect evaporative air coolers with four air flow patterns: parallel-flow, counter-flow, cross-flow and regenerative. The numerical simulation was performed on the basis of original two-dimensional heat and mass transfer models (in the case of cross-flow heat exchanger the model was 3-D). In Heidarinejad and Moshari (2015), a mathematical model for an IEC system with consideration of wall longitudinal heat conduction and effect of spray water temperature variation along the exchanger surface in a cross-flow configuration is presented. The numerical results show applicability of the model for both sub-and above-wet bulb cooling applications. In Liberati et al. (2017), a phenomenological IEC system model, based on a cross flow heat exchanger, has been developed and validated in typical summer operating conditions, at different air streams temperature, humidity ratio and flow rates. Thanks to the model the IEC system performance has been analyzed in different working conditions. The results confirm that IEC systems can be an effective technology to achieve significant primary energy savings in HVAC operating in summer conditions. In Wan et al. (2017), various models of different complexity (1-D and 2-D) for indirect evaporative cooling with counter flow configurations are presented for performance analysis and for control design purposes. In Chen et al. (2017) an experimental-validated IEC model emphasizing on condensation condition is used to conduct a parameter sensitivity analysis which is useful for the optimization of IEC design and operation in hot and humid regions. In Zhu et al. (2017), the performance of a counter-flow regenerative heat and mass exchanger (RHMX) for indirect evaporative cooling is analyzed by means of an ANN model: the effects of various operating conditions, namely the inlet air temperature, relative humidity and air velocity as well as the extraction air ratio on the performance of the RHMX are investigated.

The mathematical model of the evaporative cross-flow air-to-air heat exchanger considered in this work is based on steady-state heat and mass transfer balances, validated using experimental data. The main improvement of this model over the ones existing in the literature is that it takes into account also:

- the spatial distribution of atomized water within the exhaust air channels;
- the condensation process that may occur in part of the supply air channels.

In order to take into account the spatial distribution of the vaporized water, both Finite Volume (FV) and Moving Boundary (MB) methods may be considered for modelling the evaporative heat exchanger. While a FV approach divides the heat exchanger into a number of fixed control volumes, the MB method segments the heat exchanger depending on the physical phenomena that occur inside it and moves the boundaries of each zone as the operative conditions change. For example, the evaporative heat exchanger may be partitioned in zones where the exhaust air channels are wetted by the water film and zones where the exhaust air channels are dry. This approach typically leads to a lower number of control volumes and so it has to solve fewer equations. As a result, this method offers a good balance between model complexity and accuracy. Since we are not interested in describing the transient phenomena of thermal processes that occur within the heat exchanger, we employ a static MB method, i.e., we neglect the dependency on time.

Following the static MB approach, we introduce the fictitious depth δ and the fictitious width $1-\xi$ of the portion of exhaust air passages reached by the vaporized water. The overall heat exchanger results divided into three parts, as shown in Fig. 5.4: the *Wet-Evap* zone is wetted by the film of water, whereas *Dry* and *Wet-Noevap* zones are dry. The static correlations that we propose for determining the control volume boundaries δ and ξ are the following:

$$\delta = \delta_{00} + \delta_{10}\dot{m}_E + \delta_{01}\dot{m}_W + \delta_{11}\dot{m}_E\dot{m}_W, \tag{5.1}$$

$$\xi = \xi_a e^{\xi_b \dot{m}_E}, \tag{5.2}$$

where coefficients δ_{00}, δ_{10}, δ_{01}, δ_{11}, ξ_a, and ξ_b are calibrated by exploiting experimental data as described in Section 5.2.1.4. In order to properly describe the heat exchange phenomena in the three different parts of the heat exchanger, we developed two different mathematical models: the *NOEVAP* model describes the *Dry* and *Wet-Noevap* zones, while the *EVAP* model represents the *Wet-Evap* zone. Fig. 5.5 depicts a schematic representation of how the exhaust airflow $\dot{m}_E$ splits in the heat exchanger depending on the fictitious parameters δ and ξ. Part of the airflow ($\dot{m}_E\,\xi$) goes through the *Dry* zone,

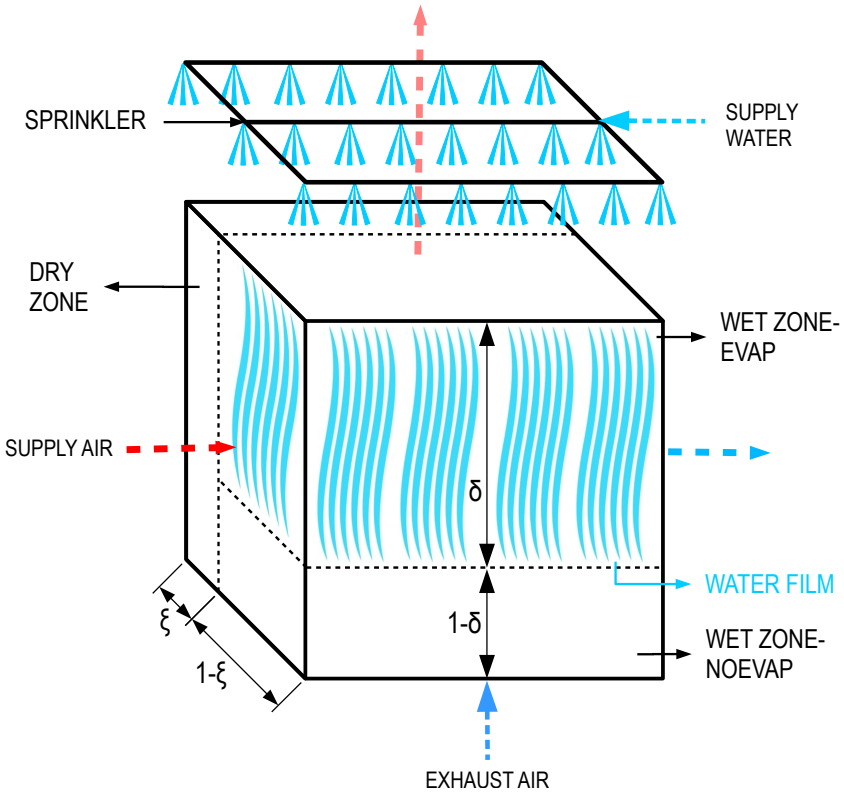


Figure 5.4: IEC: air-to-air cross-flow heat exchanger.

where it warms up by exchanging heat with the supply air, and the exhaust air output temperature $t_{E,o,Dry}$ and humidity $x_{E,o,Dry}$ are computed by using the *NOEVAP* model. The remaining part of the inlet airflow $\dot{m}_E(1-\xi)$ first enters the *Wet-Noevap* zone and then goes through the *Wet-Evap* zone. The intermediate temperature $t_{E,\delta}$ and humidity $x_{E,\delta}$ are given by the *NOEVAP* model, whereas $t_{E,o,Wet}$ and $x_{E,o,Wet}$ are provided by the *EVAP* model. Analogously, Fig. 5.6 shows a schematic representation of how the supply airflow splits into three parts.

5.2.1.1 NOEVAP Model

The *NOEVAP* model represents portions of the heat exchanger where exhaust air channels are not wetted by the water film. A schematic representation of the *NOEVAP* model is shown in Fig. 5.7.

The *NOEVAP* model also takes into account condensation phenomena that may occur in the supply air channels. To this purpose, we exploit a fictitious dimensionless coordinate γ that represents the portion of the channels where the condensation phenomenon occurs and which is defined as function of the inlet and outlet sections wall mean temperature

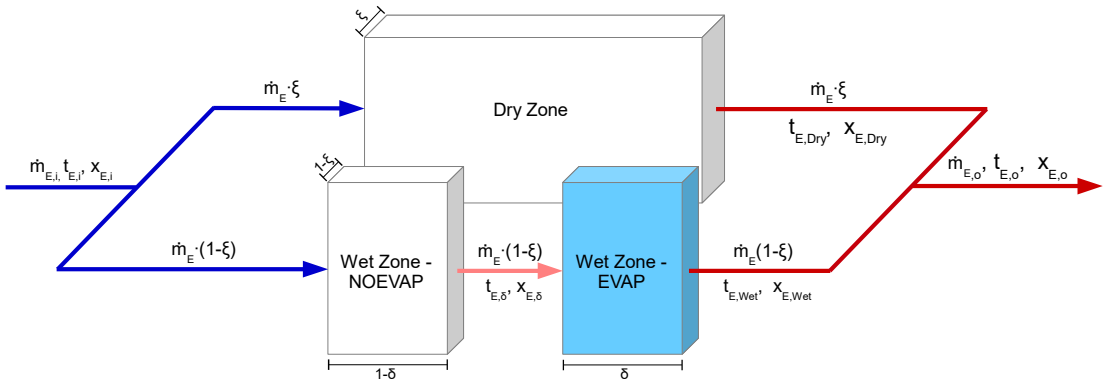


Figure 5.5: Exhaust air flow.

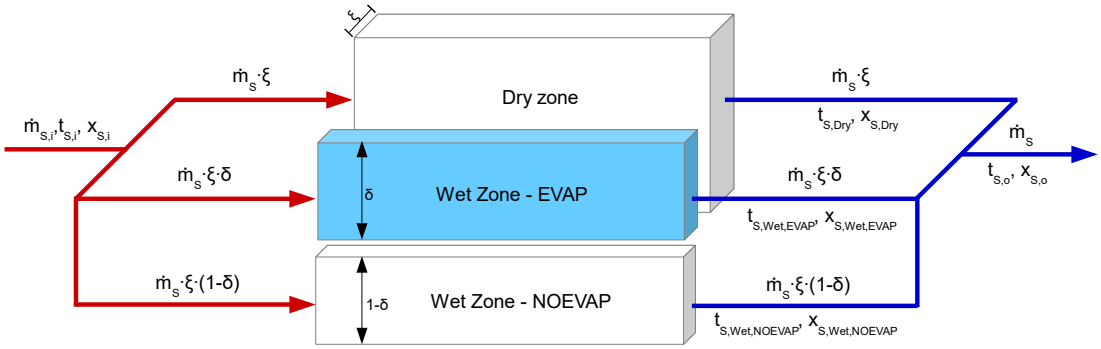


Figure 5.6: Supply air flow.

$t_{w,A}$ and $t_{w,B}$:

$$\gamma = \frac{t_{w,B} - t_{dp,S,i}}{t_{w,B} - t_{w,A}}. \tag{5.3}$$

A set of weights w_{dry} and w_{wet} is defined depending on the value of γ. If $\gamma \geq 1$ then there is no condensation and $w_{dry} = 1$, $w_{wet} = 0$, while if $\gamma \leq 0$ then supply air channels are completely wet and $w_{dry} = 0$, $w_{wet} = 1$. Otherwise, there is a mixed situation and $w_{dry} = \gamma$ and $w_{wet} = 1 - \gamma$.

Then, the total, sensible, and latent heat flow rates exchanged by the supply air result as follows:

$$\begin{cases} q_{S,tot} = q_{S,sens} + q_{S,lat} \\ q_{S,sens} = w_{dry} q_{dry} + w_{wet} q_{wet,sens} \\ q_{S,lat} = w_{wet} q_{wet,lat} \end{cases}, \tag{5.4}$$

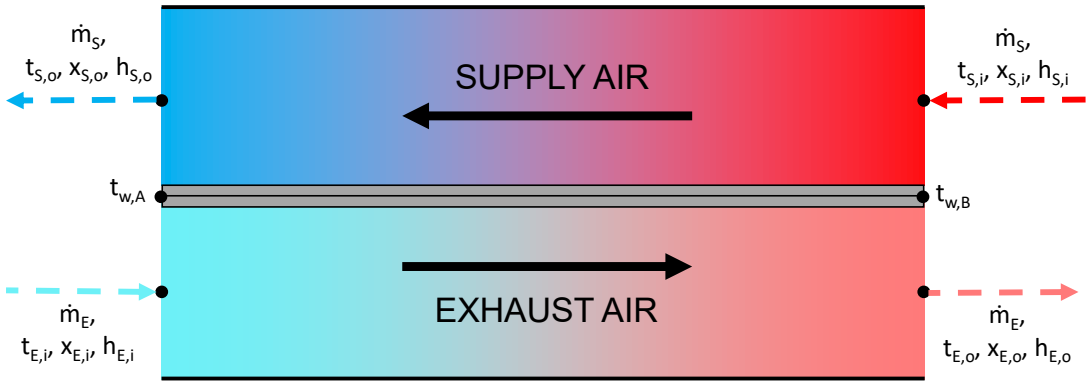


Figure 5.7: The *NOEVAP* model.

while those exchanged by the exhaust air are:

$$\begin{cases} q_{E,tot} = q_{E,sens} = q_{S,tot} \\ q_{E,lat} = 0 \end{cases}, \tag{5.5}$$

where q_{dry} and q_{wet} are the heat flow rates in dry and wet supply air channels conditions respectively.

When condensation phenomenon does not occur in the supply air channels, the heat exchanged between the two airflows is:

$$q_{dry} = P_{S,dry} \dot{m}_S c_{p,S,i} (t_{S,i} - t_{E,i}). \tag{5.6}$$

The temperature effectiveness $P_{S,dry}$ is computed as a function of the two airflows thermal capacities ratio $R_{S,dry}$ and the number of transfer units $NTU_{S,dry}$:

$$P_{S,dry} = 1 - \exp\left\{ NTU_{S,dry}^{0.22} R_{S,dry}^{-1} \left[\exp\left(-R_{S,dry} NTU_{S,dry}^{0.78} \right) - 1 \right] \right\}, \tag{5.7}$$

with:

$$NTU_{S,dry} = \frac{K_{dry} A_{plates}}{\dot{m}_S c_{p,S,i}}. \tag{5.8}$$

The overall heat transfer coefficient K_{dry} is given by:

$$K_{dry} = \left(K_{dry_S}^{-1} + K_{dry_E}^{-1} \right)^{-1}, \tag{5.9}$$

where K_{dry_S} and K_{dry_E} are the heat transfer coefficients referred to the supply and exhaust side, respectively. In the following, we report the equations for only the supply air side, since the exhaust air ones can be obtained in the same way:

$$K_{dry_S} = \left(\frac{A_{plates}}{A_S \, \alpha_{dry,S} \, \Omega_{dry,S}} + \frac{\frac{s_{plates}}{2}}{\lambda_m} \right)^{-1}, \tag{5.10}$$

where Ω_{dry_S} is the dry finned surface efficiency:

$$\Omega_{dry_S} = 1 - \frac{A_{finned}}{A_S} \left(1 - \frac{\tanh\left(m_{dry_S} L\right)}{m_{dry_S} L} \right), \tag{5.11}$$

with

$$m_{dry_S} = \sqrt{\frac{2 \, \alpha_{dry_S}}{\lambda_m s_{fin}}}. \tag{5.12}$$

The convective heat transfer coefficient α_{dry_S} for dry-fin conditions is given by:

$$\alpha_{dry_S} = \frac{\lambda_i Nu}{d_h}, \tag{5.13}$$

where the Nusselt number Nu can be described by using the correlations proposed by Polley and Abu-Khader in Polley and Abu-Khader (2005):

$$Nu = \left(Nu_{turb}^{n_{10}} + Nu_{lam}^{n_{10}} \right)^{\frac{1}{n_{10}}}. \tag{5.14}$$

Terms Nu_{turb} and Nu_{lam} are respectively the turbulent and laminar components of the Nusselt number:

$$Nu_{turb} = n_3 \sqrt{f} Re^{n_4} Pr^{n_5}, \tag{5.15}$$

$$Nu_{lam} = \left[n_6^3 + n_7^3 + \left(n_8 Gz^{n_9} - n_7 \right)^3 \right]^{\frac{1}{3}}, \tag{5.16}$$

where f is the Fanning friction factor:

$$f = \left[\left(\frac{16}{Re} \right)^3 + \left(n_1 Re^{n_2} \right)^3 \right]^{\frac{1}{3}}, \tag{5.17}$$

and parameters $n_1, n_2, \ldots, n_{10}$ are calibrated by exploiting experimental data, as described in Section 5.2.1.4.

As in the dry supply channels case, the exchanged heat flow rate when condensation

occurs can be obtained by applying the P-NTU method:

$$\begin{cases} q_{wet} = P_{S,wet}\,\dot{m}_S(h_{S,i} - h_{sat,E,i}) \\ q_{wet,lat} = \dot{m}_S(h_{S,i} - h_{S,int,wet}) \\ q_{wet,sens} = q_{wet} - q_{wet,lat} \end{cases}, \tag{5.18}$$

where $h_{sat,E,i}$ is the saturated air specific enthalpy at $t_{E,i}$ and $h_{S,int,wet}$ is the specific enthalpy in the intermediate point, assuming to divide the supply air transformation in two parts: first the specific humidity is isothermally reduced, then the temperature is decreased keeping the humidity constant. To compute the enthalpy effectiveness $P_{S,wet}$, we use Eq. (5.7) as previously described but correcting the overall and the convective heat transfer coefficients by exploiting the correlations proposed in Kuehn et al. (1998):

$$K_{wet} = \left(\frac{b_w\, A_{plates}}{A_S\, \alpha_{wet,S} \Omega_{wet,S}} + \frac{b_E\, A_{plates}}{A_E \alpha_{dry,E} \Omega_{dry,E}} + \frac{b_w\, s_{plates}}{\lambda_m} \right)^{-1}, \tag{5.19}$$

$$\alpha_{wet,S} = \alpha_{dry,S} \frac{b_w}{c_{p,S,i}}, \tag{5.20}$$

with

$$b_w = \left. \frac{d\, h_{sat}}{d\, t} \right|_{t_w}, \tag{5.21}$$

$$b_E = \frac{h_{sat}|_{t_w} - h_{sat}|_{t_{E,mean,dry}}}{t_w - t_{E,mean,dry}}, \tag{5.22}$$

where t_w is the mean of inlet and outlet walls temperature $t_{w,A}$ and $t_{w,B}$ and $t_{E,mean,dry}$ is the mean of inlet and outlet exhaust air temperature $t_{E,i}$ and $t_{E,o}$.

5.2.1.2 EVAP Model

The *EVAP* model represents portions of the heat exchanger where the film of water wets the exhaust air channels. Fig. 5.8 depicts a schematic representation of the *EVAP* model.

The surface of the exhaust air passages is supposed to be uniformly wetted by a water film, so that heat and mass transfers take place between the wet surface and the airflow. The wet surface is considered at the same temperature t_{water} of the water film, that is assumed to be constant and that is computed by using an iterative procedure discussed later. For the sake of simplicity, the thermal resistance of the plate that separates the flows is neglected, so that also the wall surface of the supply airflow passages is at temperature t_{water}. Also, we assume that heat and mass transfers occur only in a

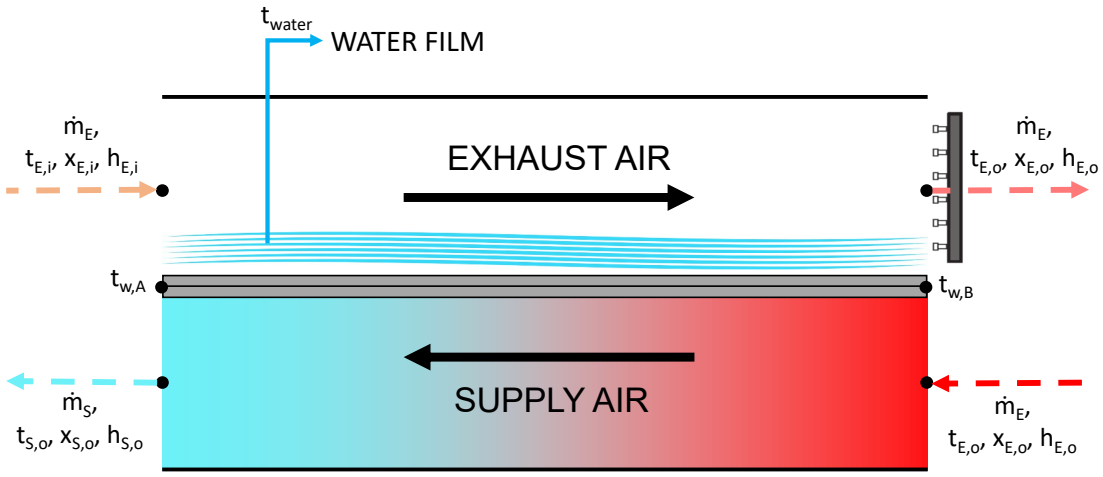


Figure 5.8: The *EVAP* model.

direction normal to flows and the separation wall is impermeable to the mass transfer and its thickness is negligible. Moreover, we suppose that the water film has infinite thermal capacity, so the energy balance is evaluated only for airflows. Thus, the $\epsilon - NTU$ method can be used to develop the *EVAP* model and the airflows energy balances are analysed readjusting the method proposed in Camargo and Ebinuma (2003).

As the *NOEVAP* model, also the *EVAP* model manages presence/absence of supply air condensation. To this purpose, it is necessary to compare the wall surface temperature t_{water} with the inlet dew point temperature $t_{dp,S,i}$. Defining a threshold Δ, we introduce a set of weights w_{dry} and w_{wet} as follows: if $t_{water} \leq t_{dp,S,i} - \Delta$, then the channels are completely wet and $w_{dry} = 0$, $w_{wet} = 1$; otherwise, if $t_{dp,S,i} \geq t_{dp,S,i}$, then there is no condensation; finally, if $t_{dp,S,i} - \Delta \leq t_{water} \leq t_{dp,S,i}$ then there are mixed conditions and $w_{dry} = 1$, $w_{wet} = 0$. Otherwise, $w_{wet} = (t_{dp,S,i} - t_{water})\Delta^{-1}$ and $w_{dry} = 1 - w_{wet}$.

Referring to the supply air side, sensible, latent, and total heat flow rates result as follows:

$$\begin{cases} q_{S,tot} = q_{S,dry} + q_{S,wet} \\ q_{S,sens} = q_{S,dry} + q_{S,wet,sens} \\ q_{S,lat} = q_{S,wet,lat} \end{cases}, \tag{5.23}$$

where $q_{S,dry}$ and $q_{S,wet}$ are the heat flow rates in dry and wet supply air channels conditions, respectively.

The exchanged heat flow rate, when condensation does not occur, results:

$$q_{S,dry} = \epsilon_{S,dry} \dot{m}_S c_{p,S,i} (t_{S,i} - t_{water}), \tag{5.24}$$

where the effectiveness is given by:

$$\epsilon_{S,dry} = 1 - \exp\left\{-w_{dry}\, NTU_{S,dry}\right\}, \tag{5.25}$$

$$NTU_{S,dry} = \frac{\alpha_{dry,S} A_{plates}}{\dot{m}_S\, c_{p,S,i}}. \tag{5.26}$$

The exchanged heat flow rate in wet supply air channels conditions is:

$$\begin{cases} q_{S,wet} = \epsilon_{S,wet} \dot{m}_S (h_{S,i} - h_{sat,t_{water}}) \\ q_{S,wet,lat} = \dot{m}_S (h_{S,i} - h_{S,int,wet}) \\ q_{S,wet,sens} = q_{S,wet} - q_{S,wet,lat} \end{cases}, \tag{5.27}$$

where $h_{sat,t_{water}}$ is the saturated air specific enthalpy at temperature t_{water}, $h_{S,int,wet}$ is the specific enthalpy in the intermediate point of the transformation, and the effectiveness $\epsilon_{S,wet}$ is:

$$\epsilon_{S,wet} = 1 - \exp\left\{-w_{wet} NTU_{S,wet}\right\}, \tag{5.28}$$

$$NTU_{S,wet} = \frac{\alpha_{m_S} A_{plates}}{\dot{m}_S}. \tag{5.29}$$

The mass transfer coefficient $\alpha_{m,S}$ is calculated as:

$$\alpha_{m,S} = \frac{\alpha_{dry,S}}{c_{p,S,i} Le_{S,i}^{2/3}}, \tag{5.30}$$

where the Lewis number is given by:

$$Le_{S,i} = \frac{\lambda_{S,i}}{D_{S,i} c_{p,S,i} \rho_{da,S,i}}. \tag{5.31}$$

The heat exchanged by exhaust air can be obtained as in (5.27), using the exhaust air effectiveness:

$$\epsilon_E = 1 - \exp\left\{-NTU_E\right\}, \tag{5.32}$$

where NTU_E can be computed using Eqs. (5.29)-(5.31).

To compute the water temperature t_{water}, we perform the following iterative procedure. First, the initial value of t_{water} is arbitrarily set, e.g., equal to the mean of the two airflows inlet temperatures. Then, the value of t_{water} is iteratively modified by using a method of False Position, i.e., Regula Falsi Burden and Faires (2001), until it satisfies

with some pre-established tolerance the following energy balance:

$$\dot{m}_S(h_{S,i} - h_{S,o}) = \dot{m}_E(h_{E,o} - h_{E,i}). \tag{5.33}$$

5.2.1.3 Model of the Pressure Drops

The proposed model takes into account also the air pressure drops Δp_S and Δp_E due to the HX passages. To this purpose, we used the correlations proposed by Kays and London in Kays and London (1984). They express the overall pressure drops in both HX channels as the sum of the following terms:

- pressure drop due to friction, given by:

$$\Delta p_f = \frac{G^2\, 4\, f\, L}{2\, d_h\, \rho_{da,m}}, \tag{5.34}$$

  with

$$\rho_{da,m} = \frac{\rho_{da,i} + \rho_{da,o}}{2}; \tag{5.35}$$

- pressure drops in input and output sections of each channel, given by:

$$\Delta p_i + \Delta p_o = \frac{G^2}{2}\left[\frac{k_c + 1 - \sigma^2}{\rho_{da,i}} - \frac{1 - \sigma^2 - k_e}{\rho_{da,o}}\right], \tag{5.36}$$

  where k_c and k_e are contraction and expansion coefficients, which are calculated as in Muzychka and Yovanovich (2001):

$$k_c = 0.42(1 - \sigma^2)^2, \tag{5.37}$$

$$k_e = (1 - \sigma^2); \tag{5.38}$$

- momentum pressure drop, given by:

$$\Delta p_m = G^2\left(\frac{1}{\rho_{da,o}} - \frac{1}{\rho_{da,i}}\right). \tag{5.39}$$

Therefore, the total pressure drop is:

$$\Delta p = \Delta p_f + \Delta p_i + \Delta p_o + \Delta p_m. \tag{5.40}$$

5.2.1.4 Model Calibration and Validation

In order to obtain a reasonable fit between the variables of interest of the IEC system and the model outputs, the calibration of the model parameters and the validation of the model are necessary. Given the basic form of the model described above, the calibration step involves estimating the values of certain parameters in the model structure. Once satisfactory estimates of the model parameters have been obtained, the model must be checked to assure that they adequately perform the functions for which they are intended, i.e., the validation process.

The calibration process is split into three phases: in *Phase 1*, parameters n_1 and n_2 in Eq. (5.17) are tuned; in *Phase 2*, parameters $n_3, n_4, \ldots, n_{10}$ in Eqs. (5.14), (5.15), and (5.16) are tuned; in *Phase 3*, parameters δ_{00}, δ_{10}, δ_{01}, δ_{11} in Eq. (5.1) and ξ_a, ξ_b in Eq. (5.2) are calibrated on experimental data. The data are provided by Vertiv and they correspond to steady-state operating conditions of the considered EFC unit. It is assumed that model inputs are pressure, temperature, specific humidity, and mass flow rate of both airflows, while model outputs are pressure drop, temperature, and specific humidity of both airflows. The inlet pressure is set equal to the atmospheric one.

In order to properly perform the calibration of the model, it is important to consider different data sets for the calibration and the validation of the model. For this reason, the set of observed data has been divided into three disjoint subsets: the first is the calibration data set DS_C for dry-fin conditions (about 100 operating points); the second is a dry-fin condition validation data set $DS_{v,1}$ (about 50 operating points); the third is a validation data set $DS_{v,2}$ (about 50 operating points) for wet-fin conditions due to supply air condensation. It is worth highlight that the support of the industrial partner and the experts' knowledge have been essential to create high-value datasets for modelling calibration and validation purposes. The deviation between the experimental data and the predictions may be ascribable to several factors such as model approximations and non-ideal experimental conditions. From a modelling point of view, we suppose that, for given boundary conditions, the control volumes in the static MB model are well defined and the physical conditions of the wet air and the water film inside each zone are uniform or nearly uniform. However, complex physical phenomena arise in the real system determining non-uniform zones. Because of the structure and the relatively small complexity of the static MB, it is not able to predict completely the system behaviour. In addition, the correlations that we have used to describe heat and mass transfer phenomena have been chosen as a compromise between complexity and overall model performance. The use of more highly structured correlations would probably give better predictions, to the detriment of the model parsimony and flexibility. Beside these

modelling limitations, it is worth highlighting that the industrial partner has properly collected the experimental data, but without resorting to the use of a climatic chamber. For this reason, the boundary conditions of the IEC system during the experimental tests and the data gathering were not actually constant. This inconsistency with respect to the uniform boundary conditions hypothesis can partially justify the variation of the results between the model and experimental data.

Phase 1 Parameters n_1 and n_2 are tuned by comparing the predicted pressure drops $\Delta\hat{p}_{S\cup E}$ obtained from Eq. (5.40) and the observed pressure drops $\Delta p_{S\cup E}$ contained in the DS_C data set. The parameters are given by the solution of the following optimization problem:

$$\text{argmin}_{n_1,n_2}\, RMSE\left(\Delta p_{S\cup E}, \Delta\hat{p}_{S\cup E}\right)|_{DS_C}, \tag{5.41}$$

where the $RMSE$ represents the Root-Mean-Square-Error, i.e., the average squared difference between the estimated values and the target. Fig. 5.9 shows the cost function considered in problem (5.41) as a function of parameters n_1 and n_2 and the solution found by performing a grid search is depicted by the circle. The correspondent optimal values of n_1 and n_2 are reported in Tab. 5.1a.

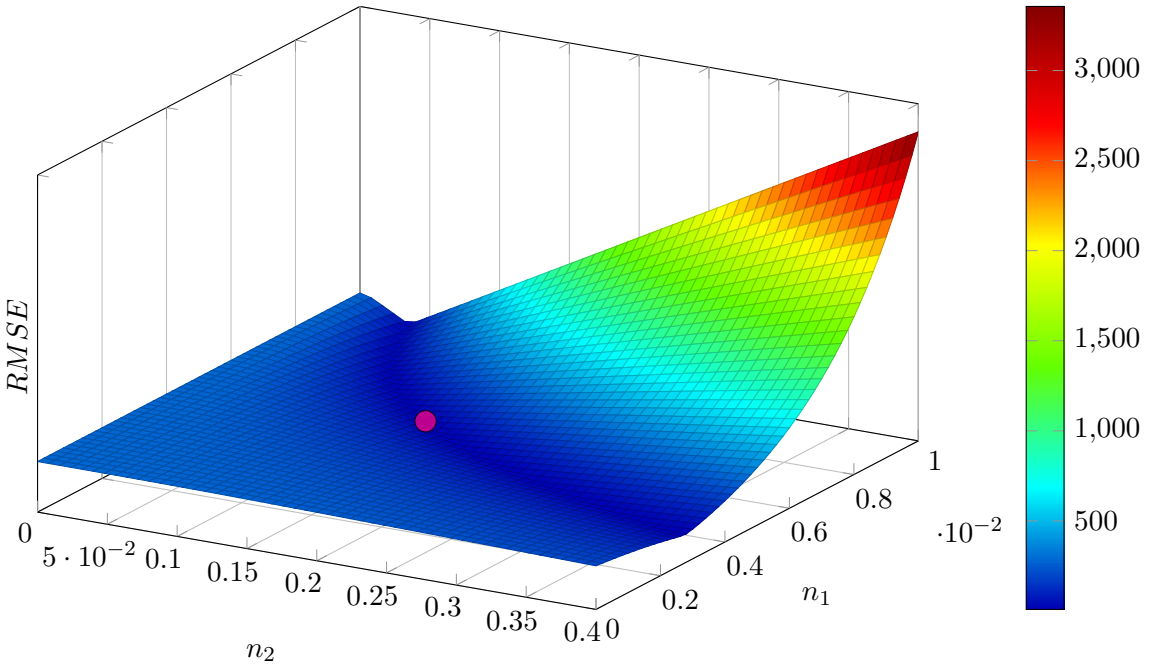
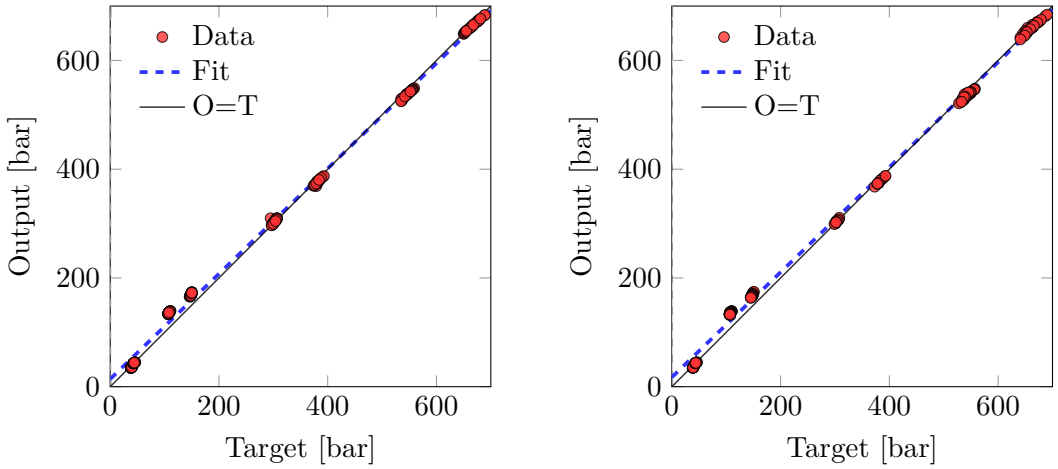


Figure 5.9: The RMSE between estimated and observed pressure drops as a function of parameters n_1 and n_2. The circle depicts optimal parameters values that minimize the RMSE.

Once known n_1 and n_2, using Eq. (5.40) it is possible to calculate the predicted pressure drops $\Delta\hat{p}_{S\cup E}$ for each data set. Note that validation data sets $DS_{v,1}$ and $DS_{v,2}$ are merged into one data set DS_v to evaluate performances of the Fanning friction factor correlation (5.17). It is possible because the supply air condensation effect on the pressure drops is negligible. Fig. 5.10 depicts the comparison between the observed $\Delta p_{S\cup E}$ (Target) and the predicted $\Delta\hat{p}_{S\cup E}$ (Output) for both DS_c (Fig. 5.10a) and DS_v (Fig. 5.10b) data sets.



(a) Calibration: RMSE = 13.93 bar **(b)** Validation: RMSE = 15.37 bar

Figure 5.10: Comparison between observed and predicted pressure drops in Phase 1.

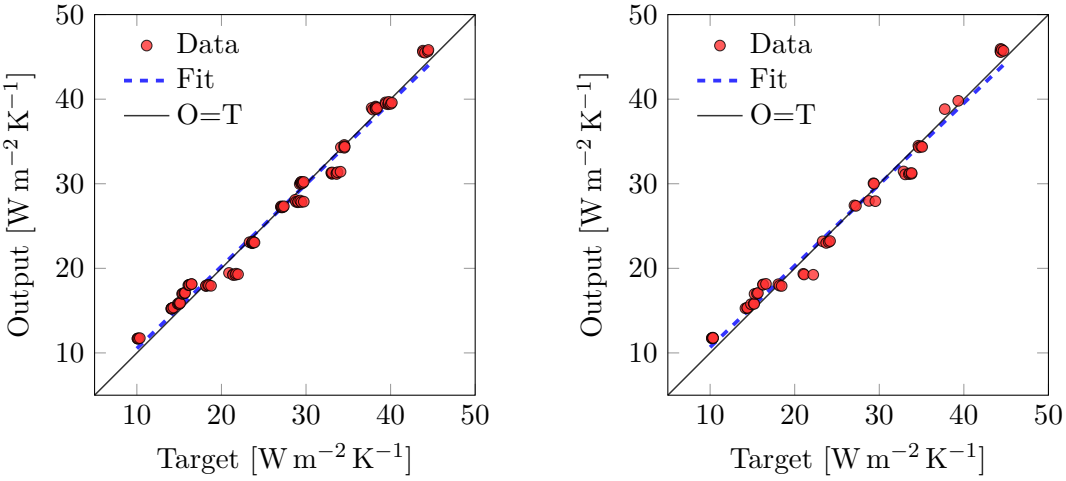
Phase 2 Parameters $n_3, \ldots, n_{10}$ in correlations (5.14), (5.15), (5.16) are calibrated by comparing the predicted overall heat transfer coefficient $\hat{K}_{dry}$ to the observed K_{dry} contained in DS_C. Tuned values of the parameters are given by the solution of the following optimization problem:

$$\text{argmin}_{n_3,\ldots,n_{10}} RMSE\left(K_{dry}, \hat{K}_{dry}\right)|_{DS_C}. \tag{5.42}$$

Due to the higher number of parameters and the complexity of the correlations involved in the computation of the overall heat transfer coefficient, it is difficult to evaluate the properties of the cost function in (5.42) which may present local minima. For this reason, problem (5.42) is solved by using a meta-heuristics approach such as genetic algorithms (GAs), Galletly (1992), and the correspondent solutions $n_3, n_4, ..., n_{10}$ are shown in Tab. 5.1a.

Once obtained parameters $n_3, n_4, ..., n_{10}$, it is possible to have the predicted $\hat{K}_{dry}$ for

DS_c and $DS_{v,1}$ data. Comparing these values to the corresponding observed K_{dry} for both data sets, one has Fig. 5.11.



(a) Calibration: RMSE = 1.25 W m^{-2} K^{-1}

(b) Validation: RMSE = 1.38 W m^{-2} K^{-1}

Figure 5.11: Comparison between observed and predicted overall heat transfer coefficient in Phase 2.

Phase 3 Parameters δ_{00}, δ_{10}, δ_{01} and ξ_a, ξ_b in correlations (5.1) and (5.2) are calibrated by comparing the predicted output supply and exhaust temperatures $\hat{t}_{S\cup E,o}$ to the observed $t_{S\cup E,o}$ contained in DS_C. The tuned values of the parameters are given by the solution of the following optimization problem:

$$\text{argmin}_{\xi_a,\xi_b,\delta_{00},\delta_{10},\delta_{01}} \, RMSE\left(t_{S\cup E,o}, \hat{t}_{S\cup E,o}\right)|_{DS_C}. \tag{5.43}$$

By using least squares regression, we get the values reported in Tab. 5.1b.

Once obtained ξ_a, ξ_b and δ_{00}, δ_{10}, and δ_{01}, it is possible to obtain the predicted $\hat{t}_{S,o}$ and $\hat{t}_{E,o}$ for both calibration and validation data. Comparing these values to the corresponding observed $t_{S,o}$ and $t_{E,o}$ for both data sets (i.e. calibration and validation), one has Fig. 5.12.

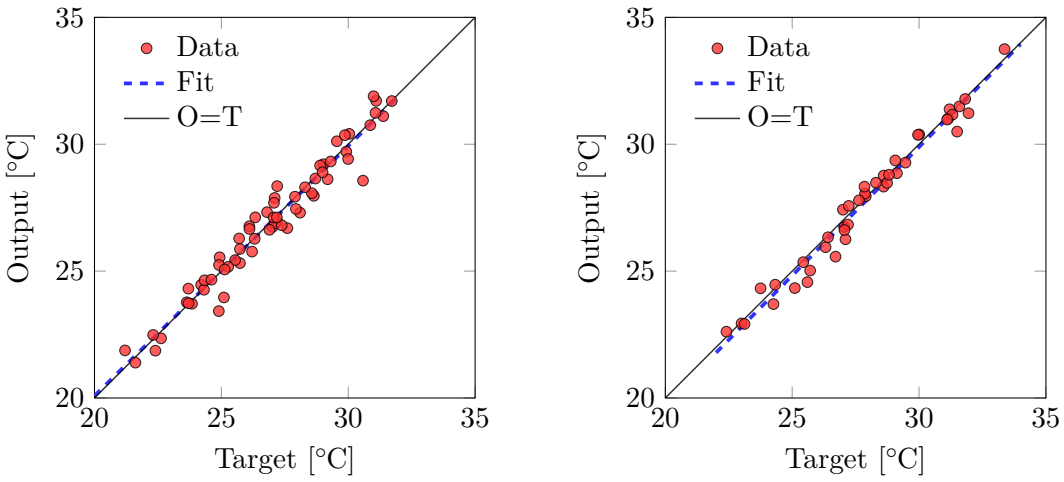
n_1	0.0040	n_6	1.0067
n_2	0.2185	n_7	0.7577
n_3	0.0579	n_8	0.5377
n_4	0.9991	n_9	0.0538
n_5	0.9084	n_{10}	19.7782

(a) Phases 1 and 2.

ξ_a	10.63	[-]
ξ_b	-0.69	$[\mathrm{kg}^{-1}\,\mathrm{s}]$
δ_{00}	0.139	[-]
δ_{10}	0.031	$[\mathrm{kg}^{-1}\,\mathrm{s}]$
δ_{01}	0.150	$[\mathrm{kg}^{-1}\,\mathrm{s}]$
δ_{11}	-0.005	$[\mathrm{kg}^{-2}\,\mathrm{s}^2]$

(b) Phase 3.

Table 5.1: Tuned model parameters.



(a) Calibration: RMSE = 0.43 °C

(b) Validation: RMSE = 0.45 °C

Figure 5.12: Comparison between observed and predicted output temperatures in Phase 3.

5.2.2 Computer Room Model

The considered server room thermal model is linear, dynamic, and lumped parameters, and it has been obtained using the electrical-thermal analogy. Most important variables are temperatures and humidities of the server room and working temperatures of electrical devices, while internal and external thermal loads are treated as disturbances. Inputs of the model are temperature and humidity of the process air.

A Resistance-Capacity (RC) electrical model has been used. Electrical resistances represent the thermal resistances between adjacent nodes, while electrical capacities represent the thermal capacities of the nodes. Nodes represent the masses composing the system. The computer room is schematically represented in Fig. 5.13, where nodes are depicted by circles and heat exchanges are denoted by arrows.

A brief description of each node is provided in Tab. 5.2. In first approximation, heat

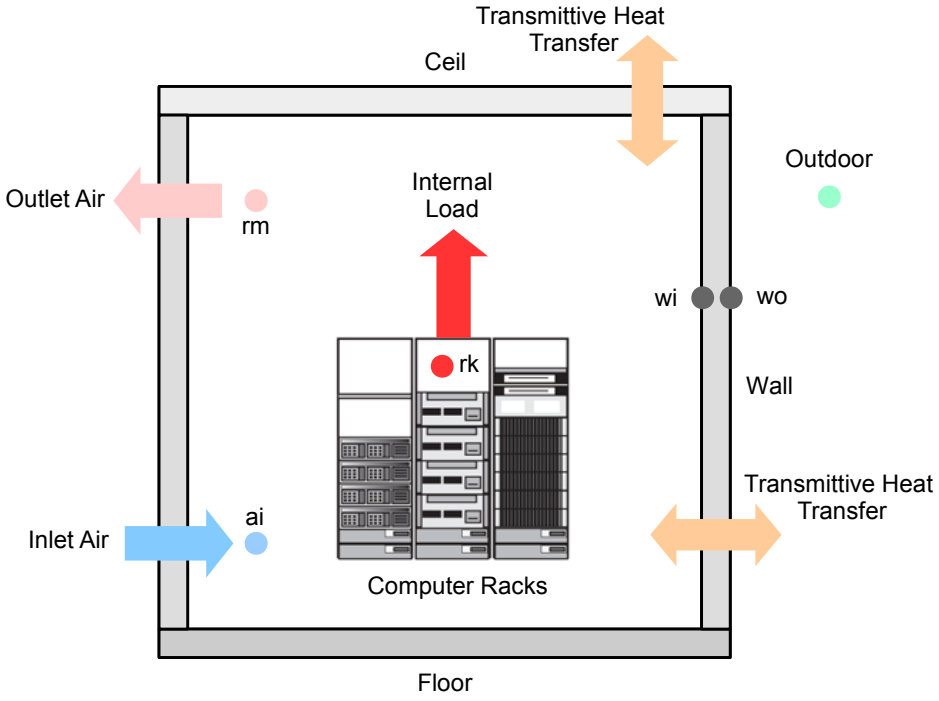


Figure 5.13: Schematic representation of the computer room.

exchanges with the floor, heath exchanges due to solar radiation incident on building exterior, and heat exchanges by radiation between the outer surface of the rack and the inner surface of the equivalent wall are neglected. The resulting equivalent RC model is reported in Fig. 5.14.

In first approximation, the model neglects also mass transfer phenomena when the air comes in contact with surfaces at a temperature lower than its dew-point temperature. This is reasonable considering ranges recommended by ASHRAE and due to the higher temperature of surfaces of the IT equipment in contact with the air. However, the model generally takes into account the dynamics of the specific moisture content of the computer

Node	Description
ai	computer room input airflow rate
rm	mass of air within the computer room
rk	all the IT devices contained in the racks
wi	internal surface conditions of the equivalent wall
wo	external surface conditions of the equivalent wall
Outdoor	outdoor air conditions, which has infinite thermal capacity

Table 5.2: Nodes in the RC model of the computer room.

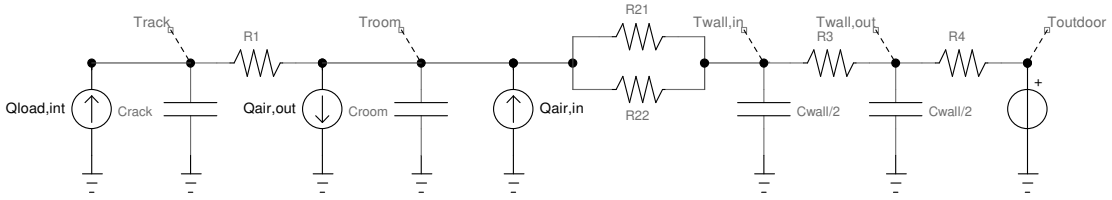


Figure 5.14: Equivalent RC model of the computer room.

room air, since the air may condense in the evaporative heat exchanger and return in the server room with a lower water vapour content.

The considered dynamic and linear model is given by a system of differential equations, called equations of state. They describe the time evolution of the states of the system undergoing the action of inputs and disturbances, starting from the considered initial time instant. The model state equations are:

$$\frac{dT_{rack}}{d\tau} = \frac{1}{C_{rack}}\left[Q_{load,int} - \frac{1}{R_1}\left(T_{rack} - T_{room}\right)\right], \tag{5.44}$$

$$\frac{dT_{room}}{d\tau} = \frac{1}{C_{room}}\left[\dot{m}_{air}\left(h_{room,in} - h_{room,out}\right) - \frac{1}{R_1}\left(T_{room} - T_{rack}\right) + \right. \tag{5.45}$$

$$\left. -\frac{1}{R_2}\left(T_{room} - T_{wall,in}\right)\right], \tag{5.46}$$

$$\frac{dT_{wall,in}}{d\tau} = \frac{1}{\frac{C_{wall}}{2}}\left[-\frac{1}{R_2}\left(T_{wall,in} - T_{room}\right) - \frac{1}{R_3}\left(T_{wall,in} - T_{wall,out}\right)\right], \tag{5.47}$$

$$\frac{dT_{wall,out}}{d\tau} = \frac{1}{\frac{C_{wall}}{2}}\left[-\frac{1}{R_3}\left(T_{wall,out} - T_{wall,in}\right) - \frac{1}{R_4}\left(T_{wall,out} - T_{outdoor}\right)\right], \tag{5.48}$$

$$\frac{dx_{room,out}}{d\tau} = \frac{\dot{m}_{air}}{M_{air,room}}\left(x_{room,in} - x_{room,out}\right). \tag{5.49}$$

It is worth highlighting that the heat flowing from the room output air to the cooling units is expressed as enthalpy flow $\dot{m}_{air}\, h_{room,out}$, where $\dot{m}_{air}$ (in $\mathrm{kg_{da}\,s^{-1}}$) is the air flow rate treated by the cooling system and $h_{room,out}$ (in $\mathrm{J\,kg_{da}^{-1}}$) is the air specific enthalpy. The latter is a function of the temperature of the room air T_{room} and its specific moisture $x_{room,out}$:

$$h_{room,out} = h\left(T_{room}, x_{room,out}\right). \tag{5.50}$$

The values of the thermal resistances and capacities in Eqs. (5.44)-(5.49) depend on geometrical quantities and material properties of the various elements within the computer room. Further details on their computation are reported in Appendix A.1. Moreover, the considered model takes into account the pressure drops through the perforated tiles

in a typical raised-floor configuration. Their computation is detailed in Appendix A.2.

The input variables of the computer room model are:

- the thermal load $Q_{load,int}$ produced by the electrical equipment contained in the racks, expressed in W;
- the outdoor air temperature $T_{outdoor}$, expressed in °C;
- the enthalpy flow $\dot{m}_{air}\, h_{room,in}$ of the air coming from the cooling unit, expressed in W. It is referred to the airflow $\dot{m}_{air}$ elaborated by the cooling system and to the air specific enthalpy $h_{room,in}$. The latter is a function of the temperature $T_{room,in}$ and specific moisture $x_{room,in}$ of the incoming air:

$$h_{room,in} = h\left(T_{room,in}, x_{room,in}\right). \tag{5.51}$$

The output variables of the model are:

- the temperature T_{rack} of the racks in the Data Center, expressed in °C;
- the temperature T_{room} of the air in the computer room, i.e., the temperature of the air returned to the cooling unit, expressed in °C;
- the temperature $T_{wall,in}$ of the internal surface of the equivalent wall, expressed in °C;
- the temperature $T_{wall,out}$ of the external surface of the equivalent wall, expressed in °C;
- the specific moisture $x_{room,out}$ of the air returned to the cooling unit, i.e., the specific moisture of the air in the computer room, expressed in $\mathrm{kg_{wv}\,kg_{da}^{-1}}$.

5.2.3 Fans Model

The considered IEC system configuration includes fans in both the supply and exhaust air sides. Two different centrifugal fan models have been taken into account, each having an impeller with six curved blades, a three-dimensional profile, and diameter $D = 0.5\,\mathrm{m}$. The fan type used in the supply air side is labelled *Fan 1*, while the one used in the secondary air side is labelled *Fan 2*. *Fan 1* moves the air in the room by forcing it through the evaporative heat exchanger; *Fan 2* elaborates the external air, which is forced through the heat exchanger and then is pushed outdoor. Both types of fan are

driven by an electric motor, which is supplied in three-phase and driven by an inverter, that allows the regulation of the rotational speed.

Typically, the main operating parameters of a fan are the following:

- the rotation speed n, expressed in min^{-1} or in $\mathrm{rad\,s}^{-1}$;
- the volumetric flow rate $\dot{V}$, expressed in m^3/h and referred to the fluid condition at the fan aspiration side;
- the fluid total pressure increase p between aspiration section (1) and delivery section (2), expressed in Pa,

$$p = p_2 - p_1; \tag{5.52}$$

- the dynamic pressure p_d, expressed in Pa and conventionally calculated referred to the average speed at the fan delivery section,

$$p_d = \frac{1}{2}\rho_2 c_2^2; \tag{5.53}$$

- the static pressure p_s, expressed in Pa,

$$p_s = p - p_d; \tag{5.54}$$

- the electric power P supplied to the electric motor, expressed in W.

Nominal specifications of the two types of fans are reported in Tab. 5.3.

Fan	Max. flow rate	Nominal electrical power	Max. rotation speed
1	$13\,650\,\mathrm{m^3\,h^{-1}}$	3.5 kW	$1860\,\mathrm{min^{-1}}$
2	$15\,500\,\mathrm{m^3\,h^{-1}}$	5.4 kW	$2135\,\mathrm{min^{-1}}$

Table 5.3: Nominal fans characteristics.

Considering as hydraulic diameter the outer diameter of the impeller and referring to the air density at the intake section, the following characteristic figures can be defined:

- flow figure,

$$\phi = \frac{\dot{V}}{nD^3}; \tag{5.55}$$

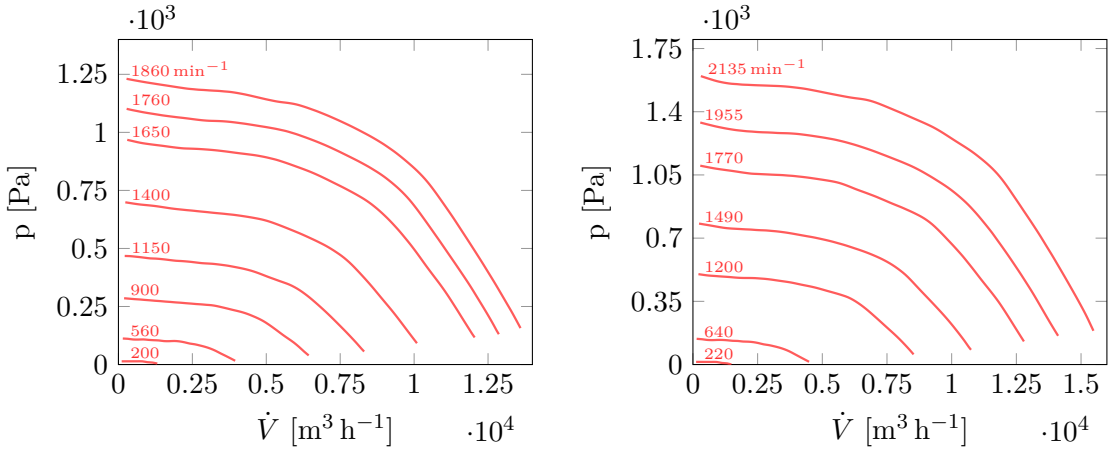
- total pressure figure,

$$\psi = \frac{p}{\rho n^2 D^2}; \tag{5.56}$$

- static pressure figure,

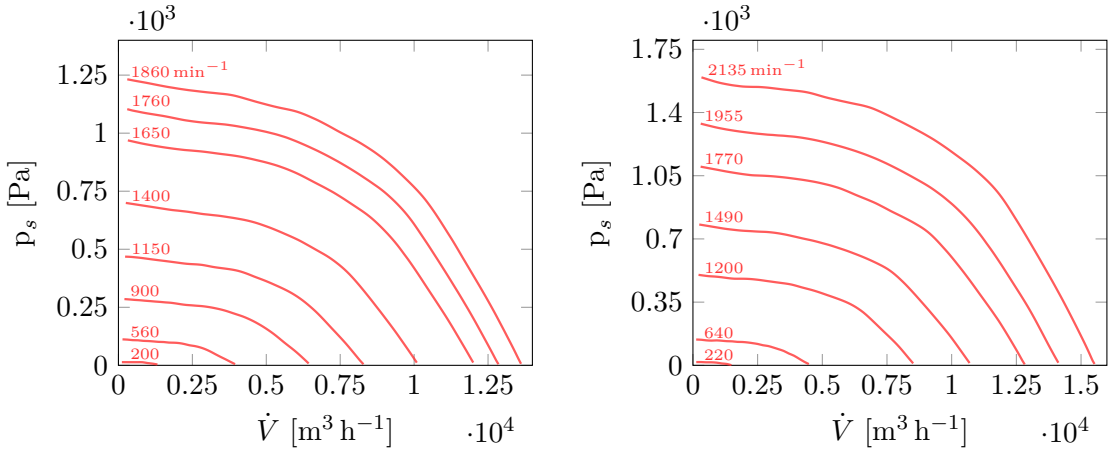
$$\psi_s = \frac{p_s}{\rho n^2 D^2}. \tag{5.57}$$

Fans performance is typically reported in $p - \dot{V}$ and $P - \dot{V}$ diagrams, called characteristic curves. The characteristic curves of the two fan models are reported in Figs. 5.15, 5.16, and 5.17.



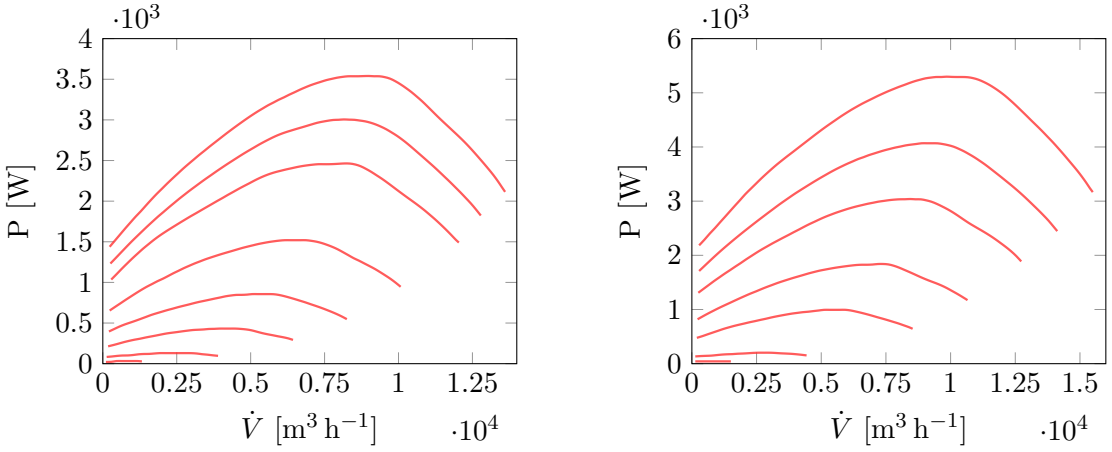
(a) Fan 1. **(b)** Fan 2.

Figure 5.15: Total pressure-flow rate characteristic curves.



(a) Fan 1. **(b)** Fan 2.

Figure 5.16: Static pressure-flow rate characteristic curves.



(a) Fan 1. **(b)** Fan 2.

Figure 5.17: Power-flow rate characteristic curves.

In order to describe the pressure contribution and the electrical power consumption of the fans, we derive a black-box model based on the correlations proposed by Wang et al. (2004). More specifically, the pressure energy contribution is modelled with polynomials based on the characteristic figures previously introduced:

$$\psi(\phi) = a_0 + a_1\,\phi + a_2\,\phi^2 + a_3\,\phi^3 + a_4\,\phi^4, \tag{5.58}$$

$$\psi_s(\phi) = b_0 + b_1\,\phi + b_2\,\phi^2 + b_3\,\phi^3 + b_4\,\phi^4. \tag{5.59}$$

Similarly, the electrical power consumed by the fans during their operation is modelled with the following polynomial:

$$\begin{aligned} P(\phi, n) = {} & p_{00} + p_{10}\,\phi + p_{01}\,\frac{n}{n_{nom}} + p_{11}\,\phi\,\frac{n}{n_{nom}} + p_{20}\,\phi^2 + p_{02}\left(\frac{n}{n_{nom}}\right)^2 + p_{21}\,\phi^2\,\frac{n}{n_{nom}} + \\ & p_{12}\,\phi\left(\frac{n}{n_{nom}}\right)^2 + p_{22}\,\phi^2\left(\frac{n}{n_{nom}}\right)^2 + p_{30}\,\phi^3 + p_{03}\left(\frac{n}{n_{nom}}\right)^3 + \\ & p_{31}\,\phi^3\,\frac{n}{n_{nom}} + p_{13}\,\phi\left(\frac{n}{n_{nom}}\right)^3 + p_{04}\left(\frac{n}{n_{nom}}\right)^4. \end{aligned} \tag{5.60}$$

All the coefficients of polynomials in Eqs. (5.58), (5.59), and (5.60) have been inferred

from experimental data. Available data regard total pressure p, static pressure p_s, and power P curves reported in Figs. 5.15, 5.16, and 5.17, and they are referred to the incoming air density measured on standard nozzle on type A installation[1] according to ISO5801 and equal to $1.16\,\mathrm{kg\,m^{-3}}$. For each fan and for each of the three considered quantities, the available data have been divided into two disjoint calibration and test data sets DS_c and DS_t. The cardinalities of all the resulting calibration and test data sets are reported in Tab. 5.4. The calibration of the polynomials coefficients have been

	$\#DS_c$	$\#DS_t$
Fan 1	369	185
Fan 2	323	162

(a) Total pressure.

	$\#DS_c$	$\#DS_t$
Fan 1	374	188
Fan 2	320	160

(b) Static pressure.

	$\#DS_c$	$\#DS_t$
Fan 1	299	150
Fan 2	260	130

(c) Power.

Table 5.4: Cardinality of all the calibration and test sets.

considered as a least squares problem. The parameters obtained from the least squares regression related to the polynomials of Eqs. (5.58)-(5.60) are reported respectively in Tabs. 5.5, 5.6, and 5.7.

	a_0	a_1	a_2	a_3	a_4
Fan 1	4.465	−2.106	6.832	−10.92	1.832
Fan 2	4.377	−1.499	4.441	−7.017	−0.3796

Table 5.5: Regression polynomials parameters of the total pressure.

	b_0	b_1	b_2	b_3	b_4
Fan 1	4.500	−2.513	7.858	−13.09	2.821
Fan 2	4.387	−1.659	4.634	−8.364	0.4513

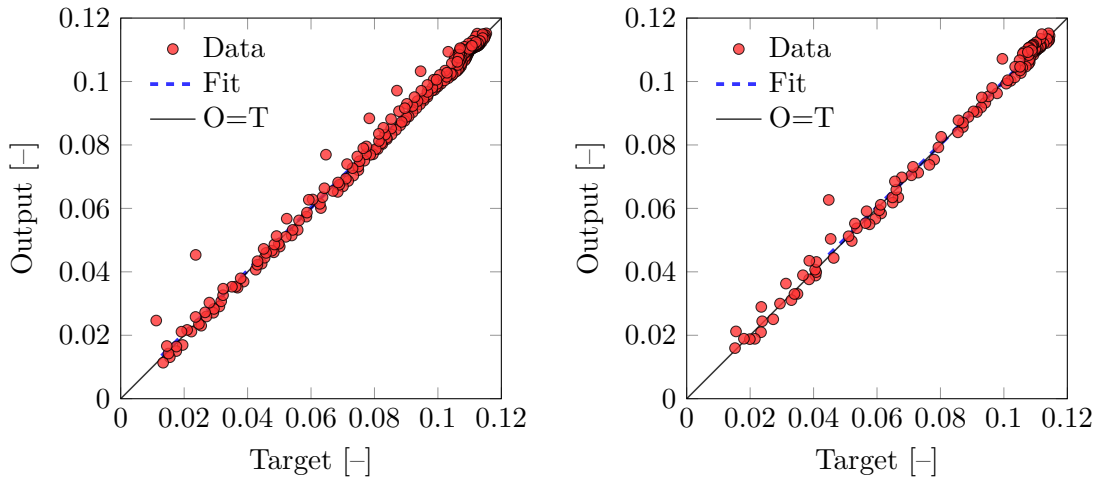
Table 5.6: Regression polynomials parameters of the static pressure figure.

Once obtained the polynomials parameters values, the corresponding $\hat{\psi}$, $\hat{\psi}_s$ and $\hat{P}$ are calculated for calibration and test sets. Fig. 5.18 reports the comparison between observed data $\hat{\psi}$ (Target) and ψ values (Output) for both calibration and test data sets. The same comparisons are reported in Figs. 5.19 and 5.20 for the quantities ψ_s and P, respectively.

[1] Installation with free incoming and outgoing section.

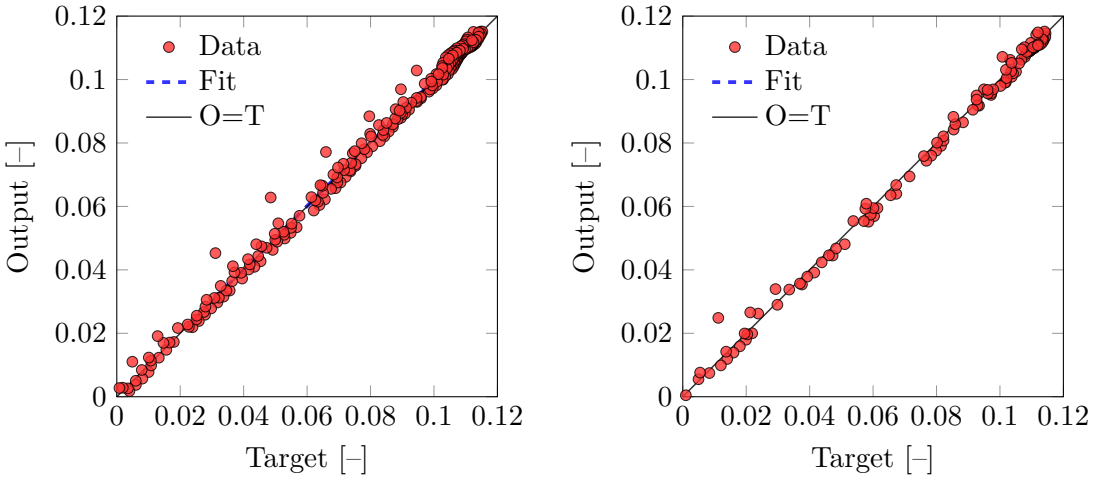
	Fan 1	Fan 2
p_{00}	−111.3	−208.7
p_{10}	2523	4461
p_{01}	327.7	739.2
p_{11}	−10060	−17010
p_{20}	5384	−9473
p_{02}	978.6	954.4
p_{21}	17860	29560
p_{12}	10640	17570
p_{22}	−12970	−20620
p_{30}	3019	5294
p_{03}	−970.1	−956.1
p_{31}	−7487	−12230
p_{13}	2351	3085
p_{04}	1161	1567

Table 5.7: Regression polynomials parameters of the static electric power.



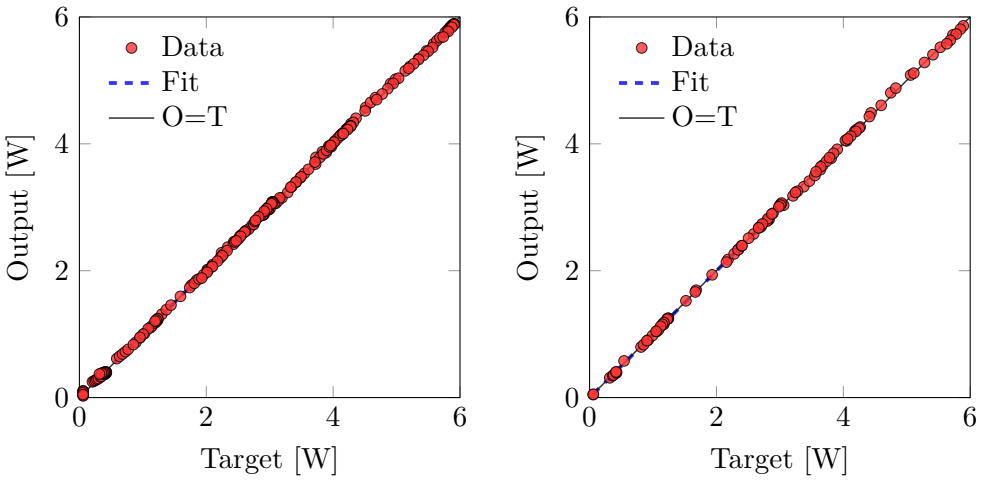
(a) Calibration: RMSE = 0.0027 **(b)** Validation: RMSE = 0.0029

Figure 5.18: Comparison between the total pressure figure provided by the model and observed data.



(a) Calibration: RMSE = 0.0023 **(b)** Validation: RMSE = 0.0028

Figure 5.19: Comparison between the static pressure figure provided by the model and observed data.



(a) Calibration: RMSE = 23.1 W **(b)** Validation: RMSE = 24.03 W

Figure 5.20: Comparison between the electrical power provided by the model and observed data.

5.2.4 Pump Model

The considered IEC unit includes a multi-stage submersible pump which takes water from a reservoir and vaporizes it into the process channels of the heat exchanger. The pump is driven by a 2-pole induction motor supplied at 50 Hz.

Typically, the main operating parameters of a pump are the following:

- the rotation speed n, expressed in min^{-1} or in $\mathrm{rad\,s}^{-1}$;
- the flow rate $\dot{V}$, expressed in $\mathrm{m}^3\,\mathrm{h}^{-1}$;
- the hydraulic head H, expressed in m;
- the electrical power consumption P, expressed in W;
- the pump efficiency η, which is dimensionless.

Nominal specifications of the considered pump are reported in Tab. 5.8. Pump per-

Flow rate range	Nominal electrical power	Nominal rotation speed
2.5 – 8 $\mathrm{m}^3\,\mathrm{h}^{-1}$	0.9 kW	2900 min^{-1}

Table 5.8: Nominal pump characteristics.

formance is typically reported in $H - \dot{V}$ and $\eta - \dot{V}$ diagrams, called characteristic curves. The characteristic curves of the considered pump at the nominal rotation speed $n_n = 2900\,\mathrm{min}^{-1}$ are reported in the pump data-sheet and they can be described by the following polynomials:

$$H(\dot{V}) = c_0 + c_1\,\dot{V} + c_2\,\dot{V}^2 + c_3\,\dot{V}^3 + c_4\,\dot{V}^4 + c_5\,\dot{V}^5, \tag{5.61}$$

$$\eta(\dot{V}) = d_0 + d_1\,\dot{V} + d_2\,\dot{V}^2 + d_3\,\dot{V}^3 + d_4\,\dot{V}^4, \tag{5.62}$$

where coefficients $c_0, \ldots, c_5$ and $d_0, \ldots, d_4$ have been inferred by fitting the pump characteristic curves. The resulting values of the coefficients are reported in Tabs. 5.9 and 5.10 and the corresponding fitted polynomials are compared with the data-sheet characteristic curves in Fig. 5.21.

c_0	c_1	c_2	c_3	c_4	c_5
12.28	21.49	−9.89	1.99	-1.9×10^{-1}	7.4×10^{-3}

Table 5.9: Coefficients of the $H - \dot{V}$ polynomial at the nominal rotation speed.

d_0	d_1	d_2	d_3	d_4
2.50	30.84	−7.99	1.08	−0.06

Table 5.10: Coefficients of the $\eta - \dot{V}$ polynomial at the nominal rotation speed.

In order to maintain the flow rate $\dot{V}$, the pump hydraulic head must be equal to the total system head loss:

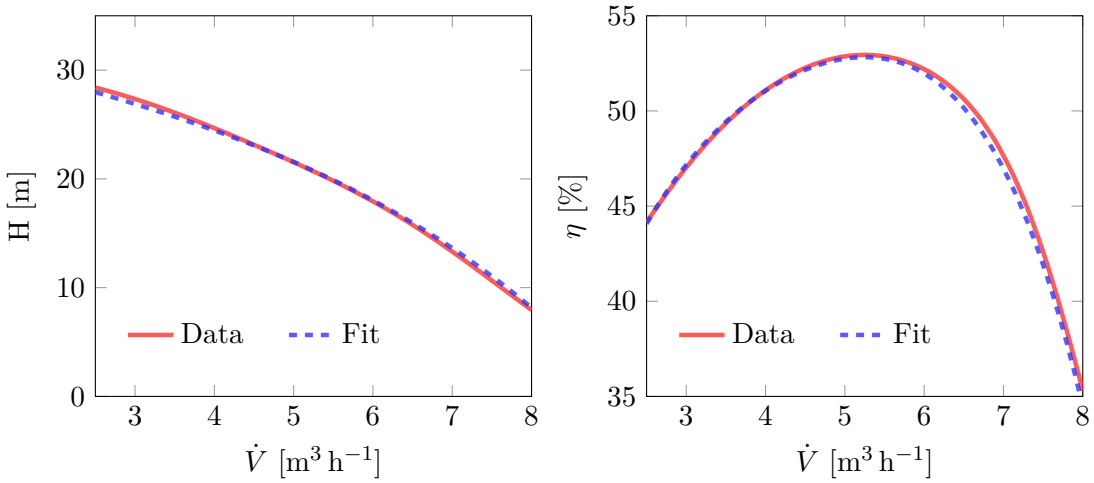
$$H(\dot{V}) = K\dot{V}^2, \tag{5.63}$$

where K is the head loss coefficient. The operating point $\dot{V}$ of the pump is given by the intersection between the pump curve in Eq. (5.61) and the system curve in Eq. (5.63). The electrical power consumed by the pump at the operating point $\dot{V}$ can be obtained as:

$$P = \frac{\dot{V}H(\dot{V})}{\eta(\dot{V}_n)}, \tag{5.64}$$

where the efficiency η is obtained by evaluating polynomial (5.62) at the nominal flow rate $\dot{V}_n$, which is given by:

$$\dot{V}_n = \frac{n_n}{n}\dot{V}. \tag{5.65}$$



(a) Experimental and fitted H-$\dot{V}$ curves. The RMSE is equal to 0.27 m.

(b) Experimental and fitted η-$\dot{V}$ curves. The RMSE is equal to 0.37 %.

Figure 5.21: Comparison between the characteristic curves reported in the data-sheet of the pump and the fitted polynomials given by Eqs. (5.61) and (5.62).

5.3 The Simulation Environment

A simulation environment based on Matlab/Simulink has been designed according to the mathematical model described in the previous section. The benefit of using the simulation environment is twofold. First, we can generate synthetic data which are related to some fundamental aspects of the thermal behaviour of the IEC system as well as understanding parameters influence during the process. Moreover, we can design and test different control strategies to efficiently manage the IEC system and evaluate their performance by means of simulations.

The inputs of the simulation environment are:

- the temperature $T_{outdoor}$ of the process air entering the heat exchanger, which is equal to the outdoor air temperature and which is expressed in °C;
- the specific humidity $x_{outdoor}$ of the process air entering the heat exchanger, which is equal to the outdoor air specific humidity and which is expressed in $\mathrm{kg_{wv}\,kg_{da}^{-1}}$;
- the thermal load Q_{th} produced by the electrical equipment contained in the racks expressed in W;
- the volumetric flow rate $\dot{V}_e$ of the process (exhaust) air expressed in $\mathrm{m^3\,h^{-1}}$, which corresponds to the mass flow rate $\dot{m}_e$ expressed in $\mathrm{kg_{da}\,s^{-1}}$;
- the volumetric flow rate $\dot{V}_s$ of the supply air expressed in $\mathrm{m^3\,h^{-1}}$, which corresponds to the mass flow rate $\dot{m}_s$ expressed in $\mathrm{kg_{da}\,s^{-1}}$;
- the water flow rate $\dot{V}_w$ sprinkled by the evaporative system in the process air channels of the heat exchanger, expressed in $\mathrm{L\,h^{-1}}$.

The main outputs provided by the simulation environment are:

- the temperature T_{supply} of the air supplied into the computer room expressed in °C;
- the specific humidity x_{supply} of the air supplied into the computer room expressed in $\mathrm{kg_{wv}\,kg_{da}^{-1}}$;
- the temperature T_{return} returning from the computer room expressed in °C;
- the specific humidity x_{return} returning from the computer room expressed in $\mathrm{kg_{wv}\,kg_{da}^{-1}}$.

Figs. 5.22, 5.23, and 5.24 show part of the IEC simulation environment developed in Simulink.

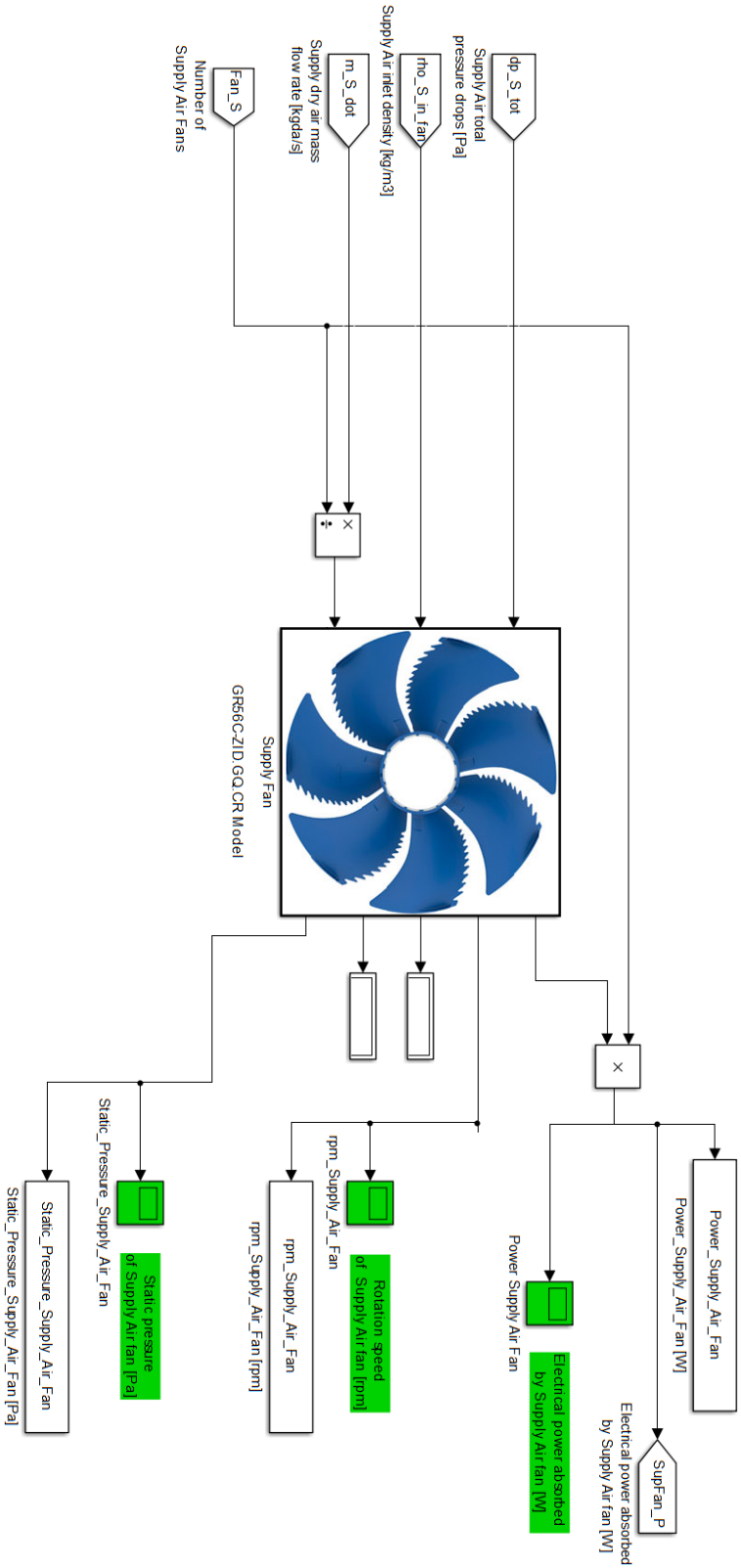


Figure 5.22: Detail of the Simulink scheme of supply air fans model.

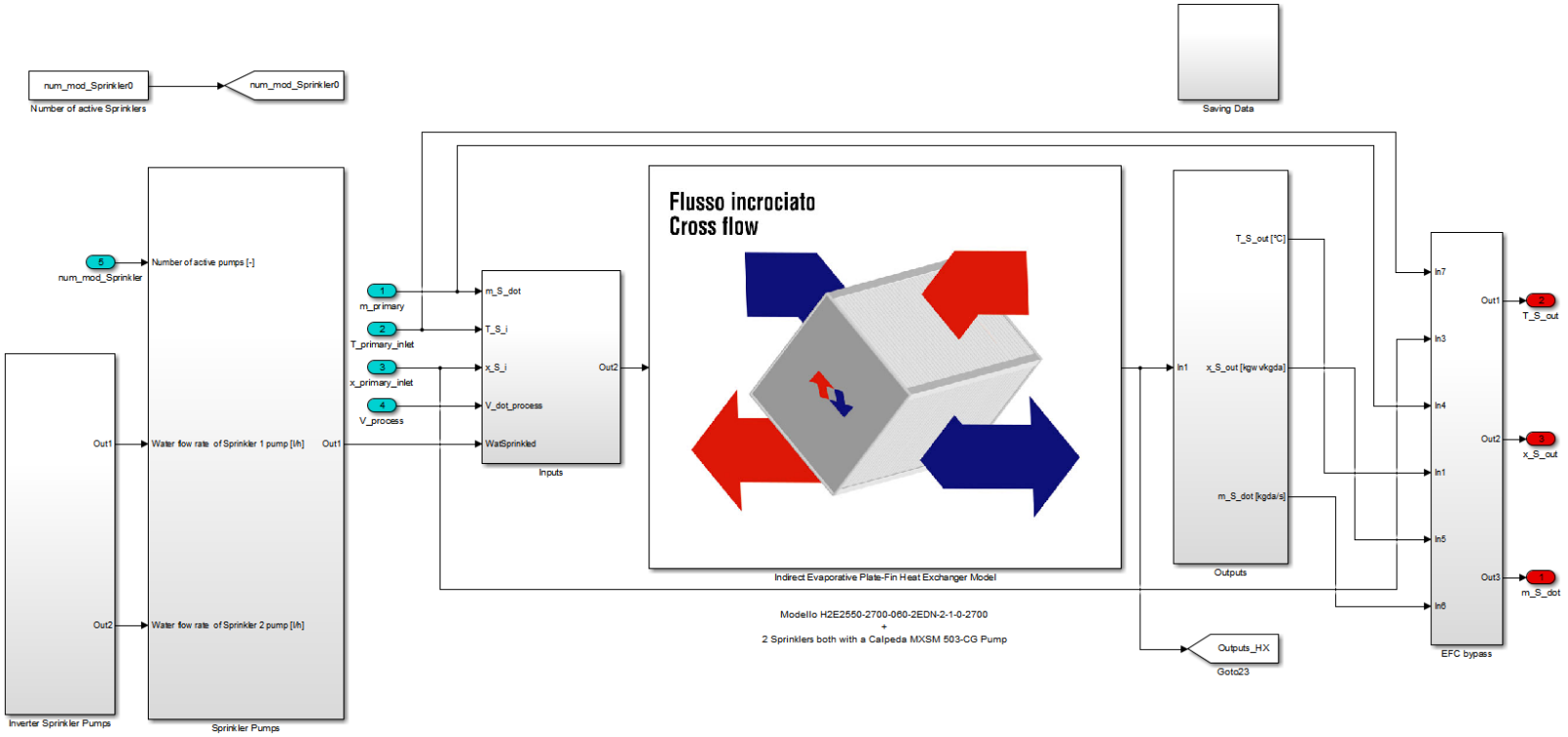


Figure 5.23: Detail of the Simulink scheme of the indirect evaporative heat exchanger model.

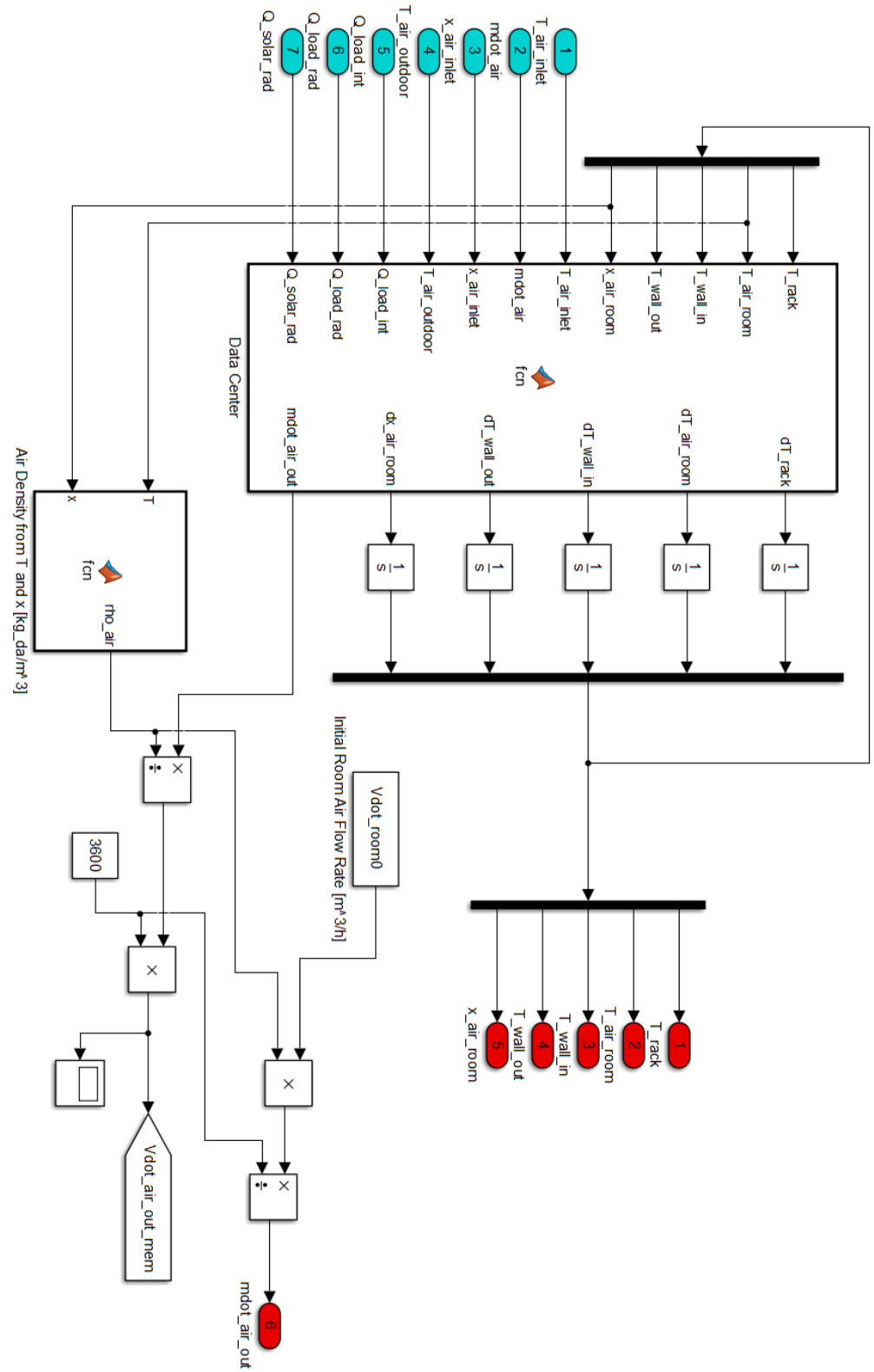


Figure 5.24: Detail of the Simulink scheme of the server room model.

5.4 Design of the Control System

In the considered scenario, optimal operation is achieved when the primary supply air temperature T_{supply} is maintained below a given safety threshold and, at the same time, the electrical power consumed by the cooling unit is minimized. The supply air temperature can be controlled by adjusting the secondary process airflow $\dot{V}_e$ and the sprinkled water flow $\dot{V}_w$ via the fans and pumps VFDs. Potentially, also the primary supply airflow $\dot{V}_s$ affects the cooling effectiveness of the targeted unit and it may be considered as a controllable input. However, a constant nominal airflow rate[2] $\dot{V}_{s,nom}$ is typically required to avoid maldistribution of the air in the computer room and hot spots on the top of the racks. This entails the reduction of the control DOF from three to two. Typically, standard regulators are used to control the supply air temperature. However, the cooling needs may be met with several different combinations of the controllable inputs values. The optimal operation is achieved by choosing the best (i.e., minimum energy consumption) combination of control inputs that guarantees matching of the temperature constraint. For this purpose, a supervisory control architecture can be designed. This kind of supervisory control system generally provides two levels of control, i.e., local and supervisory levels. In the considered scenario, the local low-level controller consists in standard regulators (i.e., PID controllers) designed to adjust the speed of fans and water pumps to keep track of the corresponding flow rate set-points in spite of the local fast process dynamics. The supervisory control is the focus of this work; it is designed to find energy-efficient flow rate set-points for all the local controllers, taking into account the slower system level dynamics and interactions and the effect of external disturbances.

The resulting supervisory control problem considers two independent manipulable variables: exhaust airflow rate $\dot{V}_e$ and the flow rate $\dot{V}_w$ of the water processed by the pumps. Operating the IT equipment safely induces a requirement on the supply air temperature:

$$T_{supply}(t) \leq \bar{T}, \tag{5.66}$$

where $\bar{T}$ is a constant threshold. In the following we consider the threshold value $\bar{T} = 22.5\,^{\circ}\mathrm{C}$, that is in the middle of the ASHRAE recommended range. A lower value would ensure an even more safe operation of the IT equipment, but requiring more power to operate the cooling unit. Conversely, a higher value of $\bar{T}$ would be more economic but more unsafe especially in case of faults or significant heat load variations. Both the

[2]The nominal airflow rate value depends, for example, on the design of floor tiles, air ducts, and servers fans and on racks positioning.

manipulable variables contribute significantly to the thermal and cost profiles of the IEC unit. This produces a need for adaptive optimizers that can reject varying heat load and environmental conditions with a satisfactory economic performance. Seeking the minimum power consumption, we consider the following absorption index:

$$P(\dot{V}_s, \dot{V}_e, \dot{V}_w) = P_{fans,s}(\dot{V}_s) + P_{fans,e}(\dot{V}_e) + P_{pumps}(\dot{V}_w), \tag{5.67}$$

and account for the constraint (5.66) by including a one-sided barrier function:

$$C(t) = (\max\{0, T_{supply}(t) - \bar{T}\})^2 \tag{5.68}$$

in the overall cost index:

$$y^o(t) = P(\dot{V}_s(t), \dot{V}_e(t), \dot{V}_w(t)) + \lambda C(t). \tag{5.69}$$

The parameter λ determines the weight of the constraint barrier function in the overall cost. In order to give the same relevance to the two terms in the cost index, we take $\lambda = 500\,\mathrm{W\,^{\circ}C^{-2}}$.

The selection of the supervisory control method for the minimization of the cost index (5.69) plays a critical role in the development of an effective control strategy. Supervisory control could be classified into two main categories: model-based and model-free supervisory control. The main tools required to perform model-based methods are models of the system components, which must be able to predict accurately the system energy and environment performance, as well as the system response to changes of the control settings. Even if we have designed an accurate model of the considered IEC system, it is very complex and non-linear. In particular, the model of the indirect evaporative heat exchanger requires the solution of equations that cannot be solved algebraically and the use of external libraries for calculating thermodynamic and transport properties of pure fluids and mixtures. Moreover, the values of model parameters must be repeatedly updated during the operation of the system, in order to keep the predictions provided by the model aligned with the real system response. For all these reasons, the design of a model-based supervisory control can require a considerable effort, especially if the computational resources available for the implementation of the controller are limited. In addition to these considerations, a model-based supervisory controller would be tailored to the specific considered configuration. If some of the components of the system are replaced, then the correspondent model must be changed accordingly. Conversely, model-free supervisory control methods do not require a model of the targeted system.

Among all the model-free supervisory techniques, the ESC method described in Chapter 4 has drawn significant attention in the control of general HVAC systems for its effectiveness and simplicity of implementation. Thus, we decided to design an ESC-based supervisory control layer for the optimization of the IEC unit operation. More precisely, we designed a two-variable Newton-like Phasor ESC scheme, which allows to exploit an estimation of both first and second derivatives of the targeted cost function to perform a Newton-like step and to drive the manipulable control variables in the direction of reduced cost index. In the considered scenario, we define the control input vector:

$$\boldsymbol{\theta} = \begin{bmatrix} \dot{V}_e \\ \dot{V}_w \end{bmatrix}. \tag{5.70}$$

Let $Q(\boldsymbol{\theta})$ denote the steady state value of the cost index y^o given the input $\boldsymbol{\theta}$. The supervisory control task is to optimize the plant operation online by steering the control inputs toward the optimal value $\boldsymbol{\theta}^*$ which minimizes the steady state value of the cost index:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \mathbb{R}^m}{\arg\min}\, Q(\boldsymbol{\theta}). \tag{5.71}$$

The control input is continuously updated as follows:

$$\boldsymbol{\theta}(t) = \widehat{\boldsymbol{\theta}}(t) + \mathbf{s}(t), \tag{5.72}$$

where $\mathbf{s}(t)$ is a vector of additive sine wave perturbation signals:

$$\mathbf{s}(t) = \begin{bmatrix} a_e \sin(\omega_e t) \\ a_w \sin(\omega_w t) \end{bmatrix}, \tag{5.73}$$

and $\widehat{\boldsymbol{\theta}}(t)$ is the current estimate of the optimal input $\boldsymbol{\theta}^*$. By exploiting the input perturbation, the Phasor-based derivative estimator provides the estimation of the gradient $\widehat{\boldsymbol{\nabla} Q}$ and the Hessian $\widehat{\mathbf{H} Q}$ of the cost function with respect to the input $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\nabla} Q} = \begin{bmatrix} \widehat{(\nabla Q)_1} \\ \widehat{(\nabla Q)_2} \end{bmatrix} = \begin{bmatrix} \widehat{\dfrac{\partial Q}{\partial \dot{V}_e}} \\ \widehat{\dfrac{\partial Q}{\partial \dot{V}_w}} \end{bmatrix}, \tag{5.74}$$

$$\widehat{\mathbf{H}Q} = \begin{bmatrix} \widehat{(HQ)_{1,1}} & \widehat{(HQ)_{1,2}} \\ \widehat{(HQ)_{1,2}} & \widehat{(HQ)_{2,2}} \end{bmatrix} = \begin{bmatrix} \widehat{\frac{\partial^2 Q}{\partial \dot{V}_e^2}} & \widehat{\frac{\partial^2 Q}{\partial \dot{V}_w \partial \dot{V}_w}} \\ \widehat{\frac{\partial^2 Q}{\partial \dot{V}_w \partial \dot{V}_w}} & \widehat{\frac{\partial^2 Q}{\partial \dot{V}_w^2}} \end{bmatrix}. \tag{5.75}$$

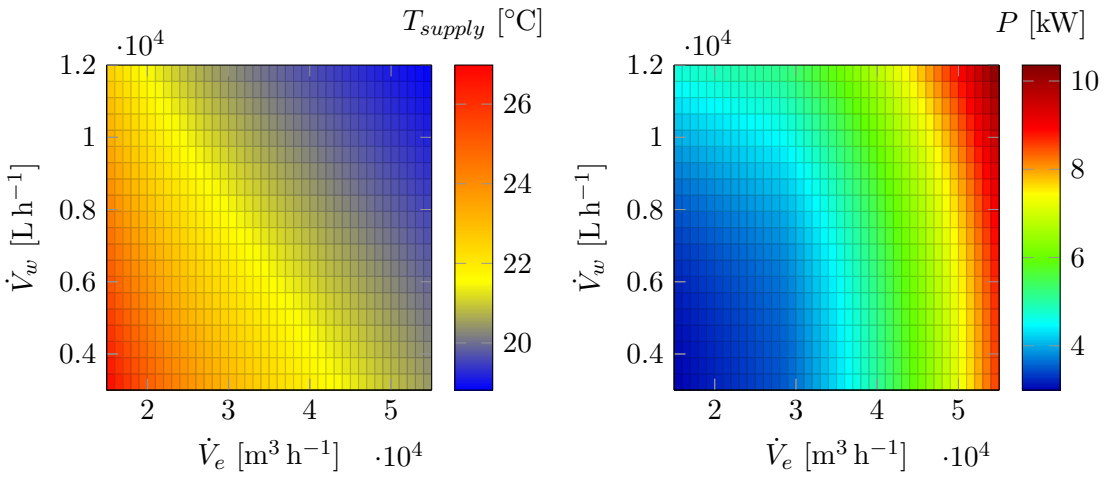
Then, by using the RDE (4.31), the inverse $\mathbf{\Gamma}$ of the estimated Hessian is computed and, finally, the Newton-like direction is obtained by using Eq. (4.33). The estimate $\widehat{\boldsymbol{\theta}}$ of the optimal input is then updated by integrating Eq. (4.14b) with integration gains:

$$\mathbf{k} = \begin{bmatrix} k_e \\ k_w \end{bmatrix}. \tag{5.76}$$

In the next Section, we exploit the simulation environment presented in Section 5.3 to analyze some of the properties of the considered system and to test in silico the effectiveness of the proposed architecture. In the simulations of the controller, we consider sinusoidal perturbations having amplitudes $a_e = 500\,\mathrm{m^3\,h^{-1}}$ and $a_w = 112.5\,\mathrm{L\,h^{-1}}$ and frequencies $\omega_e = 0.0209\,\mathrm{rad\,s^{-1}}$ and $\omega_w = 0.0524\,\mathrm{rad\,s^{-1}}$. Also, we consider the integrator gains $k_e = -1 \times 10^{-3}$ and $k_w = -1.67 \times 10^{-4}$.

5.5 Simulation Examples

The simulation environment presented in Section 5.3 offers a great support in the design of the control system for the considered IEC unit. In particular, it allows to analyze properties of the targeted system, obtain a first tuning of the designed control architecture, and assess its effectiveness in different operative conditions, without testing it directly in the real system which can be time and cost consuming and unsafe. Let us consider a first numerical scenario where the heat load to reject from the computer room is $Q_{th} = 90\,\mathrm{kW}$ and the outdoor air temperature and relative humidity are $T_{outdoor} = 27\,^{\circ}\mathrm{C}$ and $RH_{outdoor} = 40\,\%$, respectively. By exploiting the simulation environment, we evaluated the steady-state value of the supply air temperature T_{supply} and the total power consumption P as functions of the two control inputs, as shown in Fig. 5.25. The value of the primary air temperature is strongly affected by both the exhaust airflow and the water flow rates and, as expected, it monotonically decreases as the values of the control inputs increases, Fig. 5.25a. Conversely, the power consumption obviously monotonically increases as the two control inputs do, Fig. 5.25b. Moreover, increasing the exhaust airflow rate $\dot{m}_e$ has a higher impact on the power consumption compared to increasing the water flow rate $\dot{m}_w$. By means of simulations, we then evaluated the value of the cost



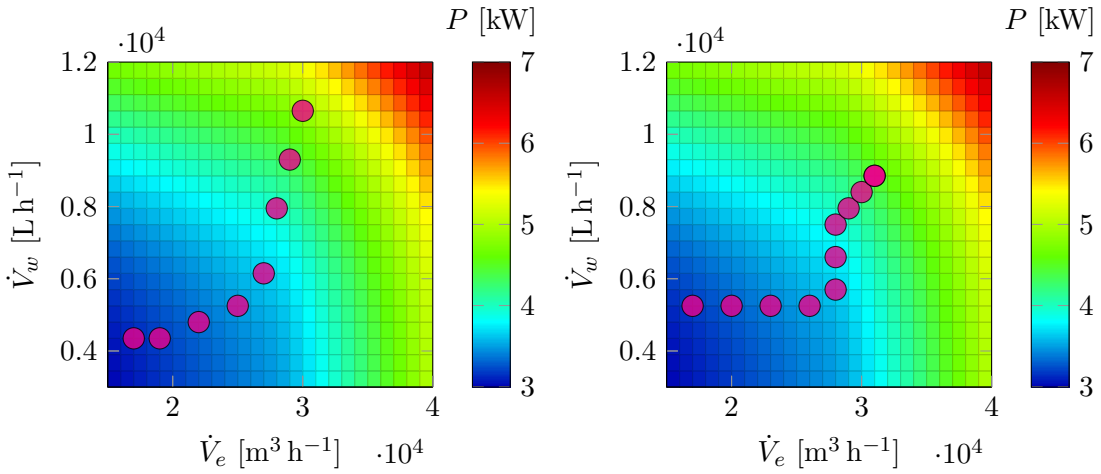
(a) Supply air temperature as a function of the two control inputs.

(b) Power consumption as a function of the two control inputs.

Figure 5.25: Power consumption and supply air temperature as functions of the two control inputs when $Q_{th} = 90\,\text{kW}$, $T_{outdoor} = 27\,^\circ\text{C}$, and $RH_{outdoor} = 40\,\%$. The supply air temperature is slightly more influenced by the airflow rate. However, it has a higher impact on the power consumption.

function (5.69) at several different operation points in order to find the optimal one that minimizes the considered cost function. In Fig. 5.26 we show how the optimal operation point changes when the outdoor air temperature and the IT thermal load varies. As it can be observed, the optimal combination of control inputs strongly depends on changes of the external disturbances. For example, in the considered case it seems to be more convenient adjusting the airflow rate for lower values of disturbances. Conversely, when the effect of the disturbances becomes more significant, increasing solely the airflow rate would imply a higher energy cost and thus it is preferable adjusting the water flow rate. Finding thorough a priori policies that optimize the operation of the cooling unit in a wide range of operative conditions results a complex task and designing automatic control strategies becomes crucial.

Before testing the proposed supervisory control architecture, we exploited the simulation environment to assess the convexity properties of the targeted cost index required for the application of the considered ESC algorithm. The cost index map evaluated as a function of the two control inputs under the operating conditions considered above is shown in Fig. 5.27. More precisely, Fig. 5.27a and Fig. 5.27b report the two terms in (5.69), i.e., the barrier constraint function and the total power consumption, while Fig. 5.27c shows the resulting overall cost index. As it can be observed, the constraint function C is equal to zero in the feasible region where the constraint (5.66) is satisfied. For lower



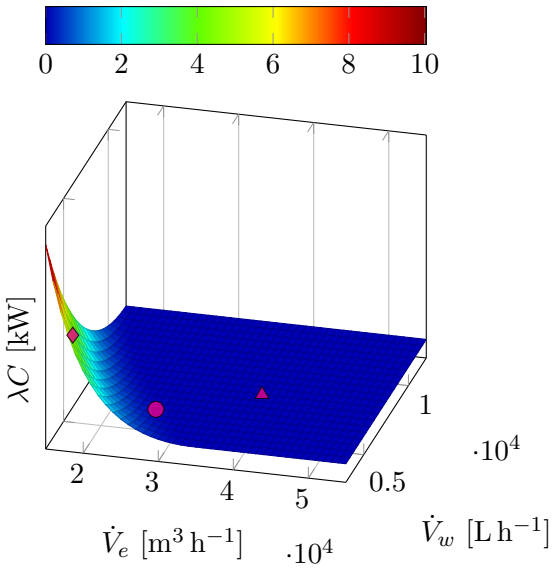
(a) Optimal operation points for increasing values of the outdoor air temperature.

(b) Optimal operation points for increasing values of the heat load.

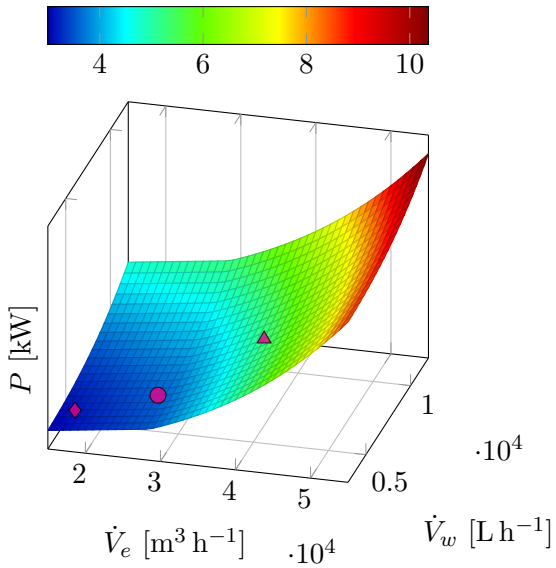
Figure 5.26: Dependence of the optimal operation point on the operating condition. In (a) the outdoor air temperature $T_{outdoor}$ is increased from 24 to 31 °C, while $Q_{th} = 90\,\text{kW}$ and $RH_{outdoor} = 40\,\%$. In (b) the heat load Q_{th} is increased from 90 to 180 kW, while $T_{outdoor} = 27\,°\text{C}$ and $RH_{outdoor} = 40\,\%$.

values of the control inputs, the cooling power provided by the system is not sufficient to overcome the cooling demand and the barrier function grows quadratically. Conversely, the power consumption index grows monotonically with the control inputs. Indeed, the fans/pump affinity laws ensure a cubic relationship between the power and the rotational speed, and they also highlight that the rotational speed is directly proportional to the flow rate. As a result, the overall cost index is the convex map shown in Fig. 5.27c. The circle represents the global minimum of the cost index function, i.e., the operating point that give the lowest power consumption among all the ones in the feasible constraint region. Also the point depicted by a triangle lies in the feasible region (Fig. 5.27a), but it gives a higher power consumption (Fig. 5.27b). Conversely, the operating point depicted by a diamond requires less electrical power than the optimal one (Fig. 5.27b), but it does not satisfy the cooling constraint (Fig. 5.27a). Fig. 5.28 shows the cost index map obtained at three different operating conditions. As it can be observed, the cost function is convex in all the three cases, but its curvature is different and, as a consequence, also the optimal operating point is different.

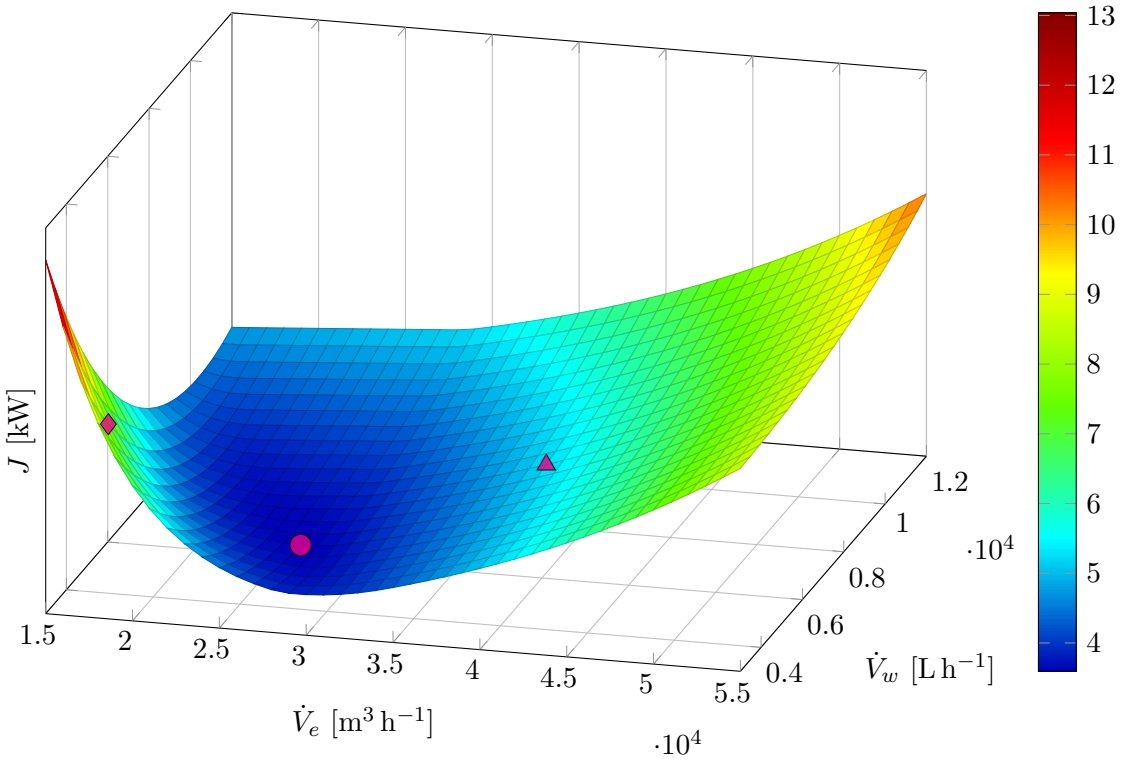
By means of simulations, we tested the effectiveness of the proposed Newton-like Phasor ESC in optimizing the operation of the system. First, we tested the algorithm considering $Q_{th} = 90\,\text{kW}$, $T_{outdoor} = 27\,°\text{C}$, $RH_{outdoor} = 40\,\%$, and two different initial



(a) Barrier constraint function weighted by $\lambda = 500\,\mathrm{W\,°C^{-2}}$.



(b) Power consumption index.



(c) Overall cost index.

Figure 5.27: The constraint function, the power consumption index, and the resulting overall cost when $Q_{th} = 90\,\mathrm{kW}$, $T_{outdoor} = 27\,°\mathrm{C}$, and $RH_{outdoor} = 40\,\%$. The overall cost index is a convex function and the minimum cost operating point is depicted by a circle. Lower control inputs (diamond) are more economic but do not satisfy the temperature constraint. Higher control inputs (triangle) lie in the feasible region but give a higher power consumption.

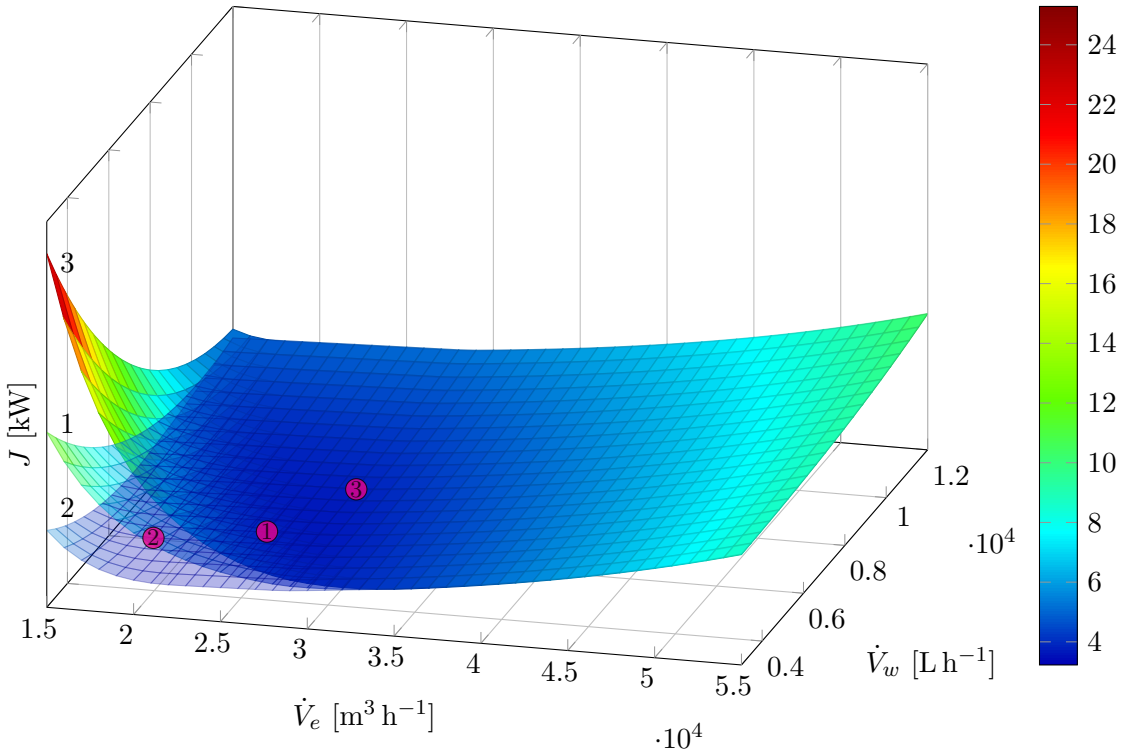
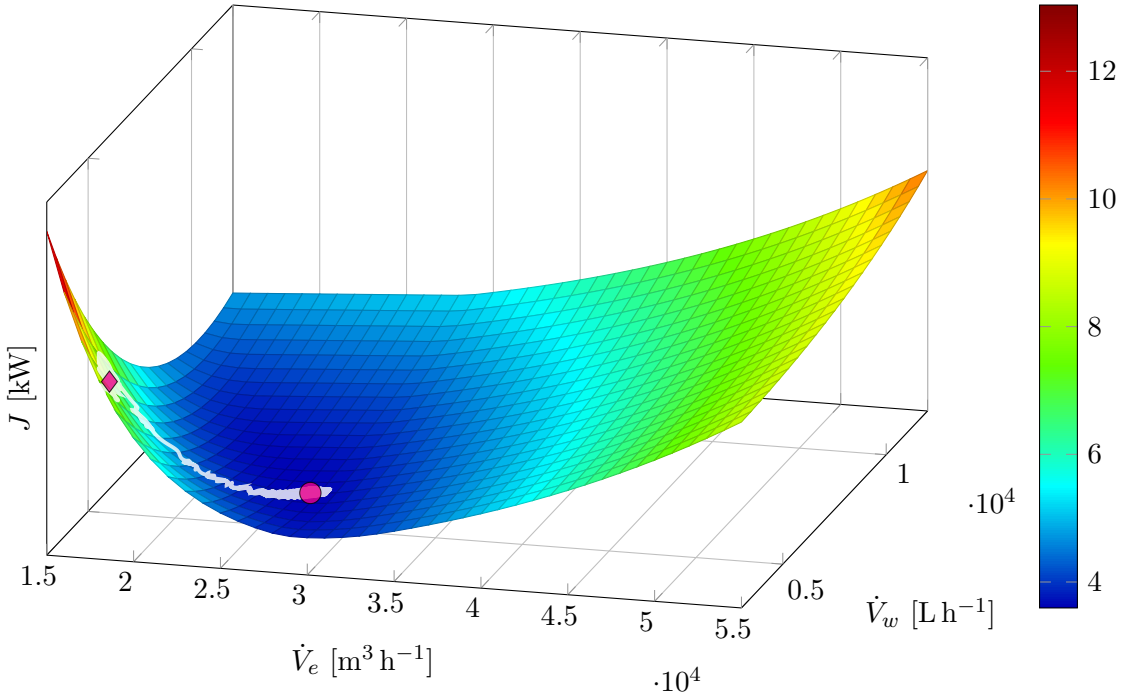


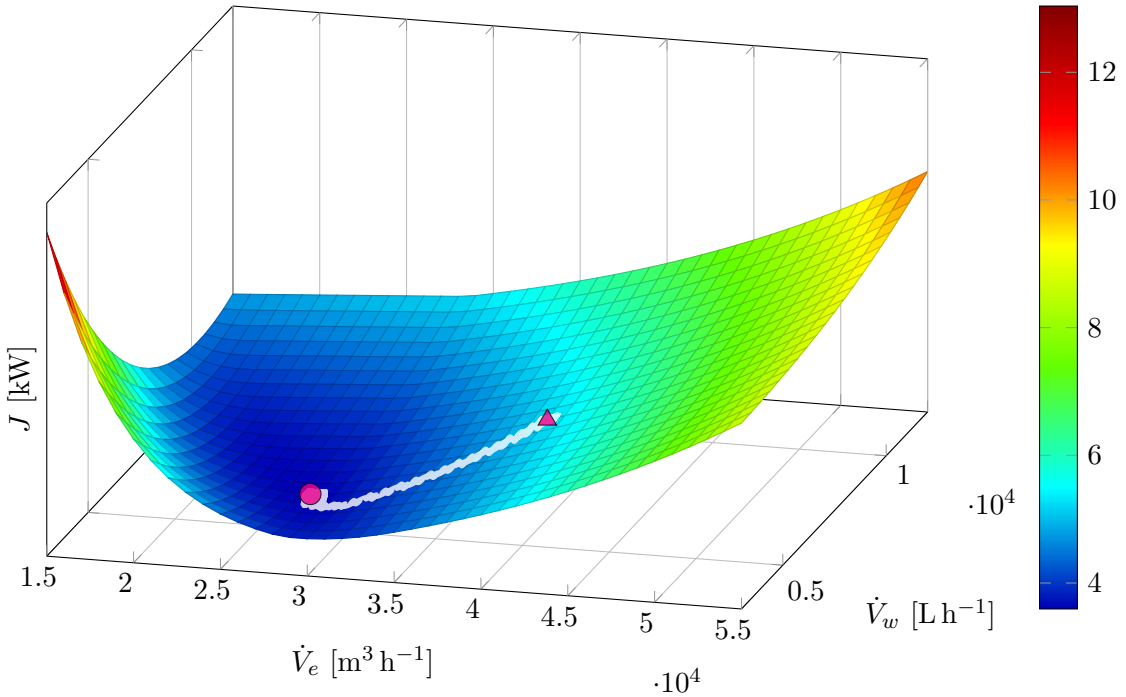
Figure 5.28: The overall cost index at three different operating conditions (1,2, and 3). The cost function is convex in all the three cases, but its shape and curvature is different. As a consequence, the optimum operating point (depicted by the circles) changes with the operative conditions.

operating points. In the simulation represented in Fig. 5.29a, the control inputs are initialized in a high-cost region where the temperature constraint is not satisfied. After the Newton-like Phasor ESC is activated, the two control inputs are driven straight to the minimum-cost point. Conversely, in the simulation represented in Fig. 5.29b, the control inputs are initialized in a feasible but energy-intensive region and the Newton-like ESC scheme is still able to drive the control inputs straight to the optimal operating point. Fig 5.29 reports the time evolution of the input and output signals during the simulation considered in Fig. 5.29a. As it can be observed, in about 2 hours the cost index is effectively reduced to its minimum value (Fig. 5.30a) and the supply air temperature is lowered to the desired threshold (Fig. 5.30b), as a consequence of the control inputs adaptation (Fig. 5.30c).

In order to assess the robustness of the proposed control scheme, we designed a numerical scenario including variations of the main disturbances. Starting from the initial operating condition $Q_{th} = 90\,\text{kW}$, $T_{outdoor} = 27\,^{\circ}\text{C}$, and $RH_{outdoor} = 40\,\%$, we inject at time $t_1 \simeq 10\,\text{h}$ a decrease in the outdoor air temperature $T_{outdoor}$ equal to $1\,^{\circ}\text{C}$. Similarly, at time $t_2 \simeq 17\,\text{h}$ we inject an increase in the IT thermal load Q_{th} equal to $10\,\text{kW}$. The simulation results are summarized in Figs. 5.31 and 5.32. Initially, the supply air temperature is higher than the considered threshold, which entails a high value of the cost function. At time $t_0 \simeq 3\,\text{h}$, the Newton-like Phasor ESC scheme is activated. Th two control inputs (Fig. 5.31c) are increased toward the optimal operating point, which minimizes the overall cost index (Fig. 5.31a). As a result, the supply air temperature is reduced to the considered threshold (Fig. 5.31b). We stress out that lower values of the supply air temperature would still satisfy the thermal constraint, but requiring more cooling effort and thus a higher electrical power consumption. Moreover, we emphasize that, among all the possible combinations of control inputs values that provide the threshold supply temperature, the algorithm finds the one which gives the minimum electrical power cost. When the outdoor air temperature is decreased at time t_1, the controller automatically reacts by decreasing both the input variables in order to reach the new optimal operating point. In fact, the new operating condition is more favourable from a cooling perspective and thus the control inputs can be decreased in order to reduce the electrical power absorption. Conversely, when the IT heat load is increased at time t_2, more cooling effort is required to the system to maintain the temperature constraint satisfied. The controller finds as the most economic solution increasing the exhaust airflow rate instead of the water flow. For completeness, we report in Fig. 5.32 the estimates of both gradient and hessian components of the cost function during the simulation. As expected, the gradient approaches to zero when the control inputs are driven to the

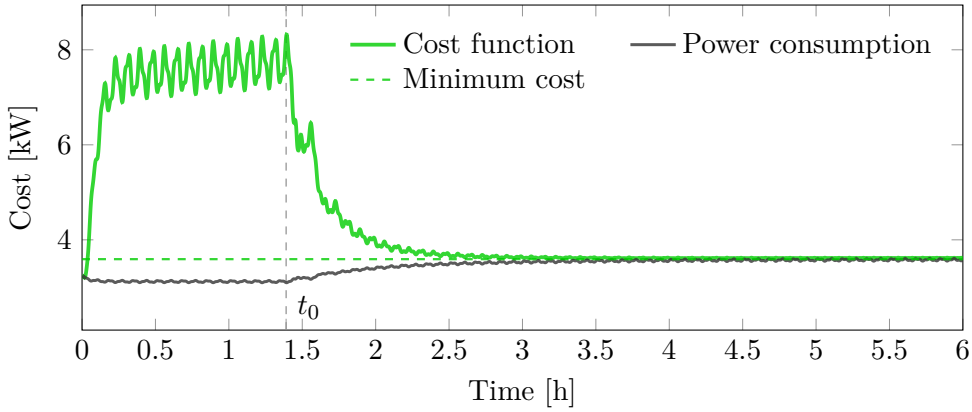


(a) Simulation example: the control inputs are initialized in a point not satisfying the constraint, but the Newton-like ESC scheme drives them to the optimal point of operation.

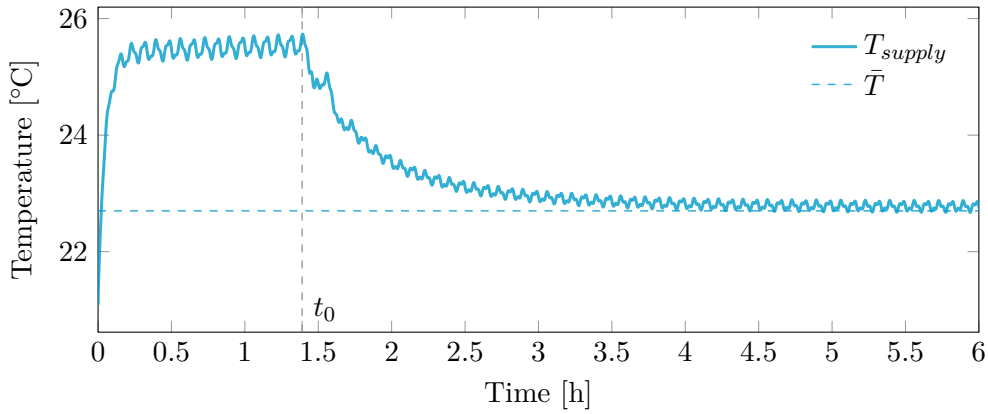


(b) Simulation example: the control inputs are initialized in an energy-intensive point, but the Newton-like ESC scheme drives them to the optimal point of operation

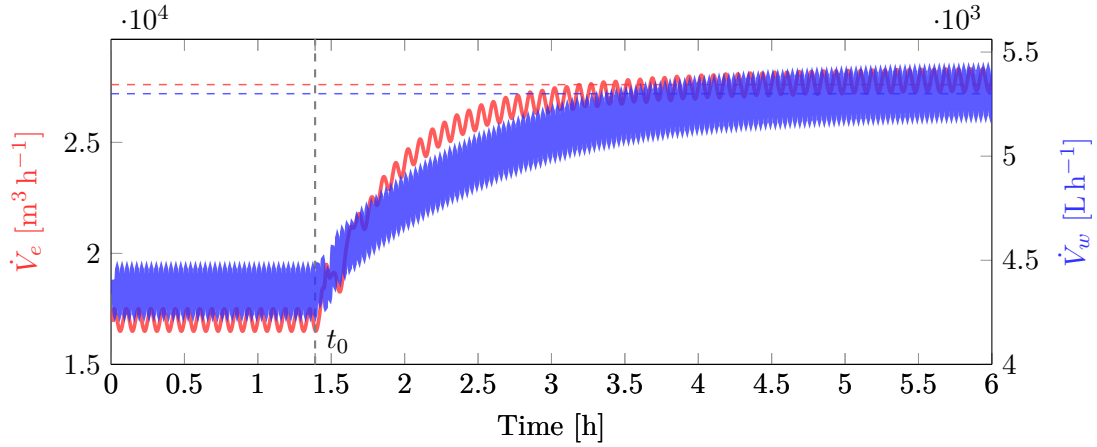
Figure 5.29: Two simulation examples of the Newton-like ESC when $Q_{th} = 90\,\text{kW}$, $T_{outdoor} = 27\,^{\circ}\text{C}$, and $RH_{outdoor} = 40\,\%$. In the two simulations, the control inputs are initialized in two different regions, but in both cases the algorithm drives them straight to the optimal operating point depicted by the circle.



(a) Time evolution of the power consumption and the overall cost function during the simulation.



(b) Time evolution of the supply air temperature during the simulation.

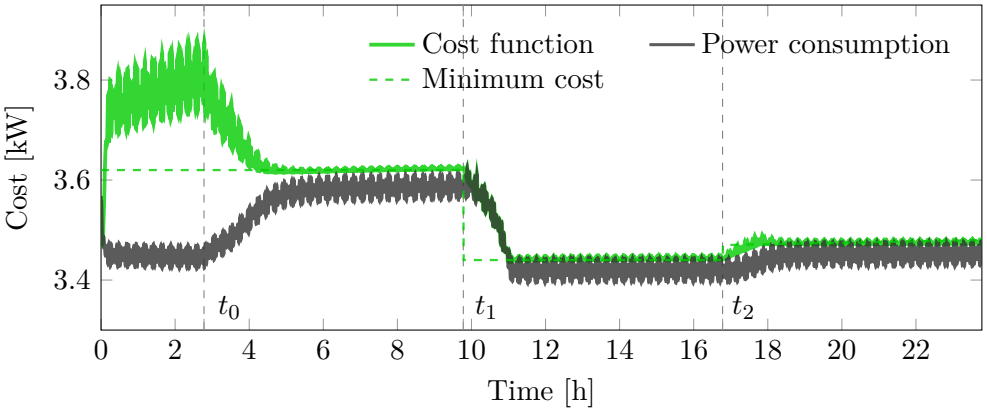


(c) Time evolution of the two control inputs during the simulation. .

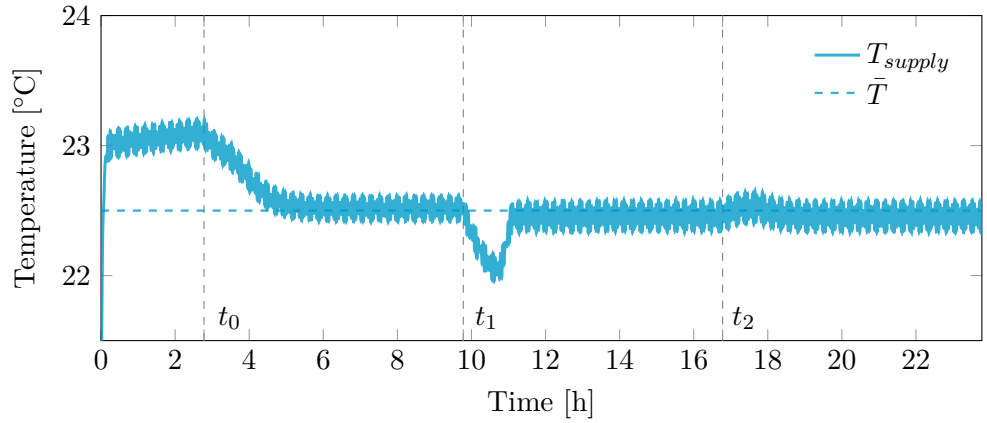
Figure 5.30: Example of simulation of the Newton-like Phasor ESC when $Q_{th} = 90\,\text{kW}$, $T_{outdoor} = 27\,^\circ\text{C}$, and $RH_{outdoor} = 40\,\%$. The control inputs are initialized in an operation point not satisfying the temperature constraint. At time t_0 the algorithm is activated and the control inputs are driven to their optimal values which minimize the cost index in about 2 hours.

minimum-cost operating point, Fig. 5.32a. The estimates of the second order derivates are shown in Figs. 5.32b and 5.32c and, as it can be observed, they converge close to the real[3] second order derivatives evaluated in the optimum operating point. In order to assess the advantages of using second order information, we compared the performance of the Newton-like ESC with that of the classic ESC scheme, which exploits only first order information. For this purpose, we simulated the classic ESC scheme in the numerical scenario described above. For the sake of comparison, we weighted the integrator gain $\mathbf{k}$ in the classic ESC scheme with the inverse of the hessian matrix $\widehat{HQ}^{-1}$ estimated at time t_0. The comparison between the Newton-like Phasor ESC and the classic ESC is shown in Fig. 5.33. As it can be observed, the first algorithm converges faster in all the three different situations. Moreover, the convergence time of the Newton-like Phasor scheme is similar in the three different cases. Conversely, the convergence time of the classic ESC is influenced by the curvature of the cost function and it is different in the three cases, which makes the tuning of the integrator gain of the classic ESC more difficult.

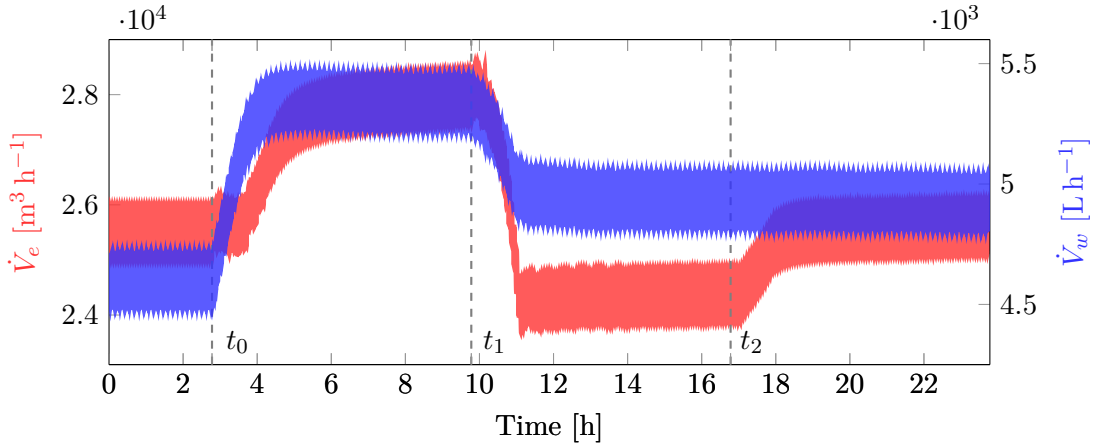
[3]For the sake of assessing the correctness of the hessian estimate, we computed the real hessian components numerically by using the Matlab command *gradient*.



(a) Time evolution of the power consumption and the overall cost function during the simulation.

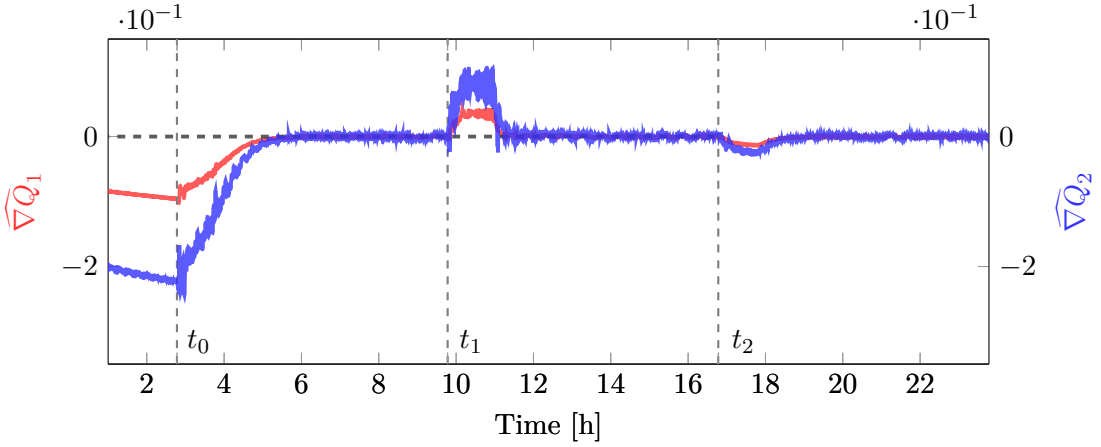


(b) Time evolution of the supply air temperature during the simulation.

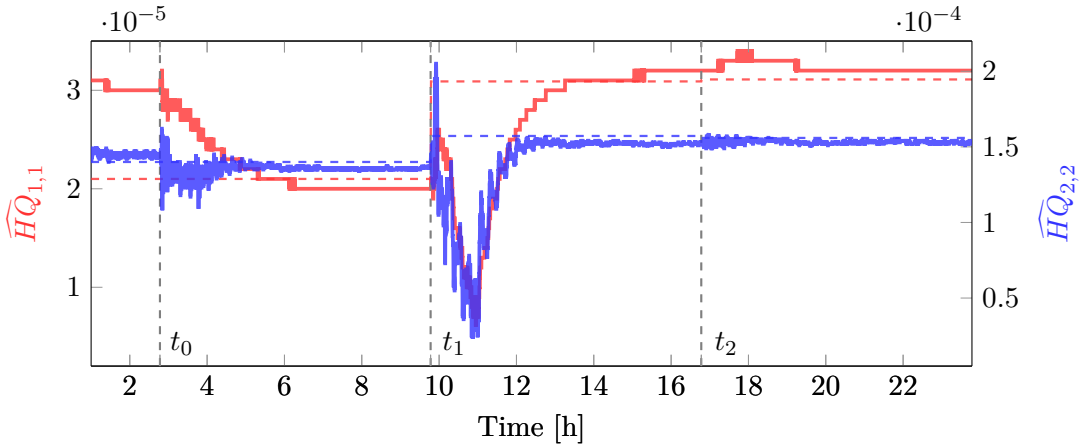


(c) Time evolution of the two control inputs during the simulation.

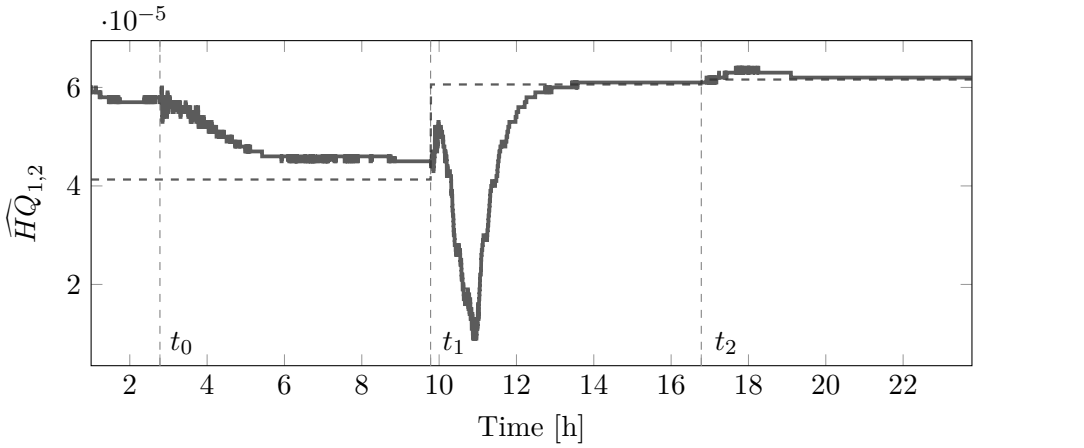
Figure 5.31: Example of simulation of the Newton-like Phasor ESC. The scheme is activated at time t_0, then at time t_1 the outdoor air temperature is decreased by 1 °C, and at time t_2 the thermal load is increased by 10 kW. Despite the variations of the disturbances, the two control inputs are adjusted in order to minimize the targeted cost function and satisfying the constraint.



(a) Estimates of the gradient components during the simulation.

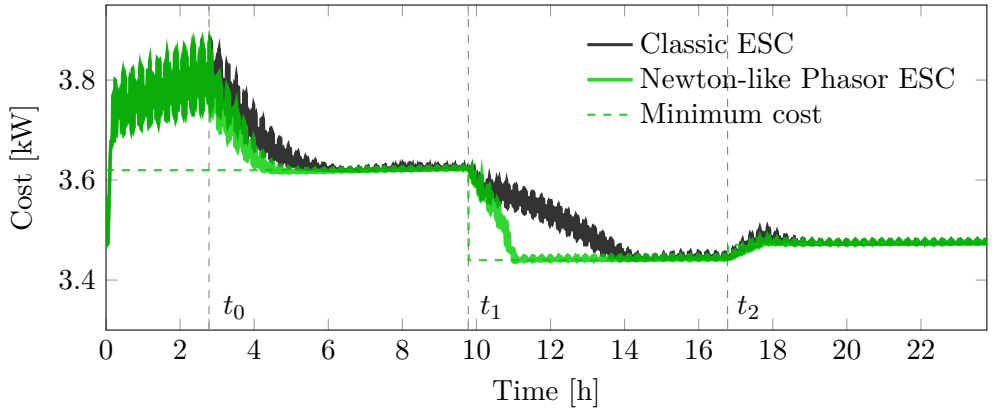


(b) Estimates of the hessian components $\widehat{HQ}_{1,1}$ and $\widehat{HQ}_{2,2}$ during the simulation compared to the corresponding steady-state real values.

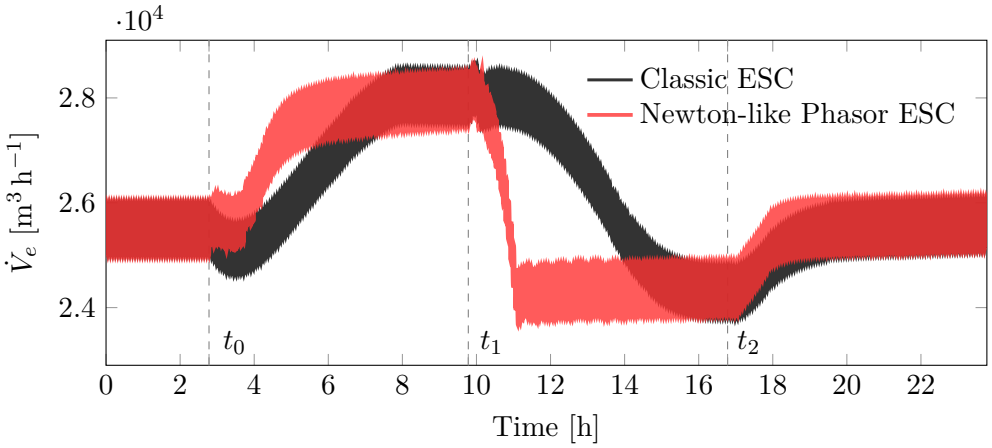


(c) Estimates of the hessian component $\widehat{HQ}_{1,2}$ during the simulation compared to the corresponding steady-state real value.

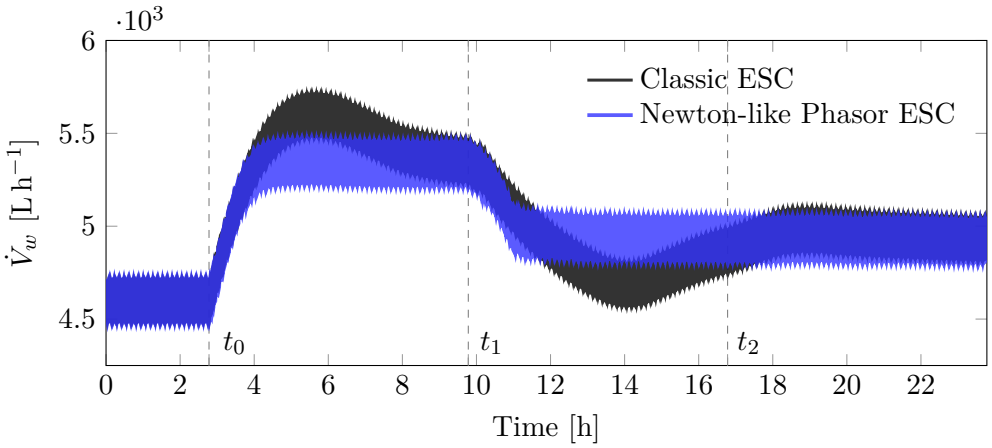
Figure 5.32: Estimates of the gradient and hessian components of the cost function during the simulation. The Newton-like ESC is activated at time t_0, then at time t_1 the outdoor air temperature is decreased by 1 °C, and at time t_2 the thermal load is increased by 10 kW. The gradient components are both driven to zero despite of the variations of the disturbances.



(a) Time evolution of the cost function during the simulation of the two control schemes.



(b) Time evolution of the exhaust airflow rate adjusted by the two control schemes during the simulation.



(c) Time evolution of the water flow rate adjusted by the two control schemes during the simulation.

Figure 5.33: Comparison between Classic ESC and Newton-like Phasor ESC performance. The two schemes are activated at time t_0, then at time t_1 the outdoor air temperature is decreased by 1 °C, and at time t_2 the thermal load is increased by 10 kW. Taking advantage of second-order information, the Newton-like Phasor scheme converges faster in all the three different situations.

6

Energy-efficient Operation of a Liquid Immersion Cooling System

This chapter considers the problem of designing a control system to efficiently operate a Liquid Immersion Cooling (LIC) unit within heat-reuse scenarios. First, in Section 6.1 the problem is formulated. Following the workflow described in Chapter 3, we first derive in Section 6.3 an energy-based First-Principle Data-Driven dynamic model of the considered cooling unit and we derive a simulation environment based on Matlab/Simulink accordingly. Then, in Section 6.4 we propose a two-level control architecture with an ESC-based supervisor designed for optimizing the operation of the LIC system. By means of simulations, the effectiveness of the proposed control strategy is assessed and the simulation results are described in Section 6.5.

6.1 Introduction

LIC means the complete immersion of electronic components in a dielectric liquid. By doing so, all the heat generated by the IT is captured in the liquid. LIC provides benefits for many layers within the Data Center industry, from the physical geographical location to the end user of a platform. First of all, suitable dielectric liquids can absorb approximately 1500 times more heat energy than air with the same volumes and temperatures. This

allows the IT equipment to operate within higher temperatures compared to air. Another benefit of LIC is the removal of cooling fans from the Data Center. Indeed, energy savings result from the removal of local fans mounted on the servers and the fans needed to distribute the cold air throughout the Data Center. Potentially, also CRAC units and chillers may be not necessary in a LIC Data Center. In LIC systems electrical power is needed only to run the pumps which move the cooling water. Another consequence of the absence of fans and chillers is that LIC systems make virtually no noise. This is not an insignificant benefit, as many modern air-cooled Data Centers operate near or above the American Occupational Safety and Health Administration's allowable limits for hearing protection. Also, since there is hardly any environmental impact like noise, Data Centers can be built in rural areas.

Another appealing advantage of the LIC approach is the possibility of implementing improved heat recovery strategies. To be precise, air cooled Data Centers may exploit heat pumps to recover the heat from the exhaust air into district heating supply flows (as being currently done in Stockholm, Sweden, Lind and Rundgren (2017)). However, since liquid coolant exhaust has a far bigger exergetic content than air coolant exhaust, recovering heat from a liquid cooling system implies smaller needs for heat pumping – i.e., a more energy efficient process. Moreover, the dielectric fluid contained within the enclosed LIC system can be used as a Thermal Energy Storage (TES): the thermal energy produced by the IT devices during their operation is captured and stored inside the dielectric liquid and it can be extracted by the closed cooling water circuit when needed. Being electrical energy consumers as well as heat producers, Liquid-cooled Data Centers can be successfully integrated into both electrical and thermal energy grids and the waste heat generated by the IT components can be effectively re-used either internally for space heating and in a multiplicity of other scenarios such as supplying heat to indoor complexes and greenhouses Paludetto and Lorente (2016), to district heating networks Brunschwiler et al. (2009), to desalination and refrigeration plants Zimmermann et al. (2012), and even to feed water pre-heaters in power plants Marcinichen et al. (2013). As a result, Data Centers will gain financial benefits from cooperating under various modalities with smart energy grids by becoming intelligent hubs at the crossroad of three energy networks: electrical energy network, heat network, and data network, Antal et al. (2018). Since electricity, heat, and information demands/productions vary in space and time, compensating local over- and under-productions requires moving and converting energies opportunely both geographically and temporally. Within this context, liquid-cooled Data Centers may help performing these movements and conversions as depicted in Fig. 6.1 by:

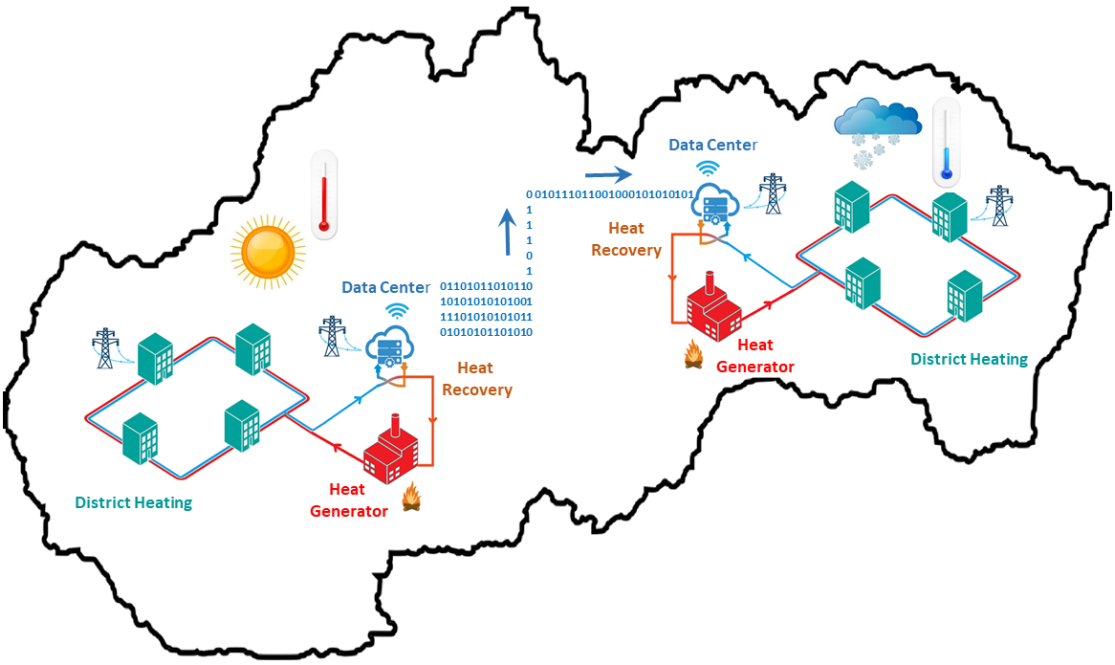


Figure 6.1: Liquid-cooled Data Centers may become intelligent hubs at the crossroad of electrical energy, heat, and data networks. The Data Center heat production can be moved in spaced by serving the IT tasks in one Data Center rather than in another and in time by using the liquid coolant as a TES.

- serving the IT tasks in one Data Center rather than in another, thus moving the heat production in space;
- using their liquid coolant as a TES to store or release heat as necessary, thus moving the heat production in time.

Within this scenario, we consider that the heat and data networks use a direct control approach for the demand-side management: they assign to each Data Center a given amount of IT tasks to perform and a thermal power (or stored energy) reference to follow according to the heat demand of the nearby buildings. At the Data Center level, the goal is thus to operate the LIC units to simultaneously maintain the servers within their thermal ranges while they are performing the required IT tasks and following the energy referenced imposed by the grid. In this chapter, we consider a lab-scale experimental setup that has been designed to capture the essence of the mimicked LIC process and that is physically located in the Infrastructure and Cloud research & test Environment (ICE) facility at Research Institutes of Sweden (RISE) Swedish Institute of Computer Science North (SICS) in Luleå, Sweden. The experimental apparatus is described in Section 6.2 and it consists of a cubic vessel that is filled with dielectric liquid, where

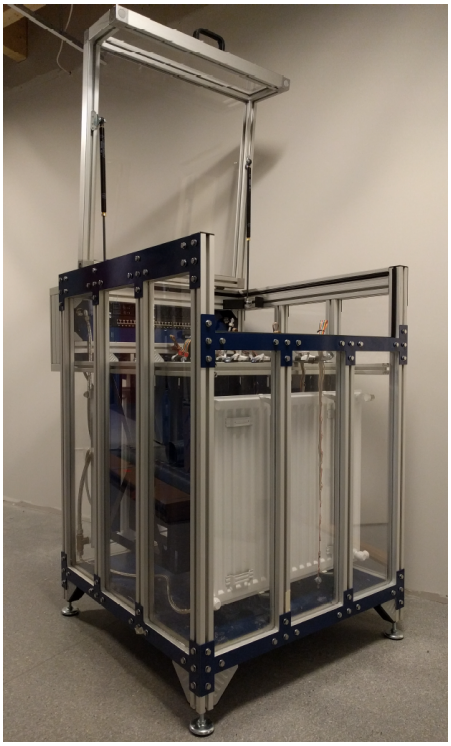
four Facebook servers are immersed. During their operation, the servers are required to serve IT tasks, which inevitably result into heat that is transferred and stored into the dielectric oil. In order to extract the thermal energy stored in the oil, the testbed also includes a closed cooling system which comprises two heat exchangers immersed in the dielectric oil and a pump circulating cooling water. To emulate a possible heat-recovery application, a water-to-air heat exchanger is then used to reject the energy extracted by the cooling water into the ambient.

The first goal of this research is to derive a mathematical model of the considered system and to design a simulation tool accordingly. This provides an invaluable support for better understanding the physical phenomena involved in the cooling process, evaluating the heat recovery capabilities of the system, and designing and testing advanced control strategies by means of simulations. CFD software tools for simulating the immersion cooling process have been proposed in the literature. They give a detailed description of the flows and temperature distributions of the oil within the LIC vessel; however, the main disadvantage of these kind of tools is the time overhead: depending on the complexity of the simulated environment, a CFD simulation can take from few minutes up to a several hours and can consume a lot of hardware resources. For this reason, we propose the lumped-parameters grey-box model described in Section 6.3 and a Matlab-based simulation environment designed accordingly. Besides being low-complexity, it is crucial that the model is able to mimic the most relevant thermal dynamics of the system and to capture the main relations between the IT equipment heat generation and the cooling system thermal energy storage and dissipation capabilities. For this purpose, we exploit energy-based modelling techniques which retain physical interpretability while ensure a good balance between complexity and accuracy. Once designed the simulation environment, we exploit it to design and test by means of simulations a control architecture to efficiently operate the LIC system within the context described above, i.e., to simultaneously maintain the servers within their thermal ranges and satisfy the heat demand required by a higher management level. Preliminary control results are reported in Section 6.4.

6.2 The Experimental Testbed

For illustration purposes we consider a testbed that is physically located in the Infrastructure and Cloud research & test Environment (ICE) facility at Research Institutes of Sweden (RISE) Swedish Institute of Computer Science North (SICS) in Luleå, Sweden, which is devoted to test innovative technologies for Data Centers ICE.

The experimental setup consists of a 0.84 m side cubic vessel, composed by an aluminium frame and acrylic glass walls and filled with a mineral oil having high dielectric strength, low viscosity, and very good heat transfer. To prevent environmental problems in case of oil leakages, a 2 mm thick metal sheet is fully welded. Pictures of the vessel before and after sealing it are reported in Figs. 6.2a and 6.2b, respectively.



(a) Vessel without the external metallic shell.

(b) Vessel with the external metallic shell.

Figure 6.2: The considered experimental testbed.

The mineral oil is used to cool four Open Compute Windmill V2 servers (Facebook), Fig. 6.3, that are directly immersed in the oil. Specifically, the servers are organized in two chassis that are fixed to the frame of the vessel and that can host a couple of servers and the corresponding power supply unit, as can be seen in Fig. 6.4. It is worth mentioning that servers cooling fans have been accurately removed and that the thermal paste between the CPU and the heat sink, commonly used in air based cooling systems, has been replaced by thin aluminium foils, which are compatible with the mineral oil.

In order to take away the waste heat transferred from the servers to the oil during their operation, a closed cooling system with water as heat transporting medium is used. In particular, water is pumped into two plate coil heat exchangers that are immersed

in the oil, so that the heat is transferred from the warm oil to the cooler water, Figs. 6.2a and 6.4. Then, waste water is supplied to an external water-to-air heat exchanger, where heat is transferred from the hot water to the environmental air thanks to a fan mounted on the heat exchanger. A schematic representation of the experimental testbed is depicted in Fig. 6.6.

In order to monitor the main physical quantities involved in the servers immersion cooling process and to be able to operate it in different operating conditions, we designed a monitoring system and a control module for the experimental testbed. The monitoring system comprises a set of sensors and a dedicated Supervisory Control And Data Acquisition (SCADA) system. Air-flow and temperature sensors are arranged in the closed water cooling circuit in proximity to the vessel inlet and outlet. Similarly, air-flow and temperature sensors are arranged at the inlet and outlet sections of the plate fin heat exchanger secondary flow. Moreover, we designed a movable matrix of thermocouples organized in 12 rows and 8 columns, that are arranged at the vessel frame in order to measure the temperature profile of the oil along a generic vertical section of the vessel. The thermocouples matrix is shown in Fig. 6.7, while in Fig. 6.4 it can be noticed that it is inserted between the servers. By using the servers' native Intelligent Platform Management Interface (IPMI), we also monitor their individual status, i.e., the temperatures of their processors and their electrical power consumption. All the measured quantities are collected by a data acquisition device with sampling time $\tau_s = 15s$ and sent to Zabbix, an open-source software installed on a remote server that offers monitoring and storage services and a simple frontend user interface, Olups (2016).

The control module comprises a set of Python scripts that allows changing remotely the water pump speed, the external fan speed, and the computational load of each servers cpu. More precisely, the scripts that communicate with the water pump and the external fan use the Modbus communication protocol, while the computational load is changed by exploiting the stress-ng tool, Stress-ng, a Linux-based stress test that has been installed in each server. It is worth specifying that the water pump is driven by an inverter,

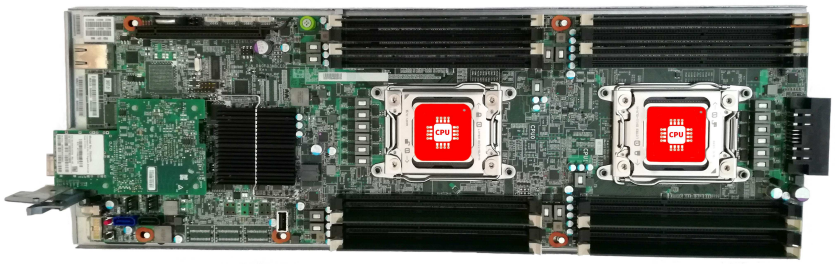


Figure 6.3: Open Compute Windmill V2 Facebook server.

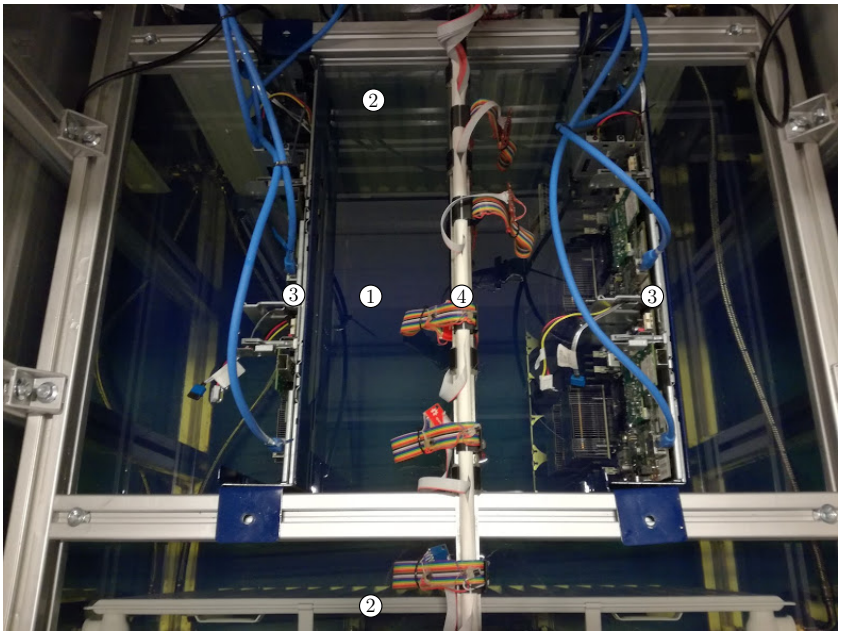


Figure 6.4: Top view of the vessel, whith: the mineral oil (1), the immersed heat exchangers (2), the servers (3), and the matrix of thermocouples (4).

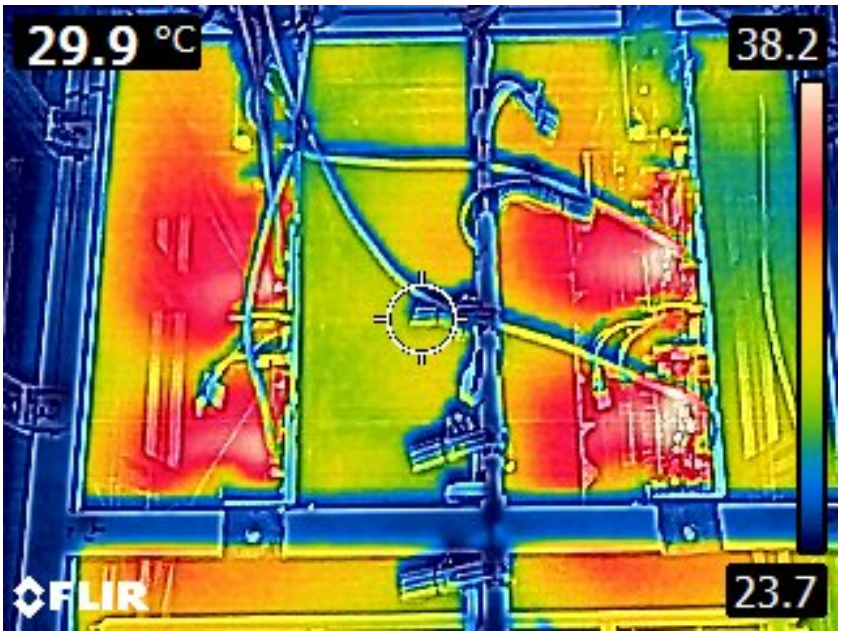


Figure 6.5: Image of the top view of the vessel acquired by a FLIR thermocamera: during servers operation, the mineral oil is warmer in the surroundings of their processors.

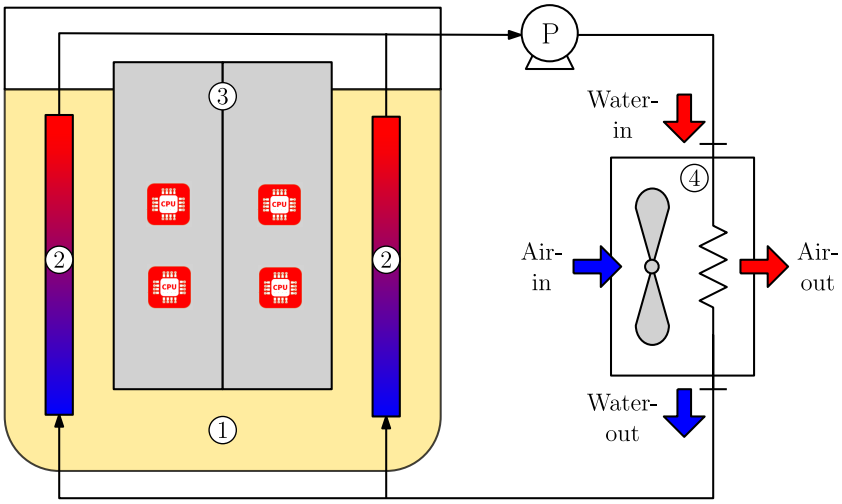


Figure 6.6: Schematic representation of the experimental setup, which comprises: the mineral oil (1), the immersed cooling plates (2), the immersed servers (3), and the external water-to-air heat exchanger (4).

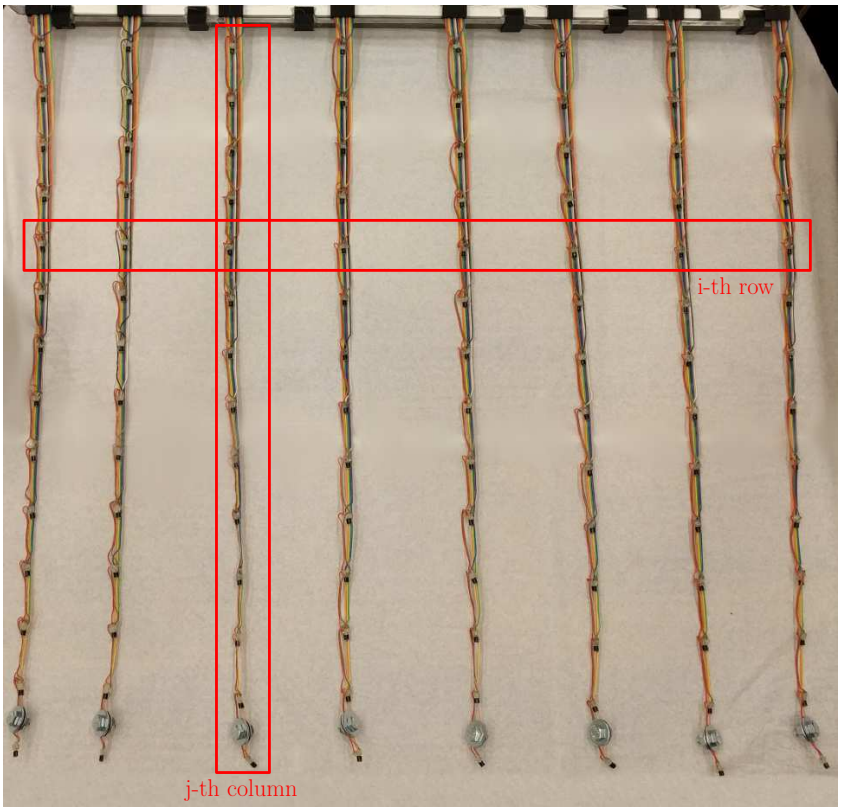


Figure 6.7: The matrix of thermocouples.

thus the pump speed can be changed continuously within its safety range. Conversely, the external fan operates at constant rotational speed. Anyway, the nominal operating conditions of the cooling system guarantee the complete rejection of the heat extracted from the vessel to the ambient environment.

6.3 Modelling the LIC System

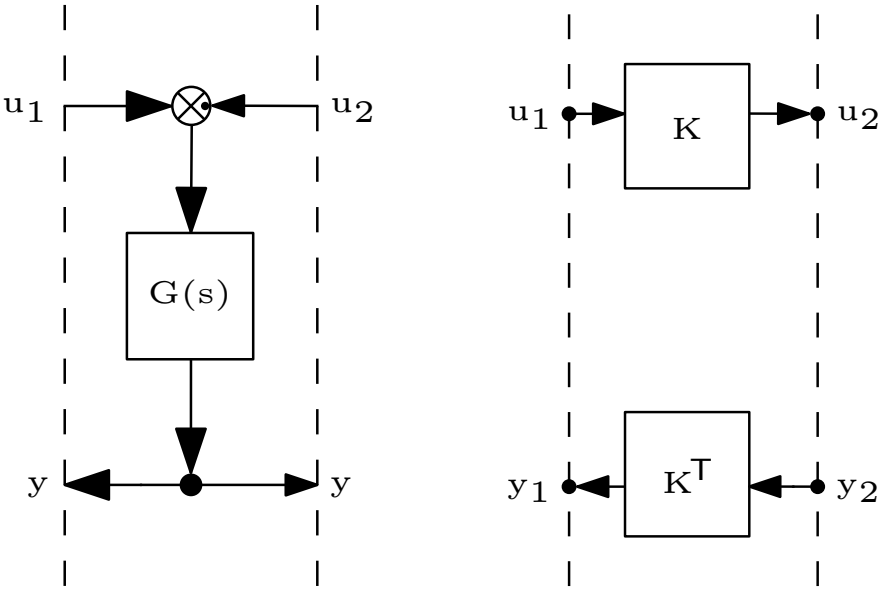
To derive the LIC thermal model, we exploit a Power Oriented Graphs (POG) approach, which is based on the use of power interaction between subsystems as a basic concept for modelling. Following this approach, the mathematical model is straightforwardly obtained by writing down the energy conservation law for each dynamic component of the considered system. The POG approach proposes a graphical notation that allows to represent the system dynamics through an easy to read block diagram, with a direct correspondence between the POG blocks and the real parts of the system. The resulting POG block scheme provides a good balance between model complexity and accuracy and it can be directly implemented in Matlab/Simulink to obtain a simulation environment. It can be useful to generate synthetic data which are related to some fundamental aspects of the LIC system thermal behaviour as well as to understand parameter influence during the process. Moreover, we can exploit the simulation environment to design and test control strategies to efficiently manage the LIC system and to evaluate their performance by means of simulations. Following this energy based approach, we first identify the main dynamics of the elements that compose the LIC system, and then we derive the POG scheme that represents the whole system. The model parameters are calibrated by exploiting real available data acquired from an experimental campaign conducted on an ad-hoc proof-of-concept LIC system based on natural convection.

6.3.1 POG Basic Features

The POG are a graphical and mathematical modelling technique introduced in Zanasi (1991), based on the same Bond Graph idea of using the power interaction between sub-systems as the key element for modelling.

This energetic approach relies on the following properties that characterize all the dissipative physical systems:

- a system always stores and/or dissipates energy;
- the dynamic model of a physical system describes how the energy moves within the system;



(a) Elaboration block. **(b)** Connection block.

Figure 6.8: The two types of blocks used in the POG formalism.

- the energy moves from point to point by means of two power variables.

The POG are essentially composed of only two types of blocks: elaboration blocks are used for modelling all the physical elements of the system that store and/or dissipate energy (i.e. springs, masses, capacities, inductances, resistances, etc.), while connection blocks are used for modelling all the physical elements that transform the power without losses. The two types of blocks are graphically represented in Fig. 6.8. The circle that is present in the elaboration block is a summation element: when a black spot is present near an entering arrow, the corresponding variable must be made negative before the sum operation.

The dashed lines in Fig. 6.8 are the power sections which connect the two POG basic blocks to the external world. There are no restrictions on the choice of the power vectors $\mathbf{u}$ and $\mathbf{y}$ involved in each dashed line, except the fact that the scalar product $\mathbf{u}^T\mathbf{y}$ must have the physical meaning of power flowing through that particular section. The consequence is that in the thermal domain we should take temperature and variation of entropy as power vectors. Since handling heat flows is more convenient than handling entropy variations, Grossi and Zanasi (2016) suggests to use temperature and heat flows as power vectors, which are conjugated with respect to the variation of entropy.

The advantages of using POG are that they are simple, modular, and easy to use. Essentially, they consist of block diagrams that can be easily implemented in

computational tools like Matlab/Simulink and there exists a direct correspondence between POG blocks and real system components. Moreover, the mathematical equations underlying the POG schemes are based on energy balances, thus energy conservation is guaranteed.

6.3.2 The LIC Thermal Model

Following the considered energy-based modelling approach, the first step consists in defining the set of dynamic elements that can store and/or dissipate energy. First, we assume that the mineral oil is divided into N^o horizontal layers, having volume V_i^{oil} and uniform temperature T_i^{oil}, $i = 1, \ldots, N^o$. Moreover, we assume that thermal properties of the oil are constant within each oil layer, thus each thermal capacity results:

$$C_i^{oil} = \rho^{oil} V_i^{oil} c_p^{oil}, \qquad i = 1, \ldots, N^o. \tag{6.1}$$

Since each oil layer can exchange heat with each adjacent layer, we denote with g_i^{oil} the heat transfer coefficient between the $i-th$ oil layer and the above one. Assuming that most of the heat transfer between adjacent oil layers is due to conduction, we write $g_i^{oil} = g^{oil} A_i^{oil} / d_i^{oil}$, where g^{oil} is the thermal conductivity of the oil and A_i^{oil} and d_i^{oil} are the contact area and the distance from the centers of the two oil layers, respectively. We also take into account the heat that flows from each oil node to the environmental air at temperature T^{env} through the vessel walls. Thus, we define the heat transfer coefficients $g_i^{env} = h^{env} A_i^{walls}$, where h^{env} is the average convective heat transfer coefficient and A_i^{walls} is the area of the vessel walls in contact with the i-th oil layer.

We also consider the processors of the N^s immersed servers as dynamic elements of the system. Considering the experimental apparatus described in Section 6.2, each server is equipped with 2 processing units located at a different height, as shown in Fig. 6.6. Each cpu is characterized by a temperature $T_{i,j}^{cpu}$ and a thermal capacity

$$C_{i,j}^{cpu} = \rho^{cpu} V_{i,j}^{cpu} c_p^{cpu} \tag{6.2}$$

where $i = 1, \ldots, N^s$ and $j = 1, 2$. During the operation of each server, most of the electrical power supplied to the processing units is converted in thermal energy $\dot{q}_{i,j}^{cpu}$. Part of this thermal energy is stored in the capacity of the cpu causing an increase of its temperature, while the remaining flows from the processing units to the adjacent oil layers. We denote with $g^{cpu} = h^{cpu} A^{cpu}$ the average heat transfer coefficient between each processing unit and the oil, where h^{cpu} is the convective heat transfer coefficient

and A^{cpu} is the contact area. It is worth noting that, due to that heat exchange, the oil in the surroundings of the processing units heats up and then, due to the buoyant force, it starts flowing toward the top of the vessel. Note also that the remaining oil below the processing units is heated up from the top, thus its temperature stratifies. In this context, we are not interested in a detailed description of the local turbulent flows that occur within the oil close to the processing units. Rather, to obtain a lumped-parameters description of the thermal phenomena, we divide the oil layers as follows: the first layer represents the oil portion in the top of the tank and that contains all the higher cpus of the servers; the second layer represents the oil portion that contains all the lower processing units of the servers; all the stratified oil that is below the cpus is divided in the remaining $N^o - 2$ layers. In order to describe the effect of the warm oil that flows from the second to the first oil layer due to convection, we introduce a fictitious heat transfer coefficient h^{conv} between each of the lower processing units and the top oil layer.

As described in Section 6.2, we also consider the presence of two heat exchangers that are immersed within the oil, through which cold water flows at the flow rate ϕ in order to cool down the surrounding oil. More specifically, we consider that water enters from the bottom of the heat exchangers at a temperature T_{in}^{water}, and exits from the top at a higher temperature T_{out}^{water}. In deriving the thermal model, we assume for simplicity that the oil is in contact with only one fictitious immersed heat exchanger having an external area equal to the total area of both the immersed heat exchangers. Then, we assume to divide the water within the equivalent heat exchanger into N^o horizontal layers, each in contact with one of the oil layers. We consider each water portion as a dynamic element of the model, having temperature T_i^{water} and thermal capacity

$$C_i^{water} = \rho^{water} V_i^{water} c_p^{water}, \tag{6.3}$$

with $i = 1, \dots, N^o$. We then denote with $g_i^{hx} = h^{hx} A_i^{hx}$ the heat transfer coefficient between the $i-th$ oil and water layers, where h^{hx} is the average convective heat transfer coefficient and A_i^{hx} is the contact area.

It is worth highlighting that the external plate-to-fin heat exchanger described in Section 6.2 is not taken into account in the development of our LIC thermal model, because LIC systems are usually coupled with heat recovery applications that exploit the energetic content of the warm water exiting the LIC unit. Thus, we aim to provide a general thermal model of LIC systems that can be coupled with different models of heat recovery units.

Once identified the main dynamic elements and the inputs of the LIC system, we derive the thermal model by writing down the energy balance equations for each element

of the system. By properly rearranging them, we straightforwardly obtain the following state-space model:

$$\begin{cases} \mathbf{L}\dot{\mathbf{x}}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}(t) + \boldsymbol{\ell}(\mathbf{x}(t), \mathbf{u}(t), \phi) & \text{(6.4a)} \\ \mathbf{y}(t) = \mathbf{C}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{D}(\boldsymbol{\theta})\mathbf{u}(t) & \text{(6.4b)} \end{cases}$$

The vector $\mathbf{x}$ in (6.4) is the state vector, which contains the temperatures of all the dynamic elements, while $\mathbf{u}$ and $\mathbf{y}$ are the input and output vectors, respectively:

$$\mathbf{x} = \begin{bmatrix} T_1^{oil} \\ \vdots \\ T_{N^o}^{oil} \\ \hline T_{1,1}^{cpu} \\ T_{1,2}^{cpu} \\ \vdots \\ T_{N^s,1}^{cpu} \\ T_{N^s,2}^{cpu} \\ \hline T_1^{water} \\ \vdots \\ T_{N^o}^{water} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} T^{env} \\ \hline \dot{q}_{1,1}^{cpu} \\ \dot{q}_{1,2}^{cpu} \\ \vdots \\ \dot{q}_{N^s,1}^{cpu} \\ \dot{q}_{N^s,2}^{cpu} \\ \hline T_{in}^{water} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \dot{q}^{env} \\ \hline T_{1,1}^{cpu} \\ \vdots \\ T_{N^s,N^c}^{cpu} \\ \hline \dot{q}^{hx} \end{bmatrix}. \tag{6.5}$$

It is worth observing that in model (6.4) input and output vectors $\mathbf{u}$ and $\mathbf{y}$ are conjugated with respect to the variation of entropy, i.e., the ratios $\dot{S}_i^y = \dot{q}_i^y / T_i^u$ and $\dot{S}_i^u = \dot{q}_i^u / T_i^y$ have the physical meaning of variation of entropy. Note also that, since all the thermal fluxes described by model (6.4) are expressed in terms of temperature differences, all the temperatures in (6.5) can be expressed either in Kelvin or in Celsius degrees.

The energy matrix $\mathbf{L}$ is diagonal and contains all the thermal capacities of the elements:

$$\mathbf{L} = \text{diag}(C_1^{oil}, \ldots, C_{N^o}^{oil}, C_{1,1}^{cpu}, \ldots, C_{N^s,2}^{cpu}, C_1^{water}, \ldots, C_{N^o}^{water}). \tag{6.6}$$

Note that matrix $\mathbf{L}$ is always symmetric and positive definite. Also, note that the total thermal energy stored in the system, which is given by the sum of all the thermal energies stored in all the thermal capacities of the system, can be obtained as:

$$E_t = \mathbf{1}^T \mathbf{L} \mathbf{x}, \tag{6.7}$$

where **1** is the all-ones column vector.

It is worth noting that matrices **A**, **B**, **C**, and **D** depend on the parameters vector $\boldsymbol{\theta}$, which contains all the heat transfer parameters of the system:

$$\boldsymbol{\theta} = \begin{bmatrix} g^{oil} & h^{env} & h^{cpu} & h^{conv} & h^{hx} \end{bmatrix}^T. \tag{6.8}$$

Matrix **A** is the power matrix that describes the power redistribution within the system and contains all the thermal conductivities. In particular, **A** can be partitioned as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} \\ \mathbf{A}_{1,2}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_{3,1} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{6.9}$$

where block $\mathbf{A}_{1,1}$ represents the power redistribution within the oil layers:

$$\mathbf{A}_{1,1} = -\sum_{i=1}^{N^o}\sum_{j=1}^{N^o}(\mathbf{e}_i - \mathbf{e}_j)g_{ij}^o(\mathbf{e}_i - \mathbf{e}_j) - \mathrm{diag}(g_1^e, \ldots, g_{N^o}^e)^T, \tag{6.10}$$

block $\mathbf{A}_{1,2}$ describes the heat fluxes from the cpus to the oil:

$$\mathbf{A}_{1,2} = g^{cpu}(\mathbf{e}_1\sum_{i=1}^{N^s}\mathbf{e}_i^T + \mathbf{e}_2\sum_{i=N^s+1}^{2N^s}\mathbf{e}_i^T) + h^{conv}\mathbf{e}_2\sum_{i=1}^{N^s}\mathbf{e}_i^T, \tag{6.11}$$

and block $\mathbf{A}_{1,3}$ describes the heat exchanged between water and oil elements:

$$\mathbf{A}_{1,3} = \sum_{i=1}^{N^o}\sum_{j=1}^{N^o}(\mathbf{e}_i - \mathbf{e}_j)g_{ij}^{hx}(\mathbf{e}_i - \mathbf{e}_j). \tag{6.12}$$

Block $\mathbf{A}_{3,1}$ will be discussed later, after introducing the other matrices of the model.

Matrix **B** is the input power matrix that describes the thermal fluxes in input and

output of the system:

$$\mathbf{B} = \left[\begin{array}{ccc|ccc|ccc} g_1^e & \dots & g_{N^o}^e & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline 0 & \dots & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ \hline 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{array}\right]^T . \tag{6.13}$$

Matrices **C** and **D** are the output power and the input-output power matrices, respectively:

$$\mathbf{C} = \left[\begin{array}{ccc|ccc|ccc} g_1^e & \dots & g_{N^o}^e & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline 0 & \dots & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ \hline g_1^{hx} & \dots & g_{N^o}^{hx} & 0 & \dots & 0 & -g_1^{hx} & \dots & -g_{N^o}^{hx} \end{array}\right]^T , \tag{6.14}$$

$$\mathbf{D} = \left[\begin{array}{c|ccc|c} -\sum_{i=1}^{N^o} g_i^e & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & 0 \end{array}\right] . \tag{6.15}$$

Note that the matrices (6.14) and (6.15) derive from the following definitions of the thermal fluxes $\dot{q}^{env}$ and $\dot{q}^{hx}$:

$$\dot{q}^{env} = \sum_{i=1}^{N^o} g_i^{env}(T_i^{oil} - T^{env}), \tag{6.16}$$

$$\dot{q}^{hx} = \sum_{i=1}^{N^o} g_i^{hx}(T_i^{oil} - T_i^{water}). \tag{6.17}$$

The non-linear vector field $\boldsymbol{\ell}$ describes the power redistribution within the portions of water in the immersed heat exchangers:

$$\boldsymbol{\ell}(\mathbf{x}, \mathbf{u}, \phi) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \\ \hline \ell_1(\mathbf{x}, \mathbf{u}, \phi) \\ \vdots \\ \ell_{N^o}(\mathbf{x}, \mathbf{u}, \phi) \end{bmatrix}. \tag{6.18}$$

In this work, we propose two different approaches to describe the heat exchanged through the immersed heat exchangers. An approach consists in using the Number of Transfer Units (NTU) method, which is used to calculate the heat transfer in heat exchangers when there is not sufficient information about the outlet temperatures. In our specific case, it leads to $\mathbf{A}_{3,1} = \mathbf{0}$ and:

$$\ell_i(\mathbf{x}, \mathbf{u}, \phi) = \epsilon_i \, C_i^{water}(T_i^{oil} - T_i^{water}), \tag{6.19}$$

where ϵ is the thermal effectiveness that can be expressed as a function of NTU:

$$\epsilon_i = 1 - \exp(-NTU_i), \tag{6.20}$$

$$NTU_i = \frac{g_i^{hx}}{\phi \, c_p^{water}}. \tag{6.21}$$

In the following of this manuscript, we will refer to the model obtained through this first approach with POG-NTU model. An alternative model, that we will call simply POG model, is obtained by writing down the energy balance of the i-th water node as:

$$C_i^{water}\dot{T}_i^{water} = g_i^{hx}(T_i^{oil} - T_i^{water}) + \phi \, c_p^{water}(T_{i+1}^{water} - T_i^{water}), \tag{6.22}$$

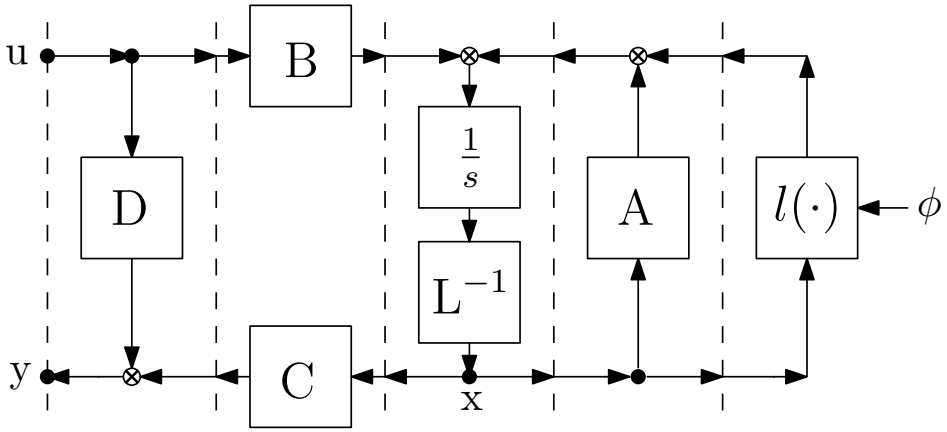


Figure 6.9: POG block scheme of the LIC system.

that results in taking $\mathbf{A}_{3,1} = \mathbf{A}_{1,3}^T$ and:

$$\ell_i(\mathbf{x}, \mathbf{u}, \phi) = \phi \, c_p^{water}(T_{i+1}^{water} - T_i^{water}), \tag{6.23}$$

where $T_{N^o+1}^{water} = T_{in}^{water}$.

The POG block diagram that represents the LIC thermal system can be straightforwardly derived from the state-space model (6.4) and it is depicted in Fig. 6.9.

We stress out that the proposed thermal model describes a generic LIC system which exploits natural convection. In fact, the geometry of the system is taken into account in the model and it can be modified by changing all the geometrical parameters. Moreover, the model takes into account a generic number of servers N^s: if we add or remove some servers from the system, we only need to augment or reduce the state, input, and output vectors and the matrices in (6.4). Also, the model structure does not depend on the number of heat exchangers, it is only required that they are placed vertically within the vessel. In this work, we assess the generalization capabilities of the proposed thermal model by considering the LIC proof-of-concept apparatus described in Section 6.2. In that specific case, we have $N^s = 4$ immersed servers each having two processors mounted at different height. The mineral oil is divided into $N^o = 3$ layers: the first one is the top layer, which contains all the higher cpus; the second one is the middle layer, which contains all the lower cpus; the third one is the bottom layer, which represents all the oil that is below the cpus. All the geometrical quantities have been measured through direct inspection, while the model parameters grouped in the vector $\boldsymbol{\theta}$ defined in (6.8) have been estimated as described in the following section.

6.3.3 Model Calibration and Validation

Once defined the model structure described in Section 6.3.2, estimation of model parameters $\boldsymbol{\theta}$ in (6.8) is carried out by considering a Fisherian system identification approach where the model parameters are considered deterministic yet unknown. By using the monitoring system and the control module described in Section 6.2, we design and perform two different experiments on the proof-of-concept apparatus considered in this work. More precisely, in the first experiment we impose step variations to the water flow ϕ across the heat exchangers around a nominal value, while the electrical power provided to each server is kept constant, Fig. 6.10a. In the first part of the second experiment, the water flow ϕ is kept constant, while the electrical power provided to all the servers is varied around a nominal value. Then, all the servers switch to idle state (i.e. their electrical power consumption is reduced to the minimum) and we start varying the water flow similarly to the previous test, Fig. 6.10b.

For each experiment, we collect the following measured data:

$$\mathbf{m}_i(t) = \left[\begin{array}{ccc|ccc|c} T_1^{oil} & \dots & T_3^{oil} & T_{1,1}^{cpu} & \dots & T_{4,2}^{cpu} & T_{out}^{water} \end{array} \right]^T, \tag{6.24}$$

where the subscript $i = 1, 2$ identifies the number of the experiment.

Following a standard identification procedure, we divide the data acquired in both experiments in training and validation sets $\mathbf{m}_i^{tr}(t)$ and $\mathbf{m}_i^{val}(t)$: the estimation step of vector $\boldsymbol{\theta}$ is performed by using all the training data $\mathbf{m}^{tr} = \mathbf{m}_1^{tr} \cup \mathbf{m}_2^{tr}$, then the two validation sets $\mathbf{m}_1^{val}(t)$ and $\mathbf{m}_2^{val}(t)$ are both used for testing the generalization capabilities of the model.

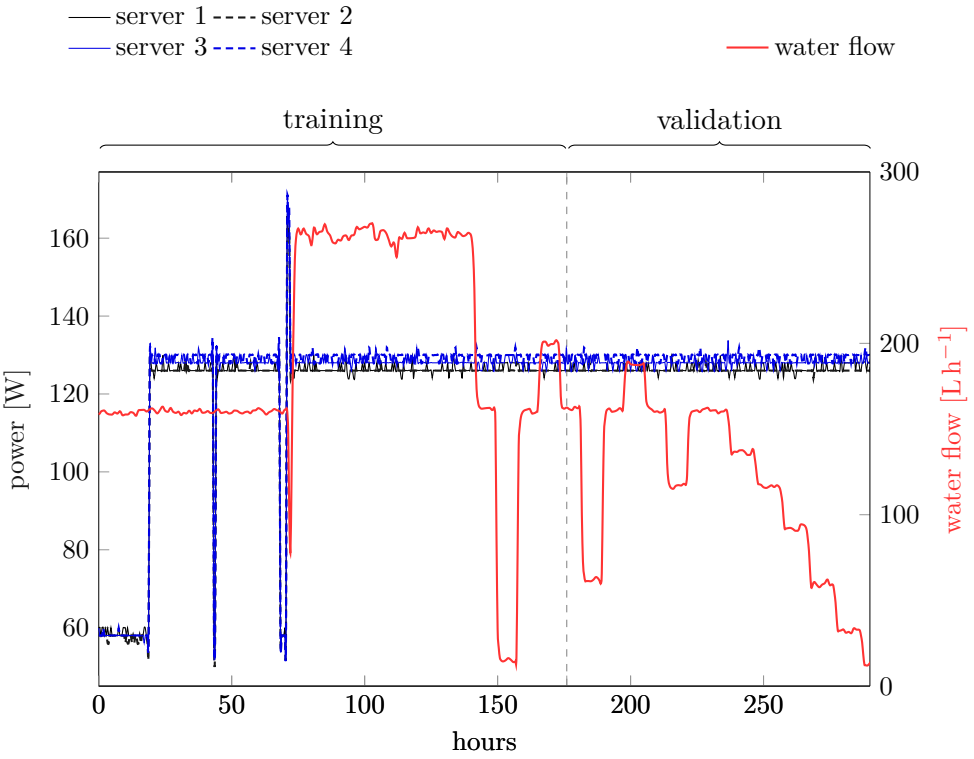
The problem of estimating vector $\boldsymbol{\theta}$ is formulated in a Least Squares (LS) fashion:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{k=1}^{K} \left\| \widehat{\mathbf{m}}^{tr}(k; \boldsymbol{\theta}) - \mathbf{m}^{tr}(k)) \right\|^2, \tag{6.25}$$

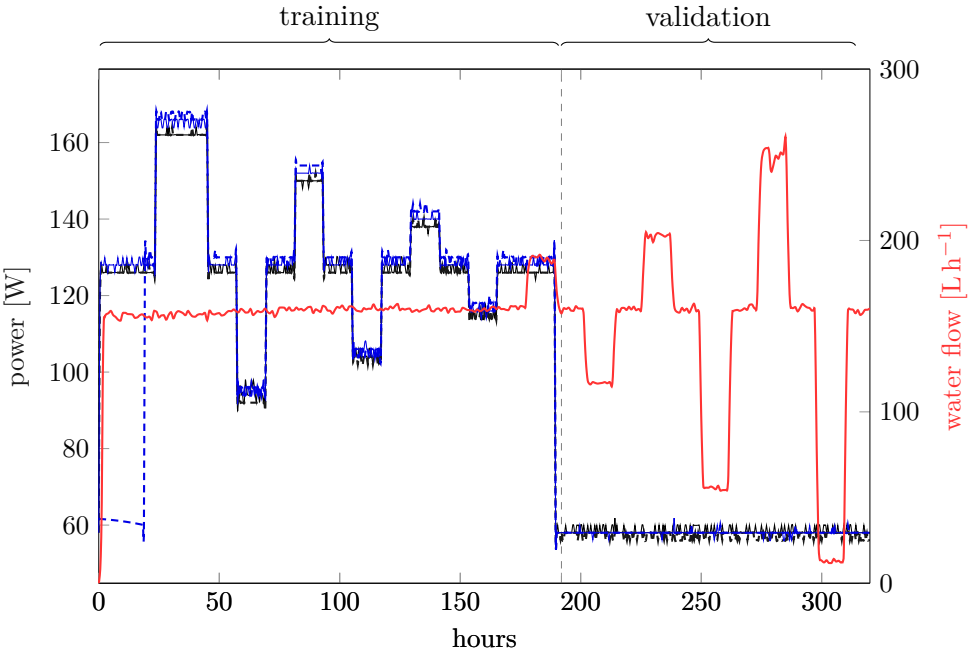
where K is the number of samples contained in the training dataset and $\widehat{\mathbf{m}}^T$ is the vector obtained at each time instant by taking $\widehat{\mathbf{m}}^{tr} = \widehat{\mathbf{C}}\hat{\mathbf{x}}^{tr}$ with

$$\widehat{\mathbf{C}} = \mathrm{diag}\{\underbrace{1, \dots, 1}_{12}, 0, 0\} \tag{6.26}$$

and $\hat{\mathbf{x}}^{tr}$ is obtained by propagating the forward-Euler discrete-time counterpart of the continuous time dynamics in (6.4a) initialized with the first measurement in the training dataset. The hypothesis space $\boldsymbol{\Theta}$ in (6.25) is opportunely constrained to guarantee the



(a) Input signals in the first experiment.



(b) Input signals in the second experiment.

Figure 6.10: Input signals designed for the two experiments.

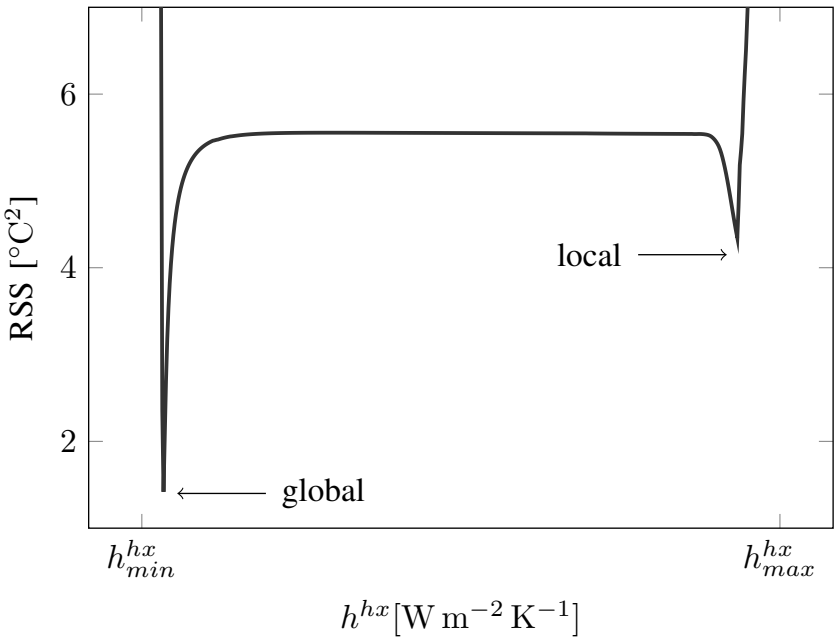


Figure 6.11: Residual Sum of Squares (RSS) error computed by varying parameter h^{hx} of the POG-NTU model within its admissible range while maintaining the other parameters fixed to their nominal values.

physical meaningfulness of the resulting estimates.

As it can be observed in Fig. 6.11, the cost function related to the optimization problem (6.25) presents local minima. The choice of the optimization algorithm is closely related to the characteristics of the cost function: a standard deterministic algorithm, such as a gradient-based method, may suffer from high sensitivity to local optima, despite it moves rapidly to the global optimum if guided accordingly; on the other hand, an Evolutionary Algorithm (EA) minimizes the risk to converge to a local optimum, but it may require high computational cost and it reduces its searching efficiency when close to the optimum. In view of these considerations, problem (6.25) has been tackled by considering the hybrid method suggested in Lissandrin et al. (2017): initially, a Gravitational Search Algorithm (GSA) is used to discover the regions that are feasible, and then the solution obtained from the GSA is employed as the initial guess for the Levenberg–Marquardt (LM) algorithm, which produces the final solution. The parameters estimates obtained by solving problem (6.25) for both the proposed POG and POG-NTU models are reported in Tab. 6.1.

To assess the generalization capabilities of the two variations of the proposed thermal model, we compare the measured data contained in the two validation datasets $\mathbf{m}^V_{1,2}$ and the simulated data $\widehat{\mathbf{m}}^V_{1,2} = \widehat{\mathbf{C}}\hat{\mathbf{x}}^V$, obtained by propagating the forward-Euler discrete-

Parameter	POG	POG-NTU	Unit
g^{oil}	1.12	1.03	$\mathrm{W\,m^{-1}\,K^{-1}}$
h^{env}	4.04	3.21	$\mathrm{W\,m^{-2}\,K^{-1}}$
h^{cpu}	3.28	3.82	$\mathrm{W\,m^{-2}\,K^{-1}}$
h^{bou}	0.40	0.38	$\mathrm{W\,m^{-2}\,K^{-1}}$
h^{hx}	8.59	8.31	$\mathrm{W\,m^{-2}\,K^{-1}}$

Table 6.1: Estimated parameters.

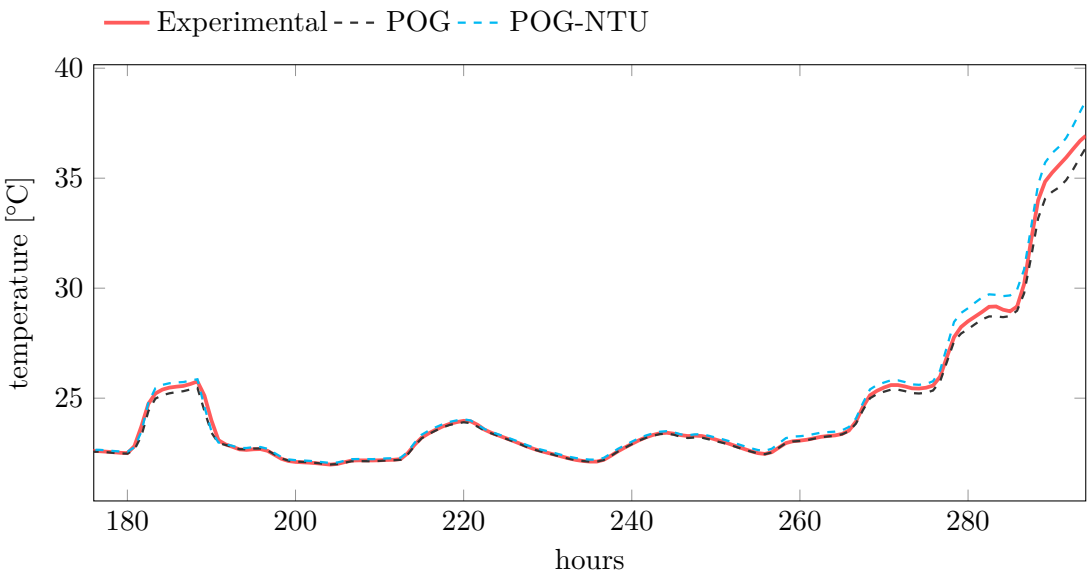
time counterpart of the continuous time dynamics in (6.4a) initialized with the first measurement in the two validation datasets. Tab. 6.2 reports the Root Mean Square Error (RMSE) between measured and simulated data for both the datasets and considering both the variations of the model. As it can be observed, the outlet water temperature is

	Dataset 1		Dataset 2		
Node	POG	POG-NTU	POG	POG-NTU	Unit
$\mathrm{T}_1^{\mathrm{oil}}$	0.40	0.46	0.94	0.91	°C
$\mathrm{T}_2^{\mathrm{oil}}$	0.39	0.70	0.40	0.50	°C
$\mathrm{T}_3^{\mathrm{oil}}$	0.12	0.19	0.27	0.29	°C
$\mathrm{T}_{1,1}^{\mathrm{cpu}}$	1.14	1.34	2.54	2.87	°C
$\mathrm{T}_{1,2}^{\mathrm{cpu}}$	0.53	0.76	2.36	2.68	°C
$\mathrm{T}_{2,1}^{\mathrm{cpu}}$	1.03	0.80	0.45	0.76	°C
$\mathrm{T}_{2,2}^{\mathrm{cpu}}$	0.82	0.63	3.79	4.12	°C
$\mathrm{T}_{3,1}^{\mathrm{cpu}}$	0.95	0.92	3.08	3.40	°C
$\mathrm{T}_{3,2}^{\mathrm{cpu}}$	0.43	0.41	1.78	2.09	°C
$\mathrm{T}_{4,1}^{\mathrm{cpu}}$	1.24	1.02	2.76	3.09	°C
$\mathrm{T}_{4,2}^{\mathrm{cpu}}$	0.86	1.10	3.11	3.44	°C
$\mathrm{T}_{\mathrm{out}}^{\mathrm{water}}$	0.26	0.32	0.46	0.28	°C

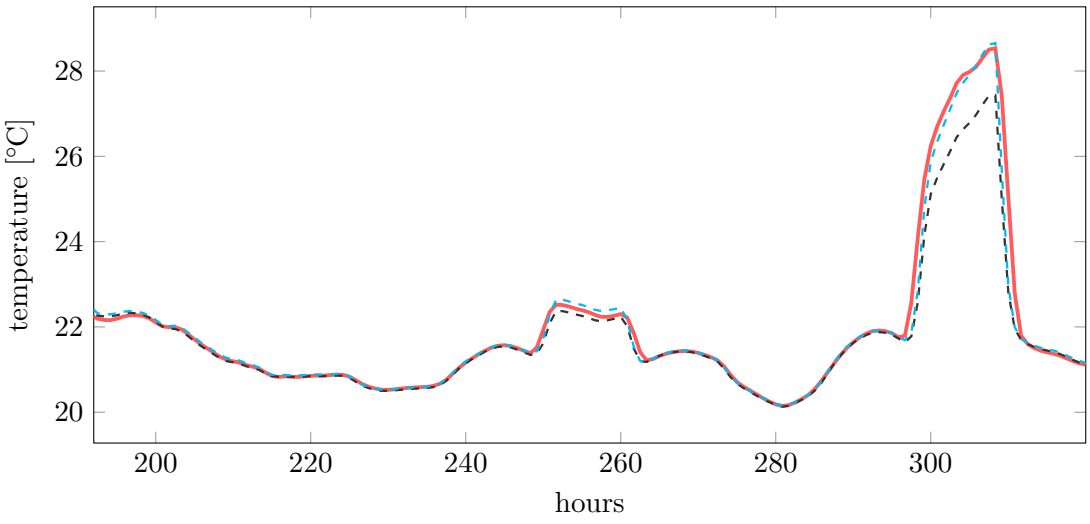
Table 6.2: RMSE on the validation sets associated to the identified gray box estimator.

well approximated by both POG and POG-NTU models, with a RMSE always below 0.5 °C. As it is evident also from Fig. 6.12, the fitting performance of the two proposed models are very similar in the first validation dataset (6.12a), while the POG-NTU model better approximates the outlet water temperature in the second validation dataset (6.12b). From Tab. 6.2, it also follows that the prediction capabilities of the oil temperature in the different layers is good enough for the considered application, with a RMSE below 1 °C in both validation datasets and for both models. Such good approximation performances can not be achieved in predicting the cpus temperatures, especially if we consider the

second validation dataset. This fact is probably due to the use of a not entirely realistic linear relation describing the heat exchange between the oil and the processing units. Also, we do not explicitly take into account the thermal capacity of the heat sinks in the model. This would increase the number of parameters of the model and it would require temperature sensors placed also on the heat sinks to calibrate those parameters. As it is evident from Figs. 6.13 and 6.14, the performance of POG and POG-NTU models are similar, since they are identical in describing the heat exchange between the oil and the cpus.

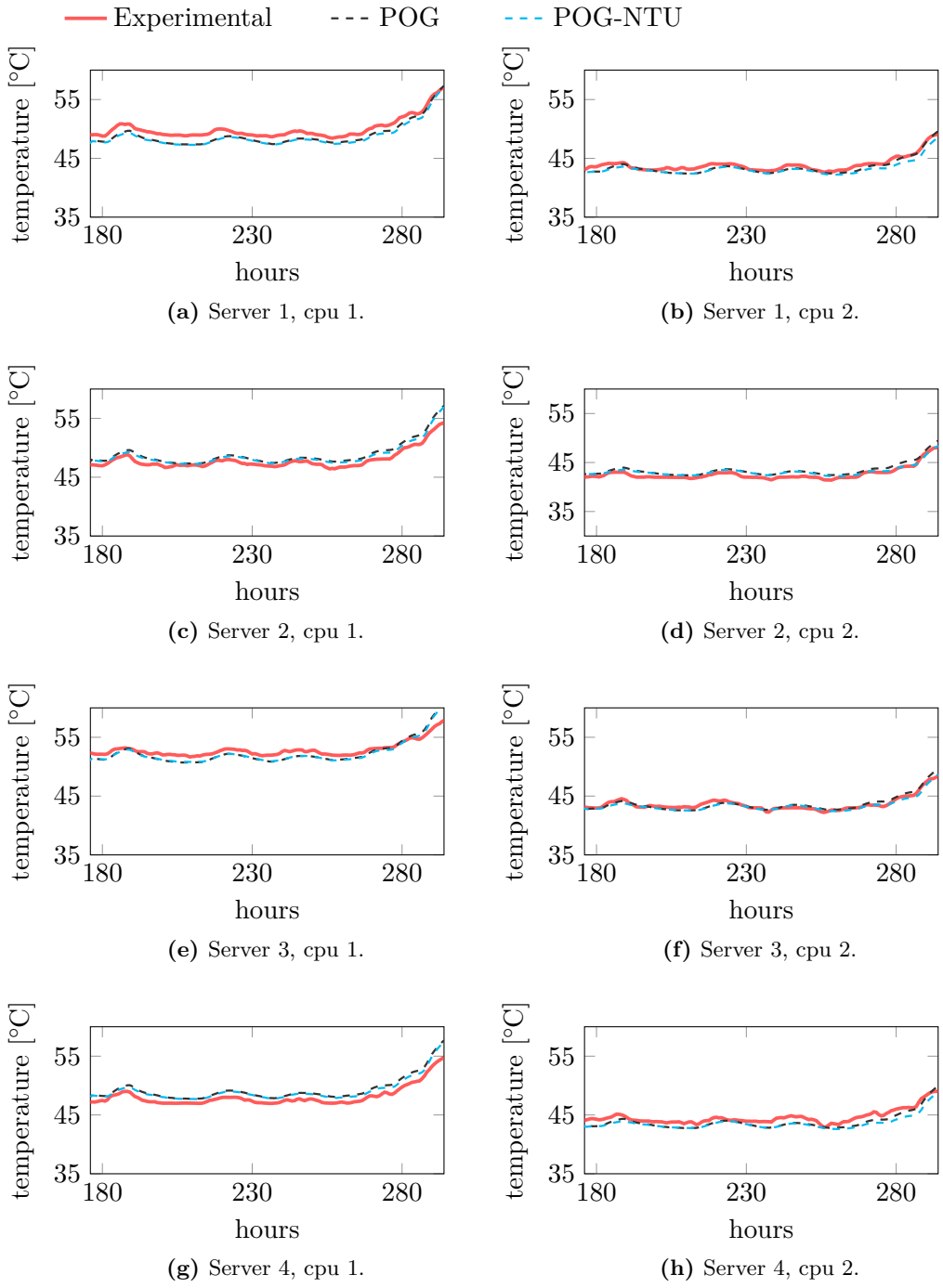


(a) Results for the first validation dataset.



(b) Results for the second validation dataset.

Figure 6.12: Simulated vs experimental outlet water temperature.



(a) Server 1, cpu 1.

(b) Server 1, cpu 2.

(c) Server 2, cpu 1.

(d) Server 2, cpu 2.

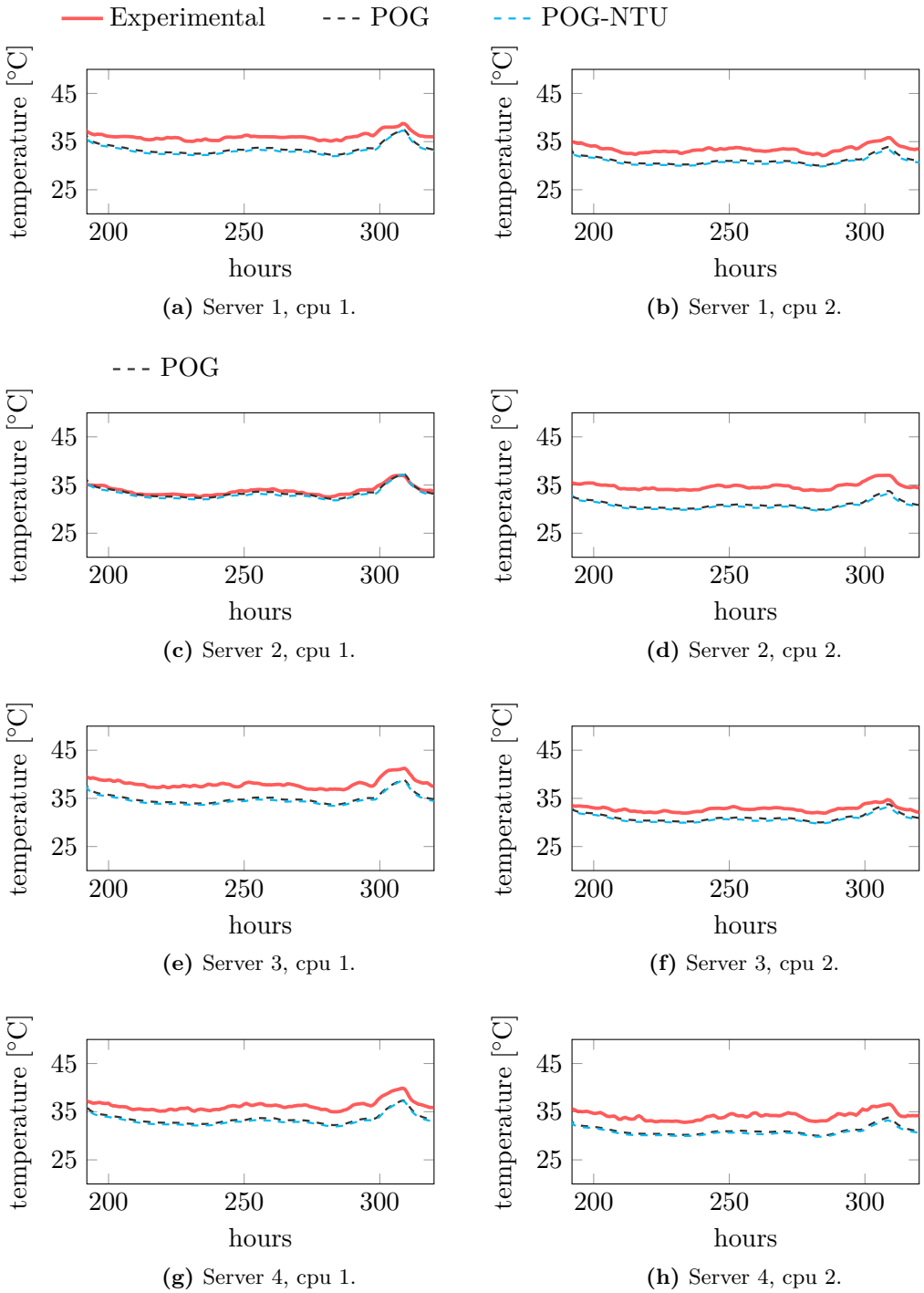
(e) Server 3, cpu 1.

(f) Server 3, cpu 2.

(g) Server 4, cpu 1.

(h) Server 4, cpu 2.

Figure 6.13: Simulated vs experimental temperatures of the processing units in the first validation dataset.



(a) Server 1, cpu 1.

(b) Server 1, cpu 2.

(c) Server 2, cpu 1.

(d) Server 2, cpu 2.

(e) Server 3, cpu 1.

(f) Server 3, cpu 2.

(g) Server 4, cpu 1.

(h) Server 4, cpu 2.

Figure 6.14: Simulated vs experimental temperatures of the processing units in the second validation dataset.

6.4 Design of the Control System

We consider the scenario where the LIC system is required to simultaneously maintain the servers within their thermal operating ranges while they are performing the assigned IT tasks and satisfying the thermal energy demand imposed by the grid. For this purpose, we propose a hierarchical control scheme with two layers:

- the supervisor (high layer), which specifies the set-points that are optimal under a certain criterion;
- the local controllers (low layer), which maintain the set-points specified by the supervisor.

More specifically, the local controller is tasked to regulate the water flow rate ϕ by acting on the pump VFD motor which is equipped with a standard PI regulator. The supervisor is the focus of this work; the aim is to determine the set-point value for the local controller that guarantees the thermal energy demand required by the grid, while satisfying the servers thermal constraint.

The thermal constraint requires that temperatures of all the servers CPUs are maintained below a safety threshold in order to avoid overheating and preserve hardware reliability. The constraint can thus be formulated as follows:

$$\max\{T_{i,j}^{cpu}\} \leq \bar{T}, \quad i = 1, \dots, N^s, \quad j = 1, 2, \tag{6.27}$$

where $\bar{T}$ is the temperature threshold and, according to the considered servers, is fixed equal to 80 °C. As regards the demand of thermal energy, we consider two different scenarios. First, we consider the case where the LIC unit is tasked to store a reference value of thermal energy E_{oil}^{ref} imposed by a higher level of management. Second, we consider the case where the LIC system is tasked to store the maximum possible amount of thermal energy E_{oil}^{max} according to the servers workload. It is worth noting that the first problem could be solved by using a simple standard regulator (e.g., a PID controller), since the desired reference value of thermal energy is known. However, in its basic form, a standard regulator could not manage the servers thermal constraint and thus the structure of the controller should be extended in order to take it into account. Conversely, an ESC-based scheme can easily tackle the considered problem accounting for the thermal constraint by defining an appropriate cost index. For example, we can consider the following cost function which accounts for the constraint (6.27) by including

a one-side barrier:

$$J_{ref}(\phi) = (E_{oil}(\phi) - E_{oil}^{ref})^2 + \lambda_1(\max\{0, (\max\{T_{i,j}^{cpu}(\phi)\} - \bar{T})\})^2, \tag{6.28}$$

where λ_1 is a weighting factor. The supervisory ESC-based controller has thus to solve the following optimization problem:

$$\phi^* = \arg\min_{\phi} J_{ref}(\phi). \tag{6.29}$$

The ESC method is suitable for solving also the second considered problem, where the maximum possible amount of energy that can be stored in the oil is unknown and its value changes with the disturbances affecting the system. Similarly to the previous case, we can consider the following performance index:

$$J_{max}(\phi) = E_{oil}(\phi) - \lambda_2(\max\{0, (\max\{T_{i,j}^{cpu}(\phi)\} - \bar{T})\})^2, \tag{6.30}$$

where λ_2 is a weighting factor. The corresponding optimization problem that must be solved by the supervisory ESC-based controller results:

$$\phi^* = \arg\max_{\phi} J_{max}(\phi). \tag{6.31}$$

To face both the problems described above, we exploit the potentials offered by the model-free Newton-like Phasor-based ESC scheme described in Chapter 4 and whose effectiveness has been shown also in Chapter 5. Thanks to the flexibility of the considered method, the same control structure can be applied in both the considered cases by properly defining an objective function that we want to maximize/minimize. It is worth stressing out that the indexes defined above could be extended to take into account also other factors affecting the efficiency of the system. In both the considered scenarios, we choose as amplitude and frequency of the sinusoidal perturbation signal the values $a = 3\,\mathrm{L\,s^{-1}}$ and $\omega = 0.0018\,\mathrm{rad\,s^{-1}}$. The integrator gain and the weighting factor are chosen differently in the two considered problems. In the first case we take $k_1 = -5 \times 10^{-13}$ and $\lambda_1 = 1 \times 10^{12}$, while in the second one we choose $k_2 = -1 \times 10^{-6}$ and $\lambda_2 = 5 \times 10^5$.

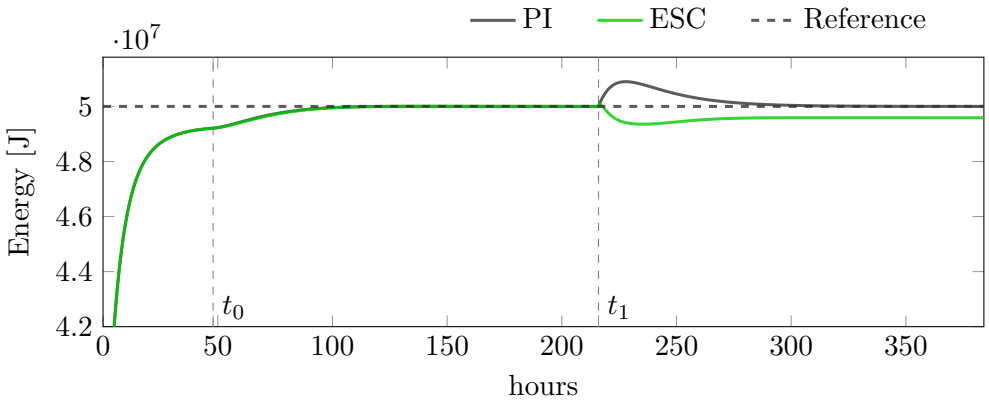
6.5 Simulation Examples

The effectiveness of the Newton-like Phasor-based ESC algorithm in solving the two considered problems has been assessed by exploiting a LIC simulation environment based

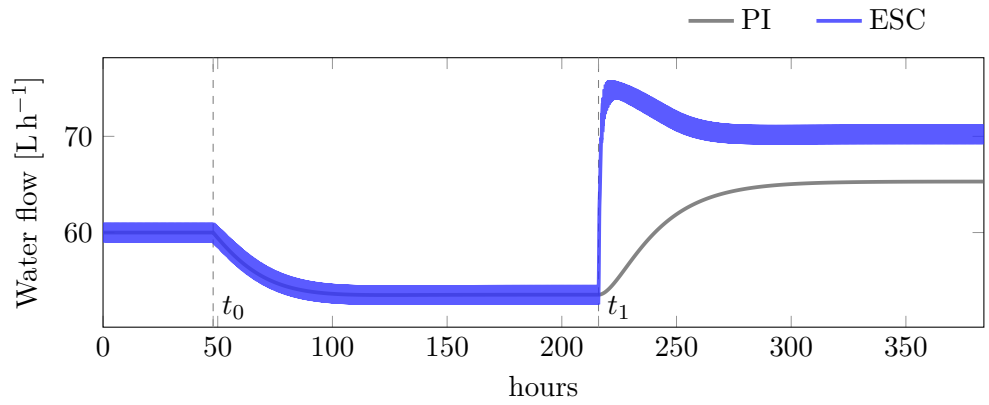
on the POG-NTU thermal model presented in Section 6.3 and developed in Matlab and Simulink.

First, we consider the problem of following a reference value for the oil thermal energy content with the safety constraint on the CPUs temperature. The proposed supervisory controller has thus to find the optimal water flow rate set-point value ϕ^* that solves the optimization problem (6.29). In a first simulation, the reference value is assumed to be constant and equal to $E_{oil}^{ref} = 5 \times 10^7\,\mathrm{J}$, while the electrical power absorbed by all the servers is increased from $\dot{q}^{cpu} = 225\,\mathrm{W}$ to $\dot{q}^{cpu} = 235\,\mathrm{W}$ after about 215 hours. The ambient air temperature is assumed to be constant and equal to $T^{env} = 20\,^\circ\mathrm{C}$ and the cooling water returns at ambient air temperature, i.e., $T_{in}^{water} = 20\,^\circ\mathrm{C}$. The performance of the Newton-like Phasor-based ESC scheme in solving the problem are compared with that of a simple standard regulator, i.e., a PI controller with gains $k_P = 1 \times 10^{-9}\,\mathrm{L\,h^{-1}\,J^{-1}}$ and $k_I = 1 \times 10^{-10}\,\mathrm{L\,h^{-1}\,J^{-1}\,s^{-1}}$ tuned by trial-and-error. The results of the simulation are summarized in Fig. 6.15. As it can be observed, both the PI and the ESC-based controllers start operating at time t_0. Initially, both the controllers are able to adapt the water flow rate value (Fig. 6.15b) in order to meet the requirement on the amount of thermal energy stored in the oil (Fig. 6.15a). Then, at time t_1, we inject a step variation of the servers workload. As it can be observed in Fig. 6.15c, which reports the temperature of the higher CPU of the first server[1], this results in an increase of the temperature of the CPU which reaches the threshold value $\bar{T}$. Since the ESC controller takes into account the temperature constraint (6.27), it adapts the water flow rate in order to maintain the CPU temperature exactly at the threshold value, in order to provide an amount of stored thermal energy as close as possible to the reference value but avoiding servers overheating. Conversely, the standard PI controller adapts the water flow rate in order to meet the thermal energy reference value but violating the temperature constraint, leading to overheating and possible damage of the hardware. We then test in another simulation the Newton-like Phasor-based ESC controller when the thermal energy reference value is varied over time. The electrical power absorbed by the servers is constant and equal to $\dot{q}^{cpu} = 190\,\mathrm{W}$ and the outdoor air temperature and the returning water temperature are as in the previous simulation. The results are summarized in Fig. 6.16. As it can be observed, after the supervisory layer starts operating at time t_0, the water flow rate is adapted (Fig. 6.17b) in order to follow the stored energy reference imposed by the grid (Fig. 6.17a). In Fig. 6.16c, it can be observed that the temperature of the CPUs remains below the considered threshold during the whole simulation.

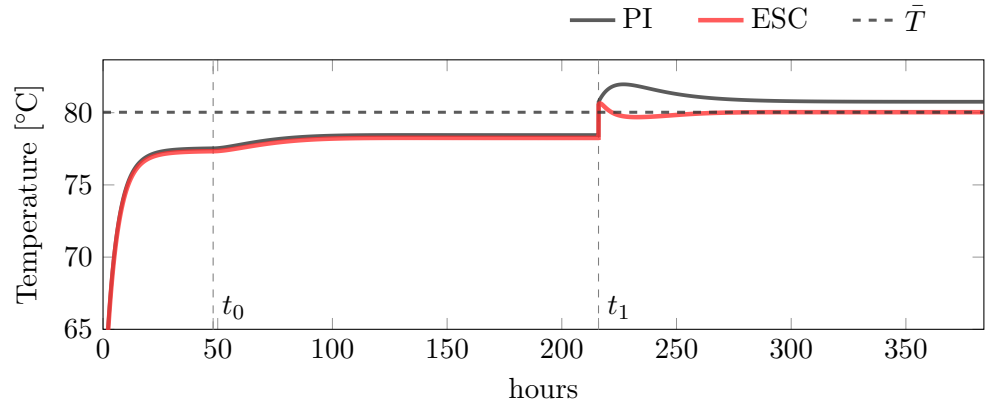
[1]Since all the servers are performing the same amount of computational load, the temperature evolution of their CPUs during the simulations is similar and thus we report only the temperature of the CPUs of the first server.



(a) Time evolution of the energy stored in the oil during the simulation.

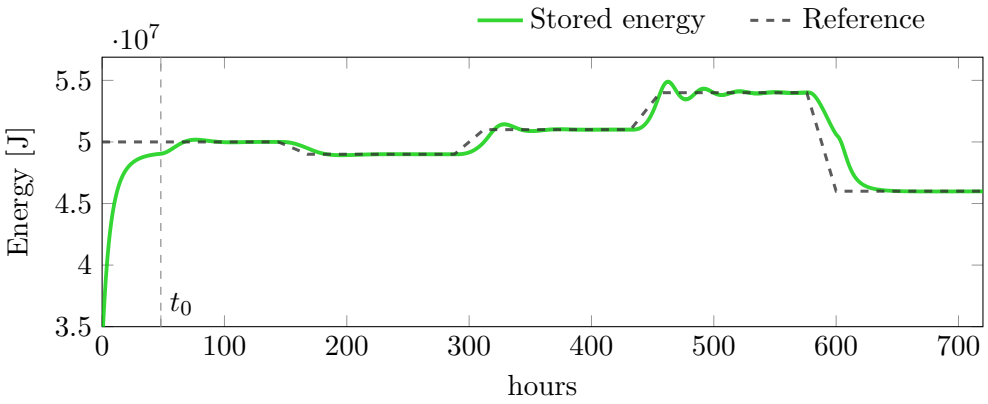


(b) Time evolution of the water flow rate during the simulation.

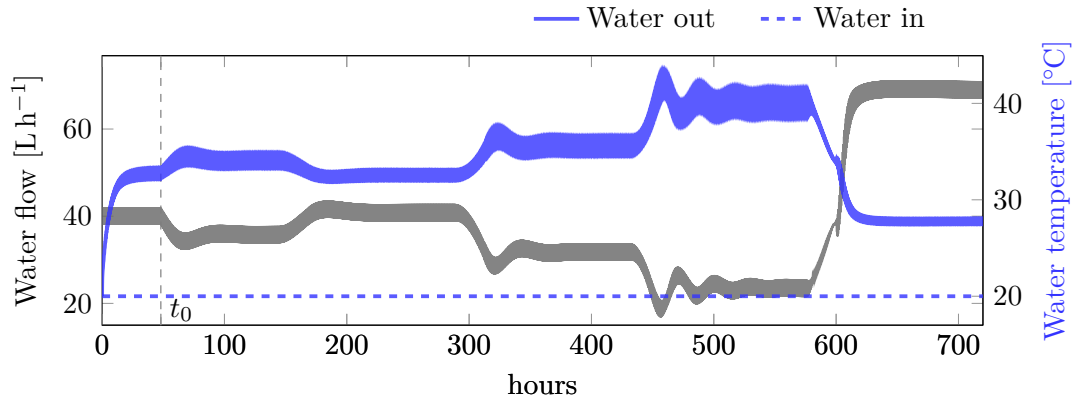


(c) Time evolution of the temperature of the first server higher CPU during the simulation.

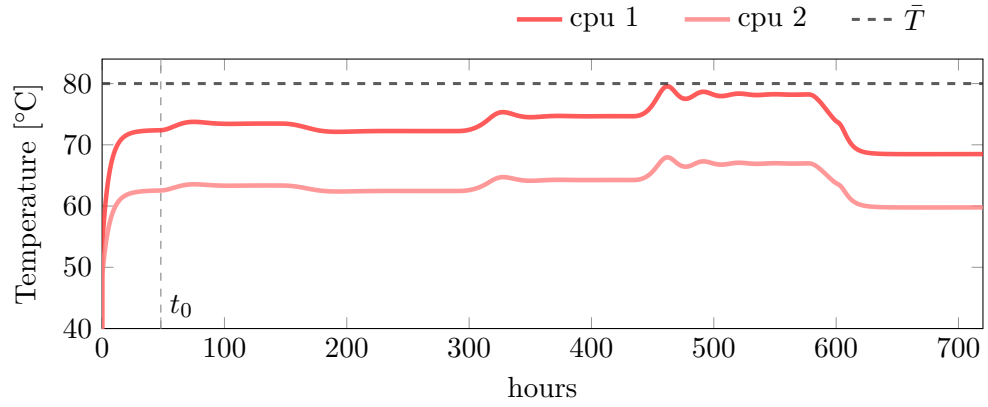
Figure 6.15: Comparison between the proposed ESC-based scheme and a standard PI controller in following a given thermal energy reference value. The controllers are activated at time t_0, while at time t_1 a step variation of the servers workload is injected. Contrary to the standard PI controller, the designed ESC scheme accounts for the constraint on the CPUs temperature, avoiding servers overheating.



(a) Time evolution of the energy stored in the oil during the simulation.



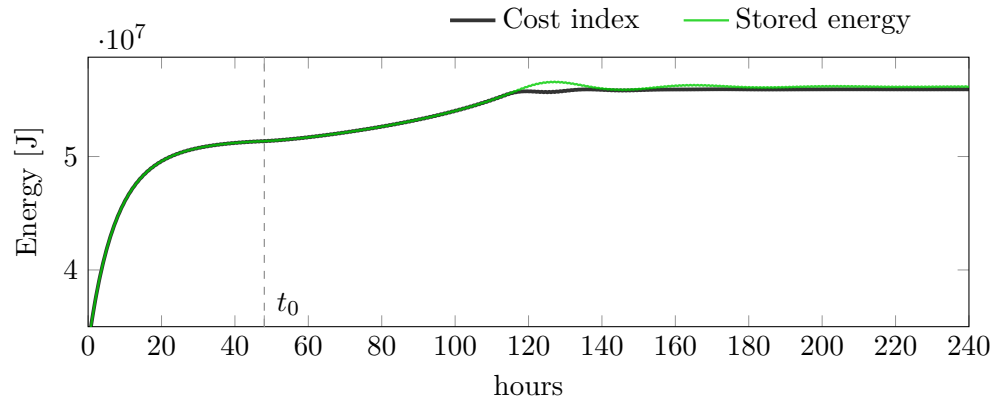
(b) Time evolution of the water flow rate and temperature during the simulation.



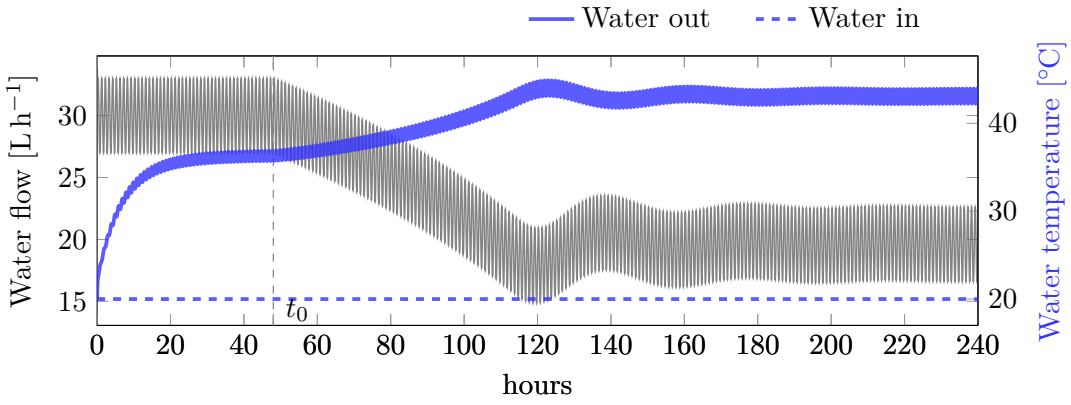
(c) Time evolution of the temperature of the first server CPUs during the simulation.

Figure 6.16: Simulation of the proposed ESC-based scheme in following a given thermal energy reference value with the CPUs temperature constraint. After the controller is activated at time t_0, the water flow rate is adapted by the controller in order meet the considered objective.

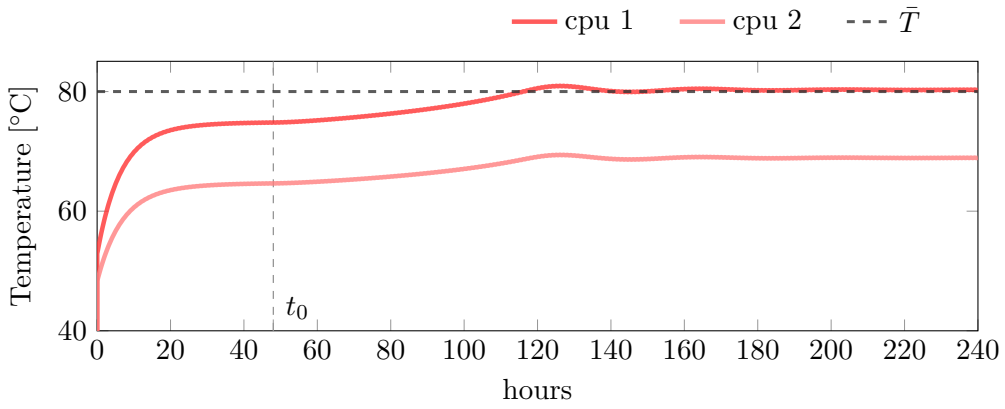
Second, we consider the problem of maximizing the amount of thermal energy contained in the dielectric oil with the safety constraint on the CPUs temperature. The supervisory level of the proposed control architecture has thus to find the optimal water flow rate value ϕ^* that solves the optimization problem (6.31). Initially, we consider that all the immersed servers are tasked to perform the same constant amount of computational load, which corresponds to a constant electrical power consumption $\dot{q}^{cpu} = 190\,\text{W}$. Moreover, we assume that the ambient air temperature is constant during the simulation and equal to $T^{env} = 20\,^\circ\text{C}$ and that the cooling water returns at ambient air temperature, i.e., $T_{in}^{water} = 20\,^\circ\text{C}$. The results of the simulation are summarized in Fig. 6.17. As it can be observed, after the supervisory layer starts operating at time t_0, the water flow rate is reduced (Fig. 6.17b) and, as a consequence, the energy stored in the oil increases (Fig. 6.17a). It is worth noting that the water flow rate is decreased as long as the temperature of the CPUs remains below the safety threshold $\bar{T}$ (Fig. 6.17c). As it can be observed, the maximum energy operating point satisfying the temperature constraint is the one that gives the maximum admissible temperature of the CPUs. For completeness, Fig. 6.17b reports also the temperature of the inlet and outlet water. Decreasing the water flow rate results in a higher temperature of the water leaving the vessel. In order to test the robustness of the proposed approach, we modify the previous numerical scenario by injecting variations of the computational load assigned to the servers, which implies variations of the corresponding electrical power consumption. More precisely, at time t_1 the electrical power absorbed by each server increases of 20 W, and then at time t_2 it decreases of 10 W. The results of the simulation are summarized in Fig. 6.18. As it can be observed, the water flow rate is adjusted by the controller in order to maintain the CPUs temperature as close as possible to the threshold $\bar{T}$, which results in storing the highest amount of energy as possible according to the temperature constraint.



(a) Time evolution of the cost function and the energy stored in the oil during the simulation.

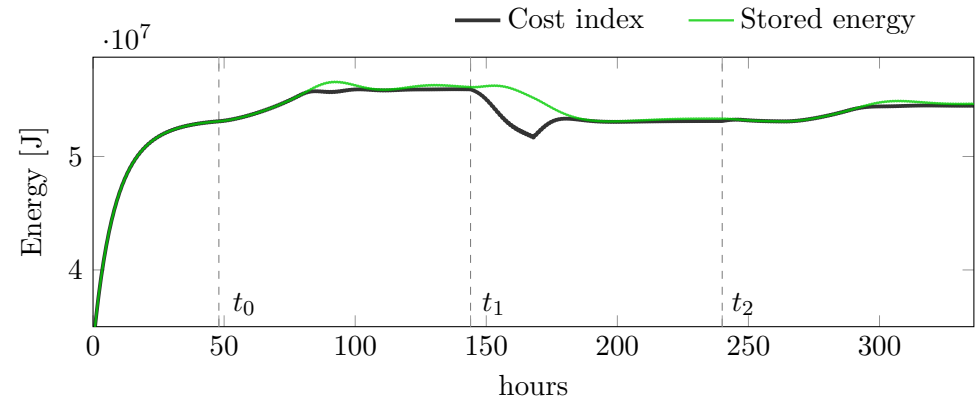


(b) Time evolution of the water flow rate and temperature during the simulation.

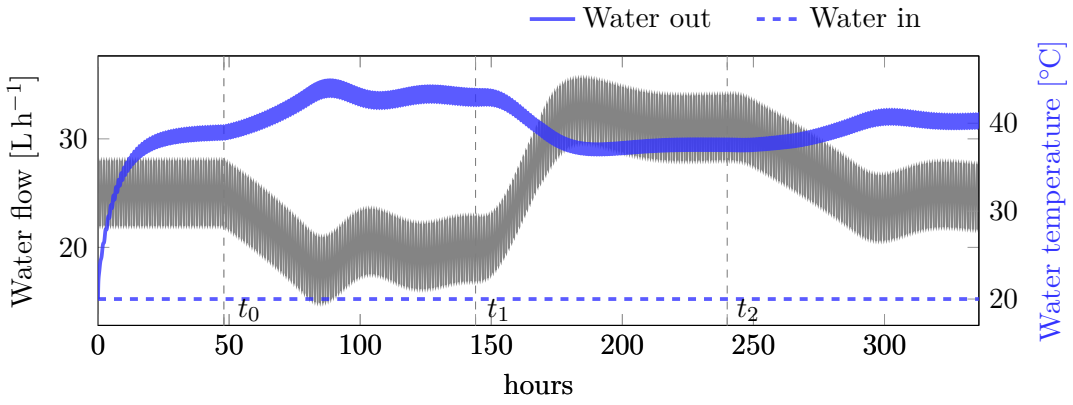


(c) Time evolution of the temperature of the first server CPUs during the simulation.

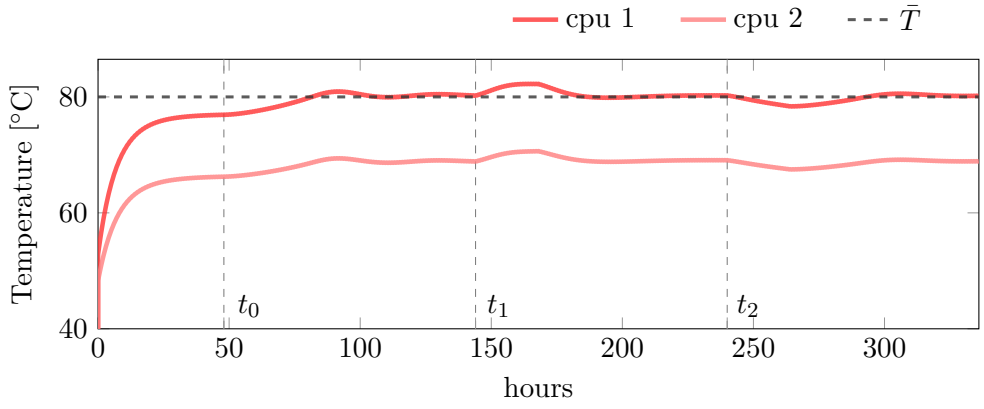
Figure 6.17: Simulation of the proposed ESC-based scheme in maximizing the amount of energy stored in the oil with the CPUs temperature constraint. After the controller is activated at time t_0, the water flow rate is decreased to the optimal value that maximizes the targeted cost index.



(a) Time evolution of the cost function and the energy stored in the oil during the simulation.



(b) Time evolution of the water flow rate and temperature during the simulation.



(c) Time evolution of the temperature of the first server CPUs during the simulation.

Figure 6.18: Simulation of the proposed ESC-based scheme in maximizing the amount of energy stored in the oil with the CPUs temperature constraint and with variations of the servers electrical power absorption $\dot{q}^{cpu}$. The controller is activated at time t_0, then at time t_1 the value of $\dot{q}^{cpu}$ increases of 20 W and finally at time t_2 it decreases of 10 W.

Conclusion

In this Thesis, the problem of optimizing the operation of cooling systems for Data Center applications has been addressed by means of analysis, modelling, and simulations. To cope with the numerous challenges of controlling this kind of complex systems, we have proposed a control architecture based on a two-layer hierarchical structure. More specifically, we have focused on the design of the supervisory layer, which specifies the set-points for the local controllers of the lower layer. To achieve optimal operation, the supervisory layer has to determine optimal set-points that minimize the power consumption (or maximize the energy efficiency) of the system, while provisioning the required cooling demand. The resulting nonlinear constrained optimization problem has been successfully solved by using a model-free Extremum Seeking optimization algorithm. There are several advantages with using this choice. First of all, by being model-free, Extremum Seeking does not require an accurate mathematical model of the system for optimization. In addition, Extremum Seeking inherits the flexibility and robustness against changes in operating conditions and technological constraints which are associated with feedback-based methods. Feedback structure of Extremum Seeking generally enables the methods to not only locate an optimum, but also to track the optimum if it moves over time, e.g., due to disturbances or process changes such as wear of the equipment. Another advantage of the algorithm is the simplicity of its control structure, which facilitates the implementation on commercial systems, with plug-and-play capabilities. In this work, we have proposed a Newton-like Phasor Extremum Seeking scheme, which enhance even more the robustness and the convergence properties of the algorithm. The performance of the algorithm has been evaluated by means of simulations by considering two different case studies.

First, we have considered the problem of optimizing the operation of a Data Center Indirect Evaporative Cooling unit in terms of minimizing its power consumption while provisioning the required cooling demand. The proposed control architecture has been tuned and tested in silico, by means of a simulation environment that is designed starting with a First-Principle Data-Driven model of the plant, which has been calibrated and

validated against experimental data acquired from a commercial cooling unit. The results show the effectiveness of the proposed approach in solving the targeted problem. Also, the results confirm the advantages of the proposed Newton-like Phasor scheme over the classic Extremum Seeking method in terms of convergence speed and robustness against variations of operating conditions. We stress out that the flexibility and plug-and-play properties of the proposed control architecture would preserve its effectiveness in the case of variations of the cooling system architecture and replacement of components, by properly retuning control parameters. Future developments include the implementation of the algorithm on a commercial Indirect Evaporative Cooling system.

Second, we have considered the problem of optimizing the heat recovery capabilities of a Liquid Immersion Cooling system. The effectiveness of the proposed control architecture in solving the problem has been assessed by means of simulations performed with an energy-based dynamic model of the cooling unit. The model has been calibrated and validated against experimental data acquired from a lab-scale experimental testbed. Future developments include coupling the thermal model with models of different heat recovery systems, so to have a complete and holistic quantitative picture of the Data Center plus heat harvester infrastructure, and validating the whole control strategy by implementing it into the real systems.

A

Appendix

A.1 Thermal Parameters of the Computer Room Model

The dynamic and linear model of the computer room described in Section 5.2.2 is given by the system of differential equations (5.44)-(5.49), which depend on the thermal parameters described in the following.

- C_{rack} is the total heat capacity of the racks and of their electrical equipment, expressed in $\mathrm{W\,K^{-1}}$:

$$C_{rack} = M_{rack}\, c_{rack}, \tag{A.1}$$

  where:

  - M_{rack} is the total mass (in kg) of the servers components:

$$M_{rack} = N_{rack}\, m_{rack}, \tag{A.2}$$

    where m_{rack} is the mass of each rack (in kg) and N_{rack} is the number of racks;

  - c_{rack} is the average specific heat capacity of the servers components, in $\mathrm{J\,kg^{-1}\,K^{-1}}$.

- C_{room} is the heat capacity of the air in the computer room, expressed in $\mathrm{W\,K^{-1}}$:

$$C_{room} = M_{air,room}\, c_{p,room}, \tag{A.3}$$

where:

  - $M_{air,room}$ is the total mass (in kg) of the air volume $V_{air,room}$ in the computer room:

$$M_{air,room} = \rho_{air,room}\, V_{air,room}, \tag{A.4}$$

assuming constant room air density $\rho_{air,room} = 1.2250\,\mathrm{kg\,m^{-3}}$ and where the air volume $V_{air,room}$ (in $\mathrm{m^3}$) is obtained by subtracting the total racks volume V_{rack} from the computer room volume V_{room};

  - $c_{p,room}$ is the specific heat capacity at constant pressure of the computer room air, expressed in $\mathrm{J\,kg^{-1}\,K^{-1}}$:

$$c_{p,room} = c_{pa,room} + x_{room,out}\, c_{pv,room}, \tag{A.5}$$

which depends on the specific humidity $x_{room,out}$ of the room air and on the specific heat capacities at constant pressure $c_{pa,room} = 1005\,\mathrm{J\,kg_{da}^{-1}\,K^{-1}}$ and $c_{pv,room} = 1875\,\mathrm{J\,kg_{wv}^{-1}\,K^{-1}}$ of the dry air and of the water vapour contained in the moist air mixture, respectively.

- C_{wall} is the heat capacity of the equivalent wall, expressed in $\mathrm{W\,K^{-1}}$ and obtained by summing the heat capacities of all the different materials constituting the walls and the ceiling of the room.

- R_1 is the thermal resistance between the electrical devices and the room air, expressed in $\mathrm{K\,W^{-1}}$ and given by:

$$R_1 = \frac{1}{\alpha_1\, A_{air,rack}}, \tag{A.6}$$

where:

  - α_1 is the convective heat transfer coefficient between the air and the IT devices, expressed in $\mathrm{W\,m^{-2}\,K^{-1}}$;

  - $A_{air,rack}$ is the area (in $\mathrm{m^2}$) of the heat exchange surface between the air and the IT devices.

- R_2 is the thermal resistance between the room air and the internal surface of the equivalent wall, expressed in $\mathrm{K\,W^{-1}}$ and given by:

$$R_2 = \frac{R_{21}\,R_{22}}{R_{21} + R_{22}}, \tag{A.7}$$

where R_{21} is the thermal resistance between the air and the internal surface of the ceiling:

$$R_{21} = \frac{1}{\alpha_{21}\,A_{wall,in_{ceil}}}, \tag{A.8}$$

and R_{22} is the thermal resistance between the air and the internal surface of lateral walls:

$$R_{22} = \frac{1}{\alpha_{22}\,A_{wall,in_{lateral}}}, \tag{A.9}$$

where:

  - α_{21} is the convective heat transfer coefficient between the room air and the ceiling, expressed in $\mathrm{W\,m^{-2}\,K^{-1}}$ and taken from ISO 6946;
  - α_{22} is the convective heat transfer coefficient between the room air and lateral walls, expressed in $\mathrm{W\,m^{-2}\,K^{-1}}$ and taken from ISO 6946;
  - $A_{wall,in_{ceil}}$ is the area (in $\mathrm{m^2}$) of the heat exchange surface between the air and the internal side of the ceiling;
  - $A_{wall,in_{lateral}}$ is the area (in $\mathrm{m^2}$) of the heat exchange surface between the air and the internal side of lateral walls.

- R_3 is the thermal resistance between the internal surface and the external surface of the equivalent wall, expressed in $\mathrm{K\,W^{-1}}$. Assuming that the walls and the ceiling are composed by n layers having the same composition and the same thickness s, the specific thermal resistance $\bar{R_3}$ expressed in $\mathrm{m^2\,K\,W^{-1}}$ is given by:

$$\bar{R_3} = s \sum_{k=1}^{n} \frac{1}{\lambda_k}, \tag{A.10}$$

where λ_k are the thermal conductivities of the different layers. Finally:

$$R_3 = \frac{\bar{R_3}}{A_{wall}}, \tag{A.11}$$

where A_{wall} is the area of the equivalent wall, obtained as the average between the area of the external surface $A_{wall,out}$ and the area of the internal surface $A_{wall,in}$.

- R_4 is the thermal resistance between the external surface of the equivalent wall and the outdoor air, expressed in $\mathrm{K\,W^{-1}}$ and given by:

$$R_4 = \frac{1}{\alpha_4 \, A_{wall,out}}, \tag{A.12}$$

  where:

  - α_4 is the convective heat transfer coefficient between the air and the external surface of the equivalent wall, expressed in $\mathrm{W\,m^{-2}\,K^{-1}}$ and taken from UNI EN ISO 6946;
  - $A_{wall,out}$ is the area (in $\mathrm{m^2}$) of the heat exchange surface between the air and the external surface of the equivalent wall.

The values of the parameters of the computer room model described above are reported in Tab. A.1.

Parameter	Value	Unit
Mass of each rack m_{rack}	113	kg
Number of racks N_{rack}	15	-
Total mass of the racks M_{rack}	1695	kg
Average specific heat of the racks c_{rack}	500	$\mathrm{J\,kg^{-1}\,K^{-1}}$
Overall heat capacity of the racks C_{rack}	847 500	$\mathrm{W\,K^{-1}}$
Density of the air in the computer room $\rho_{air,room}$	1.225	$\mathrm{kg\,m^{-3}}$
Computer room volume V_{room}	137.5	$\mathrm{m^3}$
Total volume of the racks V_{rack}	21.4	$\mathrm{m^3}$
Air volume in the computer room $V_{air,room}$	116.1	$\mathrm{m^3}$
Air total mass in the computer room $M_{air,room}$	116.03	kg
Equivalent wall heat capacity C_{wall}	3.8582×10^7	$\mathrm{W\,K^{-1}}$
Air-rack heat transfer coefficient α_1	100	$\mathrm{W\,m^{-2}\,K^{-1}}$
Air-rack heat exchange area $A_{air,rack}$	117.72	$\mathrm{m^2}$
Thermal resistance R_1	8.4947×10^{-5}	$\mathrm{K\,W^{-1}}$
Air-ceiling heat transfer coefficient α_{21}	7.69	$\mathrm{W\,m^{-2}\,K^{-1}}$
Air-ceiling heat exchange area $A_{wall,in_{ceil}}$	38.70	$\mathrm{m^2}$
Air-lateral walls heat transfer coefficient α_{22}	10	$\mathrm{W\,m^{-2}\,K^{-1}}$
Air-lateral walls heat exchange area $A_{wall,in_{lateral}}$	74.65	$\mathrm{m^2}$
Thermal resistance R_2	9.5767×10^{-4}	$\mathrm{K\,W^{-1}}$
Area of the internal surface of the building $A_{wall,in}$	113.35	$\mathrm{m^2}$
Area of the external surface of the building $A_{wall,out}$	128.58	$\mathrm{m^2}$
Area of the surface of the equivalent wall A_{wall}	120.96	$\mathrm{m^2}$
Thermal resistance R_3	0.0110	$\mathrm{K\,W^{-1}}$
Outdoor air-external wall heat transfer coefficient α_4	25	$\mathrm{W\,m^{-2}\,K^{-1}}$
Thermal resistance R_4	3.1108×10^{-4}	$\mathrm{K\,W^{-1}}$

Table A.1: Values of the parameters of the computer room model.

A.2 Model of the Pressure Drops Through the Perforated Tiles

In most of air-cooled Data Centers, the chill air provided by the cooling unit flows through a plenum placed under a raised floor and then it enters in the computer room through perforated tiles. The fans which move the air have to provide enough static pressure to overcome the pressure losses caused by the variation of section. To compute these losses, we consider the following model.
The pressure drop Δp_{tile} through the tiles is given by:

$$\Delta p_{tile} = R\,\dot{V}_{air}\,|\dot{V}_{air}| = R\,\frac{\dot{m}_{air}}{\rho_{room,in,tile}}\left|\frac{\dot{m}_{air}}{\rho_{room,in,tile}}\right|, \tag{A.13}$$

where:

- R is the resistance factor, given by:

$$R = \frac{1}{2}\,\frac{\rho_{room,in,tile}}{A^2}\,K, \tag{A.14}$$

  where K is the loss coefficient obtained as in Idelchik and Steinberg (1996):

$$K = \frac{1}{f_{tile}^2}\left(1 + 0.5\,(1 - f_{tile})^{0.75} + \sqrt{2}\,(1 - f_{tile})^{0.375}\right), \tag{A.15}$$

  where the porosity coefficient f_{tile} represents the ratio of the perforated area to the total area of the tile and is equal to 0.5, while A is the total surface of the floor occupied by the tiles and is equal to $2.88\,\mathrm{m}^2$;

- $\dot{V}_{air}$ is the volumetric flow rate of the air flowing through the tiles, in $\mathrm{m}^3\,\mathrm{s}^{-1}$;

- $\rho_{room,in,tile}$ is the density of the air before passing through the perforated tiles, expressed in $\mathrm{kg_{da}}\,\mathrm{m}^{-3}$ and considered equal to the density of the air leaving the heat exchanger and entering in the computer room.

References

ABB . Abb review - datacenters. *The corporate technical journal*, page 1–84, 2013.

Akhremenkov A. A. and Tsirlin A. M. Mathematical model of liquid immersion cooling system for supercomputer (in russian). *Program Systems: Theory and Applications*, 7(1):187–199, 2016.

Almaneea A., Thompson H., Summers J., and Kapur N. Cooling system analysis for a data center using liquid immersed servers. *International Journal of Thermal Technologies*, 4(3):200–207, 2014.

Alonso J. S. J., Martínez F. R., Gómez E. V., and Plasencia M. A.-G. Simulation model of an indirect evaporative cooler. *Energy and Buildings*, 29(1):23 – 27, 1998.

Anisimov S. and Pandelidis D. Theoretical study of the basic cycles for indirect evaporative air cooling. *International Journal of Heat and Mass Transfer*, 84:974 – 989, 2015.

Antal M., Cioara T., Anghel I., Pop C., and Salomie I. Transforming data centers in active thermal energy players in nearby neighborhoods. *Sustainability*, 10 (4):939, 2018.

Ariyur K. B. and Krstic M. Analysis and design of multivariable extremum seeking. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, volume 4, pages 2903–2908. IEEE, 2002.

ASHRAE T. Environmental guidelines for datacom equipment - expanding the recommended environmental envelope. *Refrigeration and Air-Conditioning Engineers Inc*, 2008.

ASHRAE T. Thermal guidelines for data processing environments - expanded data center classes and usage guidance. *Refrigeration and Air-Conditioning Engineers Inc*, 2011.

Astrom K. J. Adaptive control around 1960. *IEEE Control Systems Magazine*, 16(3): 44–49, 1996.

Atta K. T., Johansson A., and Guay M. On the generalization and stability analysis of pareto seeking control. *IEEE Control Systems Letters*, 2(1):145–150, 2017.

Atta K. T., Johansson A., and Gustafsson T. Extremum seeking control based on phasor estimation. *Systems & Control Letters*, 85:37–45, 2015.

Barnett Jr T. The zettabyte era officially begins (how much is that?). *Cisco, September*, 9, 2016.

Beiranvand V., Hare W., and Lucet Y. Best practices for comparing optimization algorithms. *Optimization and Engineering*, 18(4):815–848, 2017.

Bramucci M., Di Santo D., and Forni D. Linee guida per la progettazione di datacenter ad alta efficienza, 2010.

Brunschwiler T., Smith B., Ruetsche E., and Michel B. Toward zero-emission data centers through direct reuse of thermal energy. *IBM Journal of Research and Development*, 53(3):11–1, 2009.

Burden R. L. and Faires J. D. *Numerical Analysis*. Brooks/Cole, 2001.

Camargo J. R. and Ebinuma C. D. A mathematical model for direct and indirect evaporative cooling air conditioning systems. In *Brazilian Congress of Thermal Engineering and Sciences, Encit*, volume 9, pages 30–34, 2003.

Chen Y., Yang H., and Luo Y. Parameter sensitivity analysis and configuration optimization of indirect evaporative cooler (iec) considering condensation. *Applied Energy*, 194:440 – 453, 2017.

Chi Y. Q., Summers J., Hopton P., Deakin K., Real A., Kapur N., and Thompson H. Case study of a data centre using enclosed, immersed, direct liquid-cooled servers. In *2014 Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, pages 164–173. IEEE, 2014.

Chioua M., Srinivasan B., Guay M., and Perrier M. Dependence of the error in the optimal solution of perturbation-based extremum seeking methods on the excitation frequency. *The Canadian Journal of Chemical Engineering*, 85(4):447–453, 2007.

Cisco V. Cisco visual networking index: Forecast and trends, 2017–2022. *White paper, Cisco Public Information*, 2019.

Corcoran P. and Andrae A. Emerging trends in electricity consumption for consumer ict. *National University of Ireland, Galway, Connacht, Ireland, Tech. Rep*, 2013.

Czop P., Kost G., Sławik D., and Wszołek G. Formulation and identification of first-principle data-driven models. *Journal of Achievements in materials and manufacturing Engineering*, 44(2):179–186, 2011.

Dai J., Ohadi M. M., Das D., and Pecht M. G. *OPTIMUM COOLING OF DATA CENTERS.* Springer, 2016.

Dayarathna M., Wen Y., and Fan R. Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):732–794, 2015.

Dixit U. S., Hazarika M., and Davim J. P. *A brief history of mechanical engineering.* Springer, 2017.

Drakunov S., Ozguner U., Dix P., and Ashrafi B. Abs control using optimum search via sliding modes. *IEEE Transactions on Control Systems Technology*, 3(1): 79–85, 1995.

Draper C. S. and Li Y. T. *Principles of optimalizing control systems and an application to the internal combustion engine.* American Society of Mechanical Engineers, 1951.

Dunlap K. and Rasmussen N. The advantages of row and rack-oriented cooling architectures for data centers. *APC White Paper*, 130, 2006.

energyliteracy.com . URL http://energyliteracy.com/.

Evans T. The different technologies for cooling data centers. *APC white paper*, 59, 2012.

Facebook . The open compute server architecture specifications. URL www.opencompute.org.

Fritzson P. *Principles of object-oriented modeling and simulation with Modelica 2.1.* John Wiley & Sons, 2010.

Galletly J. E. An overview of genetic algorithms. *Kybernetes*, 21(6):26–30, 1992.

Ghaffari A., Krstić M., and NešIć D. Multivariable newton-based extremum seeking. *Automatica*, 48(8):1759–1767, 2012.

Grossi F. and Zanasi R. Dynamic models of thermal systems using an energy-based modeling approach. *Journal of Heat Transfer*, 138(10):102801, 2016.

Group I.-T. R. and others . Top 10 energy-saving tips for a greener data center. *Info-Tech Research Group.*, 2007.

Hartman T. All-variable speed centrifugal chiller plants. *ASHRAE journal*, 43(9): 43–53, 2001.

Hasan A. Going below the wet-bulb temperature by indirect evaporative cooling: Analysis using a modified ϵ-ntu method. *Applied Energy*, 89(1):237 – 245, 2012. Special issue on Thermal Energy Management in the Process Industries.

Haskara I., Özgüner U., and Winkelman J. Extremum control for optimal operating point determination and set point optimization via sliding modes. *Journal of Dynamic Systems, Measurement, and Control*, 122(4):719–724, 2000.

Heidarinejad G. and Moshari S. Novel modeling of an indirect evaporative cooling system with cross-flow configuration. *Energy and Buildings*, 92:351 – 362, 2015.

Hou Z.-S. and Wang Z. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.

ICE R. S. Swedish institute of computer science, infrastructure and cloud datacenter test environment. https://ice.sics.se/.

Idelchik I. and Steinberg O. Handbook of hydraulic resistance, begell house. *New York*, 1996.

Jaber S. and Ajib S. Evaporative cooling as an efficient system in mediterranean region. *Applied Thermal Engineering*, 31(14-15):2590–2596, 2011.

Kang S., Schmidt R. R., Kelkar K. M., Radmehr A., and Patankar S. V. A methodology for the design of perforated tiles in raised floor data centers using computational flow analysis. *IEEE Transactions on Components and Packaging Technologies*, 24(2):177–183, 2001.

Karki K. C., Radmehr A., and Patankar S. V. Use of computational fluid dynamics for calculating flow rates through perforated tiles in raised-floor data centers. *HVAC&R Research*, 9(2):153–166, 2003.

Kays W. M. and London A. L. *Compact heat exchangers*. McGraw-Hill, New York, NY, 1984.

Khalaj A. H. and Halgamuge S. K. A review on efficient thermal management of air-and liquid-cooled data centers: From chip to the cooling system. *Applied energy*, 205:1165–1188, 2017.

Khalil H. K. Nonlinear systems. *Upper Saddle River*, 2002.

Kiran T. R. and Rajput S. An effectiveness model for an indirect evaporative cooling (iec) system: Comparison of artificial neural networks (ann), adaptive neuro-fuzzy inference system (anfis) and fuzzy inference system (fis) approach. *Applied Soft Computing*, 11(4):3525 – 3533, 2011.

Koomey J., Berard S., Sanchez M., and Wong H. Implications of historical trends in the electrical efficiency of computing. *IEEE Annals of the History of Computing*, 33 (3):46–54, 2010.

Korovin S. and Utkin V. Using sliding modes in static optimization and nonlinear programming. *Automatica*, 10(5):525–532, 1974.

Krstić M. Performance improvement and limitations in extremum seeking control. *Systems & Control Letters*, 39(5):313–326, 2000.

Krstic M. and Wang H.-H. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36(4):595–601, 2000.

Kuehn T. H., Ramsey J. W., and Threlkeld J. L. *Thermal environmental engineering*, volume 188. Prentice Hall Upper Saddle River, NJ, 1998.

Lääkkölä R. and others . Data center degrowth-an experimental study. 2015.

Landauer R. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961.

Lazzarin R. Informatory note on refrigeration technologies. Technical report, International Institute of Refrigeration, January 2015.

Leblanc M. Sur l'electrification des chemins de fer au moyen de courants alternatifs de frequence elevee. *Revue générale de l'électricité*, 12(8):275–277, 1922.

Lee K.-P. and Chen H.-L. Analysis of energy saving potential of air-side free cooling for data centers in worldwide climate zones. *Energy and Buildings*, 64:103–112, 2013.

Liberati P., Antonellis S. D., Leone C., Joppolo C. M., and Bawa Y. Indirect evaporative cooling systems: modelling and performance analysis. *Energy Procedia*, 140:475 – 485, 2017. Beyond NZEB Buildings (AiCARR 50th International Congress, Matera (I), 10-11 May 2017).

Lind J. and Rundgren E. Industrial symbiosis in heat recovery collaborations between data centers and district heating and cooling companies, 2017.

Lissandrin M., Rampazzo M., Cecchinato L., and Beghi A. Optimal operational efficiency of chillers using oil-free centrifugal compressors. *International Journal of Refrigeration*, 82:83–96, 2017.

Liu S.-J. and Krstic M. Newton-based stochastic extremum seeking. *Automatica*, 50 (3):952–961, 2014.

Liu Z., Allen W., and Modera M. Simplified thermal modeling of indirect evaporative heat exchangers. *HVAC&R Research*, 19(3):257–267, 2013.

Lucchese R. and Johansson A. On energy efficient flow provisioning in air-cooled data servers. *Control Engineering Practice*, 89:103 – 112, 2019. ISSN 0967-0661. URL http://www.sciencedirect.com/science/article/pii/S0967066119300747.

Malinowski K. and others . *Computer Aided Control System Design: Methods, Tools and Related Topics*. World Scientific, 1994.

Manzie C. and Krstic M. Discrete time extremum seeking using stochastic perturbations. In *2007 46th IEEE Conference on Decision and Control*, pages 3096–3101. IEEE, 2007.

Manzie C. and Krstic M. Extremum seeking with stochastic perturbations. *IEEE Transactions on Automatic Control*, 54(3):580–585, 2009.

Marcinichen J. B., Olivier J. A., Lamaison N., and Thome J. R. Advances in electronics cooling. *Heat Transfer Engineering*, 34(5-6):434–446, 2013.

Matsuoka M., Matsuda K., and Kubo H. Liquid immersion cooling technology with natural convection in data center. In *2017 IEEE 6th International Conference on Cloud Networking (CloudNet)*, pages 1–7. IEEE, 2017.

Milad M. A. *UPS system: how current and future technologies can improve energy efficiency in data centres*. PhD thesis, 2017.

Moase W. H., Manzie C., and Brear M. J. Newton-like extremum-seeking part i: Theory. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 3839–3844. IEEE, 2009.

Moase W. H., Manzie C., and Brear M. J. Newton-like extremum-seeking for the control of thermoacoustic instability. *IEEE Transactions on Automatic Control*, 55(9): 2094–2105, 2010.

Moore G. E. and others . Cramming more components onto integrated circuits, 1965.

Muzychka Y. S. and Yovanovich M. M. Modeling the f and j characteristics for transverse flow through an offset strip fin at low reynolds number. *Journal of Enhanced Heat Transfer*, 8(4), 2001.

Nada S., Said M., and Rady M. Cfd investigations of data centers' thermal performance for different configurations of cracs units and aisles separation. *Alexandria Engineering Journal*, 55(2):959–971, 2016.

Nasiri A., Nie Z., Bekiarov S. B., and Emadi A. An on-line ups system with power factor correction and electric isolation using bifred converter. *IEEE Transactions on Industrial Electronics*, 55(2):722–730, 2008.

Nešić D. Extremum seeking control: Convergence analysis. *European Journal of Control*, 15(3-4):331–347, 2009.

Ogawa M., Fukuda H., Kodama H., Endo H., Sugimoto T., Kasajima T., and Kondo M. Development of a cooling control system for data centers utilizing indirect fresh air based on model predictive control. In *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 132–137. IEEE, 2015.

Oliveira F. and Ukil A. Comparative performance analysis of induction & synchronous reluctance motors in chiller systems for energy efficient buildings. *IEEE Transactions on Industrial Informatics*, 2019.

Olups R. *Zabbix Network Monitoring.* Packt Publishing Ltd, 2016.

Oró E., Allepuz R., Martorell I., and Salom J. Design and economic analysis of liquid cooled data centres for waste heat recovery: A case study for an indoor swimming pool. *Sustainable cities and society*, 36:185–203, 2018.

Paludetto D. and Lorente S. Modeling the heat exchanges between a datacenter and neighboring buildings through an underground loop. *Renewable Energy*, 93:502–509, 2016.

Pan Y., Özgüner Ü., and Acarman T. Stability and performance improvement of extremum seeking control with sliding mode. *International Journal of Control*, 76 (9-10):968–985, 2003.

Parolini L., Garone E., Sinopoli B., and Krogh B. H. A hierarchical approach to energy management in data centers. In *49th IEEE Conference on Decision and Control (CDC)*, pages 1065–1070. IEEE, 2010a.

Parolini L., Sinopoli B., Krogh B. H., and Wang Z. A cyber–physical systems approach to data center modeling and control for energy efficiency. *Proceedings of the IEEE*, 100(1):254–268, 2011.

Parolini L., Tolia N., Sinopoli B., and Krogh B. H. A cyber-physical systems approach to energy management in data centers. In *Proceedings of the 1st acm/ieee international conference on cyber-physical systems*, pages 168–177. ACM, 2010b.

Poess M. and Nambiar R. O. Energy cost, the key challenge of today's data centers: a power consumption analysis of tpc-c results. *Proceedings of the VLDB Endowment*, 1(2):1229–1240, 2008.

Polley G. T. and Abu-Khader M. M. Interpreting and applying experimental data for plate-fin surfaces: problems with power law correlation. *Heat transfer engineering*, 26(9):15–21, 2005.

Prucnal D. Doing more with less: Cooling computers with oil pays off. *The Next Wave*, 20(2):20–29, 2013.

Rasmussen N. The different types of air distribution for it environments. *Schnider electric, APC cooling*, 2012.

Richert F., Rückert J., Schloßer A., and Abel D. Comparison of modelica and matlab by means of a diesel engine model. *IFAC Proceedings Volumes*, 37(22):287–292, 2004.

Rotea M. A. Analysis of multivariable extremum seeking algorithms. In *Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No. 00CH36334)*, volume 1, pages 433–437. IEEE, 2000.

Schmidt R. R., Karki K. C., Kelkar K. M., Radmehr A., and Patankar S. V. Measurements and predictions of the flow distribution through perforated tiles in raised floor data centers. *Proceedings IPACK2001-15728*, 2001.

Shah J. M., Eiland R., Siddarth A., and Agonafer D. Effects of mineral oil immersion cooling on it equipment reliability and reliability enhancements to data

center operations. In *2016 15th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, pages 316–325. IEEE, 2016.

Simon K. Project natick-microsoft's self-sufficient underwater datacenters. *IndraStra Global*, 4(6):4, 2018.

Simonazzi E., Galrinho M. R., Varagnolo D., Gustafsson J., and Garcia-Gabin W. Detecting and modelling air flow overprovisioning/underprovisioning in air-cooled datacenters. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 4893–4900. IEEE, 2018.

Spall J. C. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on aerospace and electronic systems*, 34(3):817–823, 1998.

Sridhar A., Sabry M. M., and Atienza D. System-level thermal-aware design of 3d multiprocessors with inter-tier liquid cooling. In *2011 17th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, pages 1–9. IEEE, 2011.

Stress-ng . URL https://kernel.ubuntu.com/git/cking/stress-ng.git/.

Sundaralingam V., Arghode V. K., Joshi Y., and Phelps W. Experimental characterization of various cold aisle containment configurations for data centers. *Journal of electronic packaging*, 137(1):011007, 2015.

Tan Y., Li Y., and Mareels I. M. Extremum seeking for constrained inputs. *IEEE Transactions on Automatic Control*, 58(9):2405–2410, 2013.

Tan Y., Moase W. H., Manzie C., Nešić D., and Mareels I. Extremum seeking from 1922 to 2010. In *Proceedings of the 29th Chinese Control Conference*, pages 14–26. IEEE, 2010.

Trollberg O., Carlsson B., and Jacobsen E. W. Extremum seeking control of the canon process—existence of multiple stationary solutions. *Journal of Process Control*, 24(2):348–356, 2014.

Varga A. Computer-aided control systems design: Introduction and historical overview.

Walsh G. C. On the application of multi-parameter extremum seeking control. In *Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No. 00CH36334)*, volume 1, pages 411–415. IEEE, 2000.

Wan Y., Ren C., and Xing L. An approach to the analysis of heat and mass transfer characteristics in indirect evaporative cooling with counter flow configurations. *International Journal of Heat and Mass Transfer*, 108:1750 – 1763, 2017.

Wang F., Yoshida H., and Miyata M. Total energy consumption model of fan subsystem suitable for continuous commissioning. *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 110(1):357–364, 2004.

Wang H.-H. and Krstic M. Extremum seeking for limit cycle minimization. *IEEE Transactions on Automatic control*, 45(12):2432–2436, 2000.

Whitehead B., Andrews D., Shah A., and Maidment G. Assessing the environmental impact of data centres part 1: Background, energy use and metrics. *Building and Environment*, 82:151–159, 2014.

Wibron E., Ljung A.-L., and Lundström T. S. Comparing performance metrics of partial aisle containments in hard floor and raised floor data centers using cfd. *Energies*, 12(8):1473, 2019.

Yeh C.-C. and Manjrekar M. D. A reconfigurable uninterruptible power supply system for multiple power quality applications. *IEEE transactions on power electronics*, 22(4):1361–1372, 2007.

Zanasi R. Power oriented modelling of dynamical system for simulation. In *IMACS Symp. on Modelling and Control of Technological System*, volume 2, pages 31–35. Lille, France, 1991.

Zhang H., Shao S., Xu H., Zou H., and Tian C. Free cooling of data centers: A review. *Renewable and Sustainable Energy Reviews*, 35:171–182, 2014.

Zhu G., Chow T.-T., and Lee C. Performance analysis of counter-flow regenerative heat and mass exchanger for indirect evaporative cooling based on data-driven model. *Energy and Buildings*, 155:503 – 512, 2017.

Zimmermann S., Meijer I., Tiwari M. K., Paredes S., Michel B., and Poulikakos D. Aquasar: A hot water cooled data center with direct energy reuse. *Energy*, 43(1):237–245, 2012.

Zupancic B., Karba R., Atanasijevic-Kunc M., and Music J. Continuous systems modelling education–causal or acausal approach? In *ITI 2008-30th International Conference on Information Technology Interfaces*, pages 803–808. IEEE, 2008.