MSc Program
Environmental Technology & International Affairs

Cloud Computing's Energy Consumption and Greenhouse Gas Emissions in the Context of Industry 4.0

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by
Dr. Klaus Rapp

Michael Toborek

00853356

Vienna, 12.10.2018

Affidavit

I, **MICHAEL TOBOREK**, hereby declare

1. that I am the sole author of the present Master's Thesis, "CLOUD COMPUTING'S ENERGY CONSUMPTION AND GREENHOUSE GAS EMISSIONS IN THE CONTEXT OF INDUSTRY 4.0", 66 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and
2. that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, 12.10.2018

Signature

Abstract

Industry 4.0 as the main driver of future industrial manufacturing is receiving considerable attention in recent years. Especially since it is viewed as technology that could decouple economic growth from resource use. However, there is no definite agreement how this technology will affect our environment, more specifically global energy demand, greenhouse gas emissions and consequently climate change. At the core of Industry 4.0 lies a vast expanding data layer that is being connected through cloud computing. Energy consumption from cloud computing and thus the number of data centres is on the rise and with it the associated greenhouse gas emissions. This is due to the increasing amount of data created in the world. Since adoption of Industry 4.0 is expected to generate additional, exponentially growing amounts of data it will impact cloud computing's energy consumption and subsequently its environmental effects. Since the magnitude of the effect is not yet determined this thesis examines it via a quantitative analysis of secondary data in a mixed approach. To this end, the total energy consumption of ICT is defined and then broken further down into energy consumption by cloud data centres which is then used to calculate the share that Industry 4.0 is responsible for. In a second step, three different growth scenarios based on two base variables of amount of data created and cloud data centre efficiency are elaborated to predict possible developments until 2025.

This paper finds that currently the amount of data generated through Industry 4.0 manufacturing processes represents a small part of overall generated data and therefore total energy demand. Moreover, Industry 4.0 is expected to deliver efficiency gains that by far outweigh the increase in energy consumed for its operation. Therefore, its use to further boost manufacturing efficiency can provide a valuable contribution to addressing global challenges such as climate change and increasing the share of renewable energy in the system which in turn will also reduce the environmental impact of the energy consumed.

Keywords:

Industry 4.0 / Cloud Computing / IIOT / IOT / Energy Efficiency

Industrie 4.0 / Cloud Computing/ IIOT/ IOT/ Energieeffizienz

Table of Contents

ABSTRACT i

TABLE OF CONTENTS ii

LIST OF ABBREVIATIONS iv

ACKNOWLEDGMENTS v

1. INTRODUCTION 1

2. GOALS AND OBJECTIVES 4

3. STRUCTURE 4

3.1. LIMITATIONS 5

4. UNDERSTANDING THE BIG PICTURE – A LITERATURE REVIEW 5

4.1. INDUSTRY 4.0 6
4.2. INDUSTRY 4.0 CORE TECHNOLOGIES ENABLED BY CLOUD COMPUTING 11
4.3. IT/OT CONVERGENCE 11
4.4. CYBER PHYSICAL SYSTEMS (CPS) 11
4.5. BIG DATA 13
4.5.1. THE DATA JOURNEY IN CPS FOR INDUSTRY 4.0 15

5. CLOUD COMPUTING - AN INTRODUCTION 17

5.1. INDUSTRY 4.0 & CLOUD – A MATCH MADE IN HEAVEN. 20
5.2. OTHER ENABLING TECHNOLOGIES 20

6. ENERGY CONSUMPTION & RELATED GREENHOUSE GAS EMISSIONS FROM ICT & CLOUD COMPUTING 22

6.1. SCOPE OF CLOUD COMPUTING SYSTEMS 25
6.2. GREEN HOUSE GAS EMISSIONS 26
6.3. ENERGY CONSUMPTION METRICS 29

7. CLOUD EFFICIENCY DISCUSSION 30

7.1. APPROACHES TO GREEN THE CLOUD. 32
7.2. STATE OF THE ART OPTIMIZATION BY GOOGLE 33
7.3. FUTURE APPROACH – EDGE COMPUTING 34

8. METHODOLOGY 36

8.1. GENERAL METHODOLOGY 36
8.2. ENERGY DEMAND OF THE CLOUD 36
8.3. DATA TRANSFER 37
8.4. DATA CENTRE STORAGE 38
8.5. DATA PROCESSING 39
8.6. CLOUD ENERGY CONSUMPTION 40

9. RESULTS - ENERGY CONSUMPTION OF THE CLOUD FOR INDUSTRY 4.0 41

9.1. FUTURE SCENARIOS 42

10. CONCLUSION 45

BIBLIOGRAPHY 47

LIST OF TABLES **61**

LIST OF FIGURES **62**

ANNEX **A1**

ENERGY CONSUMPTION OF DATA TRANSFER A1
ENERGY CONSUMPTION OF DATA CENTRE STORAGE A1
FORECAST OF THE AMOUNT OF DATA CREATES BY INDUSTRY 4.0 ..... A1
GLOBAL IP DATA TRAFFIC A2
GLOBAL DATA CENTRE TRAFFIC A2
SCENARIO I A3
SCENARIO II A3
SCENARIO III A3
DEVELOPMENT OF VARIOUS SCENARIOS UNTIL 2025 ..... A4

List of Abbreviations

AI	Artificial Intelligence
BDA	Big Data Analytics
CPS	Cyber Physical Systems
DPM	Dynamic Power Management
DT	Digital Twin
DVFS	Voltage Frequency Scaling
ICT	Information and Communication Technology
IIoT	Industrial Internet of Things
IoP	Internet of People
IoS	Internet of Services
IoT	Internet of Things
M2M	Machine 2 Machine
PUE	Power Usage Effectiveness
PDU	Power Distribution Unit
SMO	Smart Manufacturing Objects
SOA	Service Oriented Architecture
UPS	Uninterruptable Powering Supply

Acknowledgments

I want to use this opportunity to express my gratitude to my family, friends, and colleagues from the ETIA9 cohort. Thank you for supporting me throughout the course of this project. I am thankful for your aspiring guidance, invaluably constructive criticism and friendly advice during the time of my studies. Moreover, my gratitude goes to two very special people without whom writing this thesis would not have been possible. I am very grateful for knowing that you have my back.

Finally, I would like to thank my supervisor, Dr. Klaus Rapp for the constructive feedback and for the encouragement to pursue this topic. His insights and guidance have allowed me to compose a very interesting approach to the questions presented in this thesis.

1. Introduction

Our ability to collaborate in large numbers with strangers is the sole reason that we have been able to rise from insignificant animals some 70,000 years ago to the rulers of the world, according to Yuval Noah Harari. (Harari 2015) Through our ability to create a dual world consisting of objective reality, on the one hand, and an imaginative one on the other, we have been able to transform our planet beyond recognition. Creating fictional entities and belief systems such as money, governmental & non-governmental organizations, human rights and god, allows us to foresee and steer human interaction by convincing a wide audience of their significance. As a result, individuals can trust complete strangers because they adhere to the same belief system and therefore can collaborate without prejudice.

While human collaboration undoubtedly remains an important factor, it is far from being the sole reason for the developments of the last century. Although collaboration in large numbers allowed Homo sapiens to rise to the top of the food chain, the key differentiator of the past 200 hundred years was economic growth. The transition from traditional labour-intensive systems to industrialisation via mechanization of production paved the way to our modern and automated society. However, successful industrialization is far from being an automated process. It depends on a variety of enabling conditions starting with the availability of excess capital, to entrepreneurs willing to take the risk of embarking on a new venture. Moreover, it needs an excess of skilled labour force, elaborated manufacturing technologies and last but not least a developed transportation and infrastructure system for distribution. Once society reaches a certain level of urbanization, such conditions are usually readily available. These preconditions coupled with the novelty of the steam engine, enabled the first industrial revolution, which marked a tipping point that introduced unprecedented growth that has since then remained unmatched. (Mathis 2016) Of course, the transformation from steel to electric power during the second industrial revolution and the automation of production through information technology during the third industrial revolution all left their own mark on history and our economic system. Nonetheless, their contribution never reached the true transformative effects of the first industrial revolution. Currently, another immense transformation is underway that might once again have the potential to alter every part of the human ecosystem. This forth-industrial revolution is supposed to bring forth a level of transformation unseen until now, and by far surpassing the effects of the previous transitional waves. (Schwab 2017) The spread and convergence of technology and industry that blurs borders and establishes simultaneous connections between the physical, digital and biological world brings us closer to another substantial leap forward. The first leap forward has happened when human beings came up with a way to support their physical ability, their muscles, with machines. The next leap forward might come when human beings discover how to support their cognitive abilities, their brain, with machines. Regardless of the time this state will be accomplished, it is fundamental to keep in mind that this exponential increase in possibilities, always comes with an equivalent increase in risk.

In this context, it is of utmost importance to carefully consider the risks and adverse effects that accompanied the previous development waves. There have certainly been stellar advancements in every area of human life that can be attributed to these changes. However, these transformations also came at a cost. A prime example is the complete neglect for impacts on the ecosystem and the environment throughout this period of economic growth. As a result, climate change is the most systemic threat to humankind that we face today, according to United Nations Secretary General Antonio Guterres. (Sengupta 2018) Nevertheless, western leaders and citizens often ignore this fact because they are incapable of seeing the adverse effects their lifestyle choices are producing. On the one hand, because active communication on these issues is limited and on the other hand, because externalization of these environmental costs and adverse effects to the global south has led to a twisted view of reality. (Demirović et al. 2011) The good news is that public and private awareness for this issue is increasing. An increasing number of people is refuting the long-presented notion that unlimited growth on a planet with limited natural resources is attainable. Additionally, the belief that technological solutions will provide all the answers is also being questioned. (Fatheuer, Fuhr, and Unmüßig 2015)

In light of these developments, Industry 4.0 has to perform a miracle. Global economic growth has been slowing for years (Haenn, Harnish, and Wilk 2016) leading to numerous recessions and the urgent need for a development to spark another upwards trend. At the same time negative environmental effects are no longer just predictions. Climate change as a result of manmade environmental pollution is real and its effects like increased number of flooding, hurricanes, heatwaves and an ever declining air quality are a reality for an increasing number of people. (Fountain and Plumer 2018) Industry 4.0 is poised to be the solution to all those challenges and much more. For the first time in centuries the impossible seems feasible. An increasing rate of digitalization is celebrated as being able to spark economic growth while at the same time solving our ecological, economic and social crisis. (M. Lee et al. 2018)

However, there are also more critical voices pointing out that numerous other technologies in their time were celebrated as the harbingers of the forth-industrial revolution. From the 1940s till today the term has been recycled numerous times. First, modern communication was endangering the current status quo. (Carr 1940) Then, after the Second World War, intra-atomic energy was rumoured to be a game changer only to be replaced by electronics and the computer age during the 70s. In the 1980s, the forth-industrial revolution was supposed to be introduced by the information revolution only to be substituted by nanotechnology in the 90s. (Edgerton 2011) Therefore, according to (Garbee 2016) every technological breakthrough of most of the past century has once been entitled as the fourth industrial revolution without being able to deliver on its promise to truly revolutionize the economic, social or political sphere. While, it is true that most technological advancements have the power to profoundly transform societies, only time can tell how profound those changes can be.

However, even in processes of transformation and change, some things remain constant. Just like the first industrial revolution built on a core technology to deliver profound societal change, so does the fourth. The critical enabler for this development is being provided over the Internet. The internet provides the core layer upon which other technologies and innovations are created and connected. Just as the first industrial revolution would not have been possible without the steam engine, the fourth industrial revolution will not be possible without cloud computing. (Bateman 2018) Therefore, this thesis wants to go beyond the hype and provide an explanation of the interplay between the technologies that make up the forth industrial revolution. Furthermore, it will explain why cloud computing is important for the fourth industrial revolution. On top of that it will examine the current energy consumption and greenhouse gas emissions of cloud computing for Industry 4.0 and attempt to predict its development until 2025. Last but not least the results will be put back into perspective of Industry 4.0 to anticipate if we can expect a total emission reduction.

2. Goals and Objectives

Industry 4.0 is celebrated as the saviour to all the present challenges in the socio-economic and ecological sphere. Possibly surpassing the scope & transformative effect of the first industrial revolution, it is supposed to spark economic growth while at the same time tackling environmental problems, like the global climate change crisis. To fulfil these great expectations, Industry 4.0 works on connecting several modern production and communication technologies to increase productivity and ensure efficiency on all levels. Thus, it is not only transforming production processes like the first industrial revolution, but it is targeting a complete and systematic transformation with changes to business models, customer value creation, operational company set up and the market as a whole.

The purpose of this thesis is to examine cloud computing and its environmental effects in the context of Industry 4.0. Since Industry 4.0 builds on various enabling technologies that are mostly at the beginning of the innovation adoption cycle, it currently remains difficult to assess its general impact. However, since cloud computing acts as enabler for most of these technologies it is an interesting candidate for examination. Hence, this paper will focus on examining the energy demand of cloud computing for Industry 4.0 applications today and in the future by trying to answer the following research:

- What is the current research landscape regarding energy efficient cloud computing solutions?
- What effect will the emergence of Industry 4.0 have on energy demand of cloud computing?
- What effect will the emergence of Industry 4.0 have on CO_2 emissions generated to power cloud computing?

3. Structure

To provide answers to the presented questions, the following combined methodology will be employed. First, the connection between Industry 4.0 and cloud computing as enabling technology will be presented. Since technologies like the internet of things, big data and cyber physical systems (CPS) built on cloud computing to enable industry 4.0 their definition and interplay will be described. Secondly, a qualitative literature review will be conducted to examine the current research landscape regarding cloud computing and its current and predicted environmental impact. For this an analysis of relevant

scientific journal articles, policy briefs, conference paper, as well as studies by various institutions and businesses will be conducted.
The second part of this study will use a secondary quantitative approach to estimate the current proportional energy demand of cloud computing for Industry 4.0 applications and its future development. This will be based on an approach that was proposed by J. Koomey (2011) and on relevant datasets from various sources. (Cisco Systems 2018; Andrae et al. 2015; Malmodin and Bergmark 2015) In the end the results will be summarized and put into perspective of the total expected environmental impact of Industry 4.0.

3.1. Limitations

Due to the complexity and the inter-connected character of the considered system, the thesis is facing numerous limitations. Therefore, the detailed results should be viewed not only with a critical eye but also with a focus on the big picture and the relationships between global energy demand, ICT energy demand and cloud energy demand in the context of Industry 4.0. The main challenge was the limited availability of current data on cloud energy demand and the amount of data produced by Industry 4.0. Due to the rapid developments in the sector, the available data sets age fast. Therefore, the thesis works with assumptions and estimates where necessary. Moreover, it only considers the operational energy demand and the accompanying GHGs. A full lifecycle analysis is not conducted. What is more also the economic impact and the impact on society are not part of the examination.

4. Understanding the big picture – a literature review

This section will first briefly present the history of Industry 4.0 and describe its core enabling technologies. Afterwards, the focus will be on elaborating the system perspective to foster the understanding of the importance of cloud computing. Finally, the conclusion of this chapter will underline the vital importance of this technology for Industry 4.0.

4.1. Industry 4.0

Although the term Industry 4.0 has recently spread around the globe, it is a relatively young invention. It originated in Germany in only 2011 as part of the future development strategy to boost economic growth in the manufacturing sector. (Kagermann, Wahlster, and Helbig 2013) Through the usage of the information and communications networks that are established around the world and constant and automated exchange of information between production and business processes, the German manufacturing and logistics industry is supposed to interconnect its complete value chain. (Kamarul Bahrin et al. 2016). In the Anglo-American world the term Internet of Things (IoT) is being used to describe the same idea and has found widespread adoption. Therefore, the interchangeable use of these terms has led to some confusion, which will be solved in the following section.

The constant advancements in Information and Communication Technology (ICT) over the past centuries have reached a point, where we are on the brink of new and unparalleled ICT usage worldwide. Nowadays, memory modules, microprocessors and sensors can be integrated into practically every object because they have become smaller, cheaper and more energy efficient. Combined with high-speed data transmission and computing power they generate additional value because now those objects possess the ability to gather data, receive information and adapt their behaviour according to the situation. What is reality today was already predicted by Steinbuch (1966) who claimed that in the span of a few decades there won't be a single object that is not operated by a computer and connected through a network like the human nervous system. Moreover, Weiser (1999) anticipated that technology will find its way into the background of everyday objects and minimize the need for conscious human steering and interaction. Today this concept is known in the academic literature as ubiquitous computing. Slightly adapted by industry practice, the concept became known as pervasive computing, where the continued processing of available information is used for web-based processes and electronic commerce. (Bohn et al. 2005) A phrase that is more common to the average user and can be used to understand pervasive computing are smart devices. Those are the manifestation of the underlying idea of creating devices that collect, understand and distribute data. Through the creation of interconnected networks, these devices become capable of understanding their surroundings and thus can reduce the necessary human interaction with them without losing functionality. The vision is to create ambient intelligence, where those devices provide unparalleled human experience and added value. Through a network capable of understanding its environment and acting accordingly without the need for human intervention it caters to the wish for an omnipresent assistant operating in the background. (Grossman 2016)

Pervasive Computing and Internet of Things (IoT) are both concepts that are very closely interlinked. However, while pervasive computing envisions a world where human interaction is reduced to the usage phase where the technology is working autonomously in the background, IoT requires more human attention. Nonetheless, IoT is a significant step forward because objects no longer relate only to the user but also to the surrounding objects and a database. The term IOT was first introduced in 1999 during a presentation at Proctor & Gamble by Ashton (2009) where he linked the concept of Radio Frequency Identification (RFID) with the internet. According to him computers and consequently the internet is limited by the fact that all available information and data on things is entered by people. The resulting problem is twofold. On the one hand, people only have a limited amount of time to capture the data. On the other hand, they are inherently bad at capturing data due to their natural lack of attention and accuracy. In addition, human beings tend to focus on ideas rather than pure data itself. As a result, the internet today holds a lot of information on human ideas while sometimes failing to produce wholesome data collections on pure facts and figures collected via our surroundings. Therefore, ICT should be set up in a way that eliminates the human factor in data entry and allows it to gather all the data about the things in the world without human interaction. This way accurate data that has not been subject to human interpretation, can be captured. (Holmes 2016)

To a certain degree, this process, of shifting from human dependence of data entry to automated data collection, has already happened. IoT today describes a state where everyday objects or things are coupled with a sensor and possess the ability to communicate their state. (Satyavolu et al. 2014) To make this data exchange and collection work, there is still a high degree of human configuration and interaction required during the set-up phase. However, once the devices are set up correctly and the connections established, the vision of a completely automated system becomes reality. Since every object is collecting and transmitting large amounts of data fully automated, it is possible to gather an unprecedented amount of data and extract knowledge at an incremental cost. This process, in turn, allows to react to changes in the real world in an automated, fast and informed way, while at the same time opening the possibility to optimize various processes. On top of that new usage scenarios become feasible that are not only able to create value for the economy but simultaneously generate benefits for the society as a whole. (Mattern and Flörkemeier 2010) One practical example of a private product that is putting exactly this theory into practice is Amazon's series of smart speakers. These speakers are able to perform a multitude of tasks only via their connection to the internet and other surrounding devices. (Rossman 2016)

While IoT is mostly understood as connecting things to each other to cater to human needs the Industrial Internet of Things (IIOT) aims at doing exactly that for industrial machines. Therefore, it should come as no surprise that General Electric (GE) introduced the term in 2012 along with the term industrial internet. Both are used interchangeably and stand for a

> *"network of a multitude of industrial devices connected by communications technologies that results in systems that can monitor, collect, exchange, analyse, and deliver valuable new insights like never before. These insights can then help drive smarter, faster business decisions for industrial companies."*

("Everything You Need to Know About the Industrial Internet of Things (IIoT)" 2016)

The World Economic Forum takes a more simplified approach by stating that the IIOT is used to describe IOT in an industrial setting. (O'Halloran and Kvochko 2015) A more comprehensive definition breaks IIOT down in the components of the system and the value that they provide. Therefore, IIOT can be described as a system that is made up of smart objects, things and physical assets bundled with information technologies like cloud and edge computing in order to allow for autonomous, real time and intelligent communication, access, collection and exchange of data that in turn can be used in an industrial setting to optimise overall production value. Value in this context may describe various things from productivity boosts, to product & service delivery improvements, to energy consumption reductions or build to order cycle shortenings. (Boyes et al. 2018)

In summary, all three introduced concepts built on the idea of transforming objects into networks of smart devices, that gather data about its surroundings and adapt to the physical world around them and to additional value. Therefore, IoT is the broadest definition as it does not differentiate according to type of smart device or provided additional value. On top of that lie the concepts of Industry 4.0 and IIOT that use the smart device in the industrial setting. The differentiation between those two concepts is not used consistently across the literature and often mixed up. Nonetheless, when examining the details, it becomes clear that Industry 4.0 has a defined focus on the manufacturing industry. On the contrary IIOT considers every sector that uses industrial & professional equipment. On top of that Industry 4.0 goes further than just creating a network of objects and their data. It emphasizes the digitalization of the complete value chain. Moreover, since the term was born out of a governmental initiative it comes as no surprise that it has spread across the public professional setting more broadly. (Sontag 2018)

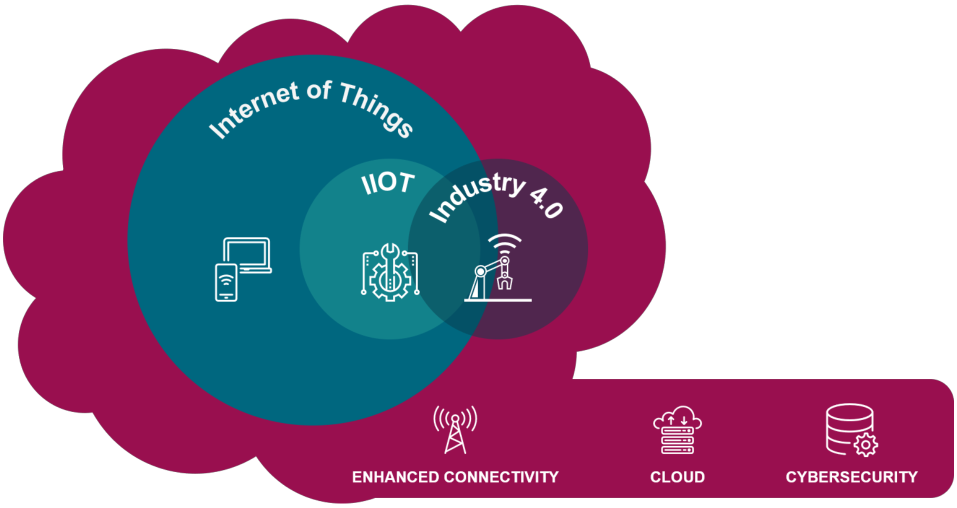

FIGURE 1 – HOW DOES IOT, IIOT, INDUSTRY 4.0 AND THE CLOUD RELATE TO EACH OTHER
(Source: own depiction after Sontag 2018)

Finally, it is important to understand one additional key differentiator between IOT and IIOT as well as Industry 4.0. It lies in the industrial use of the devices in the latter case which carries numerous additional challenges. First, the conditions that those devices need to operate in are several magnitudes more challenging than the average usage scenario of the apple watch. Industrial use means extreme conditions like corrosive or combustible environments, high and low temperatures and numerous other harsh surroundings that consumer devices are usually never confronted with. Second, there is the issue of scalability which is much more complex with IIOT. Usually IIOT applications have more devices involved and as a result, they create and collect more datapoints than their consumer targeted counterparts. Therefore, the amount of data generated is vastly larger than with IOT. As a result, transmission of the large amounts of data from the sensors to the control system can become challenging. This is where new powerful and large-scale computing infrastructures with technologies such as cloud, fog or edge become interesting. These three types enable high-performance data processing and storage on different levels of the IIOT system enabling the proper functioning of the overall scheme.

Third, most industrial devices in operation today were not built with modern sensor technology in mind. Those legacy devices (older industrial equipment) often stem from the pre-internet protocol era (IP). As a result, connecting them to the internet is a difficult and expensive endeavour. On top of that, they often need to be installed in remote regions, which in turn results in increased maintenance requirements and reduced energy efficiency. What is more, this also translates into adapted network requirements leading to the development of specialized solutions for the sector.

Last but not least, connecting these devices to the internet also raises the issue of cyber security. According to recent industry reports IOT devices are not well equipped to handle threats generated by cyber-attacks and therefore pose a significant security risk. (Spring 2017) Since huge amounts of sensitive and potentially security-critical data is exchanged via both IOT and IIOT devices, they are an interesting target for cyber-attacks. However, the big difference lies in the effects and impacts that the hacking an IOT device produces compared to that of an IIOT devices. While a malfunction of a smart speaker might be perceived as a major inconvenience for its owner, the impacts on society negligible. Yet, the potential effects of a hacked IIOT device are worse by several orders of magnitude and have possible impacts on numerous different levels of society. Especially, when considering that IIOT can be employed in critical infrastructure resources like for example water treatment plants or power plants to name only provide a few examples. Cyberattacks on such facilities do not only disturb normal operation and therefore service to the society, but also have the potential to severely harm human lives or cause substantial damage to the environment and the ecosystem. (Sadeghi, Wachsmann, and Waidner 2015)

4.2. Industry 4.0 core technologies enabled by cloud computing

The following chapter focuses on explaining the core necessities for making the concept of IIOT/ Industry 4.0 work. After briefly presenting the underlying convergence requirement, it will introduce and describe some of the most important systems needed for proper functioning of IIOT.

4.3. IT/OT Convergence

One the one hand, information technology (IT) is defined as “the use of computers and telecommunications equipment to send, receive, store and manipulate data.” (Daintith 2009) On the other hand, operational technology (OT) is used to monitor and control the performance of physical devices in the manufacturing or utilities sector. Traditionally, IT and OT have been two areas of technology with little to no connections between each other. However, the rise of IIOT and Industry 4.0 has changed this drastically. For its performance monitoring OT has always used extensive supervisory control and data acquisition systems (SCADA), as well as programmable logic controllers (PLC) and distributed control systems (DCS) to name only a few. Usually these legacy systems were not connected to the internet and only used proprietary communication protocols. However, in our newly connected world a functioning link to IT is indispensable. Yet, the overall legacy character of OT makes it difficult to establish this necessary and vital connection to ensure Industry 4.0 integration. Therefore, IT/OT convergence is an important underlying requirement to ensure the overall functioning of the system. Once the convergence is established the resulting product is a cyber physical system (CPS). (Bloem et al. 2014)

4.4. Cyber Physical Systems (CPS)

The above mentioned convergence of IT/OT into CPS is not only a remarkable development in information and communication technology as well as computer science but more importantly, another elemental driver for Industry 4.0. (Jeschke et al. 2017; Pereira and Romero 2017). The term CPS was first introduced at the National Science Foundation (NSF) around 2006 and refers to the integration of computation processes with physical processes. Most typical setups also include the network or communication aspect due to integrated network connectivity that allows to establish a connection with a server or a cloud. (Baheti and Gill 2011) To put it in simpler words CPS are devices or machines that have a main goal that is not associated with computing. Therefore, CPS encompass a mechanical device that is paired with an algorithm run on a computer. This algorithm controls, monitors and steers the machine without the need of any human interaction in the best-case scenario. However, what makes it special is not the lack of human interaction but the fact that it is connected to its surroundings and can therefore make decisions autonomously to adapt and optimize its current task according to varying conditions. To enable this constant adaptation to changing preconditions, it continuously accesses and processes data that is transmitted via the cloud.

The concept of CPS is often falsely used interchangeably with Industry 4.0. Consequently, it is important to note that applications for CPS go beyond the manufacturing industry and expand into a wide range of disciplines such as healthcare, autonomous driving, public transport and many others. At the same time, Industry 4.0 does not just denote the integration of physical and computational systems but the integration and interaction of the entire value chain from sourcing, through production, assembly, delivery and finally customer relationships. Additionally, it needs to be clarified that CPS are not a new technological field as such since some have been around for quite a while. A modern car for example can be viewed as an application of a CPS. Still, the discourse about this topic in the theoretical and academic area has largely been lacking and is only gaining traction now with the multitude of use cases that are being developed. Practical examples of real-life applications where CPS are at the core range from autonomous vehicles to smart grid and factories. The novel character of their interactions stems from their constant real-time interactions with their surrounding physical world, especially when compared to traditional information and communication systems of the past. While traditional ICT systems were only used as an extension of human steered actions, CPS act as complete substitutions lacking human interference. At least that is the vision, today human beings are still present to fulfil the role of a fail safe in case that the CPS malfunctions.

Consequently, the highly autonomous CPS enable a seamless communication process between products, producing machines and humans along and across the complete value chain. No matter the specific use case, they are able to interpret the received sensory data correctly and react accordingly. Considering that, CPS utilized in an Industry 4.0 setting allow the manufactured product to be aware of its necessary manufacturing steps. What is more this also means that the connected manufacturing equipment that is necessary for production, reacts accordingly. This production process generates an incredible amount of data which in turn can be analysed and used for continuous optimization of the functioning of the overall system (Brettel et al. 2014). If observed on its own the amount of data generated by this system is not significant. However, when looking at it from a much broader perspective it quickly becomes evident that through the rapid increase of devices collecting data in an industrial setting the amount of collected data also surges exponentially. When comparing the amount of data among industries, the manufacturing sector by far produces and stores the greatest amounts (Chen et al. 2016). This surge in the overall amount of data has given rise to big data, a concept that will be explained in the next paragraph.

4.5. Big Data

It should not be a surprise that a constant increase in created data in the world poses opportunities but at the same time comes with challenges and limitations. This observation is not new, since the evident consequences of the ever-growing amount of available data volume was first described in the 1940s. During that period, that is today known as the time of information explosion, scholars in the US have started to consider the doubling time of data in their libraries. At that time, data was still exclusively stored in books and consequently faced it own inherent space and time limitations. However, even back then it was expected that the amount of data would grow exponentially at a doubling rate of 16 years. Spinning this scenario until 2040, this growth rate would have resulted in 200 million books. (Rider 1944) A number that was incomprehensible at that time. When looking at the current global book stock, it quickly becomes evident that these estimates from the 1940s were rather on the conservative side. A practical example to illustrate the mismatch in scale and dimension between theory and reality, is the largest library operated today. The library of the U.S Congress currently has 164 million items in its catalogue. (Fischer 2016)

The first time that this challenge between availability of resources and the amount of data created, was addressed with regard to computational data sets in the late 90s. Back then, especially NASA scientists were struggling with data visualization for data sets larger than 100 gigabytes. Since memory and storage capacities of computer systems of that time were rather limited they were struggling with this issue. As a result the phrase “the problem of big data” was coined (Cox and Ellsworth 1997).

Today, big data is used as an all-encompassing term to describe datasets which have reached a size and complexity that makes them difficult to process with a standard grade computer in an acceptable amount of time. Operations that are being considered within the big data nexus consist of data capture, data storage, data transfer, data analysis, data curation, data visualization, data sharing, data security & data privacy. However, this very general definition also comes with its limitations since both the size of big data and the computational capacity of computers are constantly evolving factors. The evolution of big data over the last 10 years perfectly illustrates this. Big data has evolved from terabytes in 2005 (1 TB = 1000 GB) to petabytes in 2010 to zettabyte in 2017 (1 ZB = 1,000,000,000,000 GB). While relating to the size of a terabyte is no challenge as that is the storage capability of modern desktop solutions, petabytes and zettabyte are more foreign concepts. According to Arthur (2011) one zettabyte is the equivalent of the storage capacity of 250 billion DVDs which is enough to cover 55% of earth's surface in DVDs. This rapid development is caused by the exponential growth rate that data has shown over the past decades. Due to this characteristic, 90% of data in the world at any given moment in time has been created over the period of the previous two years. (Dragland 2013)

In general, Big Data is typically characterized by the three factors volume, variety and velocity. (Sagiroglu and Sinanc 2013; De Mauro, Greco, and Grimaldi 2016) Volume is used to describe the amount of data that is created and that can be stored. It is a central feature because the more data that is available the more reliable estimations can be calculated. In turn, variety relates to the nature and heterogeneity of the data created. Since high heterogeneity with several types of data require the system to be robust and adaptable in its operations. This is important, since the various perspectives (data types) of the same data set allow for enhanced pattern recognition capabilities. Lastly, velocity describes the speed with which the data is being created. To ensure full usage of the generated data, the techniques for data processing must be on par with the data generation. The moment that data creation exceeds the analytical capabilities of the system, a new type of data called "dark data" is created. This term is used to refer to the increasing amount of unstructured data that is being generated by everyone and everything without ever being utilized. (Kambies et al. 2017)

4.5.1. The data journey in CPS for Industry 4.0

The real-life use of Big Data for CPS in Industry 4.0 can be divided into two major applications. On the one hand, there is the important part of system architecture that describes the underlying connectivity of devices and facilities as well as the way that they are creating data. On the other hand, there is the aspect of data analytics which is used to make sense of the created data to optimize production and thus increase resource efficiency. The underlying system architecture is mainly in charge of enabling real-time communication of devices within the system. In order to achieve this constant flow of communication, the first step it fulfils is facilitating the collection of data. Data does not only serve as the basis for decision making but can also be described as the raw material of the process. In the context of Industry 4.0, the data collected at this stage includes mainly data from sensors distributed in the system ranging from RFID records, camera images, GPS devices and enterprise resource planning systems. (Li et al. 2017; Posada et al. 2015; Xu and Duan 2018) In order to reap the full benefits of this process the degree of automation should be kept at the highest level possible. This is especially important considering that any human interference results in not only increased cost but also significantly reduced efficiency. Once the data is successfully captured, it needs to be stored and made available for retrieval. However, the very definition of big data makes it evident that no single computer system is capable of storing this amount of data. Therefore, over the years two major processes have been developed. The traditional relational database system and the non-traditional database. These systems differ in their level of accuracy, redundancy and consistency as well as their inherent energy consumption. For Industry 4.0 the option of non-traditional database systems is the better choice, since it comes with significant scalability and velocity benefits. (Gölzer, Cato, and Amberg 2015)

The next step in the data journey is the one where cloud computing plays a major role. Since it was already established that data constantly needs to be processed and a single local computer does not have the necessary computing power to do so, a cluster of computers is the most viable option to fulfil this job. Depending on the requirements of the individual use case there is a variety of possible distributed computing systems that can handle the task. (He and Xu 2014) Processing is the crucial step in value creation from the collected data, since a pool of raw data does not provide any additional value. It is achieved through the application of analytical methods to the available data lake (collection of data) to identify patterns that humans are not able to see. Data analytics can be broken down into three types, everyone more complex than the previous. (Delen and Demirkan 2013) Firstly, descriptive analytics is used to describe past events. Secondly, predictive analytics try to anticipate future events considering that situations from the past will repeat in the same or in a similar manner. Finally, prescriptive analytics is used to come up with actions to react to a possible future and influence it.

Descriptive analytics is the least complex method of data analytics as it mainly utilizes statistical functions to foster basic insights into data sets for example to understand trends. However, descriptive analytics has also more complex approaches like correlations and clusters. Correlation analysis allows to identify characteristics which are changing at the same time and thus the strength of the relationship between two variables. Clustering on the other hand allows to group records of similar nature together and thus allows the creation of homogeneous groups that can be treated in a similar way. In industry 4.0, clustering is a vital tool. Grouping sensors in the appropriate sensor network allows to reduce messaging overhead and moreover, allow for better detection of hardware and software failures. (Younis and Fahmy 2004; Lapira 2012)

Predictive analytics goes one step further and focuses on past patters to predict what can or will happen in the future. This type of data analytics method makes predictions assuming that events that happen in the past will repeat or happen in a similar way in the future. To do so it uses data mining and machine learning techniques to automate the way that an analytical model is built. Usually, predictive analytics involves a dataset with many normal attributes and one target attribute for prediction. The newest discipline of data analytics, called prescriptive analytics, goes even one step further and proposes solutions to a future scenario and the necessary actions to avoid or achieve it. Prescriptive analytics is for example what allows self-driving car development. In the cloud context and its application in data centres it allows to predict load and react accordingly to increase energy efficiency. In order to be able to make those very complex decisions, advanced analytical methods and artificial intelligence are used. (J. Lee, Kao, and Yang 2014)

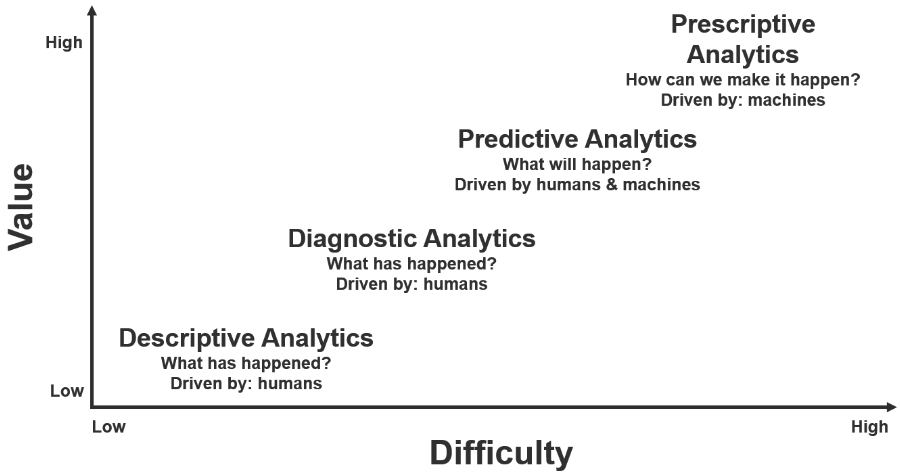

FIGURE 2 – THE DIFFERENT TYPES OF DATA ANALYTICS
(Source: own depiction after "2017 planning guide for data analytics" accessed October 15th, 2018 https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/2017_planning_guide_for_data_analytics.pdf)

5. Cloud Computing - an introduction

One key platform technology that plays an essential role in connecting the various available technologies for Industry 4.0 is cloud computing. Cloud computing in general describes the possibility of sourcing computing resources over the internet. This virtualization of resources and services serves as technical backbone to most of the technologies employed in Industry 4.0. Moreover it enables data exchange and network establishment between Industry 4.0 objects and delivers the computational power necessary to make sense of the collected information. (Candel Haug, Kretschmer, and Strobel 2016) When aiming to connect the whole value chain, from a couple machines to the entire plant and ultimately including the complete supply chain, it becomes evident how important the real-time character of these connections is. Only then it becomes possible to use the generated insights from the collected data in a value adding way. The resulting requirements for data access and sharing only underline the complex nature of cloud computing in this context. (Vaidya, Ambad, and Bhosle 2018)

In order to fulfil those requirements, cloud computing has a set of essential characteristics defined by Mell and Grance (2011) that are outlined below:

- ➔ On demand self service

  Computing capabilities, for example network storage can be supplied to users without the need for human interaction. This service is usually delivered through a self-service portal.

- ➔ Broad network access

  Resource access is possible over the network, which is usually the internet. Heterogenous platforms (e.g. mobile devices) have access due to standard connection mechanisms.

- ➔ Resource pooling

  The available physical computing resources (e.g. storage, processing power, memory, …) are pooled to provide multiple users with the demanded services. This is accomplished by dynamically assigning various available physical and virtual resources to the users. On average users have no influence on the physical properties of the accessed resources although with recent policy making practice the localization of the pooled computing resources is becoming a topic.

- ➔ Rapid Elasticity

  Increase and decrease in the scale of required computing capabilities is done automatically according to the requirements. This ensures that the exact amount of capabilities is available without employing excess computing power.

- ➔ Measured Service

  The usage of resources is metered continuously. It is monitored, measured & reported providing transparency to provider a consumer. Moreover, only used resources are billed.

The above described cloud services are usually delivered in three distinct models as described by Hogan et al. (2011) :

- ➔ Cloud Software as a Service (SaaS)

  The user can use the applications running on the cloud as developed by the cloud provider. The user has no influence on the type of applications. However, he can access the data through a web interface. The provider is in charge of managing the software on the cloud and the underlying hardware infrastructure (e.g. network, storage, servers, …) This delivery model is the most widely distributed in the end user space as it makes installation of applications on your own hardware obsolete. Moreover, the defined use cases make operation easy and reduce the need for in-house maintenance and support. H

- ➔ Cloud Platform as a Service (PaaS)

  The user can deploy applications that he created or acquired onto the cloud infrastructure if he obeys the limitations regarding programming languages and available tools imposed by the provider. The user has no influence on the underlying hardware infrastructure and its operating system (e.g. network, storage, servers, …) however he can influence the applications and their hosting environment. This delivery model is mostly targeted at developers and builds on top of the Infrastructure as a Service Model.

➔ Cloud Infrastructure as a Service (IaaS)

The user can deploy and run software that includes operating systems and applications capable of influencing and steering fundamental computing resources on the cloud (e.g. processing, storage, networks). The user does not control the underlying cloud infrastructure however he can be granted limited control of selected network components. This delivery model is of major interest for Industry 4.0 applications as it allows them to manage desired applications and middleware with the possibility of immediate scaling. Examples for this model include Amazon Web Services (AWS) and Microsoft Azure.

The described cloud services can also be deployed in different forms:

➔ Public Cloud

The cloud infrastructure is available for the public for open use. The infrastructure is on the premise of the provider but can be accessed and operated by whoever wants to make use of it (e.g. businesses, governments, academic institutions. The advantages are low operational and implementation cost however also limited level of security.

➔ Private Cloud

The cloud infrastructure is operated exclusively for one entity. Still, it does not mean that the infrastructure is owned by the entity or hosted on premise. It just ensures a higher security level and an increased level of reliability & scalability that comes at a higher cost.

➔ Hybrid Cloud

The hybrid cloud tries to leverage the best of both worlds. Through bundling together unique entities and allowing for data and application portability they provide the benefits associated with private clouds at a percentage of the cost.

5.1. Industry 4.0 & Cloud – a match made in heaven.

The characteristics of cloud computing are vital because they provide the ideal background to the requirements of the globalized connected production system that Industry 4.0 aims at creating. With cloud computing traditional system boundaries blur due to the natural convergence between IOT needs and inherent cloud characteristics. The perfect match character becomes visible when examining the requirements one by one. The requirement of ubiquitous accessibility, comprising of the need to facilitate the highest level of connectivity between the heterogenous objects and users is facilitated through the broad network access and the integration of application programming interfaces (API). This allows the users to access the required services from anywhere over the internet regardless of the device chosen. Dynamic management and orchestration require adaptability to a constantly changing amount of processing power, users and amount of data. The rapid elasticity of the cloud allows to deploy the necessary resources for a chosen period whenever a specific need arises. This efficient resource pooling characteristic of cloud computing is the basis that enables maximum resource utilization. Through virtualization, meaning the creation of a virtual version of something instead of an actual one e.g storage devices or virtual computer hardware platforms, users can easily share the available resources. Last but not least the personalization according to user profiles and requested/provided services calls for a customized approach to the available software. This personalization is realised through the measured service approach described above. (Biswas and Giaffreda 2014)

5.2. Other enabling technologies

The above described technologies were presented in a more detailed manner to allow a better understanding of cloud application in an Industry 4.0 scenario. Since they represent a core part of the whole system, it is difficult to understand the importance of cloud computing without a basic understanding of these enabling technologies. Nevertheless, there are other technological enablers of the system that also deserve a brief mention at this point. This is important because to leverage the full potential attributed to Industry 4.0, it is vital that every piece in the system plays its part. However, since there is no agreed upon definition on what constitutes a key technological enabler for Industry 4.0, the scope in the literature often varies. According to (Pilgrim, Groneweng, and Reckordt 2017) there are four core technologies. Those include the previously described CPS, Big Data and Cloud as well as additive manufacturing processes. Additive manufacturing processes or 3D printing as it is more widely known, changes the paradigm from taking resources and applying subtractive steps to arrive at the product to taking resources and applying additive steps for product creation. The drawback of a subtractive approach is the waste inherently produced in the process. With 3D printing only the needed resource amount is utilized which in turn leads to major efficiency improvements and thus a less resource intensive production. Bechtold et al. (2014) expands the list of core technology enablers to seven by adding mobile technologies, machine to machine communication as well as community platforms. The term mobile technologies include all forms of wireless communication technologies that allow us to access information with any device, at any time and from everywhere. This increased accessibility constitutes a tremendous shift, especially considering that until recently most data was only available at fixed locations. Machine to machine communication is a term that closely interlinks with CPS as it is the core technology that allows for exchange of information between CPS. The six enabling technologies described until now are mostly technological nature, while community platforms – the last ones in the list - have humans at its core. They provide an environment that fosters connections between humans and the creation of a personal network where information is exchanged, and knowledge shared, not only between devices but especially between humans. On top of that list Gerbert et al. (2015) added two additional technologies namely augmented reality and cybersecurity to arrive at a new total of nine enabling technologies. Augmented reality will give human users the chance to exceed previously known cognitive barriers by interlinking action of humans and machines on a completely different level. On the other hand, Cybersecurity will be a condition “sine qua non” that will make or break the approach, due to the previously explained possible implications cyber-attacks could have on the system and society.(Thames and Schaefer 2017)

These enabling technologies are not only at the core of the Industry 4.0 system but are more importantly fuelling a variety of important real-life applications in the overall system. Figure 3. below illustrates the main enabling technologies and their interplay with relevant value drives for digitalization of Industry 4.0. For simplicity reasons, the table focuses on the list of seven enabling technologies by Bechtold et al. (2014) described above to outline their importance for making the eight most important value drivers of Industry 4.0 function. Value drivers are the key enablers for future-proofing the industrial sector and driving digitalization of the manufacturing process. They are often grouped into the four key pillars of Smart Solutions, Smart Innovation, Smart Supply Chains and Smart Factory. While the first two mainly contribute to enhanced growth potential, the latter two contribute to efficiency gains. Smart Solutions are very much focused on the market side of the manufacturing process and contain both, Smart Products, such as the NEST thermostat, and Smart Services such as innovative implementing technologies. Smart Innovation translates into distributing innovative ideas and solutions and is done on the level of connected lifecycles and extended innovation. On the other hand, Smart Supply Chains are horizontal and vertical integration via agile collaboration networks and connected supply chains. Lastly, Smart Factory mainly refers to decentralized production control and data-driven operational excellence. What stands out in this table is that cloud computing and advanced analytics are at the core of the system and crucial enablers for almost all value drivers. Another major trend that can be observed is that Mobile and M2M are especially important for Smart Solutions and Smart Factory but also for connected lifecycle innovation and connected supply chains. On the other hand, it becomes obvious that the overall system enabler effect of Community Platforms, 3D Printing and Advanced Robotics are limited to certain value drivers. It can be observed that Community Platforms are especially crucial for realizing extended innovation and agile collaboration networks, while 3D printing and advanced robotics are mainly involved in enabling decentralized production control.

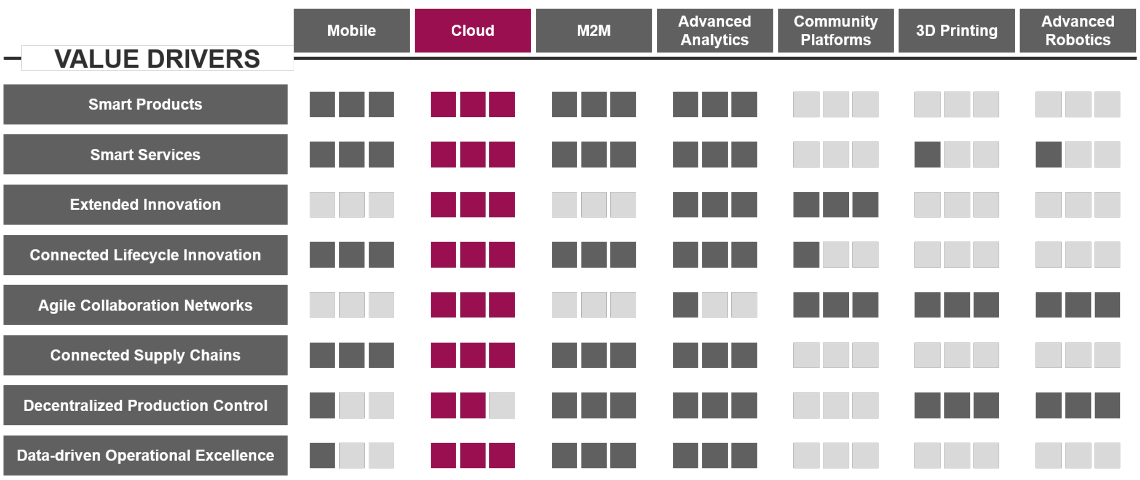

FIGURE 3 – CLOUD AS CORE ENABLING TECHNOLOGY
(Source: own depiction after Bechtold et al. 2014)

6. Energy Consumption & related Greenhouse Gas Emissions from ICT & Cloud Computing

Assessing the (net impact) on energy consumption by information and communication technology is a topic that that has been gaining significant academic attention in recent years. Especially, since several technologies that are currently on the rise are driving substantial disruptive change and their lasting effects are yet to be determined. The most prominent among these technologies, besides cloud computing, is the upsurge in high-speed wireless access network that is promising to connect the other 4 billion people on our planet to the internet (Finley 2015). The other one is the rise in thin clients. The term thin client describes devices like tablets and smart phones and while they are more energy efficient in certain use cases than laptops and desktop computers, their total effect on the system is yet to be determined. (Pattinson, Cross, and Kor 2015; Miettinen and Nurminen 2010) At the same time, older technology that is used for processing and storage of information is becoming more efficient in its total energy consumption. (Erol-Kantarci and Mouftah 2015) Moreover, usage patterns are being transformed as ICT is finding its way into our daily life and transforming almost every aspect of it.

Since the introduction of the green computing movement, which in its core describes the environmentally responsible use of technology, over two decades ago, a lot has happened. One of the first major initiatives that was born out of this movement was the ENERGY STAR initiative that introduced specifications for energy efficient computers and monitors. (Harmon and Auseklis 2009) Over the years, as the topic matured, there have been several studies trying to estimate the global energy demand of ICT as well as the accompanying amount of CO2 generated. However, these studies have been carried out with varying approaches and have thus presented very heterogeneous results. This lack of uniform methodology makes it not only difficult to compare results but also difficult to predict trends. Nonetheless, this lack never stopped scientists from conducting their research albeit with varying degrees of success. A study that raised huge controversy among the scientific community was conducted by Mills (1999). According to him, the electricity demand of the United States of America was going to rise by about 30% to 50% of total electricity demand over the following years mainly fuelled by increased internet usage. Shortly afterwards, numerous scientific rebuttals were published which served to put the results of Mill's study into perspective. (Baer, Hassell, and Vollaard 2002; Kawamoto et al. 2002)

Today, considering the global scale, the direct electrical energy consumption, which describes the electricity consumed to run ICT devices, of the ICT sector is relatively small. In the year 2015, the share of ICTs in total electricity consumption amounted to 1700 TWh out of the approximately 22000 TWh electricity that was used on the global scale. That only amounts to approximately 7% of electricity demand of the ICT sector. (Malmodin and Bergmark 2015) Another study with comparable scope conducted by Van

Heddeghem (et al. 2014) estimated that ICT is responsible for 5% of electricity use and that it is growing with more than double the pace at 7% compared to the overall global electricity consumption growth rate. In the global context electricity consumption increased by 3,1% to 25 570 TWH in 2017. This growth is predominantly fuelled by the recent increased economic output in Asia especially China and India which are responsible for 70% of this development. At the same time, this growth does not go hand in hand with an increase in energy efficiency as. Actually, improvements for overall energy efficiency of ICT equipment have declined due an overall slowdown in implementation of energy efficiency policies. (IEA 2018)

The first comprehensive review of electricity consumption of data centres focussing on the US market and estimation of the size of their contribution to the global ICT demand was done by Koomey (2007, 2008). It has proven to be a challenging task since studies and reports with that focus were scarce at that point. Moreover, most companies were not providing access to this data and rapid technological developments in the area were only adding additional difficulty to the estimations. Therefore, Koomey devised a projection approach that multiplied the number of estimated installed servers with their average power consumption. The average power consumption was calculated through addition of the average power consumption of the mostly utilized individual parts that make up a server. This approach allowed to arrive at the total electricity consumption. Following this approach, about 0,6% of total electricity consumption in the US could be attributed to US based data centres in 2005. On the global scale, the study showed that data centres would consume 1% of the total electricity demand in the same year. Five years later, Malmodin et al. (2010) calculated that the ICT sector as a whole demands about 3,9% of global electricity and that data centres account for a mere 1% of that. However, since forecasts were predicting a rise of data centre electricity demand by 76% from 2005 to 2010, the question arises if there was no growth in the installed server base in the timeframe. The answer to this question was provided by J. Koomey (2011) a couple of years later. When he reviewed his earlier assumptions, he discovered that the overall increase in data centre electricity demand, during the above-mentioned timeframe, was 50% lower than previously anticipated. As a result, the amended study calculates that worldwide data centre electricity demand is around 1,1% and therefore in line with the results presented by Malmodin et al. (2010). The moderate rise in electricity demand of data centres can be attributed to overall efficiency gains in the hardware used and other introduced energy efficiency measures. Those measures have proven to be able to make up for a big part of the additional demand created by the growth in data amount. While electricity consumption for data centres increased by about 90% from 2000 to 2005, the increase from 2005 to 2010 only amounted to approximately 24%. Most recently, from 2010 to 2014, the increase amounted to only 4% (Bein 2018).

It is important to note that cloud storage and computing can be considered a subsystem in the data centre world. Therefore, when further splitting up the system into smaller pieces to only consider electricity consumption of cloud computing, the same challenges that were presented by J. G. Koomey (2008) also apply. Both, the lack of available data and uncertainty regarding the appropriate scope of analysis are posing significant challenges. Nevertheless, this document presents an approach that allows an estimation of the electricity demand of cloud computing for Industry 4.0.

6.1. Scope of Cloud Computing systems

To understand & quantify the electricity consumption of cloud computing it is important to quantify the system in question. While the characteristics of the cloud have been described in the previous chapters, the components that make up the cloud will be described below. This will allow to better understand why there are differences in the previously outlined calculations as well as the electrical efficiency of various data centres.

At the heart of cloud computing are connected data centres. A data centre is in its core a collection of servers organized in clusters where data is stored and organized. Those clusters are usually mounted onto racks which are usually organized in rows. While in traditional data centres servers usually are isolated and run only a predefined set of applications, servers in the cloud are all connected and thus create one large computer. However, this computer has not much in common with a computer that a typical end-user is accustomed to. One key difference is that information is stored on numerous servers in different data centres to ensure the information's safety, even in the case of hardware failure. (Barroso, Clidaras, and Hölzle 2013) Besides servers, a typical data centre consists of switches, Power Distribution Units (PDU), Uninterruptable Powering Supply (UPS) and a cooling and ventilation system. Regarding electricity consumption the cooling and ventilation system is the biggest consumer as can be seen in Figure 4. below. The reason for that is that servers produce lots of excess heat during operation. Since servers perform best and life longest at around 22 degrees Celsius, the data centre needs to be constantly cooled. (Moore et al. 2005) This is also the reason why most data centres can be found in climate regions that are naturally colder. As such the largest data centre in Europe and the fifth largest in the World is currently being in the final stages of construction in Norway. The data centre is projected to initially draw in 70MW of power, and at full build will measure as much as 0,6km² which equals 112 football fields and have the ability to process 1,000 MW by the year 2027 when it is estimated to reach full capacity. At the same time it is expected to be fully powered by renewable energy sources .(Marques Lima 2017)

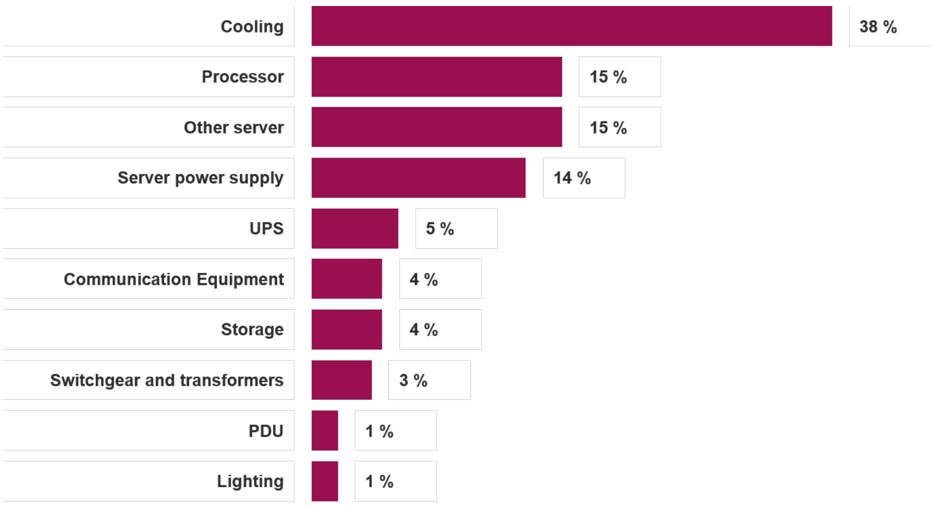

FIGURE 4 – ENERGY DEMAND ACCORDING TO EQUIPMENT TYPE
(Source: own depiction of information provided by Clarke Energy. Energy demand distribution in data centres as of 2015, by use. https://www.statista.com/statistics/602629/worldwide-data-centers-energy-demand-distribution/ (accessed 10/6/18, 12:24 PM).)

6.2. Green House Gas Emissions

There is no doubt that the topic of climate change and its implied negative effects on the environment and human health are major global challenges that are currently being widely addressed in both governance and business as well as among society as a whole. According to the latest IPCC report (IPCC 2014), rising GHG emissions, are the major cause of anthropogenic driven climate change. Unfortunately, the production of electricity is still inherently coupled with the production of a significant amount of GHG emissions. According to the IEA (2018) global electricity consumption was 25 570 TWH in 2017 as previously mentioned.. Since the electricity demand of the developing world will only rise in the coming years, the scientific and academic community predicts a significant upward trend for this figure. Especially, since these countries are still in earlier stages on their way to industrial transformation. These developments in countries from the so-called Global South are fuelling estimates that show an increase of up to 75% in electricity demand is likely to occur.

Due to the significant amount of electricity consumed and its continuous expansion, the ICT sector is counted among the contributors to rising global GHG emissions. This is especially concerning when considering the fact that most data centres today are not powered by renewable energy. Therefore, cloud computing can be counted to the larger contributors of GHG emissions among the total ICT sector. According to Pettey (2007) the ICT Industry emits around 2% of worldwide CO2 emissions. However, data centres are responsible for almost 23% of these. On the other hand, Malmodin et al. (2010) estimates that the ICT sector contributes 1,3 % of global greenhouse gas emissions. Both studies used a similar base methodology, by taking 2007 as their base year but estimated different metric tons of carbon dioxide (MTCO2) with 860 MtCO2 emissions for the former and 620 MTCO2 for the latter. Still, both studies agreed on the fact that of their respective MTCO2 result data centres were responsible for 0,5% of GHG consumption. While the estimations are in the same order of magnitude, a detailed review shows that the base definition of ICT equipment varied and that the lifecycle assessment also differed, making a comparison flawed at the very least. More recent studies have been dealing with the same challenges of limited data availability. Nevertheless, most recent estimations display that globally between 1 and 1.5 gigatons equivalent of greenhouse gas emissions are associated with Cloud Computing activities. These amounts account for approximately 2.5 percent of the global greenhouse emissions. To put it into perspective, these emissions equal to the global share of greenhouse gas emissions from Germany. (Andrae et al. 2015)

A major role for affecting the amount of GHG that are emitted due to powering data centres is the energy mix that is used to power them. Since this strongly corelates with their localization on the world map a closer examination is necessary. Currently, most data centres are located in North America and Western Europe Estimations. Both regions combined are hosts to more than one thousand data centres. Meanwhile, Asia has close to 400 data centres, while Africa and South America have 80 data centres altogether. (Gold 2014) Taking Europe as an example and comparing the CO2 emission intensity for electricity generation it the scope of the difference becomes evident and can be examined in Figure 5. The sample countries have been chosen according to the amount of present cloud data centres. (“Colocation Western Europe” 2018) It becomes evident that powering your data centre in the UK results in a CO2 emission intensity that is numerous orders of magnitude higher compared to for example Sweden. Therefore, it should be expected that in the future data centre operators will opt for colder regions with the necessary focus on renewable energy resources. However, organizations such as Greenpeace have concluded that the biggest data centre growth globally will happen in

China. Hence, they urge IT companies to raise awareness and support renewable energy sources as this growth happens. In addition, China produces its electricity mainly from coal and natural gas. On the other hand, there are diverging opinions as of the location of new future data centres. Some expect them to happen in Europe, North America, and Asia, while Africa and south America are expected to have new small data centres built by 2020. (Oscarsson 2014)

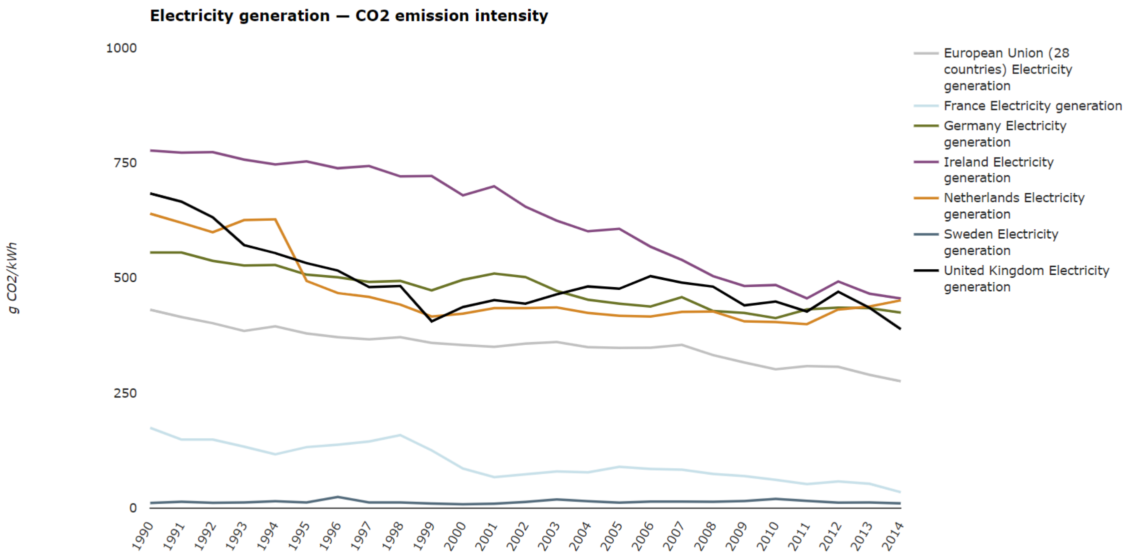

FIGURE 5 – ELECTRICITY GENERATION CO2 EMISSION INTENSITY ACCORDING TO TOP CLOUD COMPUTING DATA CENTRES

(Source: own depiction created with European Environment Agency CO2 emission intensity data visualization tool. Accessed October 10th

Source: https://www.eea.europa.eu/data-and-maps/daviz/co2-emission-intensity-3#tab-googlechartid_chart_11_filters=%7B%22rowFilters%22%3A%7B%7D%3B%22columnFilters%22%3A%7B%22pre_config_ugeo%22%3A%5B%22European%20Union%20(28%20countries)%22%3B%22France%22%3B%22Germany%22%3B%22Ireland%22%3B%22Netherlands%22%3B%22Sweden%22%3B%22United%20Kingdom%22%5D%7D%7D

6.3. Energy Consumption Metrics

As shown in the table below there is a number of metrics that can be used to describe electricity consumption.

TABLE 1 – ENERGY CONSUMPTION METRICS

Benchmark	Metric	Level	Domain
Total Power Consumption	$ cost of power consumed Kilowatts used	Data centre	Enterprise
Green Grid PUE	Ratio of facility power to It equipment power	Data centre	Enterprise
Green Grid Die	Percent of power that reaches IT equipment	Data centre	Enterprise

(Source: own depiction of information provided by Oscarsson 2014; (Belady et al. 2008)

Power Usage Effectiveness (PUE) is a thermodynamic indicator introduced by Malone and Belady (2006). It compares energy consumption to energy efficiency during the operations of a facility. This means that it compares the power that is used to run the IT equipment in the facility to the power consumption of the facility as a whole. The IT equipment power is then defined via the consumption of the installed computers, network equipment and peripherals. The power consumption of the whole facility can be measured at a utility meter. (Belady et al. 2008) Since it is useful to provide a high-level insight into the data centre the metric has found the widest level of adoption to measure energy consumption in the ICT industry. (Kant 2009) Another metric that has found wide adoption is the Data Centre Infrastructure Efficiency (DCiE) that is just the inverse function of the PUE. Therefore, it allows to denote the IT electricity consumption as a percentage of the total electricity consumption of the data centre. Both metrices are considered useful tools as they allow an easy assessment of the overhead of power at any given data centre. At the same time, when considered over an extended timeframe they provide insights into variations of usage effectiveness. (Brady 2016)

Although widely adopted, both metrices have their limitations. To calculate PUE, energy consumption of parts, that are fundamentally different in the way they are built and operated from each other, must be calculated. Therefore, to arrive at plausible data, accurate measurement of the actual load of the IT equipment is crucial. However, this is often ignored. Instead the power use of the equipment is estimated and calculated which often leads to inaccurate results. (Itoh et al. 2010) Nevertheless, due to the wide availability of the PUE metric measurement for various types of data centres, it will be considered in the upcoming calculations and used to describe the efficiency of cloud data centres.

7. Cloud Efficiency Discussion

Before talking about energy efficiency, it is imperative to understand the concept. According to Moisan and Bosseboeuf (2010)

> "Energy efficiency can be defined as a reduction of energy used for a given service or level of activity"

The efficiency improvements that have happened in the computational sector are impressive to say the least. Nevertheless, the total amount of energy needed for its operations have been continuously growing. Such a development is no surprise as it can be observed on numerous occasions in the past. It was first conceptualized by Jevons (1906) in his work "The coal question" where he explained how technical progress aimed at increasing coal mining efficiency would lead to an increase in its overall demand that in turn would lead to a coal shortage. The public response at that time was disbelief, however, history has proven him right. Even though no coal shortage ever occurred, and coal was later replaced by oil, efficiency gains in coal mining inevitably lead to an increase in coal demand. Jevons' theory on energy efficiency effects was later transformed into what is known as the Jevons Paradox or the Rebound Effect. Over the years the theory has been supported with a lot of experiential evidence. (Alcott et al. 2012; Bauer et al. 2009)

Although the rebound effect has been primarily described in the context of psychology it also finds its application in the ICT sector. A prime example for it is Moore's law that states that the microchip performance and thus efficiency per cost unit doubles every year and a half. Since this theory was introduced in the 1960s and held true for almost 50 years we energy consumption in the ICT sector should be at an all-time low. However, the opposite is true. The growth mindset that is fuelling the globalized economy, led to an increase of applications for ICT technology and consumption patterns. This can be observed when examining the installed computer base between 1980 and 2008 which doubled every three years. Ergo, the observed use of electricity for computing has steeply increased. Thus, energy consumption has increased event though todays transistors are approximately 30 million times more efficient than 50 years ago. (Brettel et al. 2014) Applying the same efficiency gains to a VW Beetle from the 1970s would result in a car that is able to travel at a speed of approximately 500 000 kilometres per hour or 418 times the speed of sound, for 3 200 000 kilometres on one gas tank and that would cost around 4 dollars. Going back to the ICT sector Andrae et al. (2015) believes that rebound effects can offset energy savings at the macro-level completely through the increase in the amount of data transferred.

Currently, it is not certain that Moore's law can continue to apply because transistor technology is reaching physical limits. On the one hand the heat that is generated by ever smaller transistors that are being bundled together in an ever-closer space is becoming an issue that cannot be addressed by traditional cooling techniques. On the other hand, fundamental engineering limits are being reached by the microprocessor industry as certain chip features have become too small to manufacture. Currently microchip circuit features are as small as 14 nanometres which makes them smaller than most viruses and they are expected to further reduce its size to a couple of nanometres. Due to this limit and the fact that there is no innovative technology to replace the silicon transistor technology in sight, the approach in the industry is shifting towards an application centric approach. Instead of developing the tech and building the applications around it, the applications (e.g. supercomputers, cloud data centres) will be at the centre and the technology will be built around it to enable it best. (Waldrop 2016)

This development does however carry implications for further energy efficiency gains that will no longer be accomplished by technological development. Overall Shehabi et al. (2016) has proposed an overview of the taxonomy of effects that can be expected in the future. As can be examined in Figure 6. the described effects are part of a bigger picture that can be divided into first, second and third order effects that have positive and negative net energy consumption effects. Taking into account this complexity it becomes evident why it is challenging to agree on the net effect that these developments will have.

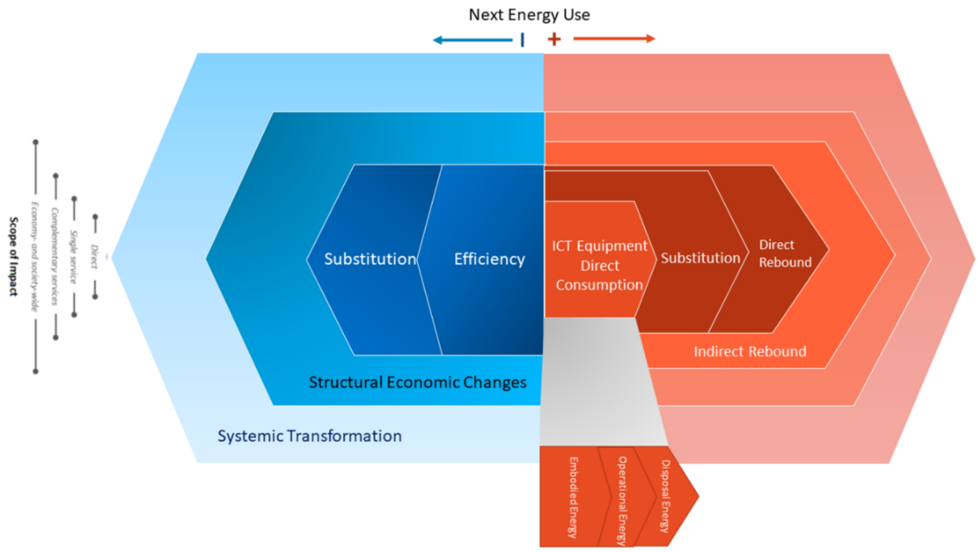

FIGURE 6 – VARIOUS ENERGY EFFECTS
(Source: own depiction created after Shehabi et al. 2016)

7.1. Approaches to green the cloud.

As a result, there are multiple approaches on how to reduce the contribution that the cloud has to GHG creation. It should not come as a surprise that the most researched topic in this area is optimization of energy consumption. After all less energy consumed has a direct monetary effect for the data centre. According to research conducted by Radu (2017) the topic of optimization of date centre energy consumption has gathered 4 times more scholarly articles than the topic of GHG emissions from data centres.

In general energy-saving solutions are designed to reduce resource consumption as far as possible while still being able to function to complete necessary tasks. In practical terms, this means putting as many currently unused devices or functions of a device into sleep mode to maximize resource efficiency. Academia has established three main categories of computing-, communicational- and infrastructure energy consumers in this context. The energy consumption of all three groups can be optimized via both software and hardware solutions. However, the most important base factor for resource efficiency is the setup of a so-called “green cloud infrastructure” including only energy efficient components from servers to storage to lighting and cooling. While hardware improvements primarily refer to measures of dynamic power management (DPM) and voltage frequency scaling (DVFS), software optimization is more complex including different high performance and efficient computing modes. Since 2009, there has been enormous academic and scientific interest in the use of algorithms based on so-called virtual machines (VM) for the optimization of cloud resource use. Especially, since the possibility for reduction of energy demand for conventional cloud applications lies at around 27%. Another important factor in the overall optimization process of cloud computing refers to creating the right environment in data centres. On the one hand this means optimizing the airflow inside the data centre, which is usually the main cooling solution and can be further improved using a smart controller. On the other it means that choosing the right physical location of data centres in temperate climates is important as it reduces the energy necessary to cool the ambient air. This again, can be counted towards hardware improvements. A good example in this regard is the GENETIC project, which used and innovative and integrated optimization strategy to reduce energy consumption. This included among others the generation of power locally but also recovery of waste heat and energy efficient data centre cooling.(Torrens et al. 2016)

7.2. State of the art optimization by Google

How far the solutions for energy savings can be pushed has been tested and implemented in numerous real-life cases. One prime example is the approach that Google and its engineers have taken. Although not primarily linked with Industry 4.0 applications, Google operates numerous hyperscale data centres to power everything from its search algorithm to YouTube. In order to ensure that operations run as energy efficiently as possible, Google's in-house developed artificial intelligence (AI) solution is being applied to cut data energy use through what is today known as prescriptive analytics. The AI was first introduced in 2016 and through predictive analytics able to produce recommendations on how to operate the data centre that were then checked and implemented by humans running the operations. This approach resulted in an efficiency gain of around 40%. However, this was just the beginning. Recently the AI took over operations and now completely steers the data centre without human interference. According to George (2018) the AI is able to predict temperature and the load of the data centre up to 60 minutes in advance, which allows it to adjusts the systems accordingly to utilize the least resources possible, This is enabled by constant data collection from a network of sensors installed in the cooling system. Every five minutes the AI updates data and runs it through its neural network. The AI neural network is in its core a statistical model that is recreating to some degree the biological neural networks of the human brain. The network then anticipates the effect on energy consumption of different configurations, consequently enabling the AI system to identify the strategy that will use the least energy possible. This policy at Google is also clearly demonstrated in the energy efficiency numbers of its data centres. On average, they ran on a PUE of 1,12 in 2017. This means that the overhead energy consumption during operation for the infrastructure is only 12% of the energy consumption that powers the IT components. The scope of this achievement becomes visible, when taking into consideration that the average PUE of data centres lies at 1,6 and 1 is considered as full efficiency. ("Efficiency: How We Do It – Data Centers – Google" 2018)

7.3. Future approach – Edge Computing

Although the energy consumption of data transmission to the cloud is not a prominent research issue, the stability of the infrastructure is. Traditional cloud architecture is facing an exponential increase in data that must be transported from physical assets to the cloud for storage and analysis. With all the advantages that cloud brings to distributed environments it has its limitations for specific use cases. Whenever the use case includes time sensitive processes, cloud solutions are less suited due to a number of reasons. First, the cloud latency, meaning the time it takes from request to response of the cloud bases service, plays an important role. Second, the time that it takes to transmit the data and get a response can be impaired by limited bandwidth support due to remote areas. Therefore, whenever the requirements of the service don't allow for cloud application or simply would not benefit from it, a new approach is needed. Over the past years two new solutions called fog computing and edge computing have been introduced. Both address the described issues by moving processing capabilities and intelligent analytics closer to the place of origin. While both terms are often being used interchangeably leading to confusion there is a way to differentiate both. Fog computing incorporates the computing capability at the level of the local network thus allowing for the data that needs processing to be processed in a Fog Node or IoT gateway close to the origin of the data. Nevertheless, Fog does not replace cloud. Instead both are connected and complement each other. Edge computing takes computational & communicational capability one step closer to the source and integrates these functionalities directly into devices like programmable automation controllers. (Shi et al. 2016)

The possibility to reduce energy demand due to convergence of the technologies can be examined across 4 parameters. In some use cases energy consumption of the involved network to transport the data is a factor that can prove to be relevant. (Jalali et al. 2014) Reducing or even eliminating the steps to move the data can therefore have an effect on reducing the total energy demand under the condition, that the fog infrastructure is already in place. Otherwise a longer timeframe needs to be considered to account for the resource consumption of installation of the fog computing capabilities. The next parameter worth examining is the idle power consumption. Depending on the time between operations that need to be carried out, fog or cloud applications might result in higher efficiency. Due to their architecture, fog data centres cannot employ the same energy saving tactics that cloud data centres can. (Jalali et al. 2016) Another parameter to consider is the type of application. While fog computing is more energy efficient in continuous use cases with low computational requirements that are close to the data source, cloud has superior energy efficiency in most other use cases. (Deng et al. 2016) Last but not least the topic of virtualization and network management are an important factor for energy demand. According to (Al-Azez et al. 2015) it is possible to optimize energy consumption of Fog infrastructure via a layered architecture model where Virtual Machines process the IoT generated data in mini clouds close to the point of the origin of data. Nevertheless, it has yet to be determined which approach will result in being the most energy efficient one. Most likely a combination of cloud, fog and edge computing will prove to deliver the desired results in the most energy efficient way.

8. Methodology

This chapter aims at outlining the quantitative analysis of cloud computing in the context of energy efficiency and Industry 4.0 carried out over the course of this paper. After a short introduction to the general mixed approach and its inherent limitations, the basic parameters of this quantitative analysis will be further elaborated.

8.1. General Methodology

The calculations on the power demand from cloud computing for Industry 4.0 carried out are based on a mixed approach and a bottom-up analysis. This is done, to better understand the individual contribution of all parts of the cloud towards the overall energy consumption. Afterwards, the results of the calculation model are compared and checked against the reasonableness of their respective underlying assumptions. To reach this goal, it is necessary to first define the total energy and data usage by the Internet (top down). In a second step, the energy consumption per GB of the underlying parts that represent a cloud data centre will be calculated (bottom up). After successfully calculating the electricity demand of cloud computing, various scenarios will be employed to estimate to which part the cloud will be used for Industry 4.0 applications until the year 2025.

The present quantitative analysis is based on set of very specific parameters and assumptions and comes with certain limitations. Today, there is only limited relevant and recent data available, as no mandatory energy consumption reporting for cloud data centre is established. Since it was outside of the scope of this thesis to collect data, the data used for this quantitative analysis comes from a variety of sources. (Brady 2016; Dayarathna, Wen, and Fan 2016; Andrae et al. 2015) Moreover, the thesis is not able to consider the whole lifecycle of the employed hardware. While estimates on resource intensity for the production of the necessary equipment for a cloud data centre are available (Pilgrim, Groneweng, and Reckordt 2017; Marscheider-Weidemann et al. 2016) the commercial end of life phase of data centres is not well documented.

8.2. Energy demand of the cloud

The energy demand of the cloud equals the electric energy consumption and can be calculated according to the energy that is consumed for data storage, -transfer and -processing. Data is usually expressed in bytes as the smallest unit of data. To make the comparison feasible and on equal terms, a functional unit is defined as number of bytes that are stored, transferred and processed during a fixed period e.g. one year.

Furthermore, energy demand of the overall infrastructure needed to make the cloud work also needs to be included in the overall demand. The overall energy demand of the cloud can be expressed in the following formula below:

$$E_{Cloud} = E_{transfer} + E_{storage} + E_{processing} + E_{infrastructure}$$

8.3. Data Transfer

As can be observed in the formula above, data transfer from and to the data centre is one key area playing into the overall energy demand of the cloud. Therefore, its electricity consumption must be defined. To achieve this goal, the IP traffic (Internet traffic) to the datacentre will be estimated by the following calculations. First, the number of bytes transferred needs to be calculated. Following, the energy demand for data transfer per byte will be defined. Since, the consumed electricity varies depending on the network connection there will be separate factors taken into consideration to account for this system irregularity. Lastly, the data growth rate until 2025 will be considered.
According to a report by Cisco Systems (2018) the IP Traffic per month in the year 2017 accounted for 0.12 Zettabytes (ZB). Scaling this monthly number up to a year, we get a total of 1,46 ZB per year. Current projections show that this number will grow up to 3,34 ZB by the year 2021 with an annual growth factor (CAGR) of 23%. In 2017, the total traffic amounted to around 9 ZB and is expected to surge up to around 20 ZB until the year 2021. This discrepancy between calculated yearly IP traffic and total traffic comes from the fact that the data that is transferred from user to data centre only accounts for about 14%. The bulk of the transferred data namely 77% in 2017 occurs within data centres.

In order to take recent developments in the connectivity sector into account, the calculations also need to consider broadband traffic, 3G connections as well as 4G connections. Although 5G connections are currently in the developing stages, their dispersion is not expected to cross the threshold of consideration until 2025. The share between data exchange via mobile network and wide area network (WAN) can be estimated to be 50:50. To account for the different types of connections considered in this calculation, their different energy demands need to be considered. This is done via an estimation of their current development. Projected development will be considered in the scenario part that follows this section. In our base year 2017, only 26% of mobile connections were 4G compared to 33% of 3G connections. At the same time 4G connections represented a share of 69% of mobile data traffic while 3G connections were responsible for 24% of data traffic. The remaining traffic is considered to stem from 2G connections. In 2021, 4G connections are expected to account for 53% of total connections. At the same time, 4G will be responsible for 79% of traffic while 3G contribution will be reduced to only 20%. The same development that is being experienced by mobile user devices described previously also applies to M2M connections which are the most relevant for Industry 4.0 use cases. This explanation leads to the following calculation formula for data transfer:

$$E_{Transfer} = P_{2G} * C_{bytes}\, P_{3G} * C_{bytes} + P_{4G} * C_{bytes}$$

In the formula above, P stands for the power demand per connection and C for the respective number of bytes. To make the results more understandable the bytes are converted to GB to provide a value that is closer to real world applications. According to Malmodin et al. (2014) the following average energy consumption depending on mode of connection applies where:

- Fixed Broadband use 0,16 kWh per GB
- 2G connections use 37,1 kWh per GB
- 3G connections use 1,65 kWh per GB
- 4G connections use 0,45 kWh per GB

Applying this to the base year 2017, the average power consumption for data transfer is 0,37 kWh per GB. To put this calculation into perspective findings of Aslan et al. (2018) are taken into consideration. That study examined 14 studies of electricity intensity of Internet data transmission networks and found that depending on the executed study the results ranged from 7,3 kWh per GB to 0,023 kWh per GB. Ten of those studies arrived at a result below 0,5 kWh. Last but not least, Koomey (2011), proposed an approach to calculate the transfer rate by simply taking 15% of the processing power consumption. This, however, would produce an estimate of 0,001 kWh per GB which does not fit to the proposed ranges by Aslan. Therefore, when taking those results into consideration, the average power consumption for data transfer would result in **0,18 kWh per GB.**

8.4. Data Centre Storage

The next important parameter for defining total energy demand of the cloud is data storage. For the purpose of this calculation, data storage equals data centre storage.

This value can be calculated by multiplying the amount of data in bytes to be stored per year (n_{bytes}) with the energy used per byte per year (C_{bytes}).

Therefore, the equation is as follows:

$$E_{storage} = n_{bytes} * C_{bytes}$$

The number of stored bytes (C_{bytes}) will be assumed at 2017 values. According to Cisco Systems (2018), there were 0,87 ZB stored in the cloud in 2017. This number is significantly lower than the amount of data created that has been discussed in a previous chapter. This effect can be largely explained by the fact that an increasing number of the created data is only temporary and is therefore discarded shortly after its creation. Moreover, most data is still stored on personal devices. (Workman 2018) The compound annual growth rate (CAGR) until 2021 is projected to be 58%. This value translates into roughly 2,3 ZB of stored data. The second part of the equation above, looks at the energy consumption of the storage devices (n_{bytes}) which are enterprise level hard drives. Currently, most data centres operate 4TB and 8 TB hard drives. The former accounts for approximately 65% and the latter for around 28%. (Klein 2017) The remaining 7% are usually 3TB drives that have been in operation the longest and are being replaced upon failure by their less energy intensive successors. The 4TB drive, immediate successor, consumes 6,7 W or 1,68 W per TB. On the other hand, the more energy efficient 8TB consumes 8 W which amounts to 1W per TB when idle (Hormann and Campbell 2014). Further, it needs to be assumed that each hard drive alternates its time 50/50 between idle and work mode and that power consumption increases by 40% when the drive is in work mode. Consequently, when calculating the energy demand these factors need to be taken into consideration. However, since data centres operate 24/7 their hard drives are also always-on. Thus, it can be assumed that they never enter sleep mode, which would harbour significant energy consumption reduction potential. Instead they only alternate between work and idle mode. The overall result of average energy demand for data centre storage as calculated via the above explained method is **0,0162 kWh** per GB per year. The table containing the pertaining detailed calculation can be found in Annex 1.

8.5. Data processing

The third parameter needed to calculate the overall energy demand of the cloud is data processing. Data processing refers to the processing of raw data into readable formats via data modification and manipulation. It heavily relies on the capabilities of the computational system, which in turn is dependent on the performance of the processors. This is measured by the clock rate given in hertz or its multiples and the instructions or computations that it can perform per second. (Oyanagi 2002) Therefore, the functional unit that was proposed at the beginning of this section is limited in this regard. As a result, calculations on a global scale have not been found. However, Koomey (2011) discovered that the relation between energy consumption for storage and energy consumption for processing is 1:4. Applying this methodology, to the previously calculated energy consumption for data storage the power consumption for data processing can be assumed as **0,065 kWh per GB per year.**

8.6. Cloud Energy Consumption

As previously shown, a major part of the energy consumption of every cloud data centre is the surrounding infrastructure that is necessary to run and cool the IT equipment. Due to the fact that there are numerous types of equipment used and also numerous approaches to cloud data centre cooling, a detailed explanation is beyond the scope of this work. While the usual approach is air cooling there are also liquid and "free" cooling approaches. On top of that it is difficult to find a direct correlation between the number of processed bytes and the infrastructure energy consumption. Therefore, the Power Usage Effectiveness (PUE) will be used to account for the data centre infrastructure. Since the number varies considerably between data centres ranging from around a PUE of 2 for regular small and medium size cloud data centres to a PUE of almost 1 in modern hyperscale data centres, an average will be considered. According to Ascierto (2018) the average PUE in 2017 was 1,58 which in turn shows that on average the efficiency improvements have been stagnating since 2013 where the average PUE was 1,65. Since the Energy Consumption for the IT-systems in the cloud has already been calculated, the total energy demand can be calculated in the following way:

$$E_{cloud} = E_{IT} + E_{Infrastructure}$$

Where:

E_{IT} … is the result of the above calculation of energy demand for storage, processing and transfer.

$E_{infrastructure}$ … is calculated through the multiplication of E_{IT} with the average PUE.

Applying this methodology, power consumption for a cloud data centre is approximately **0,413 kWh per GB per year.**

9. Results - Energy consumption of the cloud for Industry 4.0

In order to consider the impact of Industry 4.0 on cloud computing the first step is to attempt to quantify the amount of data that is being produced in this particular context.
According to a projection made by Oracle (2015) the number of cellular enabled factory devices reached 610.000 units at the end of 2017. This point marked a growth streak that has been sustained since 2012 with an average CAGR of 52% and that is expected to continue. In the same year, the sales of industrial robots amounted to 387.000 units. The sustained 10 percent year on year growth in the sales volume of industrial robots across the globe has led to an installed base of 1.800.000 units. This number is expected to double until to 3.053.000 units in 2020 and keep growing at a CAGR of 14% which would result in 5.878.291 Industrial robots in the world. Considering the predictions from the business side the forecasted sales volume in 2025 is estimated at $18.620.000.000. (Tractica 2018) Taking into consideration that according to (Mahto and Hemnabh 2018) the average price for an industrial robot is $ 46.000, this approach would estimate only 404 800 which is only a relatively small increase in sales per year. This can be most likely attributed to the fact that by the time, the market will have matured and two-digit growth will be difficult to accomplish.
To understand the approach of those calculations it is important to know that industrial robots can be classified as representatives of a cyber physical system and are therefore an important enabler of Industry 4.0. As a result, they can be used as one approach to derive the amount of created data from Industry 4.0. Especially since there are estimates by (Craig 2018) that a smart industrial machine creates an average of 5GB performance data per week, resulting in 260 GB of data per year. Taking that as the base and multiplying the number by the installed base of industrial robots, the expected amount of data created reaches 468.000.000 GB which translates to 486 Petabyte. This is a relatively small number compared to the amount of gathered data that Intel reported in their case. The microchip producer introduced sensors and data analytics into one of their microchips with the goal to anticipate equipment failure and thus react proactively. They were able to leverage an impressive optimization potential by reducing spare parts cost by 20%, cutting maintenance time in half and achieving a 25% increase in yield. In order to accomplish that, 5 terabytes of machine data were captured during every hour the machine was in operation. Assuming operational time of 18 hours a day the machine produced around 33 petabytes in one year. Further estimates expect a smart factory to produce 5 petabytes of data that needs to be analysed per week. For a year this again

results in 240 Petabytes. Considering that the World Economic Forum has identified 1000 fully enabled smart factories, the creation of 240 000 petabytes can be expected. (Twentyman 2018) The energy demand of cloud computing for Industry 4.0 applications can be therefore calculated the following way:

$$E_{Industry\ 4.0} = E_{cloud} * n_{bytes}$$

Where:

N_{bytes} … is the amount of data created through Industry 4.0 (displayed in petabytes for better understanding)

Applying this methodology, the cloud energy demand that is used for Industry 4.0 is approximately 99 TW/H in 2017. Considering that previously estimated total ICT energy demand of 1700 TW/H then Industry 4.0 only contributed to around 6% of the cloud energy demand.

9.1. Future Scenarios

Predicting future scenarios in the ICT sphere is comparable to looking into a crystal ball. Most of the ICT-infrastructure and hardware that is regarded as standard today, was not available a mere 10 years ago. To put it even further into perspective, the first iPhone was launched in 2007 and the Amazon Web Service one year prior to that. Therefore, it is possible, that the next disruptive innovation in the cloud space will again transform the system beyond recognition. Nevertheless, there are forecasts by leading industry experts that will be used to get an idea of the scope of the future developments.

In order to model the future scenarios two variables will be used that will be bundled to scenarios which are most relevant. The first changing variable is the amount of data created by Industry 4.0. For this, on the one hand the approach that has stood the test of time in regard to data growth will be applied. The amount of data has been doubling consistently in very short cycles in the past and will be assumed to increase at an annual rate of 50%. (Fettweis and Zimmermann 2008). This represents the exponential scenario. On the other hand, Industry 4.0 is expected to grow at a CAGR of 23,1% till 2023 which will represent the more conservative approach. (Sullivan 2016) Since the scenarios are being considered until 2025 the same CAGR that was calculated by Sullivan will be applied to the remaining two years that go beyond the scope of the used report. The amount of data produced is therefore expected to grow to 1.266.000

Petabytes in the conservative approach and to 6.150.937 Petabytes for the exponential scenario.

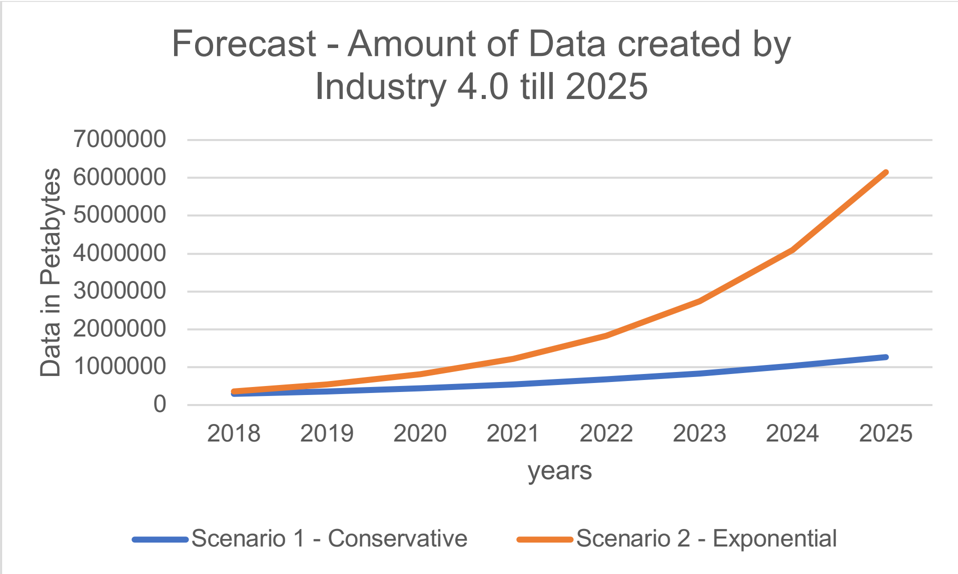

FIGURE 7 – CREATED DATA BY INDUSTRY 4.0 APPLICATION
(Source: own depiction for illustrational purposes of the applied scenario calculation methodology)

The second variable that is considered is the cloud centre energy efficiency. Since optimization of the installed IT Hardware is slowing on average the optimization potential needs to be leveraged on the infrastructure side which is only possible with a wide-reaching migration to hyperscale data centres. Therefore, the base scenario assumes that current PUE will remain as a constant until 2025 representing no move to hyperscale data centres. On the other hand, the hyperscale scenario assumes that a rapid increase in hyperscale data centres will drive PUE significantly down. The PUE for Hyperscale Datacentres will be used from Google's insights and amount to 1,12. This is in line with the expected hyperscale growth rate of almost 100% from 2016 levels until 2021 and another doubling until 2025. Leading to an increased coverage of hyperscale data centres making them responsible for 53% of the installed server base in 2025. When applying these assumptions, the average PUE in 2025 results in being 1,35. Putting the 2 variables into the scenario logic, the following 3 scenarios will be examined.

Scenario I - The worst-case

Considers exponential data growth rate with the base PUE factor. Upon entering this input into the calculation model the expected energy consumption from cloud computing for Industry 4.0 is 2540 TW/H which is more than the total considered ICT demand in 2017 and approximately 10% of global energy consumption.

Scenario II - Business as usual

Considers slow growth and a base PU factor. Upon entering the Input in the Model, the expected energy consumption for Industry 4.0 results in 522 TWH which amounts to a third of the ICT demand in 2017 and approximately 2% of global energy consumption.

Scenario III - Most likely

Considers exponential data growth rate with the Hyperscale PUE factor. Upon entering the input in the model, the expected energy consumption for Industry 4.0 results in 2168 TWH which is more than the whole ICT sector in 2017.

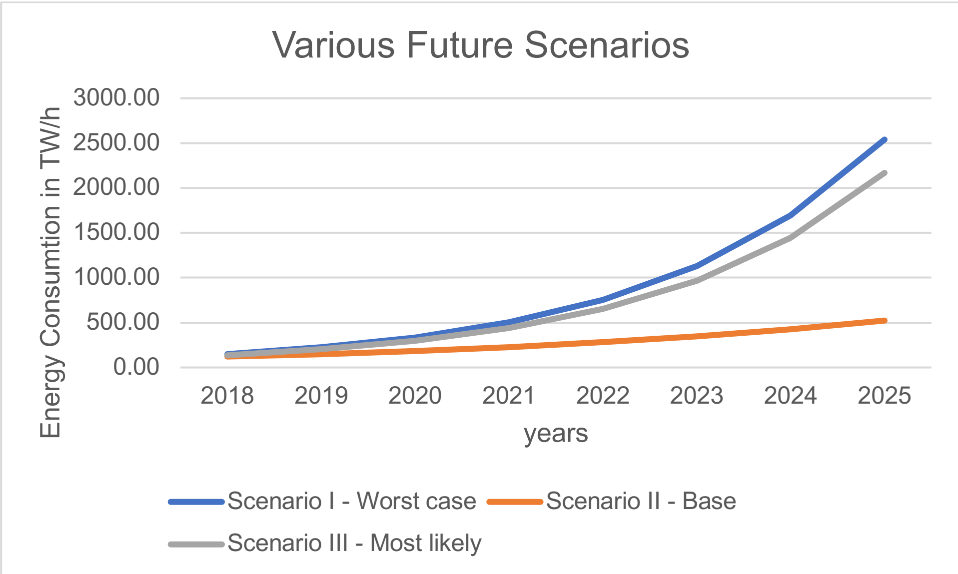

FIGURE 8 – CALCULATED SCENARIOS OF INDUSTRY 4.0 CLOUD ENERGY CONSUMPTION TILL 2025

(Source: own depiction for illustrational purposes of the applied scenario calculation methodology)

10. Conclusion

In this investigation, the aim was to assess the overall energy consumption of cloud computing and the greenhouse gas emissions caused in the context of Industry 4.0 via a mixed approach quantitative analysis of secondary data from various sources. The results of this investigation show that currently the ICT sector is responsible for around 1,3 % of global greenhouse gas emissions with cloud computing contributing approximately 25% of the total. Out of those 25% energy consumption for cloud computing in the context of Industry 4.0 accounts for only 10%.

However, generally, the study found that the impact of data generated by Industry 4.0 seems relatively insignificant compared to internet video streaming services like YouTube or Netflix to name a few. Against this backdrop of data generation via an avalanche of cute cat videos, latest movie blockbusters and compelling TV-series, every other type of application seems to be irrelevant. The true motor of rapid data increase lies with video streaming services like Netflix & Co, sending thousands of bytes though the internet every second. According to Cisco Systems (2018), video streaming already makes up about 73% of all IP traffic generated. This trend is only predicted to continue, with an expected increase to 82% of total IP traffic by the year 2021. The magnitude and extent of this development is especially evident in North America, where approximately 1/3 of IP traffic today is exclusively used for video streaming services. Since it is safe to assume, that society will not give up their favourite TV-show in favour of a book in the near future, the solution needs to be increased energy efficiency via Industry 4.0.

Therefore, the relevance of increasing overall energy efficiency in Industry 4.0 by running large data centres on low-GHG emission energy sources and further boosting efficiency is clearly supported by the current findings. According to Greenpeace (2017) the latter is already being done by numerous leading internet companies. Since Industry 4.0 provides a feasible way of addressing global challenges such as climate change while also positively contributing to meeting commitments under several international treaties and fora. All aiming for increased energy efficiency and reduced GHG-emissions to transition to a low-carbon economy in which growth is completely decoupled from emissions. The results of this research support the idea that the positive effects will outweigh the negative ones linked to the increased data amount especially in the industrial context. The potential for optimization and better decision making is numerous orders of magnitude greater.

The generalisability of these results is subject to certain limitations. For instance, today most of the world's leading data centre companies, do not disclose details about their operation. While Google and Microsoft both recently built new data centres in northern Europe where they can run almost exclusively on renewable energy sources, other companies like Amazon are less transparent about their operations. Most of its large data centres that are powering the Amazon Web Services are located in Northern Virginia in the USA. The state currently uses renewable resources to generate only 3% of its electricity. The main share is occupied by natural gas, which is responsible for satisfying half of the state's electricity demand. Second come the two nuclear power plants that supply one third of the demand while the remaining part is covered by coal. (US Energy Information Administration 2018) Due to increasing public pressure, most of the internet behemoths including AWS have committed to powering their data centre with 100% renewable energy. However, those commitments are usually fulfilled by carbon offsetting measures, which cannot be attributed directly to the energy and can be used for green washing practices instead of fostering sustainable change.

The findings provide valuable insights for future research in the area of the total environmental impact that Industry 4.0 has. As such the results can be used to establish a first idea, on how much the data explosion that is being attributed to Industry 4.0 actually contributes to the cloud workload. Especially, the issue of a detailed system analysis & examination that encompasses a digitalized production facility and its correlating cloud environment is an intriguing one which could be usefully explored in further research. However, if the debate is to be moved forward, a better understanding & transparency of current cloud data energy consumption needs to be developed.

Bibliography

Al-Azez, Z. T., A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani. 2015. “Virtualization Framework for Energy Efficient IoT Networks.” In *2015 IEEE 4th International Conference on Cloud Networking (CloudNet)*, 74–77. https://doi.org/10.1109/CloudNet.2015.7335284.

Alcott, Blake, Mario Giampietro, Kozo Mayumi, and John Polimeni. 2012. *The Jevons Paradox and the Myth of Resource Efficiency Improvements*. Routledge.

Andrae, Anders, Tomas Edler, Anders S. G. Andrae, and Tomas Edler. 2015. “On Global Electricity Usage of Communication Technology: Trends to 2030.” *Challenges* 6 (1): 117–57. https://doi.org/10.3390/challe6010117.

Arthur, Charles. 2011. “What’s a Zettabyte? By 2015, the Internet Will Know, Says Cisco.” *The Guardian*, June 29, 2011, sec. Technology. https://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco.

Ascierto, Rhonda. 2018. “Uptime Institute Global Data Center Survey.” Seattle: Uptime Insitute. https://uptimeinstitute.com/uptime_assets/f7bb01a900c060cc9abe42bb084609f63f02e448f5df1ca7ba7fdebb746cd1c4-2018-data-center-industry-survey.pdf.

Ashton, Kevin. 2009. “That ‘Internet of Things’ Thing.” *RFID Journal* 22 (7): 97–114.

Aslan, Joshua, Kieren Mayers, Jonathan G. Koomey, and Chris France. 2018. “Electricity Intensity of Internet Data Transmission: Untangling the Estimates.” *Journal of Industrial Ecology* 22 (4): 785–98. https://doi.org/10.1111/jiec.12630.

Baer, Walter S., Scott Hassell, and Ben A. Vollaard. 2002. *Electricity Requirements for a Digital Society*. Rand Corporation.

Baheti, Radhakisan, and Helen Gill. 2011. “Cyber-Physical Systems.” *The Impact of Control Technology* 12 (1): 161–66.

Barroso, Luiz André, Jimmy Clidaras, and Urs Hölzle. 2013. “The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines.” *Synthesis Lectures on Computer Architecture* 8 (3): 1–154.

Bateman, Kayleigh. 2018. “Industry 4.0 and the Cloud: A Match Made in Heaven.” *Information Age* (blog). May 4, 2018. https://www.information-age.com/industry-4-0-cloud-match-heaven-123471806/.

Bauer, Diana, Kathryn Papp, John Polimeni, Kozo Mayumi, Mario Giampietro, and Blake Alcott. 2009. *John Polimeni, Kozo Mayumi, Mario Giampietro & Blake Alcott, The Jevons Paradox and the Myth of Resource Efficiency Improvements*. Taylor & Francis.

Bechtold, Jochen, Andreas Kern, Christoph Lauenstein, and Lena Bernhofer. 2014. “Industry 4.0 - The Capgemini Consulting View.” Capgemini Consulting. https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/capgemini-consulting-industrie-4.0_0_0.pdf.

Bein, Wolfgang. 2018. “Energy Saving in Data Centers.” *Electronics* 7 (1): 5. https://doi.org/10.3390/electronics7010005.

Belady, Christian, Andy Rawson, John Pfleuger, and Tahir Cader. 2008. “Green Grid Data Center Power Efficiency Metrics: PUE and DCIE.” Technical report, Green Grid.

Biswas, A. R., and R. Giaffreda. 2014. “IoT and Cloud Convergence: Opportunities and Challenges.” In *2014 IEEE World Forum on Internet of Things (WF-IoT)(WF-IOT)*, 375–76. https://doi.org/10.1109/WF-IoT.2014.6803194.

Bloem, Jaap, Menno van Doorn, Sander Duivestein, David Excoffier, Rene Maas, and Erik van Ommeren. 2014. “The Fourth Industrial Revolution - Things to Tighten the Link Between It and Ot.” 3. Groningen: Sogeti. https://www.sogeti.com/globalassets/global/special/sogeti-things3en.pdf.

Bohn, Jürgen, Vlad Coroamă, Marc Langheinrich, Friedemann Mattern, and Michael Rohs. 2005. “Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing.” In *Ambient Intelligence*, 5–29. Springer.

Boyes, Hugh, Bil Hallaq, Joe Cunningham, and Tim Watson. 2018. “The Industrial Internet of Things (IIoT): An Analysis Framework.” *Computers in Industry* 101 (October): 1–12. https://doi.org/10.1016/j.compind.2018.04.015.

Brady, Gemma Ann. 2016. “Energy Efficiency in Data Centres and the Barriers to Further Improvements: An Interdisciplinary Investigation,” January, 286.

Brettel, Malte, Niklas Friederichsen, Michael Keller, and Marius Rosenberg. 2014. “How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective.” *International Journal of Mechanical, Industrial Science and Engineering* 8 (1): 37–44.

Candel Haug, Katharina, Tobias Kretschmer, and Thomas Strobel. 2016. “Cloud Adaptiveness Within Industry Sectors - Measurement and Observations.” *Telecommun. Policy* 40 (4): 291–306. https://doi.org/10.1016/j.telpol.2015.08.003.

Carr, Albert. 1940. *America’s Last Chance.* First Edition, First Printing edition. Thomas Y. Crowell Co.

Chen, Yong, Hong Chen, Anjee Gorkhali, Yang Lu, Yiqian Ma, and Ling Li. 2016. “Big Data Analytics and Big Data Science: A Survey.” *Journal of Management Analytics* 3 (1): 1–42.

Cisco Systems. 2018. “Global IP Data Traffic from 2016 to 2021 (in Petabytes per Month).” Cisco. https://www.statista.com/statistics/499431/global-ip-data-traffic-forecast/.

“Colocation Western Europe.” 2018. May 22, 2018. https://www.datacentermap.com/western-europe/.

Cox, Michael, and David Ellsworth. 1997. “Application-Controlled Demand Paging for Out-of-Core Visualization.” In *Proceedings of the 8th Conference on Visualization ’97*, 235–ff. VIS ’97. Los Alamitos, CA, USA: IEEE Computer Society Press. http://dl.acm.org/citation.cfm?id=266989.267068.

Daintith, John. 2009. “A Dictionary of Physics.” In *A Dictionary of Physics*. Oxford University Press.

http://www.oxfordreference.com/abstract/10.1093/acref/9780199233991.001.0001/acref-9780199233991-e-1592.

Dayarathna, Miyuru, Yonggang Wen, and Rui Fan. 2016. “Data Center Energy Consumption Modeling: A Survey.” *IEEE Communications Surveys & Tutorials* 18: 732–94. https://doi.org/10.1109/COMST.2015.2481183.

De Mauro, Andrea, Marco Greco, and Michele Grimaldi. 2016. “A Formal Definition of Big Data Based on Its Essential Features.” *Library Review* 65 (3): 122–35. https://doi.org/10.1108/LR-06-2015-0061.

Delen, Dursun, and Haluk Demirkan. 2013. *Data, Information and Analytics as Services*. Elsevier.

Demirović, Alex, Julia Dück, Florian Becker, Pauline Bader, and Wissenschaftlicher Beirat von Attac. 2011. *VielfachKrise: Im finanzmarktdominierten Kapitalismus*. 1 edition. Hamburg: VSA.

Deng, Ruilong, Rongxing Lu, Chengzhe Lai, Tom H. Luan, and Hao Liang. 2016. “Optimal Workload Allocation in Fog-Cloud Computing toward Balanced Delay and Power Consumption.” *IEEE Internet of Things Journal* 3 (6): 1171–81.

Dragland, Åse. 2013. “Big Data, for Better or Worse: 90% of World’s Data Generated over Last Two Years.” *SINTEF*, May 22, 2013. www.sciencedaily.com/releases/2013/05/130522085217.htm.

Edgerton, David. 2011. *The Shock of the Old: Technology and Global History Since 1900*. Oxford University Press, USA.

“Efficiency: How We Do It – Data Centers – Google.” 2018. 2018. https://www.google.com/about/datacenters/efficiency/internal/#measuring-efficiency&tab0=4.

Erol-Kantarci, Melike, and Hussein T. Mouftah. 2015. “Energy-Efficient Information and Communication Infrastructures in the Smart Grid: A Survey on Interactions and Open Issues.” *IEEE Communications Surveys & Tutorials* 17 (1): 179–97.

"Everything You Need to Know About the Industrial Internet of Things (IIoT)." 2016. GE Digital. October 24, 2016. https://www.ge.com/digital/blog/everything-you-need-know-about-industrial-internet-things.

Fatheuer, Thomas, Lili Fuhr, and Barbara Unmüßig. 2015. *Kritik Der Grünen Ökonomie*. Oekom Verlag.

Fettweis, Gerhard, and Ernesto Zimmermann. 2008. "ICT Energy Consumption-Trends and Challenges." In *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, 2:6. (Lapland.

Finley, Klint. 2015. "Here's the Real Way to Get Internet to the Next 4 Billion People." *Wired*, September 16, 2015. https://www.wired.com/2015/09/heres-real-way-get-internet-next-4-billion-people/.

Fischer, Audrey. 2016. "The Library of Congress by the Numbers in 2015." January 2, 2016. https://www.loc.gov/item/prn-16-023/the-library-of-congress-by-the-numbers-in-2015/2016-02-01/.

Fountain, Henry, and Brad Plumer. 2018. "What the Climate Report Says About the Impact of Global Warming." *The New York Times*, January 20, 2018, sec. Climate. https://www.nytimes.com/2017/11/03/climate/climate-change-impacts.html.

Garbee, Elizabeth. 2016. "This Is Not the Fourth Industrial Revolution." *Slate*, January 29, 2016. http://www.slate.com/articles/technology/future_tense/2016/01/the_world_economic_forum_is_wrong_this_isn_t_the_fourth_industrial_revolution.html?via=gdpr-consent.

George, Sarah. 2018. "Google 'taking AI to the next Level' to Cut Data Centre Energy Use." *Euractiv.Com* (blog). August 22, 2018. https://www.euractiv.com/section/energy/news/google-taking-ai-to-the-next-level-to-cut-data-centre-energy-use/.

Gerbert, Philipp, Markus Lorenz, Michael Rüßmann, Manueala Waldner, Jan Justus, Pascal Engel, and Michael Harnisch. 2015. "Industry 4.0: The Future of Productivity Growth in Manufacturing Industries.," September 4, 2015.

https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries.aspx.

Gold, Jon. 2014. “Mapping the Cloud: Where Does the Public Cloud Actually Live?” 2014. https://www.networkworld.com/article/2600185/cloud-computing/cloud-computing-mapping-the-cloud-where-does-the-public-cloud-actually-live.html.

Gölzer, Philipp, Patrick Cato, and Michael Amberg. 2015. “Data Processing Requirements of Industry 4.0 - Use Cases for Big Data Applications,” 14.

Grossman, Gary. 2016. “The next Stop on the Road to Revolution Is Ambient Intelligence.” *TechCrunch* (blog). 2016. http://social.techcrunch.com/2016/05/07/the-next-stop-on-the-road-to-revolution-is-ambient-intelligence/.

Haenn, Nora, Allison Harnish, and Richard Wilk. 2016. *The Environment in Anthropology (Second Edition): A Reader in Ecology, Culture, and Sustainable Living*. NYU Press.

Harari, Yuval Noah. 2015. “Why Humans Run the World.” *Ideas.Ted.Com* (blog). June 16, 2015. https://ideas.ted.com/why-humans-run-the-world/.

Harmon, Robert R., and Nora Auseklis. 2009. “Sustainable IT Services: Assessing the Impact of Green Computing Practices.” In *PICMET ’09 - 2009 Portland International Conference on Management of Engineering & Technology*, 1707–17. Portland, OR, USA: IEEE. https://doi.org/10.1109/PICMET.2009.5261969.

He, Wu, and Li Da Xu. 2014. “Integration of Distributed Enterprise Applications: A Survey.” *Ieee Transactions on Industrial Informatics* 10 (1): 35–42. https://doi.org/10.1109/TII.2012.2189221.

Hogan, Michael, Fang Liu, Annie Sokol, and Jin Tong. 2011. “Nist Cloud Computing Standards Roadmap.” *NIST Special Publication* 35: 6–11.

Holmes, Rachel. 2016. “We Let Technology into Our Lives. And Now It’s Starting to Control Us | Rachel Holmes.” *The Guardian*, November 28, 2016, sec. Opinion. https://www.theguardian.com/commentisfree/2016/nov/28/technology-our-lives-control-us-internet-giants-data.

Hormann, Peter, and Leith Campbell. 2014. “Data Storage Energy Efficiency in the Zettabyte Era.” *Australian Journal of Telecommunications and the Digital Economy* 2 (3).

IEA. 2018. “Global Energy & CO2 Status Report 2017.” International Energy Agency. https://www.iea.org/publications/freepublications/publication/GECO2017.pdf.

IFR. Worldwide sales of industrial robots from 2004 to 2017 (in 1,000 units). https://www.statista.com/statistics/264084/worldwide-sales-of-industrial-robots/ (accessed 10/7/18, 9:22 PM).

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Itoh, Satoshi, Yuetsu Kodama, Toshiyuki Shimizu, Satoshi Sekiguchi, Hiroshi Nakamura, and Naohiko Mori. 2010. “Power Consumption and Efficiency of Cooling in a Data Center.” In *Grid Computing (GRID), 2010 11th IEEE/ACM International Conference On*, 305–12. IEEE.

Jalali, Fatemeh, Rob Ayre, Arun Vishwanath, Kerry Hinton, Tansu Alpcan, and Rod Tucker. 2014. “Energy Consumption of Content Distribution from Nano Data Centers Versus Centralized Data Centers.” *SIGMETRICS Perform. Eval. Rev.* 42 (3): 49–54. https://doi.org/10.1145/2695533.2695555.

Jalali, Fatemeh, Kerry Hinton, Robert Ayre, Tansu Alpcan, and Rodney S. Tucker. 2016. “Fog Computing May Help to Save Energy in Cloud Computing.” *IEEE Journal on Selected Areas in Communications* 34 (5): 1728–39. https://doi.org/10.1109/JSAC.2016.2545559.

Jeschke, Sabina, Christian Brecher, Tobias Meisen, Denis Özdemir, and Tim Eschert. 2017. “Industrial Internet of Things and Cyber Manufacturing Systems.” In *Industrial Internet of Things: Cybermanufacturing Systems*, edited by Sabina Jeschke, Christian Brecher, Houbing Song, and Danda B. Rawat, 3–19. Springer Series in Wireless Technology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-42559-7_1.

Jevons, William Stanley. 1906. *The Coal Question*. Рипол Классик.

Kagermann, Henning, Wolfgang Wahlster, and Johannes Helbig. 2013. “Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0.” *acatech* (blog). 2013. https://www.acatech.de/Publikation/umsetzungsempfehlungen-fuer-das-zukunftsprojekt-industrie-4-0-abschlussbericht-des-arbeitskreises-industrie-4-0/.

Kamarul Bahrin, Mohd Aiman, Mohd Fauzi Othman, Nor Hayati Nor Azli, and Muhamad Farihin Talib. 2016. “INDUSTRY 4.0: A REVIEW ON INDUSTRIAL AUTOMATION AND ROBOTIC.” *Jurnal Teknologi* 78 (6–13). https://doi.org/10.11113/jt.v78.9285.

Kambies, Tracy, Sandeep Kumar Sharma, Nitin Mittal, and Paul Roma. 2017. “Dark Analytics: Illuminating Opportunities Hidden within Unstructured Data,” July 2, 2017. https://www2.deloitte.com/insights/us/en/focus/tech-trends/2017/dark-data-analyzing-unstructured-data.html.

Kant, Krishna. 2009. “Data Center Evolution: A Tutorial on State of the Art, Issues, and Challenges.” *Computer Networks* 53 (17): 2939–65.

Kawamoto, Kaoru, Jonathan G. Koomey, Bruce Nordman, Richard E. Brown, Mary Ann Piette, Michael Ting, and Alan K. Meier. 2002. “Electricity Used by Office Equipment and Network Equipment in the US.” *Energy* 27 (3): 255–69.

Klein, Andy. 2017. “2017 Hard Drive Reliability by Drive Size and Brand.” *Backblaze Blog | Cloud Storage & Cloud Backup* (blog). October 26, 2017. https://www.backblaze.com/blog/hard-drive-failure-rates-q3-2017/.

Koomey, Jonathan. 2011. “Growth in Data Center Electricity Use 2005 to 2010.” *A Report by Analytical Press, Completed at the Request of The New York Times* 9.

Koomey, Jonathan G. 2007. “Estimating Total Power Consumption by Servers in the US and the World.” Stanford University. http://www-sop.inria.fr/mascotte/Contrats/DIMAGREEN/wiki/uploads/Main/svrpwrusecompletefinal.pdf.

———. 2008. “Worldwide Electricity Used in Data Centers.” *Environmental Research Letters* 3 (3): 034008.

Lapira, Edzel R. 2012. “Fault Detection in a Network of Similar Machines Using Clustering Approach.” University of Cincinnati.

Lee, Jay, Hung-An Kao, and Shanhu Yang. 2014. “Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment.” *Procedia Cirp* 16: 3–8.

Lee, MinHwa, JinHyo Yun, Andreas Pyka, DongKyu Won, Fumio Kodama, Giovanni Schiuma, HangSik Park, et al. 2018. “How to Respond to the Fourth Industrial Revolution, or the Second Information Technology Revolution? Dynamic New Combinations between Technology, Market, and Society through Open Innovation.” *Journal of Open Innovation: Technology, Market, and Complexity* 4 (3): 21. https://doi.org/10.3390/joitmc4030021.

Li, Xiaomin, Di Li, Jiafu Wan, Athanasios V. Vasilakos, Chin-Feng Lai, and Shiyong Wang. 2017. “A Review of Industrial Wireless Networks in the Context of Industry 4.0.” *Wireless Networks* 23 (1): 23–41. https://doi.org/10.1007/s11276-015-1133-7.

Mahto, Monika, and Varia Hemnabh. 2018. “The Internet of Things: A Technical Primer,” 2018. https://www2.deloitte.com/insights/us/en/focus/internet-of-things/technical-primer.html.

Malmodin, Jens, and Pernilla Bergmark. 2015. “Exploring the Effect of ICT Solutions on GHG Emissions in 2030.” In .

Malmodin, Jens, Dag Lundén, Åsa Moberg, Greger Andersson, and Mikael Nilsson. 2014. “Life Cycle Assessment of ICT: Carbon Footprint and Operational Electricity Use from the Operator, National, and Subscriber Perspective in Sweden.” *Journal of Industrial Ecology* 18 (6): 829–45.

Malmodin, Jens, Åsa Moberg, Dag Lundén, Göran Finnveden, and Nina Lövehagen. 2010. “Greenhouse Gas Emissions and Operational Electricity Use in the ICT and Entertainment & Media Sectors.” *Journal of Industrial Ecology* 14 (5): 770–90.

Malone, Christopher, and Christian Belady. 2006. “Metrics to Characterize Data Center & IT Equipment Energy Use.” In *Proceedings of the Digital Power Forum, Richardson, TX*. Vol. 35.

Marques Lima, Joao. 2017. “Norway Lands ‘World’s Largest’ Data Centre at 1,000 MW, 6.5 Million Sqf.” *Data Economy* (blog). August 15, 2017. https://data-economy.com/norway-lands-worlds-largest-data-centre-1000-mw-6-5-million-sqf/.

Marscheider-Weidemann, Frank, Sabine Langkau, Torsten Hummen, and et al. 2016. *Rohstoffe Für Zukunftstechnologien 2016*. DERA Rohstoffinformationen. Berlin: Deutsche Rohstoffagentur.

Mathis, Franz. 2016. “No Industrialization without Urbanization: The Role of Cities in Modern Economic Development.” In *Globalization and the City : Two Connected Phenomena in Past and Present*, edited by Günter Bischof, Andreas Exenberger, James Mokhiber, and Philipp Strobl, 23–38. Interdisziplinäre Forschung. Innsbruck: innsbruck university press. http://books.openedition.org/iup/1357.

Mattern, Friedemann, and Christian Flörkemeier. 2010. “Vom Internet der Computer zum Internet der Dinge.” *Informatik-Spektrum* 33 (2): 107–21. https://doi.org/10.1007/s00287-010-0417-7.

Mell, Peter, and Tim Grance. 2011. “The NIST Definition of Cloud Computing.”

Miettinen, Antti P, and Jukka K Nurminen. 2010. “Energy Efficiency of Mobile Clients in Cloud Computing,” 7.

Mills, Mark P. 1999. “The Internet Begins with Coal: A Preliminary Exploration of the Impact of the Internet on Electricity Consumption.” *VA: The Greening Earth Society*.

Moisan, Francois, and Didier Bosseboeuf. 2010. “Energy Efficiency: A Recipe for Success.” *World Energy Council, London, UK, Tech. Rep*.

Moore, Justin D., Jeffrey S. Chase, Parthasarathy Ranganathan, and Ratnesh K. Sharma. 2005. “Making Scheduling" Cool": Temperature-Aware Workload

Placement in Data Centers." In *USENIX Annual Technical Conference, General Track*, 61–75.

O'Halloran, Derek, and Elena Kvochko. 2015. "Industrial Internet of Things: Unleashing the Potential of Connected Products and Services." World Economic Forum. http://www3.weforum.org/docs/WEFUSA_IndustrialInternet_Report2015.pdf.

Oracle. 2015. "Java and the Internet of Things: Automating the Industrial Economy." Oracle Corporation. http://www.oracle.com/us/solutions/internetofthings/java-iot-industrial-automation-2430562.pdf.

Oscarsson, Ingela. 2014. "A Forecast of the Cloud."

Oyanagi, Yoshio. 2002. "Future of Supercomputing." *Journal of Computational and Applied Mathematics*, Scientific and Engineering Computations for the 21st Century - Me thodologies and Applications Proceedings of the 15th Toyota Conference, 149 (1): 147–53. https://doi.org/10.1016/S0377-0427(02)00526-5.

Pattinson, Colin, Roland Cross, and Ah-Lian Kor. 2015. "Chapter 14 - Thin-Client and Energy Efficiency." In *Green Information Technology*, edited by Mohammad Dastbaz, Colin Pattinson, and Babak Akhgar, 279–94. Boston: Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-801379-3.00014-0.

Pereira, A. C., and F. Romero. 2017. "A Review of the Meanings and the Implications of the Industry 4.0 Concept." *Procedia Manufacturing*, Manufacturing Engineering Society International Conference 2017, MESIC 2017, 28-30 June 2017, Vigo (Pontevedra), Spain, 13 (January): 1206–14. https://doi.org/10.1016/j.promfg.2017.09.032.

Pettey, Christy. 2007. "Gartner Estimates ICT Industry Accounts for 2 Percent of Global CO2 Emissions." 2007. https://www.gartner.com/newsroom/id/503867.

Pilgrim, Hannah, Merle Groneweng, and Michael Reckordt. 2017. "Ressourcenfluch 4.0 - Die Sozialen Und Ökologischen Auswirkungen von Industrie 4.0 Auf Den Rohstoffsektor." Berlin: Powershift e.V. https://power-shift.de/wp-content/uploads/2017/02/Ressourcenfluch-40-rohstoffe-menschenrechte-und-industrie-40.pdf.

Posada, Jorge, Carlos Toro, Iñigo Barandiaran, David Oyarzun, Didier Stricker, Raffaele De Amicis, Eduardo B. Pinto, Peter Eisert, Jürgen Döllner, and Ivan Vallarino. 2015. “Visual Computing as a Key Enabling Technology for Industrie 4.0 and Industrial Internet.” *IEEE Computer Graphics and Applications* 35 (2): 26–40.

Radu, Laura-Diana. 2017. “Green Cloud Computing: A Literature Survey.” *Symmetry* 9 (12): 295.

Rider, Fremont. 1944. *The Scholar and the Future of the Research Library. A Problem and Its Solution.* Hadham Press.

Rossman, John. 2016. *The Amazon Way on IoT: 10 Principles for Every Leader from the World’s Leading Internet of Things Strategies*. Clyde Hill Publishing.

Sadeghi, A., C. Wachsmann, and M. Waidner. 2015. “Security and Privacy Challenges in Industrial Internet of Things.” In *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 1–6. https://doi.org/10.1145/2744769.2747942.

Sagiroglu, S., and D. Sinanc. 2013. “Big Data: A Review.” In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. https://doi.org/10.1109/CTS.2013.6567202.

Satyavolu, Prasad, Badrinath Setlur, Prasanth Thomas, Ganesh Iyer, and Cognizant Technology Solutions. 2014. “Designing for Manufacturing’s ‘Internet of Things,’” 15.

Schwab, Klaus. 2017. *The Fourth Industrial Revolution*. New York: Currency.

Sengupta, Somini. 2018. “Biggest Threat to Humanity? Climate Change, U.N. Chief Says.” *The New York Times*, March 29, 2018, sec. Climate. https://www.nytimes.com/2018/03/29/climate/united-nations-climate-change.html.

Shehabi, Arman, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, and William Lintner. 2016. “United States Data Center Energy Usage Report.”

Shi, Weisong, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. 2016. “Edge Computing: Vision and Challenges.” *IEEE Internet of Things Journal* 3 (5): 637–46.

Sontag, Daniel. 2018. “Industrial IoT vs. Industry 4.0 vs. … Industry 5.0?” *Medium* (blog). March 8, 2018. https://medium.com/the-industry-4-0-blog/industrial-iot-vs-industry-4-0-vs-industry-5-0-a5f9541da036.

Spring, Tom. 2017. “IoT Is Insecure, Get Over It! Say Researchers.” *The First Stop for Security News | Threatpost* (blog). November 8, 2017. https://threatpost.com/iot-is-insecure-get-over-it-say-researchers/128826/.

Steinbuch, Karl. 1966. *Die informierte Gesellschaft. Geschichte und Zukunft der Nachrichtentechnik.* Auflage unbekannt edition. Stuttgart, dva, 1966.

Sullivan, Michael. 2016. “Industry 4.0 Technologies: Global Market Through 2023.” bcc Research. https://www.bccresearch.com/market-research/manufacturing/industry-4-technologies-global-market-through-2023-mfg042a.html.

Thames, Lane, and Dirk Schaefer, eds. 2017. *Cybersecurity for Industry 4.0: Analysis for Design and Manufacturing*. Springer Series in Advanced Manufacturing. Springer International Publishing. //www.springer.com/de/book/9783319506593.

Torrens, J. Ignacio, Deepak Mehta, Vojtech Zavrel, Diarmuid Grimes, Thomas Scherer, Robert Birke, Lydia Chen, Susan Rea, Lara Lopez, and Enric Pages. 2016. “Integrated Energy Efficient Data Centre Management for Green Cloud Computing.” In *Proceedings of the 6th International Conference on Cloud Computing and Services Science-Volume 1 and 2*, 375–86. SCITEPRESS-Science and Technology Publications, Lda.

Tractica. 2018. “Robotics Market Forecasts | Tractica.” 2018. https://www.tractica.com/research/robotics-market-forecasts/.

Twentyman, Jessica. 2018. “Nine World-Leading Smart Factories Identified by World Economic Forum.” *Internet of Business* (blog). September 11, 2018. https://internetofbusiness.com/nine-world-leading-smart-factories-identified-by-world-economic-forum/.

US Energy Information Administration. 2018. “Virgina - State Profile & Energy Estimates.” August 16, 2018. https://www.eia.gov/state/analysis.php?sid=VA#67.

Vaidya, Saurabh, Prashant Ambad, and Santosh Bhosle. 2018. “Industry 4.0 – A Glimpse.” *Procedia Manufacturing*, 2nd International Conference on Materials, Manufacturing and Design Engineering (iCMMD2017), 11-12 December 2017, MIT Aurangabad, Maharashtra, INDIA, 20 (January): 233–38. https://doi.org/10.1016/j.promfg.2018.02.034.

Van Heddeghem, Ward, Sofie Lambert, Bart Lannoo, Didier Colle, Mario Pickavet, and Piet Demeester. 2014. “Trends in Worldwide ICT Electricity Consumption from 2007 to 2012.” *Computer Communications* 50: 64–76.

Waldrop, M. Mitchell. 2016. “The Chips Are down for Moore’s Law.” *Nature News* 530 (7589): 144. https://doi.org/10.1038/530144a.

Weiser, Mark. 1999. “The Computer for the 21st Century.” *Mobile Computing and Communications Review* 3 (3): 3–11.

Workman, Mike. 2018. “How Many Gigabytes of Storage Are in the Whole World?” September 10, 2018. https://www.quora.com/How-many-gigabytes-of-storage-are-in-the-whole-world.

Xu, Li Da, and Lian Duan. 2018. “Big Data for Cyber Physical Systems in Industry 4.0: A Survey.” *Enterprise Information Systems* 0 (0): 1–22. https://doi.org/10.1080/17517575.2018.1442934.

Younis, Ossama, and Sonia Fahmy. 2004. “HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks.” *IEEE Transactions on Mobile Computing* 3 (4): 366–79.

List of Tables

TABLE 1 – ENERGY CONSUMPTION METRICS 29

List of Figures

FIGURE 1 – HOW DOES IOT, IIOT, INDUSTRY 4.0 AND THE CLOUD RELATE TO EACH OTHER 9
FIGURE 2 – THE DIFFERENT TYPES OF DATA ANALYTICS 17
FIGURE 3 – CLOUD AS CORE ENABLING TECHNOLOGY 22
FIGURE 4 – ENERGY DEMAND ACCORDING TO EQUIPMENT TYPE 26
FIGURE 5 – ELECTRICITY GENERATION CO2 EMISSION INTENSITY ACCORDING TO TOP CLOUD COMPUTING DATA CENTRES 28
FIGURE 6 – VARIOUS ENERGY EFFECTS 32
FIGURE 7 – CREATED DATA BY INDUSTRY 4.0 APPLICATION 43
FIGURE 8 – CALCULATED SCENARIOS OF INDUSTRY 4.0 CLOUD ENERGY CONSUMTION TILL 2025 44

Annex

Energy Consumption of Data Transfer

Data transfer rates		
Connection Type	% in 2017	kW/h per GB
Wide Area Network (WAN)	50%	0,17
3G Connection	14%	1,65
4G Connection	35%	0,15
Total estimated energy consumption for data transfer		**0,368**

Energy Consumption of Data Centre Storage

Data Centre Storage								
Storage Capacity (TB)	Distribution	Idle consumption (W)	Idle Consumption per TB	40% more consumption when busy (W)	Total Consumption (W)	kW (1000W)	kW/h	kW/h per GB
4	65%	6,7	1,675	9,38	8,04	0,00804	70,43	0,0176076
8	28%	8	1	11,2	9,6	0,0096	84,096	0,010512
3	7%	7,7		10,78	9,24	0,00924	80,942	0,0269808
Total average of kW/h per GB of stored data (considering the distribution as outlined above)								**0,0162**

Forecast of the amount of data creates by Industry 4.0

Forecast - Data created by Industry 4.0 in petabytes from 2018 till 2025								
Scenario	2018	2019	2020	2021	2022	2023	2024	2025
Scenario 1 - Conservative	295.440,00	363.686,64	447.698,25	551.116,55	678.424,47	835.140,53	1.028.057,99	1.265.539,38
Scenario 2 - Exponential	360.000,00	540.000,00	810.000,00	1.215.000,00	1.822.500,00	2.733.750,00	4.100.625,00	6.150.937,50

Global IP data traffic

GLOBAL IP DATA TRAFFIC in zetabytes		
year	IP Traffic per month	IP traffic per year
2017	0,12	1,460328
2018	0,15091	1,81092
2019	0,186453	2,237436
2020	0,228411	2,740932
2021	0,278108	3,337296

Global data centre traffic

Global data center IP traffic from 2012 to 2021, by type in zetabyte			
year	Data center to user	Data center to data center	Within data center
2017	1,28	0,97	6,83
2018	1,61	1,35	8,60
2019	2,02	1,75	10,36
2020	2,50	2,25	12,37
2021	3,06	2,80	14,70

Scenario I

Scenario I - Worst Case		
Variable	Amount	Unit
Ecloud	0,413	kW/h per GB
nbytes	6.150.937.500.000,00	GB
E Industry 4.0	2540337187500,00	kW/h
E Industry 4.0	2540,34	TW/h

Scenario II

Scenario II - Base		
Variable	Amount	Unit
Ecloud	0,413	kW/h per GB
nbytes	1.265.539.384.207,13	GB
E Industry 4.0	522667765677,55	kW/h
E Industry 4.0	522	TW/h

Scenario III

Scenario III - Most likely		
Variable	Amount	Unit
Ecloud	0,35262	kW/h per GB
nbytes	6.150.937.500.000,00	GB
E Industry 4.0	2168943581250,00	kW/h
E Industry 4.0	2168,94	TW/h

Development of various scenarios until 2025

Forecast - Energy Consumption trough Cloud Computing for Indusrey 4.0 in TW/h								
Scenario	2018	2019	2020	2021	2022	2023	2024	2025
Scenario I - Worst case	148,68	223,02	334,53	501,80	752,69	1129,04	1693,56	2540,34
Scenario II - Base	122,02	150,20	184,90	227,61	280,19	344,91	424,59	522,67
Scenario III - Most likely	135,59	200,80	297,30	440,11	651,41	963,97	1445,96	2168,94