Institute of Architecture of Application Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit

Air Cooling vs. Liquid Immersion Cooling: Can Liquid Immersion Cooling Improve the Energy and Space Efficiency of Data Centres?

Yannis Blosch

Course of Study:	Informatik
Examiner:	Prof. Dr. Marco Aiello
Supervisor:	Prof. Dr. Marco Aiello
Commenced:	December 1, 2020
Completed:	June 1, 2021

Abstract

The demand for cloud services is rising continuously, inflating the energy demand for data centres exponentially. The purpose of this literature review is to discover the positive impact liquid immersion cooling can have on data centre efficiency and density. Utilizing a systematic literature review, this paper illustrates the research already conducted on immersion cooling. The calculated results, using the available data, shows that by switching from air to liquid immersion cooling, about half the energy used by a data centre and two thirds of the space occupied by a data centre can be freed for additional hardware. This review succeeds in showing the capabilities of liquid immersion cooling in data centres. Further research can be conducted on the actual improvement when a data centre is converted to immersion cooling.

Kurzfassung

Der Bedarf an Online-Diensten steigt kontinuierlich an, das lässt den Energieverbrauch von Rechenzentren exponentiell ansteigen. Die Absicht der Literaturrecherche ist es, zu entdecken, wie die Effizienz und Dichte von Rechenzentren durch Flüssigkeitstauchkühlung erhöht werden kann. Es wurde eine systematische Literaturrecherche gewählt, um die bisher geleistete Forschungsarbeit zum Thema aufzuzeigen. Die mit den gefundenen Daten berechneten Ergebnisse haben gezeigt, dass ein Umstieg von Luft zu Flüssigkeitstauchkühlung die verwendete Energie halbieren und den Platzverbrauch sogar auf ein Drittel reduzieren kann. Das ermöglicht eine Erweiterung der Rechenkapazität. Die Literaturrecherche hat die Möglichkeiten von Flüssigkeitstauchkühlung erfolgreich gezeigt. Weitere Forschungsarbeit könnte geleistet werden, indem man eine Umrüstung eines luftgekühlten Rechenzentrums begleitet und die Größe und den Energieverbrauch sowohl vorher als auch nachher misst.

Contents

List of Figures

List of Tables

Glossary

air cooling In this paper air cooling refers to cooling of IT equipment using air as medium. 13

CDC Cloud data centre. Data Centres with main focus on cloud computing. 15

cold plate A plate attached to a heat producing part, allowing for heat transfer into water. 7

computing power Amount of useful work a computer can accomplish. It is measured in FLOPS in this paper. 13

cooling tower A cooling tower uses the outside air to cool water that is hotter than the outside climate, and hence makes use of free cooling. 15

CPU Central processing unit, the central computing and control unit of a computer. 16

CRAC Computer room air conditioning. Equipment responsible for air cooling in a data center, for example, a chiller . 23

data centre A room or facility, dedicated to IT equipment. 7

density For this review, density stands for the amount of computing power that runs in a specific volume. 13

energy efficiency The relation between computing power and energy used. 15

ENIAC Electronic Numerical Integrator and Computer. Used by the United States Army to calculate artillery firing tables, it was the first programmable computer. The ENIAC was completed in 1945. 7

facility level the whole facility considered. 17

FLOPS Floating point operations per second is a metric for measuring computing performance . 17

free cooling Use of free resources for cooling, for example, cold air. 15

GFLOPS GigaFLOPS = 1 000 000 000 FLOPS. 35

HPC High-performance computing. This refers to servers with the sole benefit of having maximum computing power. 16

HPDC High-performance data centre. A data centre focusing on maximum computing power. 25

hybrid cooling A cooling technique, that uses indirect water cooling for high powered components and air cooling for the rest. 16

hyperscale Hyperscale is about achieving massive scale in computing. This is done to reduce cost. 13

ICT Information and communication technology. An extensional term of IT, contains some devices not included in IT. 15

immersion cooling Cooling a server by submersion into a dielectric fluid. 7

IT Information technology. A generic term for everything related to electronic data processing. 15

PUE Power usage effectiveness = $\frac{facilityenergy}{ITequipmentenergy}$, a value illustrating how efficiently energy is used in a data centre. 9

state-of-the-art Is used here with the same meaning as best practice to describe an installation with maximum possible efficiency. 32

TFLOPS TeraFlops = 1 000 000 000 000 FLOPS . 35

1 The Problem of Modern Data Centres

1.1 An Introduction

A data centre is everything from a server room designated for a company's server equipment up to a gigantic building only built to provide storage and computing power for other companies or individual people, namely a hyperscale data centre. Data centres are among the most significant contributors to global energy consumption and need much space not to be extra inefficient. They are designed to ensure high availability, reliability and security. For this paper, all types of data centres are relevant, from smaller centres up to hyperscale ones. Due to the fact that information technology equipment produces heat under load, data centres have to be artificially cooled. Air cooling is the most prominent cooling technique used. As an example, the hyperscale data centres of *Google*, *Facebook*, *Amazon* etc. also mainly rely on air cooling their servers and achieve good efficiency with it [Nic18].

However, air is not the most efficient cooling method available. Instead, efficiency could further be improved by switching to immersion cooling. Improving the efficiency of big data centres by a small percentage would significantly impact global energy consumption given their size. In addition, such a switch could influence other operators of smaller data centres in trusting immersion cooling [EEV+17]. Smaller air cooled data centres are on average much more inefficient than hyperscale ones. Because of that, especially operators of smaller data centres could increase their efficiency and power density dramatically. Regardless of their size, smaller data centres outnumber hyperscale ones and, hence, make up a significant percentage of the global energy consumption of data centres [Nic18].

In order to combat the described problem, this scientific literature review tries to gather as much information as possible regarding immersion cooling. The paper aims at calculating the possible increases in energy and space efficiency by switching to immersion cooling. Can data centre operators improve their efficiency without loss of computing power, power density or high costs? If this criteria is not given, data centre operators might be hesitant to apply liquid immersion cooling.

To answer the question, if liquid immersion cooling can improve the energy and space efficiency of data centres, data is gathered from different databases. The gathered data is then presented and used to calculate eventual efficiency gains. This should allow a well-founded answer to the research question.

The following steps are taken to analyse the possible impact of immersion cooling for data centres. First of all, current data centres are analysed, and problems that are widely agreed on are located. The technique of liquid immersion cooling and the chances associated with it are presented. Chapter 2 expounds the methodology regarding the research. Then the dominant cooling solutions will be presented in Chapter 3, air cooling and hybrid/indirect liquid cooling, also explaining their

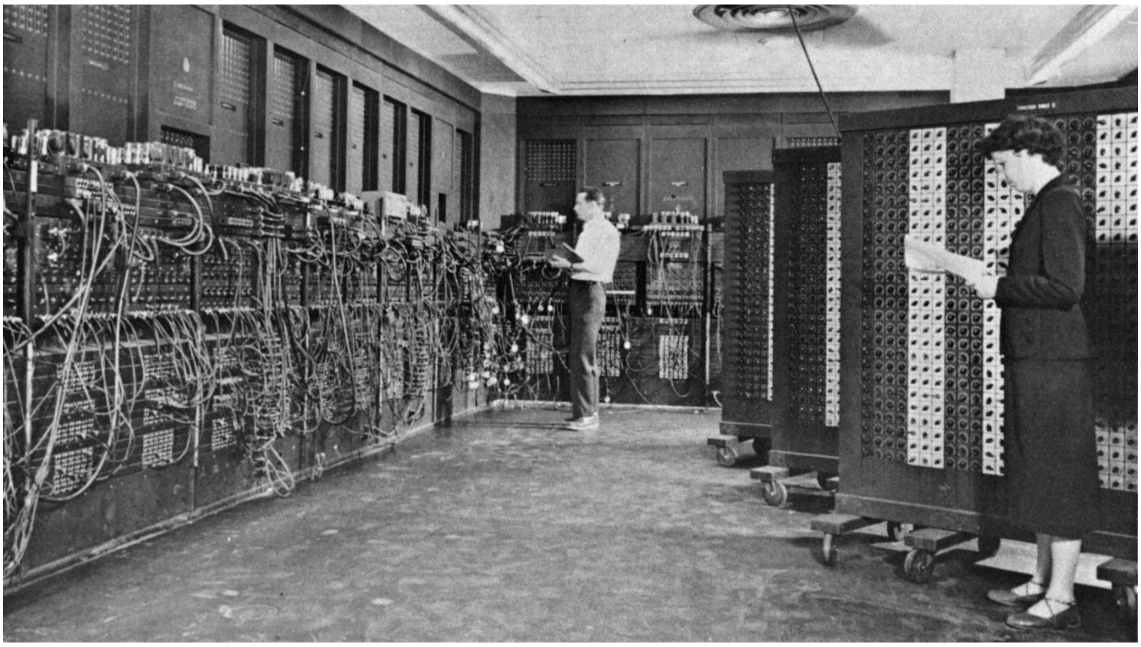

Figure 1.1: The ENIAC [ENIAC21]

limitations. They are followed by the contender liquid immersion cooling and insights into why it has the potential to supersede the other cooling solutions. Chapter 4 presents the calculation of energy and space efficiency that were performed to analyze the potential of liquid immersion cooling. Although immersion cooling is not yet dominant, a few companies make use of it already. Some of those serve as an example of immersion cooling in practice in Chapter 5. Based on the calculations, a prognosis will be given if immersion cooling is the future of data centre cooling.

1.2 Limitations of Modern Data Centres

Information technology has come a long way since its beginning. Looking back, the first data centre-like structure is probably the Electronic Numerical Integrator and Computer (ENIAC). The ENIAC was presented to the public a little more than 75 years ago. Build and operated by the US-army; it was the first programmable electric computer. Still operating with 17.468 vacuum tubes, the ENIAC looked a little like a modern server room, weighing 27 tonnes and needed 174kW to run. While different from modern data centres, the ENIAC had some similarities and almost identical requirements as modern data centres. It had a high amount of computing hardware located in one facility, downtimes were avoided, and 20kW of cooling were needed. The 40 components the ENIAC was made of even had server rack like dimensions [ENIAC21]. Figure 1.1 shows a picture of the ENIAC. The wiring and the dimensions remind of a modern data centre.

More modern data centres began gaining popularity around 1990. The demand for connectivity was rising and companies were increasingly relying on the internet. The service model of providing online storage and computing power began gaining popularity. Because of this, more and bigger data centres were built and until today this trend has not yet stopped [DC-HISTORY18].

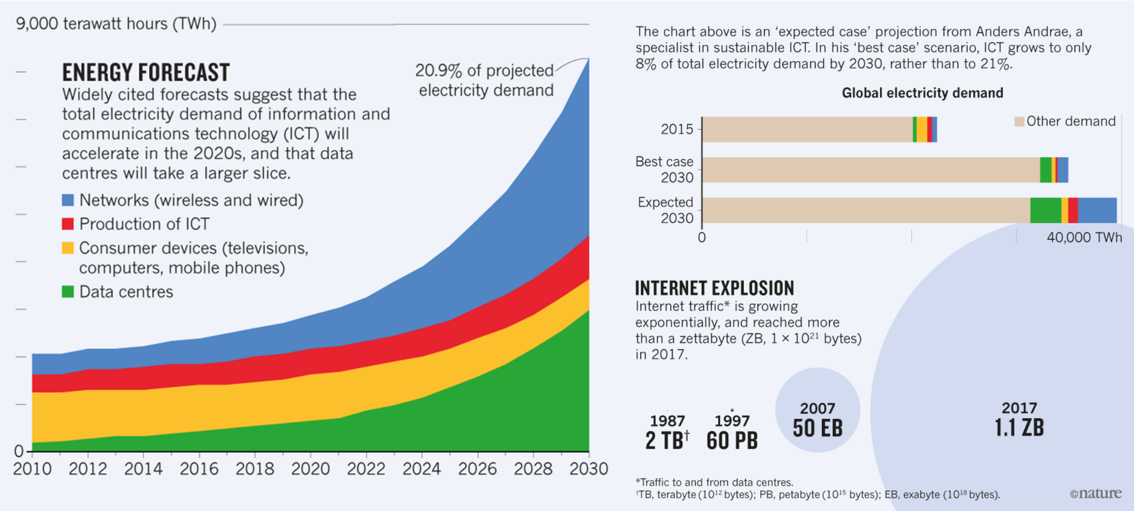

Figure 1.2: Energy forecast for information communication technology [Nic18]

The energy consumption of information communication technology (ICT), as shown in Figure 1.2, is rising exponentially. Figure 1.2 presents the expected energy usage in 2030 highlighting the consumption of data centres and other ICT equipment, the best case scenario is also presented. Today, 46% of the worlds population are internet users and they use 8 zettabytes of data daily [I W19]. This comes at the cost of 1500TWh of energy in 2014, making up 10% of the total global electricity usage [GBR+14]. 20% of the 10% global electricity usage is coming from data centres alone [360View18; MMK17]. These numbers sound already quite impressive, but usage of cloud services is increasing exponentially. Cloud data centre (CDC) energy consumption is rising 20-25% each year [360View18]. Currently, CDCs are responsible for 78.7 million tons of CO2. This is about the same as global aviation and makes up 2% of the global CO2 emissions. To give an example of a smaller size, a data centre requires between 100 and 200 times more energy than a similar-sized office building [WTKD19].

Data centre operators are already trying to improve their energy efficiency by switching to hyperscale data centers that are highly optimized. Hyperscale data centres have no other purpose than to provide compute power and digital storage. These data centres achieve higher energy efficiency than smaller ones that are not so specialized, but at the cost of space efficiency. Such centres are truly gigantic buildings which are not suitable for non-remote locations [Nic18].

Free cooling is another way of improving the energy efficiency of a data centre. This technique can be used without the expense of energy, a good example of this is a cooling tower. A cooling tower uses the outside air to cool water that is hotter than the outside climate, and hence makes use of free cooling. One fatal flaw of air cooling in combination with free cooling, is that air coming from the information technology (IT) equipment is not very hot and air going to the IT equipment needs to be fairly cold. This allows for very little free cooling in hotter climates [360View18; KG16]. Only a few spots located in colder climates allow for high amounts of free cooling in air cooled data centres. However, it is obvious that not all data centres can be moved to colder climates [360View18].

Furthermore, waste heat utilization is also possible. Waste heat utilization describes the concept of using heat generated by servers to heat other facilities, buildings or glasshouses for example. Data centre energy efficiency is not directly improving with waste heat utilization, but the overall global efficiency is. Unfortunately this concept is hard to implement with air cooling because the exhaust air is not that warm and air as a medium is hard to transport [360View18]. Thus, air does not seem to be the best medium to interact with productive methods like free cooling and waste heat utilization. Those methods can be utilized to enhance the global efficiency, but their effect would be much greater with a different medium.

In addition to the energy efficiency concerns amplified by the presence of global warming and the cost of energy, there is another concern for data centre operators, namely space efficiency. Modern data centres can only be cooled by implementing cooling techniques that take up a lot of space. Airstreams, for example, are routed through the data centre in a way that airstreams coming from IT equipment and flowing to it never mix. The mixing of air currents must be avoided in order to operate an air cooled server centre efficiently. In Chapter 3 more of these techniques are presented. These factors all increase the size of air cooled data centres and require significant engineering cost [Tum10b].

Summing up, the main problems of modern data centres are energy and space efficiency. Both of these sooner or later result in expenses. It seems impossible to build an air cooled data centre which is energy and space efficient at the same time. Thus, an energy efficient data centre cannot be space efficient and vice versa. With the high amount of energy and space being used for cooling in the average data centre, it seems obvious to try and save energy there. At this point, one can think of liquid cooling. On paper, liquids are superior to air in multiple ways. For example, the heat transfer coefficient of liquid compared to air is greater, the heat capacity is 1200 times higher, fluids can be pumped easier, and if mineral oil is used, it can even act as an insulator and corrosion protection [EFV+14; SESA16; Tum10b].

There are two different kinds of liquid cooling, namely indirect/hybrid cooling and immersion cooling. Indirect liquid cooling is used already, preferably in high-performance computing (HPC). HPC data centres are optimized only for a maximum of computing power. So they have high power central processing units (CPUs) which cannot be cooled with air easily. However, in air cooled data centres where kilowatts are still manageable, indirect liquid cooling or hybrid cooling is not good enough to replace air cooling due to high risk of leakage and the cost for tubing and cold plates. Taking together the advantages, it would be desirable to combine the performance of indirect liquid cooling with the installation cost of air cooling without the risk of leakage. This is where immersion cooling comes into play, being the answer to the limitations of air cooling described above.

2 Setup of the Systematic Literature Review

As an approach towards the topic of immersion cooling, a systematic literature review is performed. The purpose is to gather information regarding the computing efficiency and power density at the facility level. Air cooled and liquid immersion cooled data centres are compared on the basis of their performance in computing efficiency and power density. The analysis aims at illustrating an increase in floating point operations per second (FLOPS) per watt and FLOPS per square meter in immersion cooled data centres compared to air cooled data centres. On the basis of these findings, the leading question to answer is to what extend data centre operators can improve energy and space efficiency by switching to immersion cooling.

2.1 Approach

To start this systematic literature review, one compiled broad information about liquid immersion cooling and cooling solutions currently deployed in data centres. This research lead to *Asperitas*, a company pioneering in liquid immersion cooling for server applications. *Asperitas* was happy to provide information about their concept during an online appointment. The company was able to provide insights into problems of immersion cooling and why operators are hesitating to adopt it. Thanks to the collaboration, the advantages and disadvantages were clear, and the research could commence.

The first step of the research included finding databases which are specialized on information communication technology. The database collection used for this paper is at the time of writing available at [Reg21], a website provided to the public by the University of Regensburg.

All databases listed in the collection that were publicly available were consecutively searched. The collection contained 26 databases, and Google Scholar was searched additionally. The key phrase “server immersion cooling“ was used for all databases. Databases that contained little to no information were searched again with a shorter key phrase, namely “immersion cooling“, to find more results. Where databases contained too many papers for a manual review, reviewing was stopped after 10 consecutive articles were proven to be irrelevant for the research. The results were excluded in a predefined order. In order to narrow down the research, papers with irrelevant titles were excluded as well as papers with an irrelevant abstract. To ensure up-to-dateness of information, only papers written in and after 2010 were included.

Table 2.1 lists all databases regarded for this review. Especially the databases of IEEE and the Library of the University of Texas provided helpful papers on the topic of immersion cooling. Additionally, the results of Google Scholar broadened the data and strengthened previous findings. In general the list shows that there is still only little research on this topic. A lot of the consulted databases did not contain any useful information at all.

After gathering the most useful papers, they were categorized regarding their main topics. Table 2.2 depicts the different foci of the papers, namely energy efficiency, power density, reliability of IT equipment immersed into dielectric fluid, performance of different liquids and the impact of heat sinks or micro porous surfaces on cooling performance. Also sustainability was an issue.

2.2 Source Overview

As outlined above, energy efficiency and power density are the most relevant topics. The papers labeled with those topics contain the data needed to calculate the difference between liquid immersion and air cooling. The papers considering reliability give insights into why trust in immersion cooling is a central issue for operators of data centres. The papers containing information about different liquids and heat sink materials/surfaces help understanding the values gathered.

Looking at the publication dates of the consulted papers as illustrated in Figure 2.1, one can notice that especially from 2016 onwards, the interest in the performance of liquid immersion cooling has been on the ascend. It is striking, that in 2020 the amount of publications is significantly lower than one would anticipate considering the years before. Looking at the global situation, the pandemic might serve as a reason for the decline in research done during the mentioned year. It is possible that the research projects could not be executed as planned due to travel restrictions and limitations enforced in the workplace.

All in all, the results are sparse which made acquiring the required information difficult. The numbers required for the calculation are often not only hard to find, but also recorded in a different metric.

Source	**Results**	**Taken into review**	**Key phrase**
arxiv.org	39	1	server immersion cooling
ieeexplore.ieee.org	15	5	server immersion cooling
research-collection.ethz.ch	23	0	server immersion cooling
sciencedirect.com	37	1	immersion cooling
rc.library.uta.edu	78	7	server immersion cooling
scholar.google.de	3100	10	server immersion cooling
aasopenresearch.org	0	0	immersion cooling
bookboon.com	560	0	immersion cooling
stabikat.de	20	0	immersion cooling
dl.gi.de	15	0	immersion cooling
opus.ostfalia.de	5	0	immersion
dspace.mit.edu	407	1(too old)	server immersion cooling
diglib.eg.org	35	0	immersion cooling
ssoar.info	0	0	immersion cooling
infodata-edepot.de	0	0	immersion cooling
ots.at	0	0	immersion cooling
link.springer.com	0	0	immersion cooling
leibniz-publik.de	2	0	immersion cooling
portal.igpublish.com	23	0	immersion cooling
nationalarchives.gov.uk	0	0	immersion cooling
scholarpedia.org	2	0	immersion cooling
spie.org	33	0	immersion cooling
papers.ssrn.com	1	0	immersion cooling
techrxiv.org	13	0	immersion cooling
onlinelibrary.wiley.com	13896	0	server immersion cooling
oerbw.de	0	0	immersion cooling
rzblx10.uni-regensburg.de	0	0	immersion cooling

Table 2.1: Searched databanks and usable results with used key phrase.

Resource	efficiency	density	reliability	fluids	heatspreader	sustainability	year
[360View18]						x	2017
[MMK17]	x	x		x			2017
[KPBB20]	x	x					2020
[QCW+17]		x					2017
[Wei19]	x	x				x	2019
[GBR+14]	x	x				x	2014
[AAH+18]	x	x		x			2018
[EFV+14]	x						2014
[SESA16]			x				2016
[Tum10b]	x	x				x	2010
[WTKD19]	x		x				2019
[CGB17]	x	x			x		2017
[GCKA19]						x	2019
[EEV+17]	x						2017
[KG16]		x					2016
[WV17]	x	x					2017
[Tum10a]	x	x					2010
[CH16]	x	x	x			x	2016
[I W19]	x			x		x	2019
[Sha18]			x				2018
[Pat13]	x						2013
[Sha+16]			x				2016
[SBS+19]	x	x					2019
[RRC+19]			x				2019
[GCC+19]	x	x			x		2019
[Gup+18]			x				2018
[BSG+18]			x				2018

Table 2.2: Resources included in this review and topics they cover.

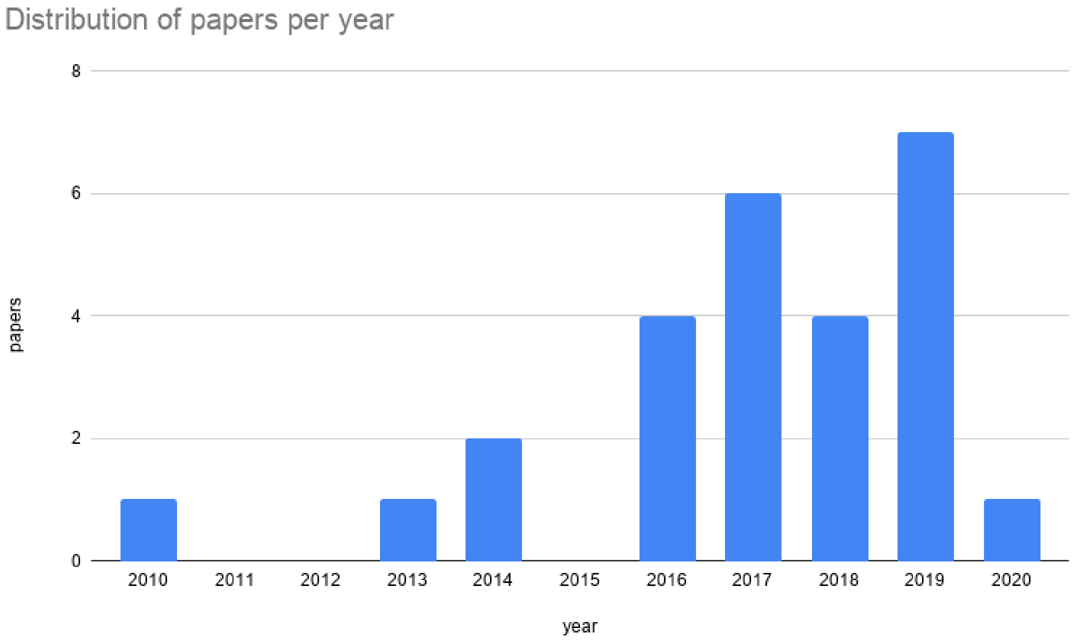

Figure 2.1: Amount of resources between 2010 and 2020

3 Current Cooling Solutions

For the purpose of comparing the different cooling methods, it is necessary to introduce the metric of power usage effectiveness (PUE). PUE is the most common metric for measuring the energy efficiency of data centres. The PUE of a data centre is the value of total facility energy divided by the energy needed for IT equipment. Since the IT energy consumption is included into the total energy, the value of PUE will always be greater or equal to one. The metric was developed and promoted by the Green Grid, a non profit group advocating for sustainability in the technology sector [PUE17]. The majority of values, revealed by the research regarding energy efficiency were available in PUE.

$$(3.1)\quad \text{PUE} = \frac{\text{total energy}}{\text{IT energy}};\ \text{PUE} \geq 1$$

3.1 Air Cooling

Air cooling is at this very moment the dominant cooling solution for data centres [RRC+19; SBS+19]. Typically servers are housed in 19-inch racks. 19-inch (48.26cm) in this case stands for the width of the front panel of each module [The21]. The racks are about 42u (units) high with one unit corresponding to 4.445cm. One unit is the minimum vertical amount a server can take up in a server rack. In addition to their standardized size, it is common for server racks in data centres to stand on a raised floor. Raised floors allow cold air to come from underneath to the front of the servers. To allow airflow out of the raised floor in front of each server, the floor tiles there are perforated [GCC+19]. Cold air is circulating through the perforated tiles up and into the front of each server [GCC+19]. The fans inside the server chassis suck the air inside the servers and then force it over heatsinks covering the processors and other heat producing hardware. Server fans are controlled by the server and change speeds according to the temperature of the critical components. After absorbing the heat of the server components, the now warm air is pushed out by new cool air coming into the server. Through forced convection driven by facility fans the warm air is then guided into the computer room air conditioning (CRAC) where it is cooled back to a temperature low enough to be used for cooling again. The equipment responsible for cooling the air back to a desirable temperature is called CRAC [KG16]. The CRAC is most of the time realized by a chiller. A chiller uses electricity and moderately cooled water to cool water under the ambient temperature via a compressor. Next, the air is forced back under the raised floor, and the cycle starts from the beginning. In air cooled data centres, servers need to be set up facing each other, this is done to prevent one server from sucking in the exhaust air of another server [RRC+19]. The typical architecture and airflow is presented in Figure 3.1. Hot air and liquid are represented in red, cold flow is represented in blue. One can see that the server racks are arranged in a way that hot and cold isles can be separated.

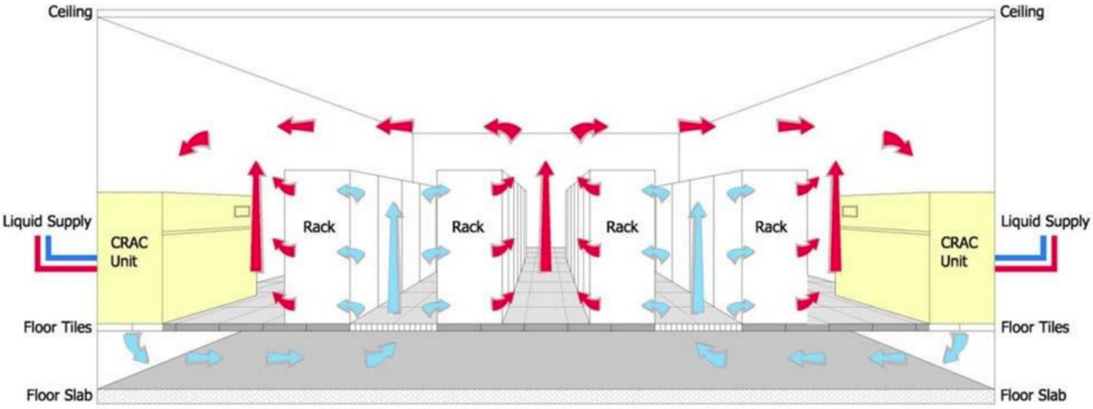

Figure 3.1: Layout of a typical raised floor data centre [RRC+19]

The question, why air cooling is the dominant cooling solution right now, can be answered by looking at the history of data centre cooling. Data centre cooling was not an issue back in the days when server rooms were still small and computing hardware did not exceed 3kW of rack power [FEELHEAT20]. These small amounts of heat were cooled by normal air conditioning and server fans, this was the easiest solution at the time. Being cheap and easy to implement air cooling quickly became the standard for server cooling and was further developed. For many years, power densities of server racks stayed the same and almost all data centre operators relied on air cooling. Only about 10 years ago, rack densities started to rise and air cooling became more challenging and cost intensive [FEELHEAT20]. Even though, the challenges of air cooling were becoming apparent, operators sticked to the standard and improved air cooling as much as possible. The initial cost for liquid cooling methods available was high, and people working on liquid cooled servers would have needed additional training. The cost and training required for employees of hybrid cooling was only acceptable for HPC, where air cooling was not longer possible due to the high power densities [LIQUIDVSAIR20]. Only now, with immersion cooling offering cheaper, more reliable and efficient liquid cooling, the standing of air cooling is threatened.

Air cooling is well standardized and there is plenty of research available for operators to start with. The cooling of low powered servers can be very simple with air cooling and maintenance is easy since no liquids are involved. It is also possible for air cooled data centres to be relatively efficient by implementing proper ducting of air, free cooling and waste heat utilization [Tum10b]. Free cooling improves the energy efficiency of a data centre by using the temperature difference of the air outside to cool down the air inside, which is then used again for cooling. In the case of waste heat utilization the heat generated by servers can be used in some other place, for example the company's office building. The concept of waste heat utilization does not improve the efficiency of the data centre itself. It can however help to keep the cost down by selling the heat to some other company or heating a building owned by the data centre operator.

There are significant flaws to air cooling, mostly because the cooling medium used is air. It is for example hard to utilize free cooling in air cooled data centres because the air temperature needed for air cooling is comparatively low. Which makes free cooling not impossible, but hard to implement, especially in warmer climates. A similar effect is hindering waste heat utilization. The air exhausted from the servers is not warm enough to heat a remote location efficiently. In addition to that, air is hard to transport because a high volume would have to be moved to allow for heating. Another problem of air cooled data centres is the quality of the air used for cooling: Certain particles in the air and moisture can bring the reliability of IT equipment down considerably. Additionally, while air cooling can be done with decent efficiency, e.g. in big hyperscale data centres, oftentimes it is not implemented this way [Sha18]. Studies underline this argument, showing that air cooled data centres need a great amount of their energy for cooling, in a range between 10% and 70% of total power consumption [360View18; CH16; GCKA19; I W19; Sha18]. This is because efficient air cooling in data centres comes at high engineering cost for successful hot and cold air separation and a very low power density. Concluding, one can say that energy efficient computing in air cooled data centres is possible, but often not implemented because of high cost, and effort as well as the high amount of space needed for such a data centre.

Speaking of power density, it is highly limited in air cooled data centres due to the fact that air cooling struggles to transport away such a high amount of heat. With efficient cooling, the densities have to be really low. In the average air cooled data centre efficiency has to be traded for a higher density. A data centre with limited space is either not efficient or sacrifices efficiency for higher computing power [Nic18].

The energy consumption of the cooling system of an air cooled data centre is composed as follows. The most energy is consumed by compressors and chillers, they consume about 41% of the cooling energy [Pat13]. Facility fans that distribute the air in the data centre need 28% of cooling energy [Pat13]. Cooling towers use about 13% and pumps only 4% [Pat13]. Server fans are responsible for only 14% of all cooling power [Pat13]. In some data centres server fans run at such high speeds that the noise they are emitting is approaching safety limitaions [KG16].

The PUE of air cooled data centres differs depending on what type of data centre one is looking at. Data centre operators that invest heavily into the efficiency of their data centre will have PUEs between 1.12 and 1.2 [Ric14]. The average data centre has PUE around 2.0 [Sha18].

The lower heat conductivity of air compared to water or oil demands bigger heatsinks and improved ducting to dissipate the heat generated in modern servers. With ducting and large heat sinks, the fans make air cooling anything but space-efficient [SBS+19]. While air cooling is still the dominant cooling technique used today, operators and especially those of an high-performance data centre (HPDC) seek more efficient cooling solutions like hybrid or immersion cooling.

3.2 Hybrid Cooling

Hybrid cooling describes a cooling technique where parts of the IT equipment producing high amounts of heat are cooled with water while the other parts remain air cooled. Thus, it is a combination of air cooling as described before and liquid cooling. For hybrid cooling, the servers are still mounted into normal server racks exactly like in air cooled data centres. There are still server fans that blow air over the equipment, but the parts that have a potential of becoming really hot are additionally cooled with cold plates [SBS+19]. Such a part is, for example, the CPU. Cold plates are made out of copper or aluminum. They have an inlet and an outlet, this allows for water to run through them and transport the heat away without the risk of a short circuit [SBS+19]. Without cold plates water could not be used for cooling of electric equipment, its electric conductivity is just too high [SBS+19]. In Figure 3.2 a board with four components cooled by four cold plates can be

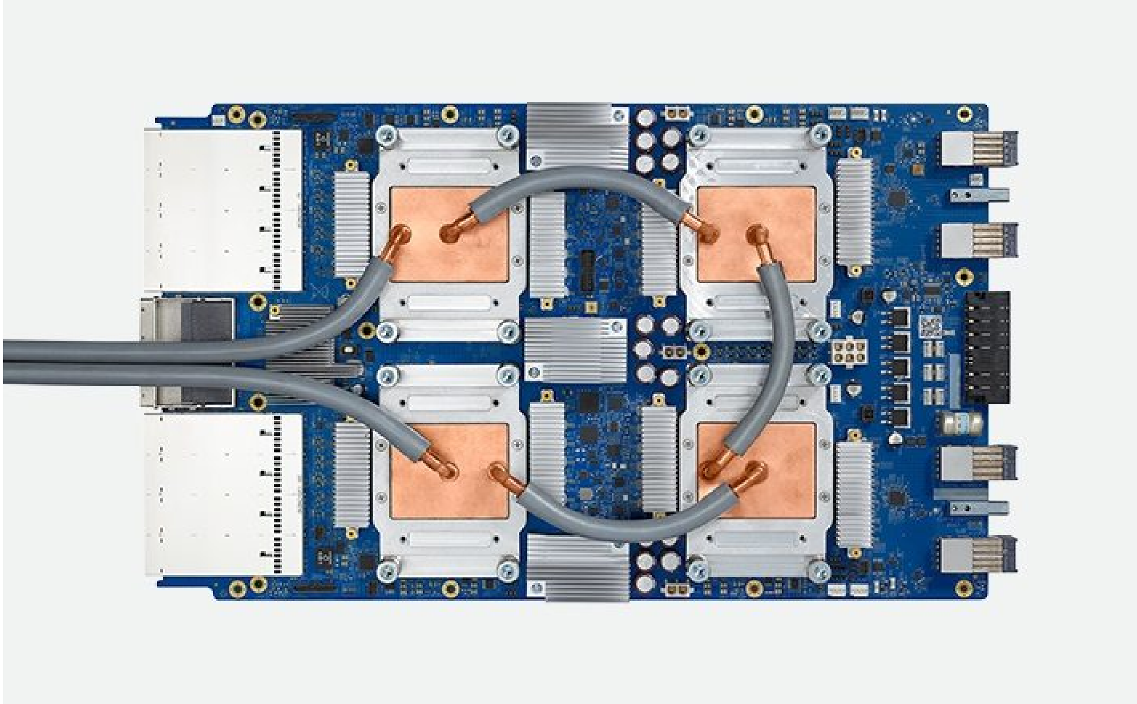

Figure 3.2: Server mainboard with copper cold plates and tubing installed [Jul19]

seen. The cold plates are made out of aluminum in this example and connected to each other using flexible tubing. After absorbing the heat from the IT equipment, the water is pumped out of the system and is then cooled actively with a chiller or passively with a cooling tower. In a cooling tower, coolant is indirectly exposed to the atmosphere and heat is transferred to it. When the water is back to a low enough temperature it can be used again for cooling [RRC+19; Tum10b].

This cooling technique is primarily used in HPDCs. HPDCs are build for maximum computing power and efficiency or cost of hardware are of lower concern. The high performance CPUs of those systems usually have a higher thermal design power, which means that they emit more heat than the average server CPU. The higher amount of heat emitted is the reason those data centres are often cooled with hybrid cooling. Heat sinks would have to be significantly bigger, airflow higher or air colder, to allow for sufficient air cooling in high performance data centres [Wyl18].

Due to the fact that water has a volumetric heat capacity more than 1000 times higher than air, hybrid cooling is much more capable of transporting heat away from server components than air cooling [RRC+19]. Volumetric heat capacity is the amount of heat a specific volume of a matter can absorb. The better heat capacity of hybrid cooling also allows for a better waste heat utilization. Another side effect of the higher heat capacity of water compared to air is that more racks with a much higher density can be cooled using the same coolant volume. Those high power densities achieved with hybrid cooling, would not be possible with air cooling alone. The efficiency of hybrid cooling is also higher compared to air cooling. Hybrid cooled data centres need only about 10% of their energy for water chilling and 5% for pumping, this leaves up 85% for IT equipment [MMK17]. A PUE below 1.1 should be easily possible with indirect water cooling [Wyl18].

Although hybrid cooling is superior in efficiency and more capable than air cooling in transferring heat away from critical components, there are downsides that prevent it from being the dominant cooling solution. One disadvantage of hybrid cooling is the remaining need for air cooling. Because of that, a lot of problems of air cooling remain relevant also for hybrid cooling. The quality of the air used for cooling is still an issue, so corrosion can still occur. Similarly to air cooling, the parts that are not covered by a cold plate can still get too hot. In addition, servers need to be changed in hybrid cooled data centres as well as in any other data centre, which is difficult because they are integrated into the cooling water circuit. To quickly connect and disconnect a server to the water loop special parts are needed, namely quick disconnects. With quick disconnects in place, hot swapping of servers is possible. Hot swapping describes the process of replacing one server inside a rack without shutting down the other ones. However, quick disconnects present potential points of failure that can cause water leakage, which can cause damage to one or multiple servers. In such a system, the connecting parts are always the most sensitive ones. In Figure 3.2 possible points of failure can be seen, if there is any tolerance of the material or if it is not installed correctly, water can drip out between the tubing and the cold plate. Lastly, costs of implementing hybrid cooling are an obvious disadvantage: Since the realization of hybrid cooling involves air cooling and liquid cooling equipment, the initial cost is much higher, compared to air cooling alone. For example, cold plates, tubing and fittings are expensive [Tum10b].

3.3 Immersion Cooling

Liquid immersion cooling is trying to achieve what cannot be achieved by air or hybrid cooling, namely space and energy efficient data centres with servers that are easy to maintain. In immersion cooled data centres, components are fully immersed into a dielectric fluid [SBS+19]. A dielectric fluid conducts heat and does not conduct electricity at all, but instead acts as an insulator. The common dielectric fluids have a heat capacity by volume about 1300 times higher than air [SBS+19], this means one liter of liquid can transport as much heat as about 1300 liters of air. Most liquids used in immersion cooling are white mineral oil, electric cooling liquids and other oils [SBS+19]. The heat of all the components is fully removed by the liquid, completely eliminating the need of air cooling. There are two types of liquid immersion cooling: single-phase and two-phase immersion cooling, both will be explained below.

In single-phase immersion cooling the liquid will stay in liquid form the entire time, there is no phase change occurring. Heat emitting components are cooled by the fluid flowing over them, and the heated fluid is transported away. The circulation of the fluid is driven by a pump or by natural convection. In natural convection driven systems, the heated fluid floats to the top of the tank because it has a higher volume than colder fluid. It then flows, due to more fluid rising to the top, to the side of the tank, where it is cooled by a heat exchanger connected to an external cooling loop. The cooled liquid is then shrinking again and pulled back to the bottom of the tank by gravity. At the bottom it then is driven back under a mainboard and circulates back to heat emitting parts where it is heated up again. In a pump driven system, the convection is driven by a pump. The pump is forcing the liquid through an inlet inside the tank and out through an outlet on the opposite side. The liquid is then cooled by flowing through a coolant-to-water heat exchanger [RRC+19]. An increase in pump capacity is also increasing the cooling capacity. This process is visualized in Figure 3.3.

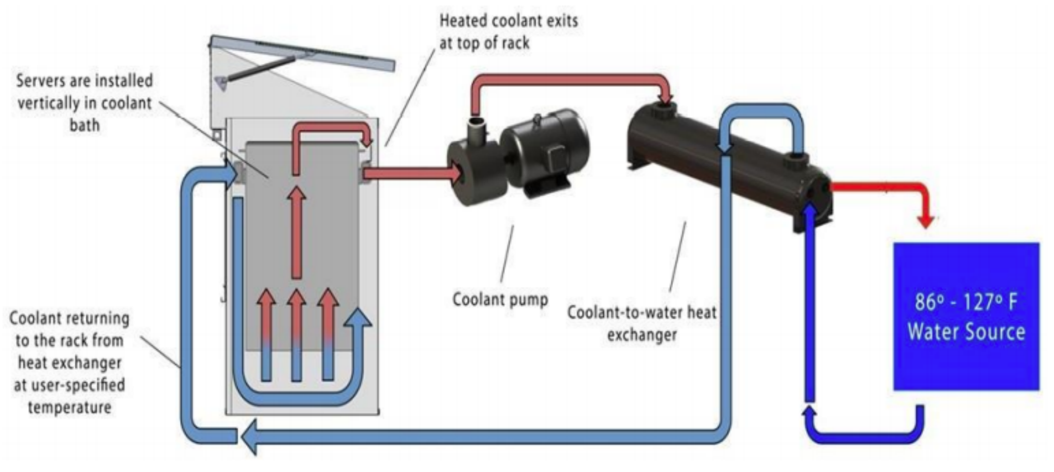
Heated coolant exits at top of rack
Servers are installed vertically in coolant bath
Coolant pump
Coolant-to-water heat exchanger
86° - 127° F Water Source
Coolant returning to the rack from heat exchanger at user-specified temperature

Figure 3.3: Pump driven liquid immersion cooling loop [RRC+19].

Two-phase liquid immersion cooling works different because the coolant will change phase whenever it gets in contact with a heat producing component [RRC+19]. In order to avoid damaging the components, the boiling point of the coolant has to be lower than the critical temperature of the to be cooled parts. The following process is illustrated in Figure 3.4. When evaporating on a hot component, the gas will float to the top of the tank and will make room for new colder coolant to absorb the component's heat. A condenser is located inside the tank above the liquid. Cooling water is flowing through a condenser to transport the heat away. The coolant condenses there and will fall back into the tank where it can absorb heat again[RRC+19].

The biggest downside of immersion cooling is, that it is not yet established. There is very little data regarding its long term reliability and because of that, data centre operators are hesitant to implement it into their data centres. Companies and universities have been researching reliability, but only for a few years now [BSG+18; CH16; Gup+18; RRC+19; SESA16; Sha+16; Sha18; WTKD19]. Concerns have been raised that data centre employees require additional training regarding the handling of cooling fluid. Spills need to be wiped instantly. In addition, servers need to be pulled out of the system vertically, which presents a problem for the employees because servers can be become quite heavy and will be soaked in immersion cooling fluid after extraction. This makes the extraction of server more difficult than in air cooled systems, where servers can just be pulled out horizontally [ASPERITAS21]. In two-phase liquid immersion cooling the tanks have to be sealed or loss of fluid needs to be prevented in order to keep costs down.

However, the advantages of immersion clearly outweigh the disadvantages. With liquid immersion cooling all the heat generated by the server is captured and transported away from it; this allows for the best waste heat utilization of all three cooling methods compared [Tum10a]. It is possible to cool an immersion cooled server with inlet temperatures of up to 45°, such a high temperature allows for a exceptionally good usage of free cooling even in hotter climates [SBS+19]. Since temperature distribution is very stable in liquid cooled servers, overheating of parts that would not be covered by heatsinks or cold plates in air or hybrid cooled data centres is prevented [WV17]. The fluid used for single-phase immersion cooling is more affordable compared to cold plates and the

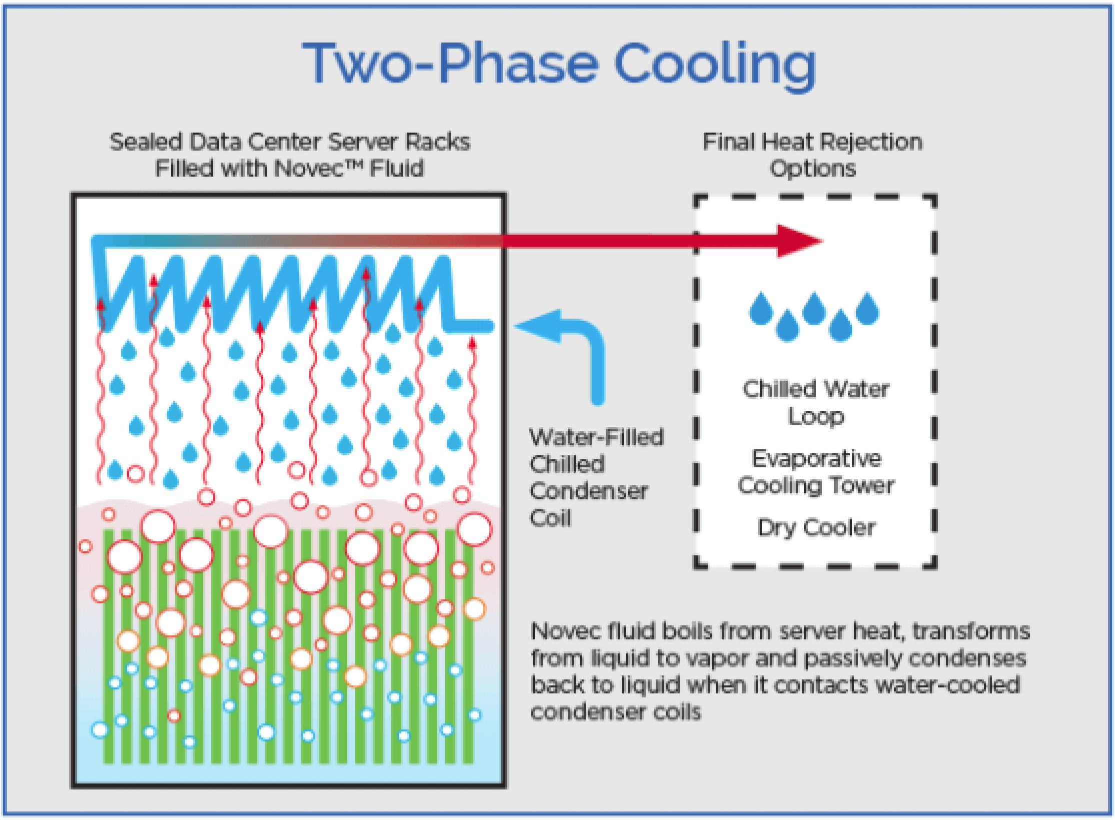
Two-Phase Cooling
Sealed Data Center Server Racks Filled with Novec™ Fluid
Final Heat Rejection Options
Water-Filled Chilled Condenser Coil
Chilled Water Loop
Evaporative Cooling Tower
Dry Cooler
Novec fluid boils from server heat, transforms from liquid to vapor and passively condenses back to liquid when it contacts water-cooled condenser coils

Figure 3.4: Two-phase liquid immersion cooling [Dhr19].

energy needed for air cooling [Tum10b]. Server chassis fans that are used in air cooled data centres cost about 0.05 USD/kW per **hour**, in contrast immersion fluid costs about one to 3 USD/kW per **year**. Furthermore, liquid immersion cooling allows for the usage of very little mechanical parts: Only a pump for the circulation of the cooling water is required [Pat13]. The pump is also the only part inside the system that needs to be upgraded to allow for a higher power density [WV17]. The reduction of mechanical parts improves the reliability of the cooling system [Pat13]. Should the pump fail, the thermal capacity of the liquid itself is enough to allow for further operation of the server until it can be replaced [ASPERITAS21].

Switching to immersion cooling helps data centre operators to increase the power density of their data centre. The higher cooling capacity of immersion cooling allows for baths with more than twice the power density compared to air cooled racks [MMK17]. 100ccm of fluid are enough to cool 1kW of IT equipment energy [Tum10a; Tum10b]. Additionally, server baths are insulated and can be put right next to each other because no air flow is needed [ASPERITAS21]. There is also much less space required for cooling hardware such as chillers and ventilation equipment. By switching to immersion cooling, up to 95% of cooling energy can be saved [GBR+14]. The overall power consumption of immersion cooled data centres is close to perfection with PUE values around 1.02 [AAH+18; CGB17; EFV+14; GBR+14; MMK17; Sha18].

4 Advantages of Immersion Cooling

As outlined in Chapter 3, air cooling is not without flaws, and immersion cooling has the potential of being the superior cooling method. The possible efficiency gains are shown in this section. Measurable increases in profitability are important for data centre operators. Data centre operators are mostly profit oriented. This means that data centres are implemented, optimizing for maximum computing output using the least amount of investment. While some operators make decisions with sustainability in mind, the majority of them, is likely to choose the most cost efficient solution.

4.1 Areas of Improvement

Viewed from the perspective of a data centre operator, there are two essential areas of improvement in a data centre, namely computing efficiency and computing density. A better computing efficiency allows for a higher amount of computing power using the same amount of energy or the same amount of computing power using less energy. One can see, there is a direct link between efficiency and profit. Higher computing density allows operators of a data centre to use their space more efficiently, meaning they can increase their computing power within their existing building. This saves cost and makes it desirable for data centre operators to switch cooling methods within an existing data centre. Furthermore, reliability is a concern. Servers, that fail more often, increase the maintenance cost of a data centre and impact the reputation negatively. It is not enough for a cooling method to be superior in computing efficiency and density, reliability is also important.

4.2 Efficiency

From a computing perspective, it is desirable to have a value resembling the computing capability in relation to the energy needed. Computing performance is traditionally measured by running a set amount of operations on a device and track the time until they are complete, this is called a benchmark. The number of floating point operations needed to complete the benchmark can then be divided by the seconds needed to complete it. This equation is illustrated in Equation (4.1). The result of the calculation is in the unit of floating point operations per second (FLOPS).

$$(4.1)\quad \text{computing performance} = \frac{\text{floating point operations in benchmark}}{\text{time to complete benchmark}}\text{FLOPS}$$

To compare the efficiency of different systems or facilities, one first has to divide the benchmark result calculated in Equation (4.1) by the peak power of the system. This would then result in the metric of floating point operations per second per watt (FLOPS/W), a metric capable of showing

Resource	reported PUE: air	reported PUE: immersion	type of data
[MMK17]	1.1	<1.04	air: minimal PUE possible immersion: experimentally proven, not optimized
[GBR+14]	1.3		air: utilizing full-time free cooling
[AAH+18]		1.02	immersion: two-phase cooling
[EFV+14]		1.03 - 1.17	immersion: experimental, not focused on efficiency
[CGB17]		1.02 - 1.03	immersion: open-bath, productive implementation
[Sha18]	1.7 - 2.9	1.02 - 1.03	air: average, North-America
[Ric14]	1.12, 1.18		air: state-of-the-art actors
[Amb13]	2.61		air: average PUE Singapore
	2.2		air: average PUE Japan
	2.42		air: average PUE Hong Kong
	2.25		air: average PUE Australia

Table 4.1: PUE values across different reference sources.

the performance in relation to the energy needed. A variable η is introduced in Equation (4.2) to describe various efficiencies from now on. In case of Equation (4.2) η resembles the overall computing efficiency.

$$(4.2)\quad \eta = \tfrac{\text{computing performance}}{\text{total equipment energy}} \text{FLOPS/W}$$

Data centre efficiency is traditionally measured with PUE. Therefore values found during research are mostly in form of power usage effectiveness (PUE). This metric is suitable for comparing efficiency of cooling and other facility equipment, but it gives no direct information about the computing efficiency of a data centre. To calculate differences in efficiency of liquid immersion cooled and air cooled data centers, PUE values have to be converted.

Although PUE was mentioned before, it is essential to explain the term in further detail now. PUE is the quotient of the total facility power divided by the power consumed by IT equipment, this is shown in Equation (3.1). The inverse of PUE, IT equipment power divided by total facility power as in Equation (4.3), represents the part of energy used for IT equipment. A data centre with a PUE of 1.5 for example uses two thirds of its energy for IT equipment. The inverse of the PUE can be used to calculate the floating point operations per second per watt; this is needed to estimate a gain of efficiency by switching to immersion cooling.

$$(4.3)\quad \text{proportion of IT energy}=\frac{1}{\text{PUE}} = \frac{\text{IT equipment energy}}{\text{total energy}}$$

Table 4.1 lists all the data gathered about PUEs of air and liquid immersion cooled data centres. In air cooled data centres PUEs range from 1.1 to 2.9 [MMK17; Sha18]. Values close to 1.1 can only be achieved by hyper scale data centre facilities, which are especially optimized for efficient cooling. For example, state-of-the-art air cooled data centres of *Google* have a full year PUE of 1.12 [Ric14]. This means that *Google's* servers use 89% of their energy for computing. However, high efficient data centres with state-of-the-art design make up for only for a small amount of the energy consumed by data centres worldwide [Nic18]. Average air cooled data centres have a much higher PUE compared to the most efficient state-of-the-art data centres. Two studies conducted in 2013 and 2014 report an average PUE of 1.7 and 2.9, respectively [Sha18]. Another source shows PUEs between 2.2 and 2.61 for the average data centre in Singapore, Japan, Hong Kong and Australia [Amb13]. Thus, the values in Table 4.1 underline clearly the significant differences in hyperscale data centres and the average data centre.

The resources containing information about the power usage effectiveness of immersion all contained similar information, ranging from 1.02 to 1.04 [AAH+18; MMK17]. The sources only vary by 0.02 PUE. A PUE of 1.02 seems to be the sweet spot for immersion cooling. Table 4.1 illustrates that the considered resources agree on the fact that PUE values around 1.02 can and will be achieved in immersion cooling. With a PUE of 1.02, about 98% of the energy consumed by the data centre go into the IT equipment. Only the remaining 2% are used for cooling. This is close to perfection and can hardly be further optimized. One of the resources shows a PUE significantly higher than the others when cooling was realized with liquid immersion cooling. The maximum PUE there is 1.17 [EFV+14]. This is because this value describes a PUE which was calculated in an experiment aiming at maximum cooling without efficiency in mind [EFV+14]. In that experiment specifically, the IT components were not cooled with maximum efficiency. This explains the higher value. For further illustration, Figure 4.1 adapts the numbers of Table 4.1 and presents them graphically.

In order to compare immersion to the average data centres, a PUE for those needs to be assumed. Based on the values in Table 4.1 a PUE of 2.0 is assumed for the average air cooled data centre. Although this value is chosen defensively, the efficiency gain by switching to immersion cooling is still significant as will be shown later.

The computing performance and the power of IT equipment are relevant for the calculation of FLOPS/W, however, they do not influence the overall efficiency difference when the cooling solution is improved. This is because a change of the cooling system is neither influencing the computing performance nor the IT equipment energy. The only value changing by replacing the cooling system is the total equipment energy.

Only knowing the changes of the PUE, the missing value, total equipment energy, can be calculated with Equation (3.1). This is illustrated in Equation (4.4)

$$\text{(4.4)}\quad \text{total equipment energy} = \text{PUE} \cdot \text{IT equipment energy}$$

To calculate the improvement in FLOPS/W, one can now combine the knowledge about PUE and FLOPS/W in Equation (4.5). The constant values are the IT energy and the benchmark. The thermal design power will not change by changing the cooling solution. The benchmark will still require the same amount of operations and because the IT equipment, first of all the CPU, does not need to be changed, the time to complete the benchmark will also be the same. Knowing the different PUEs

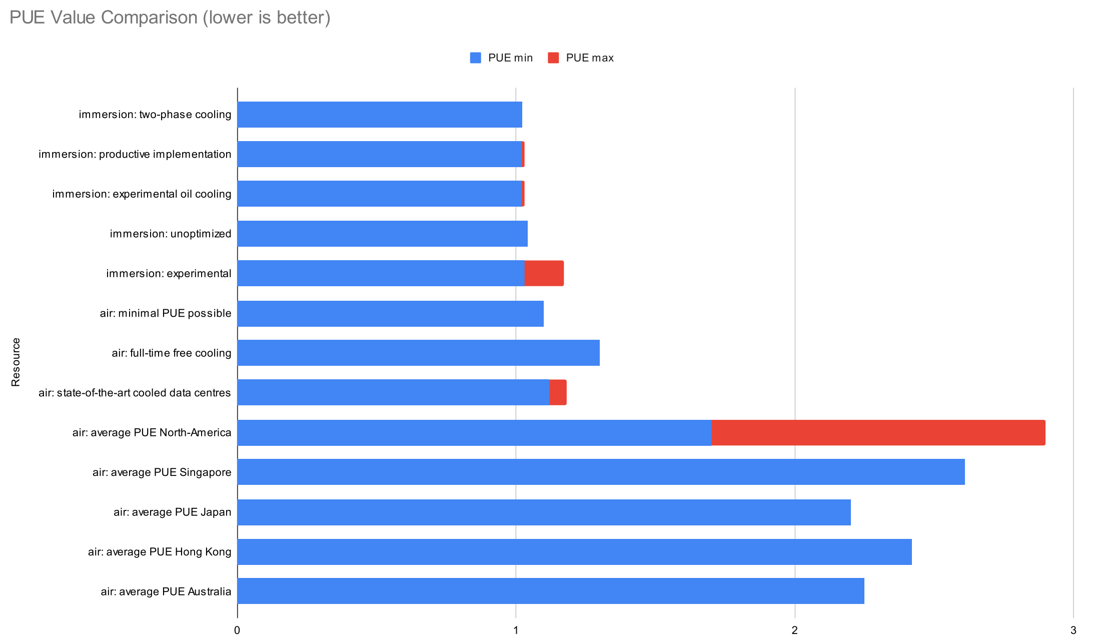

Figure 4.1: Efficiency differences between air and liquid cooling.

for efficient air cooled data centres, not so efficient air cooled data centres and for immersion cooled ones, these PUE values can be used for the calculation in Equation (4.5). The PUE can be used to calculate the total facility energy with just the IT equipment energy given. IT equipment energy multiplied with PUE equals the total facility energy as in Equation (4.4).

$$\eta = \frac{\frac{b_o}{b_t}}{e_{IT} \cdot \mathrm{PUE}} \tag{4.5}$$

IT equipment energy = e_{IT}, benchmark operations = b_o, benchmark time = b_t

The fraction can then be disassembled into the following.

$$\eta = \frac{\frac{b_o}{b_t}}{e_{IT}} \cdot \frac{1}{\mathrm{PUE}} \tag{4.6}$$

The values on the left side of the multiplication are all a constant. This means, computing power in relation to energy used depends directly on the fraction of power used for IT equipment. Furthermore, percentage differences in FLOPS/W can be calculated having only the PUE.

Knowing this, the performance increase to be expected by switching to immersion cooling is the following: For the calculation the inverse can be taken of both immersion cooling and another cooling solution, for example $\frac{1}{1.12}$, the PUE of best-practice air cooling, divided by $\frac{1}{1.02}$, the PUE

	Standard air cooling	**State of the art air cooling**	**Immersion cooling**
Standard air cooling	100%	178.57%	196.08%
State of the art air cooling	56%	100%	109.8%
Immersion cooling	51%	91.07%	100%

Table 4.2: Efficiency by switching from the cooling technique on the left to the one in the first row.

of immersion cooling, equals an efficiency of about 109.8% [AAH+18; Ric14]. The same can be done by dividing the bigger PUE with the smaller one, this is shown in Equation (4.7). Dividing 2.0, the PUE of an average north American data centre, by the PUE of immersion cooling equals an efficiency of about 196%. The respective values are shown again in Table 4.2. As expected, switching from state-of-the-art air cooling to immersion cooling results in an efficiency increase of about 10%. While this is already a significant increase that would result in much lower energy expenses for a data centre operator, the increase caused by a switch from average air cooling, to immersion cooling is almost ten times as significant. Such a conversion would almost shorten the energy used by half. From the perspective of a data centre operator, these values are impressive.

$$\frac{\frac{\frac{b_o}{b_t}}{e_{IT}} \cdot \frac{1}{\text{PUE}_2}}{\frac{\frac{b_o}{b_t}}{e_{IT}} \cdot \frac{1}{\text{PUE}_1}} = \frac{\frac{1}{PUE_2}}{\frac{1}{PUE_1}} = \frac{PUE_1}{PUE_2} \tag{4.7}$$

Although the performance improvement is not dependent on the processor used, a sample processor is used in the following calculation for illustration purposes. Instead of the percentage improvement as illustrated in Table 4.2, the implementation of a sample processor allows for exact numbers. One of the most powerful server CPUs currently on the market is the AMD EPYC 7742 [AMD21]. The processor runs with 225-watt maximum power consumption and achieves around 3.48 TeraFLOPS [Tif19]. According to that, the processor can calculate 15.6 GigaFLOPS per watt, this is calculated in Equation (4.8).

$$\eta_{EPYC7742} = \frac{3.48 \text{ TFLOPS}}{225W} = \frac{3.48}{225}\text{TFLOPS/W} \approx 15.6 \text{ GFLOPS/W} \tag{4.8}$$

This is the peak performance divided by the wattage needed at peak performance. The power used by the IT equipment is not entirely used for the processor, which is calculating the FLOPS, therefore one needs to know the proportion of power consumed by the EPYC 7742 in relation to all IT equipment. This relation is shown in Equation (4.9).

$$\eta_{IT equipment} = \frac{3.48 \text{ TFLOPS}}{225W} \cdot \text{proportion of power for processor} \tag{4.9}$$

cooling method	**GFLOPS/W**
standard air cooling	3.87
state-of-the-art air cooling	6.9
immersion cooling	7.58

Table 4.3: Computing power per watt using the different cooling methods

Unfortunately, there is no data available at the moment of writing that shows this relation for the used processor. In order to complete the calculation 50% can be assumed [360View18; I W19].

$$(4.10)\quad \eta_{ITequipment} = \frac{3.48\text{TFLOPS}}{225W} \cdot 0.5 = \frac{3.48}{450}\ \text{TFLOPS/W} \approx 7.73\ \text{GFLOPS/W}$$

As shown in Equation (4.11) the FLOPS/W for the data centre can then be calculated using the part of energy used for IT equipment, namely the inverse of the PUE.

$$(4.11)\quad \eta_{coolingmethod} = \frac{3.48\ \text{TFLOPS}}{225W} \cdot 0.5 \cdot \frac{1}{PUE}$$

In the following, Equation (4.11) is used to compute the floating point operations per watt. This allows for a comparison of immersion cooling, best-practice air cooling and average air cooling using a uniform value.

For immersion cooling, 98% of data centre power go into IT equipment.

$$(4.12)\quad \eta_{immersion} = \frac{3.48\ \text{TFLOPS}}{225W} \cdot 0.5 \cdot \frac{50}{51} = \frac{29}{3825}\ \text{TFLOPS/W} \approx 7.58\ \text{GFLOPS/W}$$

Of this 98%, 50% go into the processor, and this equals a compute performance per watt of 7.58 GFLOPS per watt . This value is calculated by multiplying the inverse of PUE with the fraction of IT power used for the CPU and the floating point operations per watt possible with the processor.

$$(4.13)\quad \eta_{best-practice-air-cooling} = \frac{3.48\ \text{TFLOPS}}{225W} \cdot 0.5 \cdot \frac{25}{28} = \frac{29}{3825}\ \text{TFLOPS/W} \approx 6.9\ \text{GFLOPS/W}$$

For state-of-the-art air cooling this value is 6.9 GFLOPS/W.

$$(4.14)\quad \eta_{average-air-cooling} \frac{3.48\ \text{TFLOPS}}{225W} \cdot 0.5 \cdot 0.5 = \frac{29}{7500}\ \text{TFLOPS/W} \approx 3.87\ \text{GFLOPS/W}$$

The results presented in Table 4.3 show that excessive amounts of energy can be saved in every air cooled data centre. Looking at the state-of-the-art air cooled hyperscale data centres, almost 10% less energy consumption with the same computing performance makes a big difference. The decrease of cost alone should be enough for operators of hyper scale data centres to consider immersion cooling in future projects. In the average data centre however, the switch to immersion cooling offers even more possibilities. Operators could decrease their power consumption by about 50% without any decrease in computing performance. These cost reductions could make even retrofitting already existing data centres with immersion cooling viable.

4.3 Density

Data centre computing density is defined as the amount of computing power that a centre can offer in relation to its size. Some big data centres trade all their computing density for efficiency and reliability [Ric14]. This is done to increase revenue, where space is not an issue. The cooling units of those centres are bigger than they need to be to allow for more efficient cooling. Power densities on the rack level have to be low in efficient data centres, otherwise either fan speeds need to be increased or air temperature needs to be lowered, to allow for sufficient cooling. Both of these measures have a lower energy efficiency as a consequence.

The metric used for computing density has to be FLOPS/m^3, with everything counted in from the server over the power management up to the cooling equipment. Oftentimes, only rack densities are considered, but when switching to immersion cooling, not only the rack level power density increases. Servers can also stand closer to each other because there is no space needed for airflow. Using immersion cooling and air cooling as an example, a comparison of rack power alone would make the advantage of immersion cooling look significantly smaller than it actually is.

During research it was evident that the main focus of immersion cooling as a cooling technique is energy efficiency at the moment. Some papers, however, provide information about the density of liquid immersion cooled data centres too. This research data is presented in Table 4.4. One resource gave information about the conversion of existing air cooled racks into immersion cooled bathtubs, the amount of bathtubs needed is about $\frac{1}{3}$ of the server racks, this can be seen in Figure 4.2 [MMK17]. Overall an 300% increase in power density is plausible according to this resource [MMK17]. Another resource claims that even more is possible, namely an increase of between 2.5 and ten times the density of air cooled data centres [Tum10b]. According to the resource, this increase is possible in both rack level density and facility level [Tum10b]. If density limitations are fully exhausted, a minimum of 100cc of fluid is needed to cool 1kW of IT energy [Tum10a; Tum10b]. Although, rack density as described before is not the important factor, immersion cooling is exceptionally efficient there. According to the research, rack level power densities of air cooled data centres are between $0.0022\frac{kW}{l}$ [Cla13] and $0.028\frac{kW}{l}$ [KG16], those of immersion cooled data centres are between $0.045\frac{kW}{l}$ [MMK17] and $0.23\frac{kW}{l}$ [GBR+14]. One resource even suggested that $4\frac{kW}{l}$ are possible with enough coolant flow, provided a compact enough IT design [Tum10a]. All research values are again presented in Table 4.4.

The increase in power density when swapping from air to liquid immersion cooling as suggested in the research is huge. Computing power could be tripled in existing data centres without the need to extend facilities, this would allow for huge cost savings. This increase in computing power density is caused by the higher rack densities as well as other factors. Servers can stand right next to each

Resource	air cooling	liquid immersion cooling	normalized to kW/l	density type
[MMK17]		14 kW/l	0.045 kW/l	rack
		density increases by factor 3 with immersion cooling		facility
[GBR+14]		400 W / 1.71 l	0.23	rack
[AAH+18]		250 kW / rack	0.14 kW/l	rack
[Tum10b]		0.1 l fluid per kW		fluid
		power density increases by factor 2.5 to 10 with immersion		facility
[WV17]		rack density scales with pump power		rack
[Tum10a]		4kW/l	4kW/l	rack
		0.1 l fluid per kW		fluid
		cooling not the limiting factor anymore		rack
[KG16]	33-50 kW/rack		0.018 - 0.028 kW/l	rack
[Smo19]	40 kW/rack		0.022 kW/l	rack

Table 4.4: Power densities presented in different resources.

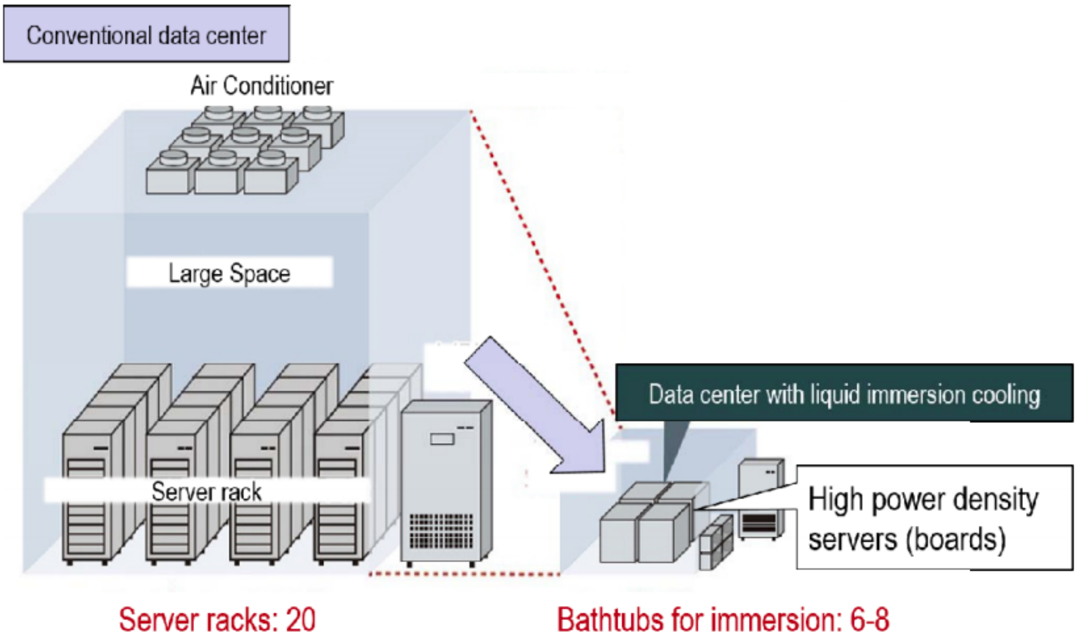
Conventional data center
Air Conditioner
Large Space
Server rack
Data center with liquid immersion cooling
High power density servers (boards)
Server racks: 20
Bathtubs for immersion: 6-8

Figure 4.2: Size comparison of two data centres with similar computing power [MMK17]

other, the only limitation being the accessibility by the service crew. Compared to air cooled racks, that need space in the front and back of them, this helps reducing the size of the facility. There is significantly less cooling equipment needed, this space can be used for servers instead. Furthermore, there is only piping needed, no raised floors or air vents, this is much more space efficient than air cooled cooling solutions [MMK17].

4.4 Interpretation of the Results

Data visualizing the efficiency of data centres is currently measured in PUE. PUE as a metric is not good enough at keeping track of data centre efficiency. Instead, it works well at illustrating the efficiency of non IT equipment in a data centre, for example cooling. It is not suited for estimating the overall efficiency of a data centre and it is even less qualified for the illustration of IT equipment efficiency. In some cases where the efficiency of the IT equipment itself is not optimal, the PUE of the less efficient system will be better than the one with the more efficient IT equipment. The metric of PUE gives insights about the relation between energy consumed by IT equipment and non IT equipment. Thus, it cannot provide information about the overall effciency of data centres because it does not directly account for computing power of the IT equipment. For this review, all values were converted to FLOPS/W. This was done to illustrate the impact immersion cooling can have on computing efficiency. Eventual inaccuracies caused by the conversion from PUE to FLOPS/W do not influence the results of this review. All results are affected in a similar manner and the percentage difference is the same. When liquid immersion cooled data centres will be established, data centre operators should measure efficiency using FLOPS/W.

Looking at the results presented in Table 4.3, one can see that the overall efficiency of data centres can be doubled by switching to immersion cooling. The increase of computing power is illustrated by the concrete values calculated above, namely 7.58 GFLOPS/W for immersion cooling, 6.9 GFLOPS/W for state-of-the-art air cooling and 3.87 GFLOPS/W for the average air cooled data centre. With an efficiency of 7.58 GFLOPS/W for immersion cooling compared to 3.87 GFLOPS/W for conventional air cooling, the efficiency of immersion cooling is about 96% higher. Compared to 6.9 GFLOPS/W for state-of-the-art air cooling, the efficiency of immersion cooling is still about 10% higher, which allows for 10% more server power using the same amount of energy. For both types of air cooling that were considered in this paper, a switch to immersion cooling would result in an significant increase in efficiency.

In order to improve the accuracy of the calculation, a first important step would be to measure the efficiency with the help of a standardized metric, for example FLOPS/W. The metric FLOPS/W allows for a more exact interpretation of efficiency accounting for IT equipment efficiency too. For liquid immersion cooling, a metric also accounting for IT equipment efficiency is necessary because energy used by cooling equipment is already insignificant. However, to increase data centre efficiency even further, the composition of the power consumption of different parts of the data centre need to be broken down and separately taken into account. In this paper, the values for the IT equipment efficiency are estimated. To further improve the accuracy of the calculation, those values need to be experimentally shown. This will help to identify inefficient components of a data centre that can further be improved. The calculation of this paper succeeds in presenting the possible efficiency improvement by switching to immersion cooling. This result can be seen as a point of reference, but it cannot predict the exact efficiency gain of a particular data centre due to the fact that they differ in numerous components such as the processors.

While research on immersion cooling is mostly targeting efficiency, the aspect of computing density should not be overlooked. The increase of computing power in a specific volume is even greater than the efficiency improvement. The increased computing density of immersion cooling allows expansion inside otherwise not expandable facilities. Looking at Table 4.4, the most conservative measurements of immersion cooling are about double the density in kW/l on rack level compared to the maximum possible in air cooled data centres. Resources estimating 4kW/l illustrate what is potentially possible with IT equipment optimized for liquid immersion cooling [Tum10a]. Oftentimes, power densities are presented in kW per rack for air cooled data centres. With immersion, there is no typical rack, therefore, values need to be converted to kW/l. This metric is most meaningful, when the whole facility is considered, as air cooling can achieve higher densities by scaling racks vertically.

With immersion cooling allowing for such high efficiency and density it offers perspectives for data centre operators looking to extend their computing power. In existing facilities, more computing power is supported without sacrificing efficiency. This should make immersion cooling an alternative to building a whole new data centre and instead upgrading step by step to immersion cooling. With immersion cooling, global data centre computing power can expand even further without the use of additional space and energy. With the focus on sustainability of customers and companies, immersion cooling has the potential to be a significant contributor to the success for data centre operators.

5 Immersion Cooling in Practice

Although air cooling is still the dominant cooling solution today, there are companies that have been offering immersion cooled server systems already. Following this direction, one of the biggest players in computing, *Microsoft*, has announced their first liquid immersion cooled data centre not long ago. Three different companies trusting in immersion cooling will be presented now.

Asperitas, for example, is a dutch company located in Amsterdam that offers complete liquid immersion cooling solutions to its customers. “We believe in order to make global data centre industry growth sustainable, it’s necessary to integrate the data centre industry within energy transition, therefore we take the lead as immersion cooling specialists“ [ASPERITAS21]. This citation addresses three main aspects when it comes to current and future cooling solutions in practice. The quotation mentions the importance of sustainability, a topic of rapidly growing relevance. In order to achieve sustainability as whole, it is crucial to also implement new and more efficient cooling systems in data centres. This is where *Asperitas* steps in by developing and distributing cooling solutions based on immersion cooling. Their server enclosures operate with single-phase immersion cooling liquids and natural convection, therefore there are no mechanical parts inside their system. As mentioned before, the lack of mechanical parts ensures a high reliability of the cooling system. Immersion cooled servers of *Asperitas* are insulated; this is done to capture all heat produced by the server in the immersion cooling fluid and to allow for maximum waste heat utilization. They face the problem of sustainable computing by utilizing a synthetic oil developed in cooperation with *Shell*, one of the well-known mineral oil companies. The oil used in their systems is fully recyclable and significantly less expensive than other fluids used for immersion cooling. *Asperitas* offer a service trolley, helping with lifting server boards out of the immersion cooling bath. By doing so, they tackle the common problem of lifting server boards out of the bath in immersion cooled systems. The trolley is also equipped with tools for cleaning and filtering liquid. Each of their AIC24 enclosures can contain up to 48 servers or 288 GPUs at a footprint of only 60cm x 120cm [ASPERITAS21]. Looking at their system as a whole, *Asperitas* addresses the main problems linked to the implementation of immersion cooled data centres.

Another company offering liquid immersion cooling enclosures in various sizes is *Submer*. The immersion cooling solutions of *Submer* are ranging from a cabinet-sized enclosure, called microPod, up to the megaPod, which is set up inside a shipping container. The small solution, the microPod, is capable of cooling 5kW of components even in direct sunlight which makes it suitable for companies who want to cool their in-house equipment efficiently. The megaPods, on the other hand, are targeted for companies with a need for higher computing power. They can be put in almost any place since there is only electricity and network connection needed. For example, it would be possible to install a megaPod onto or near a building which then supplies the building with heat. Therefore, their products are excellent for waste heat utilization. Their micro- and megaPods include equipment for warm water cooling, which makes them all-in-one cooling solutions with no further installation required. Appart from that, the company is also offering traditional immersion cooled server bathtubs that require installation and need to be connected to the cooling water circuit of the building. All of their products use single-phase immersion cooling and their synthetic fluid can be filtered and used for more than 20 years. Additionally, it is biodegradable and has zero global warming potential, an important feature for companies that keep track of their environmental impact [Sub21]. Summing up, the cooling solutions of *Submer* are lucrative for companies that are interested in cooling solutions that can be set up easily in almost any location.

Only recently, a big player in cloud computing joined the pioneers of liquid immersion cooling. *Microsoft* just announced that it is testing two-phase liquid immersion cooling in one of their data centres in Washington, USA. The data centre is used for *Microsoft Teams*, a service enabling remote collaboration. The service offers tools for video conferencing, chatting and file sharing. Collaboration services like *Microsoft Teams* became much more prevalent from 2020 onwards because of the high amount of home office deployed by companies around the globe. When the pandemic began pushing the need for online collaboration platforms, a lot of them were overwhelmed and could not satisfy all customer's demands. "Air cooling is not enough", is a statement made by Christian Belady, which illustrates the need and interest for alternative cooling solutions [Sab21]. The increased demand for collaboration services made the switch to immersion cooling necessary for *Microsoft*. The servers even became more reliable because the heat capacity of the fluid allows for better dampening of spikes in demand. Servers used for collaborating often see spikes at the begin of every hour, this is when meetings usually start [Sab21]. *Microsoft's* servers are able to run at elevated power since they are cooled with two-phase liquid immersion, allowing for more computing power using the same hardware. In addition to the energy efficiency, reliability and performance increases, the switch to two-phase immersion cooling also allows *Microsoft* to meet their goal of replenishing more water than the company consumes [Sab21]. This is possible because of the higher coolant temperatures in two-phase liquid immersion cooling, which allow for little to no water consumption of cooling towers.

As the presented examples illustrate, there are already not only various sizes of immersion cooled solutions, but also the different types, such as single and two phase immersion cooling are in use. The different types and sizes present cooling solutions suitable for all kinds of businesses. Summing up, the developments presented in this section, such as the switch to immersion cooling by *Microsoft* as well as the solutions designed by *Asperitas* and *Submer*, display that there is a dynamic and ongoing development in the field of immersion cooling.

6 Is Immersion Cooling the Future: A Concluding Statement

At this moment, most of the data centres around the world are relying on the medium air for cooling of their IT equipment. The IT equipment in data centres is getting more and more powerful and rack densities are rising because of the increasing demand for cloud services and computing capacity. With air cooling, a minimum of about 10% and often more than 50% of the energy is used for cooling. In addition to its poor efficiency, the maximum power density in air cooled data centres is often limited. Furthermore, air also impacts the reliability of servers. Contaminated air can be the reason for failure due to corrosion. Looking at the fast development in the field of information technology, the long established usage of air cooling does not correspond to the stage of development in the 21st century.

After closely inspecting the flaws of air cooling, it seems obvious to search for a better cooling solution for data centres. Hybrid cooling is not a viable alternative, it is too expensive. Additionally, it would be perfect to eliminate air from servers to combat reliability issues. This is where liquid immersion cooling comes into play.

During the research on liquid immersion cooling, several categories that are of relevance for data centre operators were identified. The relevant categories are: Efficiency of the data centre as a whole, density of a data centre together with all of its supporting equipment and reliability. In addition, waste heat utilization and expenses are important factors too. As shown in Chapter 4, data centres cooled with liquid immersion cooling have been proven to be much more efficient than their air cooled counterparts [GBR+14; MMK17; Sha18]. The results show an increase from 3.87 GFLOPS/W to 7.8 GFLOPS/W by converting an average air cooled data centre to immersion cooling. The efficiency of data centres can be doubled using immersion cooling and even the most efficient air cooled data centres of the world can increase their efficiency by about 10%.

On top of this, immersion cooled data centres are far more suitable for waste heat utilization, something that is not possible in air cooled data centres, where the temperature of the waste heat is relatively low and the properties of air make it hard to transport that air away. Better efficiency equals less money for the same amount of data centre. Immersion cooled data centers allow for much more compact designs, more than three times the density of their air cooled counterparts. In liquid immersion cooled data centres there is no trade-off between efficiency and density. Conversely, air-cooled data centres can only be either-or. As seen in Chapter 5, data centres can fit into small rooms with no need for a highly specialized environment. This could allow for computing-driven heating in spaces that had been to densely packed before [Sub21].

The amount of research published has been increasing over the last 5 years. This trend is likely be continued looking beyond the pandemic. Not least of all, looking at the outstanding performance of immersion cooling, proven by different resources in the areas of efficiency, density and reliability it is hard to explain why it is not widely used already.

Considering the advantages of immersion cooling within the mentioned categories, immersion cooling has the potential to be the future of data centre cooling. With higher powered servers, even in regular data centres and not only in HPDCs, that cannot be cooled sufficiently by air, the demand for liquid immersion cooling will rise. The demand for sustainable cloud computing, water conservation and emerging carbon taxes, will push immersion cooling to success in the coming decade [MSEW11].

Even with big advantages in efficiency and density by choosing liquid immersion cooling, not many companies have switched their cooling solution to immersion cooling yet. This could be because there was not much information about reliability [SESA16]. Without data about the reliability of the systems and the liquid, data centre operators are hesitant to implement liquid cooling. However, these concerns will slowly become alleviated with studies in the recent years showing results of equal or even better reliability [SESA16]. In addition, there are now companies offering immersion cooled server solutions with specifically designed fluid [ASPERITAS21; Sub21] and big contenders in the data centre market are starting to invest in liquid immersion cooling. This could convince other data centre operators of the benefits of liquid immersion cooling.

While data centre efficiency can be significantly increased by switching to immersion cooling, efficient cooling alone is not the answer to all problems. Of course it would be desirable if data centre operators were continuing to improve their facilities, showing interest in optimizing parts other than the cooling equipment. An example can be seen in the implementation of waste heat utilization in data centres. Since data centres will still consume high amounts of energy, waste heat utilization will be needed to further decrease cost and the environmental impact. There is no metric yet describing efficiency together with waste heat utilization. A potential metric could measure the amount of energy saved in locations utilizing the waste heat produced by a data centre, this value could then be added to a data centres efficiency index. It is also important to reward data centres for waste heat utilization, as it effectively enables locations to use even less energy.

Another idea for improvement is to minimize the usage of the metric PUE. While PUE has been a good measurement for the efficiency of non IT parts of data centres, the metric will no longer be adequate when liquid immersion cooling is the standard for data centres. PUE values will only differ slightly and less efficient IT hardware will even be rewarded [ASPERITAS21]. To prevent stagnation in data centre efficiency, the metric of FLOPS/W should be used and energy consumption of different parts of a data centre should be tracked. This could help track down other potential areas of improvement.

To allow for a perfectly accurate comparison, a data centre could be observed before and after the switch to immersion cooling. Such an experiment would allow for a direct comparison regarding efficiency, density and reliability. The values gathered by this, could be used to encourage other data centre operators to convert their centre to immersion, lowering the global power consumption of data centres. This may be the solution, how data centres can contribute to the stop of global warming, without the need to give up computing.

Bibliography

[360View18] S. S. GILL, R. BUYYA. *A Taxonomy and Future Directions for Sustainable Cloud Computing: 360 Degree View*. 2018. URL: https://arxiv.org/ftp/arxiv/papers/1712/1712.02899.pdf (cit. on pp. 15, 16, 20, 25, 36).

[AAH+18] X. An, M. Arora, W. Huang, W. C. Brantley, J. L. Greathouse. “3D Numerical Analysis of Two-Phase Immersion Cooling for Electronic Components”. In: *2018 17th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*. 2018, pp. 609–614. DOI: 10.1109/ITHERM.2018.8419528 (cit. on pp. 20, 29, 32, 33, 35, 38).

[Amb13] Ambrose McNevin. *APAC data center survey reveals high PUE figures across the region*. 2013. URL: https://www.datacenterdynamics.com/en/news/apac-data-center-survey-reveals-high-pue-figures-across-the-region/#:~:text=In%5C%20Japan%5C%20the%5C%20study%5C%20found,15%5C%25%5C%20operate%5C%20six%5C%20or%5C%20more. (cit. on pp. 32, 33).

[AMD21] AMD. *AMD*. 2021. URL: https://www.amd.com/ (cit. on p. 35).

[ASPERITAS21] Asperitas. *Apache ODE™ – The Orchestration Director Engine*. 2021. URL: https://www.asperitas.com/ (cit. on pp. 28, 29, 41, 44).

[BSG+18] P. V. Bansode, J. M. Shah, G. Gupta, D. Agonafer, H. Patel, D. Roe, R. Tufty. “Measurement of the Thermal Performance of a Single-Phase Immersion Cooled Server at Elevated Temperatures for Prolonged Time”. In: *International Electronic Packaging Technical Conference and Exhibition*. Vol. 51920. American Society of Mechanical Engineers. 2018, V001T02A010 (cit. on pp. 20, 28).

[CGB17] S. Chandrasekaran, J. Gess, S. Bhavnani. “Effect of subcooling, flow rate and surface characteristics on flow boiling performance of high performance liquid cooled immersion server model”. In: *2017 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*. 2017, pp. 905–912. DOI: 10.1109/ITHERM.2017.7992582 (cit. on pp. 20, 29, 32).

[CH16] H. Coles, M. Herrlin. *Immersion Cooling of Electronics in DoD Installations*. Tech. rep. CALIFORNIA UNIV BERKELEY BERKELEY United States, 2016 (cit. on pp. 20, 25, 28).

[Cla13] J. Clark. *Raising datacenter power density*. 2013. URL: https://www.riello-ups.com/de/blog/571-raising-datacenter-power-density (cit. on p. 37).

[DC-HISTORY18] Martin Pramatarov. *The History of Data Centers*. 2018. URL: https://blog.cloudware.bg/en/the-history-of-data-centers/ (cit. on p. 14).

[Dhr19] Dhruv Varma. *Two-Phase Versus Single-Phase Immersion Cooling*. 2019. URL: https://www.grcooling.com/blog/two-phase-versus-single-phase-immersion-cooling/ (cit. on p. 29).

[EEV+17] R. Eiland, J. Edward Fernandes, M. Vallejo, A. Siddarth, D. Agonafer, V. Mulay. "Thermal Performance and Efficiency of a Mineral Oil Immersed Server Over Varied Environmental Operating Conditions". In: *Journal of Electronic Packaging* 139.4 (Sept. 2017). 041005. ISSN: 1043-7398. DOI: 10.1115/1.4037526. eprint: https://asmedigitalcollection.asme.org/electronicpackaging/article-pdf/139/4/041005/6047057/ep_139_04_041005.pdf. URL: https://doi.org/10.1115/1.4037526 (cit. on pp. 13, 20).

[EFV+14] R. Eiland, J. Fernandes, M. Vallejo, D. Agonafer, V. Mulay. "Flow Rate and inlet temperature considerations for direct immersion of a single server in mineral oil". In: *Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*. 2014, pp. 706–714. DOI: 10.1109/ITHERM.2014.6892350 (cit. on pp. 16, 20, 29, 32, 33).

[ENIAC21] The Apache Software Foundation. *Missing Link: Hallo, ENIAC – der erste programmierbare Computer vor 75 Jahren*. 2021. URL: https://heise.de/-5054278 (cit. on p. 14).

[FEELHEAT20] Voices of the Industry. *Data Centers Feeling the Heat! The History and Future of Data Center Cooling*. 2020. URL: https://datacenterfrontier.com/history-future-data-center-cooling/ (cit. on p. 24).

[GBR+14] J. Gess, S. Bhavnani, B. Ramakrishnan, R. W. Johnson, D. Harris, R. Knight, M. Hamilton, C. Ellis. "Investigation and characterization of a high performance, small form factor, modular liquid immersion cooled server model". In: *2014 Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*. 2014, pp. 8–16. DOI: 10.1109/SEMI-THERM.2014.6892208 (cit. on pp. 15, 20, 29, 32, 37, 38, 43).

[GCC+19] D. Gandhi, U. Chowdhury, T. Chauhan, P. Bansode, S. Saini, J. M. Shah, D. Agonafer. "Computational analysis for thermal optimization of server for single phase immersion cooling". In: *International Electronic Packaging Technical Conference and Exhibition*. Vol. 59322. American Society of Mechanical Engineers. 2019, V001T02A013 (cit. on pp. 20, 23).

[GCKA19] L. Gibbons, B. Coyne, D. Kennedy, S. Alimohammadi. "A Techno-Economic Analysis of Current Cooling Techniques in Irish Data Centres". In: *2019 25th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*. 2019, pp. 1–6. DOI: 10.1109/THERMINIC.2019.8923482 (cit. on pp. 20, 25).

[Gup+18] G. Gupta et al. "Experimental Analysis of A Single-Phase Direct Liquid Cooled Server Performance at Extremely Low Temperatures for Extended Time Periods". PhD thesis. 2018 (cit. on pp. 20, 28).

[I W19] I W Kuncoro and N A Pambudi and M K Biddinika and I Widiastuti and M Hijriawan and K M Wibowo. "Immersion cooling as the next technology for data center cooling: A review". In: *Journal of Physics: Conference Series*

1402 (Dec. 2019), p. 044057. DOI: 10.1088/1742-6596/1402/4/044057. URL: https://doi.org/10.1088/1742-6596/1402/4/044057 (cit. on pp. 15, 20, 25, 36).

[Jul19] Julius Neudorfer. *Liquid Cooling Moves Upstream to Hyperscale Data Centers*. 2019. URL: https://www.missioncriticalmagazine.com/articles/92044-liquid-cooling-moves-upstream-to-hyperscale-data-centers (cit. on p. 26).

[KG16] A. C. Kheirabadi, D. Groulx. "Cooling of server electronics: A design review of existing technology". In: *Applied Thermal Engineering* 105 (2016), pp. 622–638. ISSN: 1359-4311. DOI: https://doi.org/10.1016/j.applthermaleng.2016.03.056. URL: https://www.sciencedirect.com/science/article/pii/S1359431116303490 (cit. on pp. 15, 20, 23, 25, 37, 38).

[KPBB20] I. W. Kuncoro, N. A. Pambudi, M. K. Biddinika, C. W. Budiyanto. "Optimization of immersion cooling performance using the Taguchi Method". In: *Case Studies in Thermal Engineering* 21 (2020), p. 100729. ISSN: 2214-157X. DOI: https://doi.org/10.1016/j.csite.2020.100729. URL: https://www.sciencedirect.com/science/article/pii/S2214157X20304718 (cit. on p. 20).

[LIQUIDVSAIR20] Robert Sheldon. *Liquid cooling vs. air cooling in the data center*. 2020. URL: https://searchdatacenter.techtarget.com/feature/Liquid-cooling-vs-air-cooling-in-the-data-center (cit. on p. 24).

[MMK17] M. Matsuoka, K. Matsuda, H. Kubo. "Liquid immersion cooling technology with natural convection in data center". In: *2017 IEEE 6th International Conference on Cloud Networking (CloudNet)*. 2017, pp. 1–7. DOI: 10.1109/CloudNet.2017.8071539 (cit. on pp. 15, 20, 26, 29, 32, 33, 37–39, 43).

[MSEW11] G. Müller, N. Sonehara, I. Echizen, S. Wohlgemuth. *Sustainable cloud computing*. 2011 (cit. on p. 44).

[Nic18] Nicola Jones. *How to stop data centres from gobbling up the world's electricity*. 2018. URL: https://www.nature.com/articles/d41586-018-06610-y (cit. on pp. 13, 15, 25, 33).

[Pat13] H. Patel. "Immersion Cooling of High End Data Center Server and Validation Through Experiments". In: (2013) (cit. on pp. 20, 25, 29).

[PUE17] Otto GeißlerUlrike Ostler. *Was ist eigentlich Power Usage Effectiveness - PUE?* 2017. URL: https://www.datacenter-insider.de/was-ist-eigentlich-power-usage-effectiveness--pue-a-663864/#:~:text=Die%5C%20Power%5C%20Usage%5C%20Effectiveness%5C%20(PUE,der%5C%20Energieaufnahme%5C%20der%5C%20IT%5C%2DInfrastruktur. (cit. on p. 23).

[QCW+17] D. Qiu, L. Cao, Q. Wang, F. Hou, X. Wang. "Experimental and numerical study of 3D stacked dies under forced air cooling and water immersion cooling". In: *Microelectronics Reliability* 74 (2017), pp. 34–43. ISSN: 0026-2714. DOI: https://doi.org/10.1016/j.microrel.2017.02.016. URL: https://www.sciencedirect.com/science/article/pii/S0026271417300410 (cit. on p. 20).

[Reg21] U. of Regensburg. *DATENBANK-INFOSYSTEM (DBIS)*. 2021. URL: rzblx10.uni-regensburg.de (cit. on p. 17).

[Ric14] Rich Miller. *Inside SuperNAP 8: Switch's Tier IV Data Fortress*. 2014. URL: https://blog.cloudware.bg/en/the-history-of-data-centers/ (cit. on pp. 25, 32, 33, 35, 37).

[RRC+19] S. Ramdas, P. Rajmane, T. Chauhan, A. Misrak, D. Agonafer. "Impact of Immersion Cooling on Thermo-Mechanical Properties of PCB's and Reliability of Electronic Packages". In: *International Electronic Packaging Technical Conference and Exhibition*. Vol. 59322. American Society of Mechanical Engineers. 2019, V001T02A011 (cit. on pp. 20, 23, 24, 26–28).

[Sab21] Sabina Weston. *Microsoft is submerging servers in boiling liquid to prevent Teams outages*. 2021. URL: https://www.itpro.co.uk/server-storage/data-centres/359129/microsoft-submerges-servers-in-boiling-liquid-to-prevent-teams?amp (cit. on p. 42).

[SBS+19] P. A. Shinde, P. V. Bansode, S. Saini, R. Kasukurthy, T. Chauhan, J. M. Shah, D. Agonafer. "Experimental analysis for optimization of thermal performance of a server in single phase immersion cooling, ASME Conference Paper No". In: *IPACK2019-6590* (2019) (cit. on pp. 20, 23, 25, 27, 28).

[SESA16] J. M. Shah, R. Eiland, A. Siddarth, D. Agonafer. "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)*. 2016, pp. 316–325. DOI: 10.1109/ITHERM.2016.7517566 (cit. on pp. 16, 20, 28, 44).

[Sha+16] J. M. Shah et al. "Reliability challenges in airside economization and oil immersion cooling". PhD thesis. 2016 (cit. on pp. 20, 28).

[Sha18] J. M. Shah. *CHARACTERIZING CONTAMINATION TO EXPAND ASHRAE ENVELOPE IN AIRSIDE ECONOMIZATION AND THERMAL AND RELIABILITY IN IMMERSION COOLING OF DATA CENTERS*. 2018. URL: https://rc.library.uta.edu/uta-ir/handle/10106/28654 (cit. on pp. 20, 25, 28, 29, 32, 33, 43).

[Smo19] M. Smolaks. *Power density – the real benchmark of a data centre*. 2019. URL: https://virtusdatacentres.com/item/389-power-density-the-real-benchmark-of-a-data-centre (cit. on p. 38).

[Sub21] Submer. *Submer*. 2021. URL: http://ode.apache.org (cit. on pp. 42–44).

[The21] The ServerRack FAQ. *Define: Rack Unit "U" or "RU"*. 2021. URL: https://www.server-racks.com/rack-unit-u-ru.html (cit. on p. 23).

[Tif19] Tiffany Trader. *AMD Launches Epyc Rome, First 7nm CPU*. 2019. URL: https://www.hpcwire.com/2019/08/08/amd-launches-epyc-rome-first-7nm-cpu/ (cit. on p. 35).

[Tum10a] P. Tuma. “Open Bath Immersion Cooling In Data Centers: A New Twist On An Old Idea”. In: *OPEN BATH IMMERSION COOLING IN DATA CENTERS KEEPING MOORESS LAW ALIVE* (2010), p. 10 (cit. on pp. 20, 28, 29, 37, 38, 40).

[Tum10b] P. E. Tuma. “The merits of open bath immersion cooling of datacom equipment”. In: *2010 26th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*. 2010, pp. 123–131. DOI: 10.1109/STHERM.2010.5444305 (cit. on pp. 16, 20, 24, 26, 27, 29, 37, 38).

[Wei19] J. Wei. “Liquid Cooling, opportunity challenges toward effective and efficient scalabilities”. In: *2019 IEEE CPMT Symposium Japan (ICSJ)*. 2019, pp. 83–84. DOI: 10.1109/ICSJ47124.2019.8998723 (cit. on p. 20).

[WTKD19] C. Wu, W. Tong, B. B. Kanbur, F. Duan. “Full-scale Two-phase Liquid Immersion Cooing Data Center System in Tropical Environment”. In: *2019 18th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*. 2019, pp. 703–708. DOI: 10.1109/ITHERM.2019.8757316 (cit. on pp. 15, 20, 28).

[WV17] B. Watson, V. K. Venkiteswaran. “Universal Cooling of Data Centres: A CFD Analysis”. In: *Energy Procedia* 142 (2017). Proceedings of the 9th International Conference on Applied Energy, pp. 2711–2720. ISSN: 1876-6102. DOI: https://doi.org/10.1016/j.egypro.2017.12.215. URL: https://www.sciencedirect.com/science/article/pii/S1876610217359490 (cit. on pp. 20, 28, 29, 38).

[Wyl18] Wylie Wong. *Liquid Cooling Can Lower Data Center PUE, But That’s Not Its Main Draw*. 2018. URL: https://www.datacenterknowledge.com/power-and-cooling/liquid-cooling-can-lower-data-center-pue-s-not-its-main-draw (cit. on p. 26).

All links were last followed on May 25, 2021.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

Heidelberg 29.05.2021

place, date, signature