Usage impact on data center electricity needs: A system dynamic forecasting model

Martijn Koot, Fons Wijnhoven [*]

University of Twente, the Netherlands

HIGHLIGHTS

- A system dynamic model for policy makers and researchers to simulate data center energy scenarios.
- A user behavior perspective on data center energy needs.
- User behavior increases data center energy needs from 292 TWh in 2016 to 353 TWh in 2030.
- The end of Moore's law and IoT combined cause data center energy needs going up to 1287 TWh in 2030.

ARTICLE INFO

ABSTRACT

This article presents a forecasting model of data center electricity needs based on understanding usage growth and we conclude that this growth is not fully compensated by efficiency gains of data center technological innovations. We predict a combined growth of data center electricity needs of 286 TWh in 2016 until about 321 TWh in 2030, if all currently known growth factors remain the same. We next run simulations for the end of Moore's law and the growth of industrial Internet of Things (IoT). The end of Moore's law results in about 658 TWh for 2030 and an increase of the share of global data center electricity consumption from about 1.15% in 2016 to 1.86% in 2030. A rise of the Industrial IoT may result into total energy consumption of about 364 TWh (about 1.03%) in 2030. Moore's law and IoT combined cause data center energy needs going up to 752 TWh in 2030, and about 2.13% of global electricity available. Our sensitivity analysis reveals that the future impact of the data centers' electricity consumption is vulnerable to behavioral usage trends, since the 95% confidence interval of [343, 1031] TWh is relatively wide. Our forecasts, however, exclude the energy needs of mobile devices, edge and fog computing. We offer a system dynamic model and simulation input data selected from the existing literature for replicating this study and applying alternative parameters to it. We further suggest multiple research directions on usage impact on data center energy consumption.

1. Introduction

The increasing usage of consumer and business applications is associated with more computational tasks and higher storage demands by data centers, resulting into higher data center electricity consumption. Masanet et al. [1] conclude that despite a massive growth of data storage (25-fold with only 3-fold increase in energy), IP traffic (10-fold growth with only a marginal increase in energy used), and data center compute instances (6.5-fold with 25% energy usage increase), the total energy consumption of data centers increased only 6% between 2010 and 2018 (from 194TWh to 205TWh). This remarkable result is nearly fully explained by the large energy efficiency gains of data processing

and data center infrastructures (mainly cooling and uninterrupted power supplies (UPS), but their predictions exclude data center external energy needs for "user-to-data center" and "data center-to-data-center" IP traffic.

The speedy growth of data center usage may result in a growth of energy needs of data centers when energy efficiency gains do not continue as well as they did in the past. Moore's law –which predicts a 25% annual energy decline per processing unit (or as often said a performance doubling each 2 years)– is expected to stop having an influence by 2021 and 2023 [2]. Given this trend, one may wonder if Internet users will be confronted with a scarcity of internet energy resources in the future [3]. Data center energy needs predictions for the next decade,

---

* Corresponding author at: University of Twente, Faculty of Behavioural, Management, and Social Sciences, Drienerlolaan 5, 7552 NB Enschede, the Netherlands.
E-mail address: a.b.j.m.wijnhoven@utwente.nl (F. Wijnhoven).

however, lack consensus as some authors suggest a stable energy need of about 200 TWh in 2030 and others forecast about 10 times as much [4]. This article, therefore, aims at answering the research question: How will future data center energy demand evolve as a consequence of its growing usage?

An extensive review of data center energy consumption models is from Dayarathna et al [5]. They identify studies that focus on hardware-centric and software-centric power models. The hardware-centric approach starts from the digital circuit level and moves on to describe higher-level energy consumption models at the hardware component level, server level, data center level, and system-of-systems level. The software-centric approach describes power models developed for operating systems, virtual machines, and software applications. This means that these models do not analyze the usage side of data center energy consumption. In our view, this is comparable to understanding the energy efficiency of a car as-a-machine versus understanding the driver's behavior that causes an even extremely efficient machine to consume much. Increasing consumer demand also will need more machines doing the processing. Our study thus delivers an important usage perspective to previous data center modelling work.

With usage here we mean both industrial and consumers requests for IT storage, networking, and processing. These requests enter data centers to a large part via "user-to-data center" IP traffic, which next causes storage, processing, and networking activities by data centers. The usage demand growth has an exponential nature because of mutual growth reinforcement effects between capabilities being offered, like hardware, storage and communication speed, and opportunities of using these for new applications and software. Such reinforcements between technological innovations have been observed frequently as explanations for the exponential growth of technological markets [6] and internet-based services [7]. Exponential growth patterns at a certain stage of industrial maturity result in a balancing force because of the development of resource scarcity and market saturation [8], which are typical system dynamic model patterns [9]. A similar pattern of resource scarcity, especially electric power, may slow down the exponential growth of the data center market as well.

This article is structured as follows. First, previous research in the field of data center energy consumption is presented. After that, the system dynamic methodology is explained by which we model and simulate user driven data center energy consumption. This is followed by two results sections, one section about the design of the model and one section about simulations until 2030. We next discuss our results with outcomes of previous studies and present a list of approaches for coping with potential problems in the future.

2. Literature review

We studied a variety of future estimations of data center usage behavior, data center energy use, and electricity production capacity via a systematic literature search in Scopus, Web of Science and Google Scholar and present them in their sequence of publication.

Hinton et al. [10] state that in 2011, the ICT sector is responsible for about 5% of the total electric power consumption in developed economies. 1% of the total electric power consumption was caused by the Internet, but this percentage was expected to rapidly increase under the influence of increasing data access rates. This may become an untenable development, because the expected 10% efficiency gains will not compensate for the predicted 40% traffic growth [10]. Consequently, they express an urgent need for reducing data center power consumption when workload is low, further improving the energy efficiency of core routers, and the deployment of more energy-efficient access network technologies.

Van Heddeghem et al. [11] analyzed the electricity consumption of ICT worldwide. They estimated that global electricity demand is growing slower than the electricity consumed by digital devices and infrastructure. They assess how ICT electricity consumption in the use phase has evolved from 2007 until 2012. Following their estimates, the yearly growth of communication networks, personal computers, and data centers is 10%, 5%, and 4% respectively, and thus is higher than the growth of worldwide electricity consumption in the same time frame (3%). The relative share of these three ICT products and services in the total worldwide electricity consumption has increased from about 3.9% in 2007 to 4.6% in 2012.

For analyzing the volume of greenhouse gas emissions (GHGE), Belkhir and Elmeligi [3] estimated the energy needed for ICT component manufacturing, i.e., production energy (PE), which is a fixed single time emission per device produced, and use phase energy (UPE) costs, which is an annual recurring volume. A third parameter they use for estimating GHGE is the useful life (UL) of devices, which if shorter causes a more frequent rebuy and therefore more production energy consumption. Belkhir and Elmeligi [3] determined the UL and the installed base of each device. The combined data resulted in estimations that the ICT GHGE relative contribution could grow from roughly 1–1.6% in 2007 to exceed 14% of the 2016-level worldwide GHGE by 2040, accounting for more than half of the current relative contribution of the whole transportation sector. Their study also highlights the contribution of smart phones and shows that by 2020, the footprint of smart phones alone would surpass the individual contribution of desktops, laptops and displays. Finally, they offer actionable recommendations for mitigating the ICT explosive GHGE footprint through a combination of renewable energy use, tax policies, managerial actions, and alternative business models.

The growing data center electricity demand is also discussed by Morley et al [12], who focused on determining the daily peak demand for data and therefore the peak of the electricity demand of data centers. These demand peaks are especially caused by the large volume of data transfer for video streaming and video interaction, i.e., user-to-data center IP traffic. Their study calculates that containing the overall growth in energy demand across digital infrastructures depends on more than technological efficiency alone; it also requires limiting the growth in traffic to at least keep in step with efficiency improvements, a balance which has not so far been the case.

Much of this discussion has also been described and analyzed in a 2015 paper of Andrae and Edler [13]. For the electricity usage of the total communication technology sector, they predicted an absolute rise from 2,000 TWh in 2010 to 8,000 TWh in 2030. Data centers' energy usage alone would consequently grow from 200 TWh in 2016 to 2,967 TWh in 2030. However, their predictions have high uncertainty and have been revised in 2019 to 974 TWh for data centers in 2030 [14]. Additionally, Andrae [15] expects 3,234 TWh in 2030 for the full Internet (i.e., fixed and mobile Internet, devices and WIFI). Andrae and Edler state that for the worst-case scenario, the different communication technology components —i.e., use and production of consumer devices, communication networks, and data centers— could use as much as 51% of global electricity in 2030 if not enough improvement in electricity efficiency of wireless access networks and fixed access networks/data centers is realized. In the worst-case scenario, communication technology electricity usage could contribute up to 23% of the globally released GHGE in 2030. These insights have been much debated by experts from the data center industry, who point at the efficiency gains by Moore's law and improvements in data center infrastructure of about 20% annually [1], but the gains from Moore's law flatten according to Shalf [2] and Markov [16], who predict Moore's law to be fully out of order by 2021, a view shared by Shahidi [17] later.

Clearly, there is much diversity on these forecasts. Hintemann and Hinterholzer [4] state that in the "best case" the energy consumption of data centers can remain constant, but if the current developments continue, the energy consumption of data centers will double by 2030 compared to 2019.

The diversity of these predictions is large and calls for a transparent and public model that researchers and policy makers can use to work with their own assumptions. Additionally, such a model should not stick

at being at a top-down level but needs to open up the black box of data centers and allow predictions for different data center components and different usage categories of data centers.

3. Methodology

Our goal is to develop an energy forecasting model for global data centers, with a focus on data center usage impact. In contrast to natural science predictions, forecasting human behavior and usage of data centers is difficult as diverse economic, cultural, and technological developments may change people's preferences [18]. The strategic management discipline, which also tries to understand longer term human behavior, therefore has developed so-called scenario development methods that are used to produce multiple competing but plausible scenarios of the future [19,20]. System dynamic models are thinking machines for answering "what if" questions about possible futures that are hard to create and experiment with in a non-virtual way [21,22] and aim at representing the nonlinear behavior of complex systems over time using variables that influence the flows (i.e., the volume changes) of stocks [23,24]. Non-linear systems have no proportionality and no simple causality between the magnitude of responses and the strength of their stimuli: small changes can have striking effects, whereas great stimuli will not always lead to drastic changes. This way of viewing and modeling is appropriate for many systems when multiple influences have reinforcing and balancing effects, which we assume in data center energy forecasting.

There are many tools for system dynamics modeling, but throughout this article we will work with Insight Maker, which is a free-ware, Web 2.0-based, multi-user, general-purpose, online modeling, and simulation environment. For introductions and illustrations of Insight Maker, the reader may be interested in [25,26] and the free interactive Insight Maker tutorial http://beyondconnectingthedots.com/.

System dynamic modeler Insight Maker presents stocks, like materials, customers, or money, graphically by rectangles. Flows are presented by bolded solid lines with arrows that give the direction of the 'material' flows. Variables are graphically portrayed by ovals; they can be dynamically calculated values that change over time or they can be constants, e.g., IP growth rates. Links are graphically shown by dashed lines with arrows that show the transfer of information between the different primitives in the model. Thus, we can express data center energy system dynamics in a general top-down way by the Insight Maker language as we do in Fig. 1. In our data center usage impact model, we see a stock volume of application behavior as initiated by users as the start for all internet traffic and data center activities.

In the top-down model of Fig. 1, we start with a certain stock of IP traffic expressed in Exabytes (EB) that correspond with different user applications. The IP traffic stock is annually increased by a growth factor and causes a certain level of electricity need not only for its own energy but also for the server workload and required storage. Additionally, this processing, traffic and storage produces heat as a by-product and thus needs cooling. Besides of this need for cooling, a data center's infrastructure has energy needs for its building, light, security, and other building-associated energy needs. Moore's law is supposed to directly reduce the energy needs of servers and this variable is presented as a converter variable in our model, i.e., we define the level of Moore's law influence differently for different years between 2016 and the end of the simulation. We next can determine the share of electricity consumption of the data centers per year by comparing the energy needs of the data centers with the total global electricity production capacity, which we believe to be growing each year with a certain percentage. In system dynamic terms, we may see a growth of energy consumption of data centers by increased server loads, traffic, and storage, but we may also see balancing mechanisms like Moore's law and infrastructure efficiency gains.

Clearly, such a top-down model is simplistic as more external factors may influence data center energy needs and different components of a data center will use different levels of energy. Consequently, in the next sections we present a more bottom-up model for data center energy needs. Such a bottom-up approach will model energy needs per data center component (i.e., server, network, storage, and infrastructure) and summarize these results for a data center total. Per component also other variables than IP traffic may cause its energy need. Besides of further elaborating on the structure of such a model, values for variables, stocks and flows will have to be found from reliable information sources, like [1]. To implement our focus on application behavior, we added Cisco Systems [27] data on workloads per application, and we added data from Aslan et al. [28]. to cover data on IP traffic between data centers and between data centers and users. Very likely, these values will be open for debate and thus a point prediction will be much under debate. Consequently, as common in system dynamic studies [9], the model will be used to calculate-through multiple usage growth scenarios. This article thus will further continue with the design of our model in Section 4, the simulations in Section 5, and finally a discussion and conclusions Sections 6 and 7.

4. Design of a system dynamic model for data center energy forecasting

For developing a bottom-up model, we further decompose the high-level data center model of Fig. 1, and further describe each of its
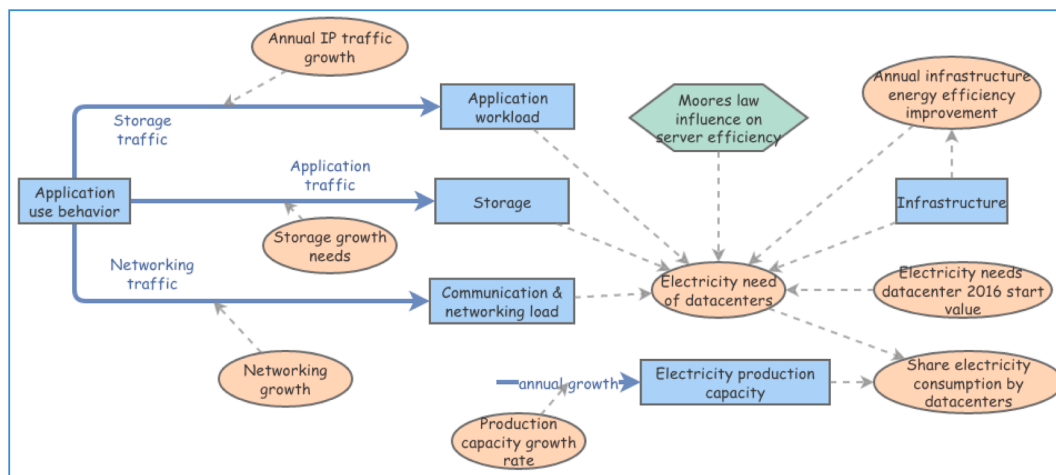

Fig. 1. A conceptual model for Internet traffic, energy needs, and CO2 emissions.

components, i.e., application behavior, servers' application workload, networking, storage, and infrastructure. At the end of this section, we integrate the resulting sub-models into one integrated model for forecasting the global data centers' electricity consumption (Fig. 2).

4.1 Application behavior parameters

Application behavior consists of activities like the use of data processing servers that cause workloads, the storage of data, and data transfers within and between data centers. Indirectly, workloads cause needs for cooling and other infrastructure services. Following Cisco Systems [27], we identify 8 types of application behavior. The first 4 are consumer-oriented behaviors consisting of search, social networking, video streaming and several other consumer apps. The last four are (smart) industry behaviors, consisting of cloud-ERP and business applications, databases, analytics and IoT, collaboration software and computations. Table 1 gives a selection of the Cisco Systems data for data center application workload, storage, and networking from 2016 until 2021 for consumer and business usage.

The annual growth rates (CAGR's) for application workloads, storage, and networking are 18.6%, 31.2% and 24.7% respectively. These parameters will be further detailed per data center component in the following subsections.

4.2 Servers' application workload parameters

Server electricity needs are caused by the volume of workloads that data centers process and the total number of servers needed for these processing workloads. Our assumption here is that a server disk runs at full capacity, but we adjust the watt used per server following [29] to compensate for idle time. To calculate the electricity consumption, the value of electricity consumption of a server disk is needed. The relation between the total volume of server power consumed depends on the total number of workloads, which is increasing as a consequence of more intensive use of data processing servers and the productivity of servers which enables more workload to be processed by the same unit of server.

Cisco Systems [27] identified 42.1 million of workloads for traditional data centers with a negative growth rate of 5% on the total amount of workloads, which is consistent with the shift away from

traditional data centers in the industry. Traditional data centers have 2.4 workloads per server disk and the efficiency growth rate of workloads per server is 6.9%. Cloud data centers had 199.4 million of workloads in 2016 and a growth rate of 22%. The number of workloads per server disk are 8.8 and efficiency growth rate is 8.5%. This would mean that cloud data centers are more efficient in their server services than traditional data centers. The number of workloads also differs between applications (see Table 2). Shehabi et al. [29] found different industry electricity consumption parameters for different server ranges, where traditional data centers, –i.e., those datacenters run internally by an organization and mostly with a size of less than 2,000ft^2 in size – on average had severs that consume 229.84 W per server in 2016, which increased by 3.6% per year since then. Non hyperscale cloud data centers, which have a size of between 2,000 and 20,000ft^2 mostly run by specialist data center services providers, have servers that use on average 302.75 W per server in 2016 and became 1.9% more efficient per year since then. For the very large scale hyperscale data centers, these number are respectively 253.49 and −0.4%.

4.3 Networking parameters

The total data center data traffic combines the traffic between data centers and users, data centers and data centers, and within data centers. The first two is named IP traffic. According to Cisco Systems [27], mobile traffic will seven-fold between 2017 and 2022, assuming that there will be 1.5 mobile devices per capita. By 2022, 5G will generate 2.6 more traffic than a 4G connection on average [27]. Using these insights, Cisco predicts a compound annual IP traffic growth (CAGR) of 22%, starting with 1,153 EB for 2016. Besides, Masanet et al. [1] base the electricity consumption of data center networking on the number of data center networking ports used. However, there is more involved in networking energy usage of data centers than port traffic power (see Table 3).

Aslan et al. [28] searched for a value for the electricity consumption of IP traffic that takes both the time used and the data volume. They conclude an electricity consumption of 0.06 TWh/EB that also decreases by half every 2 years [28]. When taking the electricity consumption of Aslan et al. and the total IP traffic according to Cisco Systems, we calculate (998 + 679) EB with 0.06 being the TWh/EB factor [28],
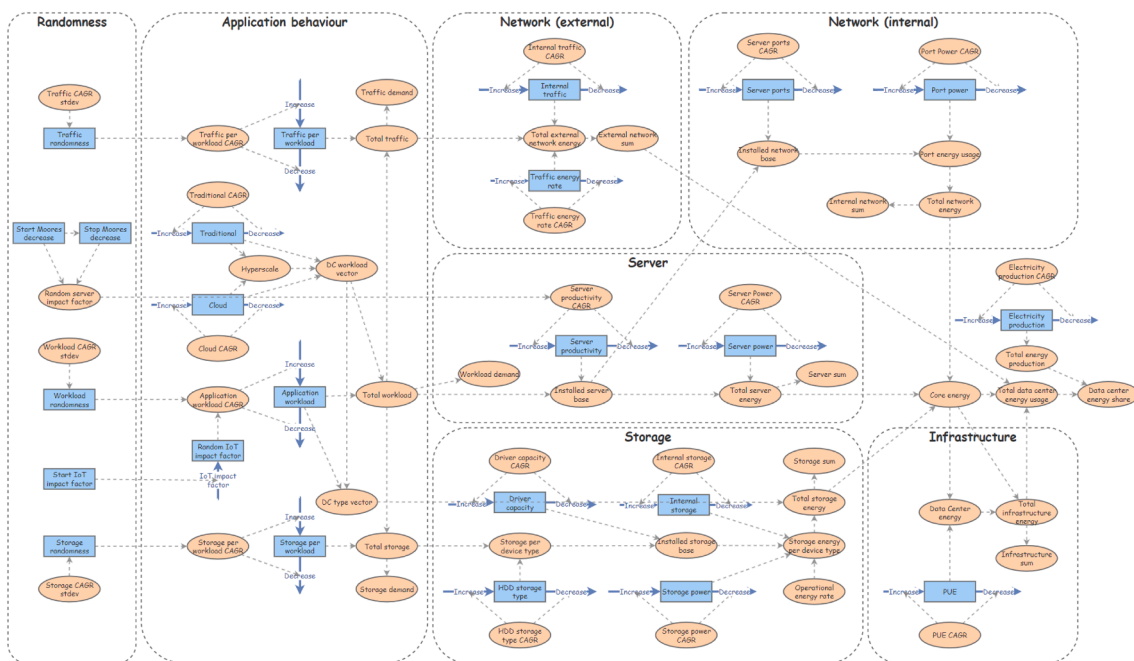

Fig. 2. Simulation model for global data center electricity consumption.

Table 1

5-year CAGR for datacenter workload, storage and networking.

Application workload in millions	2016	2017	2018	2019	2020	2021	5-year CAGR
Consumer total	58	77	98	115	133	152	21.3%
Business total	184	226	274	318	362	416	17.7%
Total	242	303	372	433	495	568	18.6%
Application storage in EB	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**5-year CAGR**
Consumer total	143	193	265	342	455	591	32.8%
Business total	520	697	915	1,156	1,516	1,982	30.7%
Total	663	890	1,180	1,498	1,971	2,573	31.2%
Application network in EB	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**5-year CAGR**
Consumer total	4,501	6,156	8,052	10,054	12,401	15,107	27.4%
Business total	2,319	2,931	3,505	4,070	4,716	5,449	18.6%
Total	6,820	9,087	11,557	14,124	17,117	20,556	24.7%

Source [27].

Table 2

Server parameters.

Class	Sub-class	Year 2016	5-year CAGR	Source
Application workload (millions)	Search	10	14.9%	[27]
	Social Networking	12	25.9%	
	Video Streaming	18	23.6%	
	Other consumer apps	18	18.5%	
	ERP & business apps	57	18.6%	
	Database, analytics, IoT	33	21.4%	
	Collaboration	48	14.4%	
	Compute	46	17.0%	
	Total	242	18.6%	
DC type workload (millions)	Traditional	42.1	−4.8%	[27]
	Cloud	103.9	12.9%	
	Hyperscale	95.5	29.2%	
	Total	242	18.6%	
DC type workload (frequency)	Traditional	17%	−19.7%	[27]
	Cloud	43%	−4.8%	
	Hyperscale	40%	8.9%	
Server productivity	Traditional	2.4	9.6%	[27]
	Cloud	8.8	8.4%	
	Hyperscale	8.8	8.4%	
Average server power	Traditional	230	3.6%	[29]
	Cloud	303	−1.9%	
	Hyperscale	253	−0.4%	

Table 3

Networking parameters.

Class	Sub-class	Year 2016	5-year CAGR	sources
Application traffic in EB	Search	776	20.7%	[27]
	Social Networking	931	32.3%	
	Video Streaming	1,397	29.9%	
	Other consumer apps	1397	24.5%	
	ERP & business apps	718	19.6%	
	Database, analytics, IoT	416	22.3%	
	Collaboration	605	15.3%	
	Compute	580	17.9%	
	Total	6,820	24.7%	
Network traffic in EB	Data center to user	998	25.2%	[27]
	Data center to data center	679	32.7%	
	Within data center	5,143	23.4%	
	Total	6,820	24.7%	
Traffic energy rate in TWh/EB	Data center to user	0.06	−28.5%	[28]
	Data center to data center	0.06	−28.5%	
	Within data center = port power.			
Average port power in W/port	Traditional	1.71	−5.9%	[30]
	Cloud	2.58	−9.5%	
	Hyperscale	3.19	−6.5%	
Ports per server	Average	4.53	1.3%	[30]

resulting in 100.62 TWh used for IP traffic towards data centers for 2016. Cisco estimates an average growth rate of 25.2% for data-to-user traffic and 32.7% annual growth rates for data center-to-data center traffic.

4.4 Storage parameters

Data center storage demand in exabytes is determined by multiplying the number of exabytes per workload with the application workloads requested and dividing this by the capacity per driver (no idle capacity assumed). Masanet et al. [1] give that different applications have different storage volume needs, where search has the lowest storage need of 23 EB with an annual growth rate of 26.7% until 2021, and social networking and video streaming with the highest growing storage needs of 35.6% and 37% annually until 2021 (Table 4). Regarding an estimate of actual energy costs of storage, the type of storage medium is important. Shehabi et al. [29] mention a 40–50% division of HDD and SSD drivers in data centers. The energy consumption of SSD drives reduces each year by about 2.3% and for HDD it reduces by 5.3% and SSD start with an average of 6.0 W per drive in 2016 and HDD starts with an average of 8.1 W per driver. The number of SSD increases with 9.5% annually and the number of HDD decreases by 2.9% annually.

Table 4

Storage parameters.

Class	Sub-class	Year 2016	5-year CAGR	Source
Application storage in EB	Search	23	26.7%	[27]
	Social Networking	29	37.0%	
	Video Streaming	48	35.6%	
	Other consumer apps	43	29.4%	
	ERP & business apps	148	31.5%	
	Database, analytics, IoT	128	30.6%	
	Collaboration	90	33.3%	
	Compute	154	28.2%	
	Total	663	31.2%	
Driver capacity in TB/drive	Solid state drives (SSD)	1.38	35.3%	[30]
	Hard disk drives (HDD)	3.78	27.0%	
Internal storage proportion	Operational energy as a fraction of storage energy	28.5%	−3.6%	[30]
Driver capacity frequency	Solid state drives (SSD) (=100%-HDD%)	19%	9.5%	[30]
	Hard disk drives (HDD)	81%	−2.9%	
Driver power in W	Solid state drives (SSD)	6.0	−2.3%	[30]
	Hard disk drives (HDD)	8.1	−5.3%	

4.5. Infrastructure parameters

Infrastructure energy needs consist of all data center energy needs that are not directly caused by server processing, storage, or network activities. Some of the main infrastructure components are cooling, light, security and building heating. The most common measure for data center infrastructure energy performance is power usage effectiveness (PUE) measured by the total amount of energy used by a data center divided by the energy used by its IT equipment [31]. Data centers differ highly per PUE, as given in Table 5. Much infrastructure energy costs reductions are gained by the generation of economies of scale, but there is no proof that this will continue in the same speed as for 2010–2018, and PUE's of 1.2 for hyperscale data centers cannot go better as 1.0, which would indicate zero infrastructure energy consumption.

4.6 Total data center energy consumption calculation

The total data center energy usage is a sum of its sub-model outcomes and their interaction, reinforcements, and balancing effects. We calculate the total energy costs with the complete model of Fig. 2. The users' behavior implicates that the data centers' electricity needs arise from the number of workloads demanded by customers and business practitioners. Therefore, we first calculate the storage capacity and corresponding IP traffic required to process one workload only. Second, we aggregate all these workloads to determine the impact in terms of server, storage, and network activities. Finally, the PUE values from Table 5 are added to conclude the total data centers' electricity needs with the infrastructures' share. The share of data center electricity consumption is calculated by dividing the predicted datacenter electricity needs for each scenario with the product of the actual electricity production volume for 2016 of 25,000TWh and the annual 2.5% growth factor of electricity production. Electricity production volumes and growth rates are taken from [32].

5. Simulations

The simulation model in Fig. 2 is used for two objectives. First, we predict how the data center electricity demand evolves if both today's technological and behavioral developments remain, i.e., the baseline model. Second, the model allows us to run a variety of scenarios and sensitivity analyzes. In Section 5.1, we first elaborate on the model's input parameters. The results of the baseline model are presented in Section 5.2, while a sensitivity analysis regarding the user's application behavior is discussed in Section 5.3. Finally, we propose two alternative scenarios referring to Moore's law and the rise of the Industrial IoT in Section 5.4.

5.1. Simulation settings

For simulating the future energy consumption of data centers, we use a time window until 2030, because the electricity suppliers' demand response time is between 5 and 8 years [33]. Additionally, we expect much of energy scarcity issues to show up after 2025, as will be shown later in the simulations. For these simulations, we present the input values for all variables of our model in Tables 2–5. Each variable is associated with an initial value originating from 2016, and a 5-year CAGR value.

Table 5
Infrastructure parameters.

Average PUE value	2016	5-year CAGR	source
Traditional	2.10	−1.0%	[27]
Cloud	1.66	−1.0%	
Hyperscale	1.21	−0.5%	

5.2. Data center energy forecast – Baseline model

We simulate the data centers' energy consumption for the period 2016–2030 with the model of Fig. 2 and assuming that future technological and behavioral trends are maintained. Fig. 3 gives the annual data center energy consumption (in TWh) for the baseline model.

The majority of the applications' workloads are processed by hyperscale servers, increasing their relative share from less than 50% in 2016 to over 80% in 2030.The total workload starts with 242 million in 2016 and ends with 2,760 million in 2030. For traditional, cloud and hyperscale, the starting values for 2016 electricity consumption are respectively 35.48, 25.21, 29.25 TWh and move to 9.18, 32.29, and 177.59 TWh in 2030.

The data centers' networking is represented by IP traffic and will become larger over time [27], starting with 5,143 EB in 2016 and growing to 97,640 EB in 2030. The data centers' external IP traffic will also accelerate growing from 998 EB for data center-user traffic and 679 EB data center-data center traffic in 2016 to 23,210 EB and 35,650 EB in 2030 respectively. Because of the declining energy costs of the data centers' external IP traffic (about 28.5% per year) and the increased energy efficiency of server ports (about 7.3% annually), the high traffic growths do not contribute much to the energy needs of data centers. The different data centers however do have different total energy costs internally due to the global workload allocations, for hyperscale we predict about 5.02 TWh in 2030, for traditional and cloud respectively 0.10 and 0.48 TWh.

The increase in data storage demand is for traditional, cloud and hyperscale data centers respectively from 118.93, 235.63, and 309.14 EB in 2016 to 368.47, 5,023.40 and 24,840.67 EB in 2030. This sharp increase in storage demand does not result into a rising electricity demand for storage devices, mainly due to the implementation of more energy efficient SSD devices. Consequently, the storage-related electricity consumption drops from 18.33 TWh in 2016 to 15.23 TWh in 2030. Because of the global workload allocations, hyperscale data centers consume the largest part of the storage related energy consumption. Hyperscale data centers will consume 12.72 TWh in 2030, while traditional and cloud data centers only consume 0.13 and 2.39 TWh respectively.

The data centers' total energy predictions thus differ for traditional, cloud and hyperscale data centers and goes from 83.73, 53.98, and 48.06 TWh in 2016 to 17.16, 50.71, and 220.32 TWh in 2030. The global data centers' energy consumption thus will grow with approximately 12% in between 2016 and 2030, if (and only if) today's technological and behavioral trends remain unchanged for the upcoming decade.

5.3. Sensitivity analysis of the baseline model

A major contribution of our simulation model is the association of the data centers' electricity consumption with the users' application behavior. A comparison of multiple Cisco Systems forecasts indicates that estimations of users' demand for data center servers depend on both technological and behavioral trends [34,34–37]. The observed stochasticity of the user's application behavior makes it hard to give a reliable estimate of the data center electricity consumption in 2030 (see Table 7). However, we can derive the sample variance and standard deviation for the model's CAGR values by comparing Cisco Systems behavioral forecasts over several years [34,34–37]. This procedure enables us to simulate the baseline model with a randomized set of growth values for the users' workloads, IP traffic, and storage activities per application.

The growth projections in Table 7 allows us to randomly generate a future scenario by slightly altering the users' application CAGR values. For each application, we add a random error value to the application's average CAGR values. The error is normally distributed with $\mu = 0\%$, and either $\sigma = 3.163\%$, $\sigma = 2.091\%$, or $\sigma = 1.469\%$ for the application's workloads, IP traffic, or storage activities respectively. The result of a
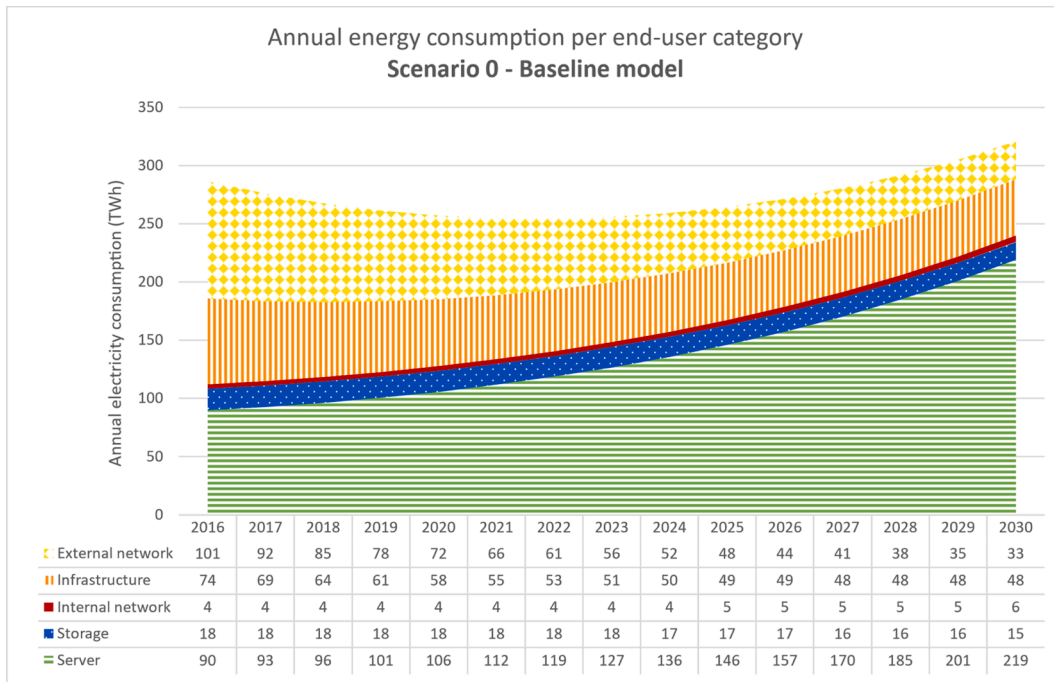
Fig. 3. Energy consumption per end-user category predicted with the baseline model.

Table 7
Cisco Systems forecasts of the users' application behavior [34,34–37].

Workload in millions	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	5-year CAGR
Forecast in 2011	57.5	70.2	85.8	107.4	131.2	154.8	–	–	–	–	–	–	21.9%
Forecast in 2012	–	71.1	86.6	108	131.6	155.1	180.6	–	–	–	–	–	20.5%
Forecast in 2013	–	–	83	99.3	119.5	140.4	163.2	188.2	–	–	–	–	14.5%
Forecast in 2014	–	–	–	108.3	125.2	143.6	163.4	184.8	211.5	–	–	–	14.3%
Forecast in 2015	–	–	–	–	129.5	161.8	194.2	232.7	276.6	319.7	–	–	19.8%
Forecast in 2016	–	–	–	–	–	175	225	290	355	410	465	–	21.6%
Forecast in 2018	–	–	–	–	–	–	242	303	372	433	495	568	18.6%
IP traffic/ workload in EB	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**5-year CAGR**
Forecast in 2011	19.8	23.5	26.1	27.5	28.5	30.7	–	–	–	–	–	–	9.1%
Forecast in 2012	–	24.7	29.5	30.1	31.3	33.8	36.8	–	–	–	–	–	8.3%
Forecast in 2013	–	–	30.9	33.6	35.3	37.1	39.1	41.1	–	–	–	–	4.8%
Forecast in 2014	–	–	–	28.3	30.6	32.9	35.5	38.3	40.5	–	–	–	7.5%
Forecast in 2015	–	–	–	–	26.6	27.3	28.9	30.0	31.0	32.4	–	–	4.0%
Forecast in 2016	–	–	–	–	–	26.7	29.0	29.7	30.3	31.5	33.0	–	4.3%
Forecast in 2018	–	–	–	–	–	–	28.2	30.0	31.1	32.6	34.6	36.2	5.1%
Storage/workload in EB	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**5-year CAGR**
Forecast in 2016	–	–	–	–	–	2.2	2.4	2.7	3.0	3.4	4.0	–	12.7%
Forecast in 2018	–	–	–	–	–	–	2.7	2.9	3.2	3.5	4.0	4.5	10.6%

Note: Users' application behavior in million workloads, storage demand per workload EB, and corresponding IP traffic per workload EB.

Monte Carlo simulation with 10,000 replications is given in Fig. 4. The sensitivity analysis results into a 95% confidence interval of [267, 422] TWh for the global data centers' electricity consumption. Our projections are similar to the estimations made by [3] while the forecasts from [38] are more fluctuating. The discrepancy between the forecast results is mainly explained by the assumptions made regarding the technological developments. Section 5.4 presents alternative scenarios.

5.4. Scenarios

The main power of our model is to run scenarios and validate several "what-if" hypotheses. In this section, we discuss two alternative scenarios for the baseline model:

• **Scenario 1:** End of Moore's law (Section 5.4.1);
• **Scenario 2:** The rise of the Industrial IoT (Section 5.4.2).

Table 8 gives the differences in data center electricity demand for the baseline model and the two scenarios, including the nonlinear trends of data centers' share of the available electricity. We will elaborate on these results in the following subsections. Note that the simulation model is published online for other researchers to use the model for other scenarios as well (e.g., the introduction of 5G networks, the rise of edge and/or fog computing, and the implementation of smaller public service providers,).

5.4.1. Impact of Moore's law

As stated earlier, Moore's law is expected to stop having an influence between 2021 and 2023 [2]. We will adjust the server productivity CAGR values to slow down the annual energy decline per same processing unit. Therefore, the possible end of Moore's law is modeled as a declining trend from 2016 (25% efficiency increase) until 2023 (zero) following [2]. The impact of ending Moore's law is given in Fig. 5, including a recalculating of data center energy needs. In 2016, servers
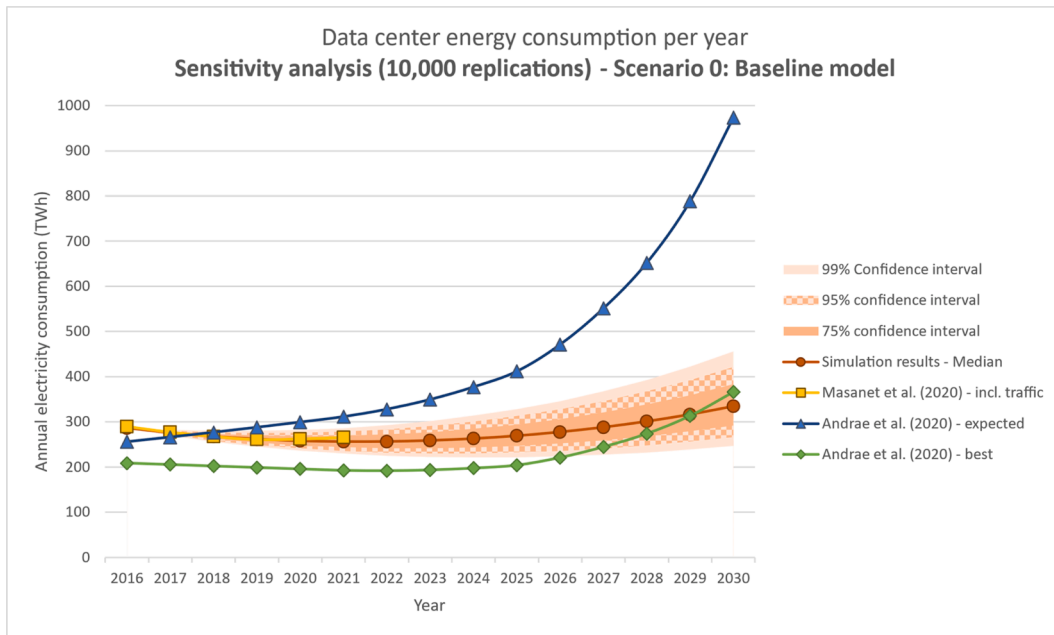
Fig. 4. A sensitivity analysis of the baseline model, including 10,000 replications. Note: Including 75%, 95%, and 99% confidence intervals for the randomized user profiles. Three benchmark values from other researchers [1,38] are added for comparison.

Table 8
Comparison of the data center electricity forecasts per scenario.

Electricity need	Baseline model	Scenario 1	Scenario 2
Electricity need 2016	286.42 TWh	286.42 TWh	286.42 TWh
Electricity need 2030	320.87 TWh	658.03TWh	364.00 TWh
Relative difference	+12.03%	+129.74%	+27.08%
Electricity share 2016	1.15%	1.15%	1.15%
Electricity share 2030	0.91%	1.86%	1.03%
Relative difference	−20.72%	+62.59%	−10.06%

are the largest energy consumers in data centers (approximately 31.40%), but the end of Moore's law will increase the server's electricity demand further. Data centers will have to expend the installed server base in order to process the increasing number of workloads. The lack of processing efficiency gains results in a total server energy consumption of 491 TWh in 2030, while only 219 TWh is required if Moore's law continues existing during the upcoming decade. Therefore, the server category represents 74.66% (or 68.27%) from the global data centers' electricity demand in case Moore's law ends (or not). Doubling the server's energy consumption will inflate the data centers' total energy need up to 658 TWh in 2030 (see Fig. 5), which will also increment the data centers' share in the global energy consumption from 1.15% up to 1.86% in 2030.

The median electricity forecast in Fig. 5 relies on the assumption that the server's productivity improvements will start declining in between 2016 and 2023, as indicated by [2]. Most scientists predict that Moore's
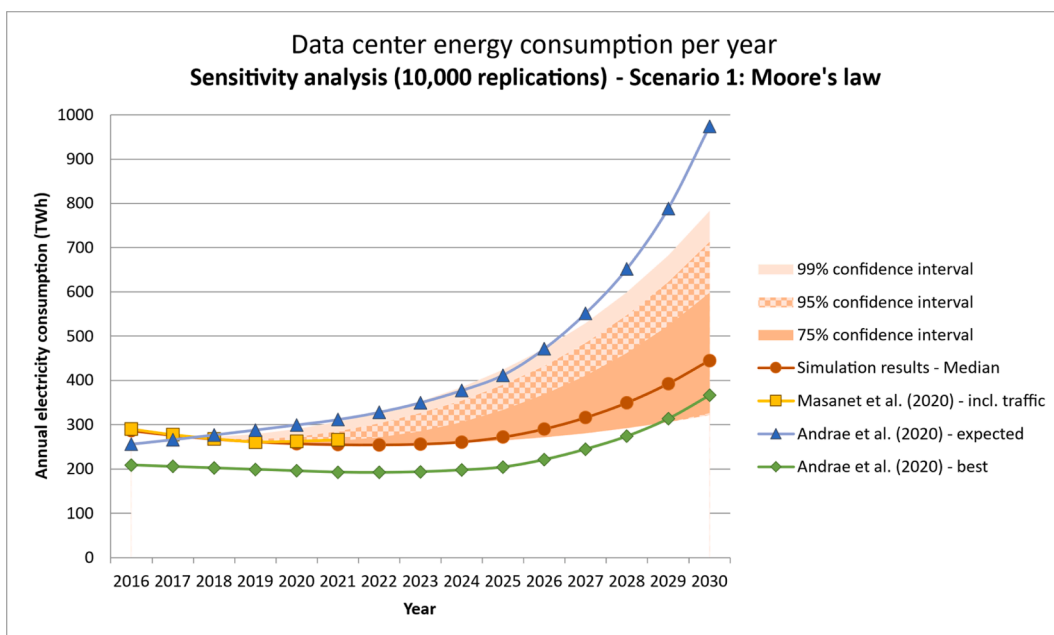

Fig. 5. Data center electricity needs effects of ending Moore's law.

law will end quite soon due to technical limitations, but it remains uncertain when the end of Moore's law will take place. Therefore, we perform a Monte Carlo simulation with 10,000 replications where the ending of Moore's law is generated randomly. This means that the first year of the server's productivity decline is uniformly distributed with U (2016, 2030), while the point of zero efficiency gains is uniformly distributed with U("Year start decline", 2030). The server's productivity will drop linearly once the decline is initiated. This procedure results into a 95% confidence interval of [326, 713] TWh for the global data centers' electricity consumption, if and only if the end of Moore's law is randomly distributed. The inclusion of the uncertainty associated with the users' application behavior will slightly alter both lower- and upper bounds (see Section 5.3), resulting into a 95% confidence interval of [307, 776] TWh. This means that the end of Moore's law has a significant worse impact on the global data centers' electricity forecast in comparison with the alternative scenarios for the users' application behavior.

5.4.2. Rise of the Industrial IoT

In 2015, only 15 billion of the world's 1.5 trillion physical objects were connected to the internet [39]. During the upcoming decade, we expect to see a sharp rise in the number of connected devices, due to Industry 4.0 IoT developments [40]. The number of devices connected to the internet, i.e., IoT, grew with 333% between 2012 and 2015 [39], which is equivalent to a CAGR value of 49.33%. This high growth percentage is consistent with recent predictions of IoT growth. For example IDC estimates data generated from connected IoT devices will grow from 18.3 ZB in 2019 to 73.1 ZB by 2025, which is about a 400% increase in 5 year and a CAGR of about 32% (https://www.idc.com/getdoc.jsp?cont ainerId = prAP46737220). In the baseline model, we mentioned a CAGR for workloads of databases, analytics, and IoT devices altogether of + 21.4% (see Table 2 from [27]). We adjust the CAGR values for database, analytics, and IoT application workloads to simulate the rise of the Industrial IoT by slowly and linearly increasing this growth percentage up to the 31.91% in 2030, which represents the growth factor of the IoT's installed base given by [39].

The workloads released by physical objects have to be processed by data centers due to the cloud-based architectures installed, resulting into a total of 249 TWh consumed by the installed server base in 2030. In comparison with the baseline scenario, the Industrial IoT paradigm will cause the global server electricity consumption to grow with 13.82% till 2030. More servers also require more internal traffic and larger supporting infrastructures, which results into an additional electricity consumption of 1 TWh (+12.39%) and 9 TWh (+19.02%) respectively. The increased use of IoT devices will also inflate the storage related energy demand for 2030 (+13.03%) because business related applications require more database capacity in comparison with user applications. The baseline scenario predicts that storage devices will consume less electricity due to energy efficient storage devices, but the rise of the Industrial IoT will cause the storage related electricity consumption to go up again, mainly due to the increased demand for storage capacity in total. Therefore, we can conclude that the Industrial IoT will result in an exponential growth of data center electricity consumption if business applications still rely on cloud or hyperscale-based architectures. These growth projections will accumulate towards a global data center energy consumption of 364 TWh in 2030 (see Fig. 6), while the data centers' share in the global energy consumption will drop from 1.15% down to 1.03% in 2030.

The benefits from the Industrial IoT seem to be very promising [39], but it remains unknown when business practitioners will implement the required platforms to support their daily operations at full scale. For example, [43] revealed that number of IoT devices grew with a CAGR of 49% between 2012 and 2015, while [41] predicts a lower CAGR value (+19%) for the number of machine to machine connections. We will embrace this uncertainty into our model by randomly generating a growth factor for the CAGR value associated with the Industrial IoT. A uniform distribution is used for the generation of a yearly growth percentage, ranging from today's CAGR value (+21%) up to a maximum CAGR value found in literature (+49%). This procedure results into a 95% confidence interval of [322, 489] TWh for the global data centers' electricity consumption, if the growth expectations for Industrial IoT applications are randomly distributed. It appears that the IoT's additional energy demand is quite modest in comparison with the 95% growth projections of the baseline model (see Fig. 4, or the discussion in Section 5.3). However, the inclusion of all uncertainty associated with the users' application behavior will significantly stretch up the gap in between the lower- and upper bounds, resulting into a 95% confidence interval of [290, 577] TWh. Therefore, the rise of the Industrial IoT could significantly speed up the global data centers' electricity needs for 2030, especially if businesses require more server and storage capacity
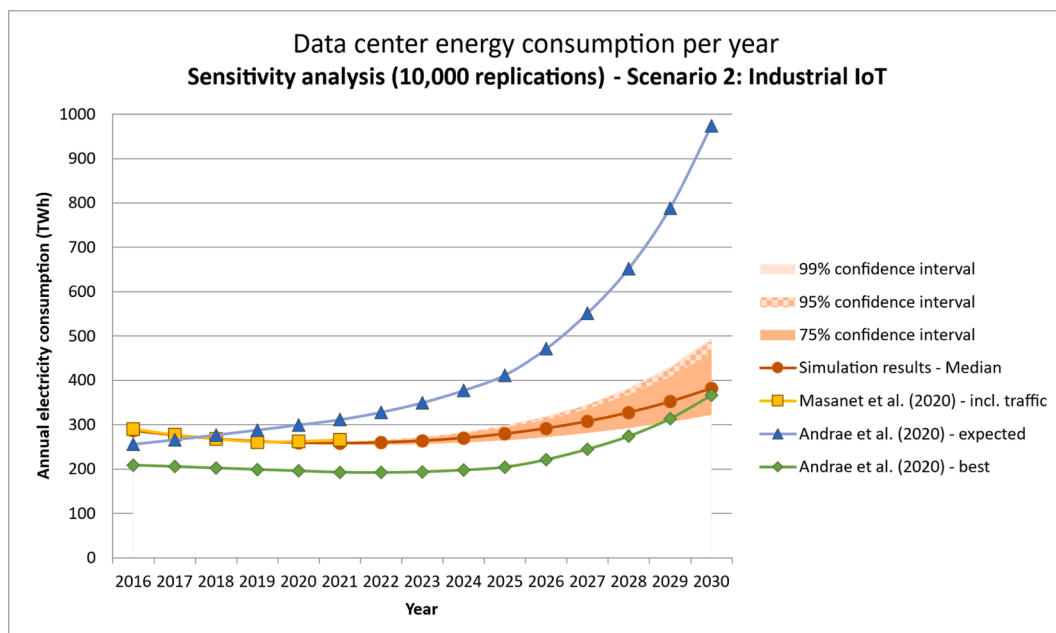

Fig. 6. Data center energy needs effects of the Industrial IoT.

for each IoT device installed.

5.4.3. Combined effects of Moore's law and IoT

The combined effects of an end to Moore's law and an increase of IoT applications for our model and using our expected behavioral usage trends is given in Fig. 7. Both scenarios will significantly increase the number of servers required to process all workloads. In 2030, 74% of all the global data centers' electricity is only consumed by servers. The infrastructure category is also responsible for a large part of the total energy consumption (17%), while the impact of all network and storage categories is quite negligible (9% only). Our sensitivity analysis reveals that the future impact of the data centers' electricity consumption is vulnerable to both technological and behavioral developments, since the 95% confidence interval of [343, 1031] TWh for all three scenarios is relatively wide. The end of Moore's law is the main cause for the exponential growth projections, while the uncertainty in the behavioral usage trends explains the discrepancies found in the literature (e.g., the benchmark values reported by [2,39]).

6. Discussion

Our study had to work with a number of assumptions and simplifications, which may be studied in further.

One assumption is our focus on global data as a relevant approach for any prediction. In reality, data centers tend to become more concentrated in very large cloud centers, named hyperscale centers, which will do most of the data center services now and especially in the future. Because hyperscale centers have very low PUE's, this may be an economically good development but also implies that energy consumption of all data centers is highly local, especially in highly industrialized geographical areas. This may make the percentage of electricity consumption in these areas very high and resulting in large energy delivery problems. For example, in the Amsterdam region, data centers' growing demand for electricity has become problematic ([42;43] Accessed: 26.01.2020.) Any percentage of electricity capacity that data centers claim make the share smaller for other increasing electrifications of life, like traffic and heating. The difference in time horizons for the rapidly developing data center industry and the less agile response of electricity distribution channels (requiring large and complex infrastructure changes that may take about 8–10 years) is a challenging practical planning problem with as yet little academic interest [29,44].

A second assumption was our limitation to server power predictions with only average server power values. Although we follow Shehabi et al. [29] in this by calculating the average sum product of all server types' wattages per data center type, server power usage may be highly different over a day. Working with usage peaks thus will be more realistic and is an important subject for further research. Morley et al. [12] argue for a focus on peak load estimations, which are stronger predictor of energy capacity needs than the annual volumes we discussed in here. In contracting with electricity suppliers, data centers have to negotiate electricity for their peak moments to remain fully operational and thus in fact needs much more than the average. Economies of scale and virtualization are important developments that reduce the total electricity demand of data centers.

A third assumption is our simplification on IoT and related to that 5 g networks. IoT is expected to further increase the usage of the Internet in an yet unpredictable way [45,46]. 5 g could result in less data travel distances and may reduce the load for data centers by edge computing, but at the same time will require data processing in more decentralized units with less efficient electricity usage. We hypothesize that data centers will remain their central position in IT infrastructures because of their ability to exploit economies of scale (especially cooling, UPS and processing virtualization) [1].

A fourth assumption is our limitation to data centers. IDC [47] predicts the size of the global datasphere – i.e., the total volume of digital content stored on core, edge and endpoint computing locations and devices – to grow from 33 Zettabyte (ZB) in 2018 to 175 ZB in 2025. Compared to data center storage volume, which is about 1.180 ZB in 2018, 31 ZB in 2030 (CGAR of 31.31%), information storage and possible related traffic seems to be just a small part of the total size of data stored, processed, and shared between all the possible electronic device we have nowadays.

7. Conclusions and further considerations

Our research question "How will future data center energy demand evolve as a consequence of its growing usage?" now is answered by a simulation model and a set of simulation parameters found from the literature. We expect a combined growth of data center electricity needs of 286 TWh in 2016 up to 321 TWh in 2030, if today's technological and
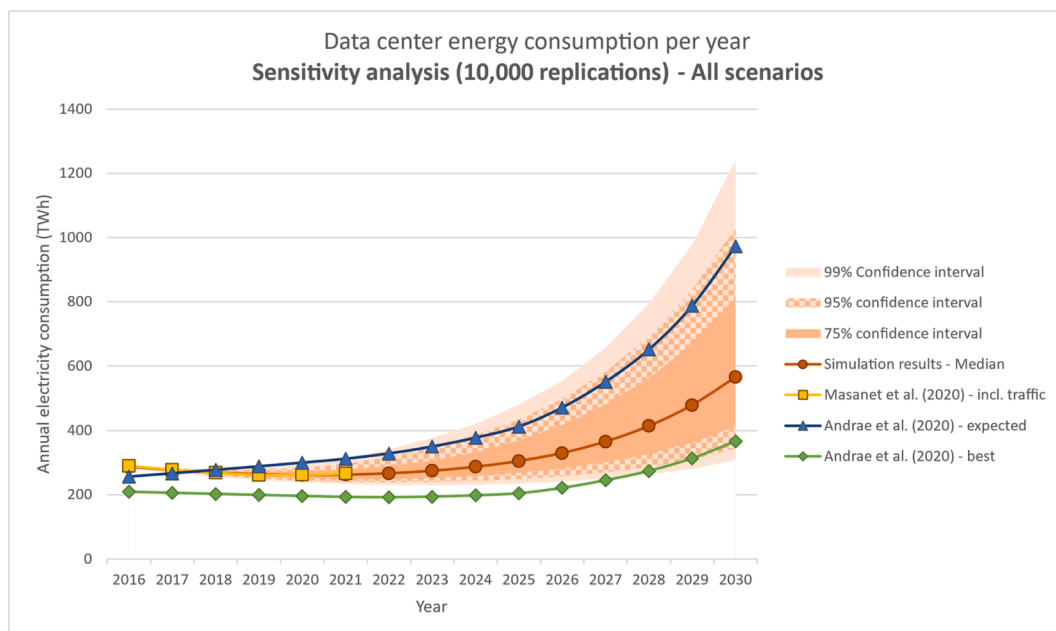

Fig. 7. Effect of both Industrial IoT and Moore's law on data center electricity consumption.

behavioral trends remain the same. The end of Moore's law results in a total of 658 TWh for 2030, and an increase of the global data centers'-share electricity consumption from 1.15% in 2016 to 1.86% in 2030. The rise of Industrial IoT applications may consume a total of 364 TWh (about 1.03%) in 2030. Moore's law and IoT combined cause data center energy needs going up to 752 TWh in 2030, and about 2.13% of global electricity available. However, looking into the future is difficult and must be accompanied with uncertainty levels, and thus we performed Monte Carlo simulations with 10,000 replications for all the baseline, Moore's law, IoT, and combined scenarios. The outcomes for the averages, lower and upper bounds of these scenarios is given in Table 9.

Our sensitivity analysis in Table 9 reveals that the future impact of the data centers' electricity consumption is vulnerable to behavioral usage trends, since the 95% confidence interval of [343, 1031] TWh is relatively wide. We conclude that the global data centers' future energy demands are kept reasonably constant due to technological innovations, even when both consumers and business workloads seem to grow exponentially during the upcoming decade. However, the end of Moore's law will most likely cause an exponential growth of the data centers' electricity consumption, while the uncertainty in both technological and behavioral developments explains the discrepancies found in today's literature (e.g., the benchmark values reported by [2,39]).

We further discuss the academic contributions of this article and next the challenges for the future from a data center supply and a data center usage perspective.

7.1. Contribution to the literature

The goal of this article is to discuss the impact of user demand growths for data center electricity needs. The simulation of this model shows exponential growths of data center usage. This results in about approximately 2 or 3 times more energy needs in 14 years. This insight is less than Andrae [14] and Belkhir and Elmeligi [3] who predicted about 2,000 TWh for 2030 but more than Borderstep who predict a bit less than 1,000 TWh for 2030. Whatever the differences in numbers, we developed a system dynamic model on basis of existing literature that defined the stock, flows and variables that determine the outcome of TWh from data center usage. This model is also publicly available via this publication [Note: to remain anonymous, we publish the link after acceptance of this article; Insight Maker file though is available] and can be used by any researcher or decision maker with alternative parameters and settings to replicate our findings. Instead of comparing estimates, our work goes into what is behind estimates by a transparent simulation model based on state-of-the-art literature.

Two alternative approaches to estimating electricity consumption of data centers and communication technology result in comparable findings but have a different dependent variable, i.e., not only electricity consumption but greenhouse gas estimations (GHGE) and a broader scope, i.e., not only data centers but also other components of the Datasphere. For example Belkhir and Emelgidi [3] have focused on the production and utilization of electronic devices. They state that the production of devices is also highly influenced by their life cycle length and thus is responsible for additional energy usage and CO_2 emissions, but the main part is the actual number of devices and its intensity of use.

Data center utilization in their model is only one part of the energy costs of use, and as addition to ecological effects they extend their impact analysis on greenhouse gas emissions (GHGE) instead of CO_2 alone. Their analysis predicts a growth of GHGE effects of communication technologies from a bit more as about 1 to 1.6% in 2007 to about 14% of the 2016 vol in 2040 in 33 years in 14 years for data center CO_2 emissions according to our prediction. These analyses may be hard to compare because CO_2 is only one part of all GHGE and Belhir and Emelgidi also include the manufacturing and usage of smaller devices, like smart phones, in their analysis. Although both analyses indicate alarming GHGE effects, we kept this out of our analysis because the calculation and simulation of GHGE implications of electricity production requires an intensive and complex additional study on a subject that is highly influenced by new energy generation technologies.

Going more in depth on user behavior, we further describe the demand and supply sides of data center services in the following sections.

7.2. Data center supply perspective implications

Data centers are highly technological and innovative organizations. Especially regarding their electricity consumption, they have high incentives to innovate as about 70% of their costs contain the electricity bill (https://info.siteselectiongroup.com/blog/power-in-the-data-center-and-its-costs-across-the-united-states, accessed April 2, 2020). This study has kept the technological side of data centers as a given but important technological innovations may make this become irrelevant. The following innovations have been mentioned in the literature may impact our forecasts of the coming decade:

1. Photonic computing is especially seen as contributing to very low energy costs of data transmission and data processing [2]. Cyber-optic transmission is already common at many places but photonic data processing is still met with controversy and doubts regarding its applicability [48]. The possible contribution of photonic computing thus needs further research.
2. Immersion computing allows all processing units to work in a basin of oil, resulting in low levels of heat production and the reduction of cooling costs of data centers by about 90% [49]. This technology is still in an experimental stage but could be fully operational within the next 5–10 years. The contribution to energy costs however is limited to reducing infrastructure cooling costs that may be less than 20% of the total energy costs of a modern data center [1].
3. Quantum computing may expand our ability to solve combinatorically complex problems in polynomial time, but it will not be much good for word processing or graphics rendering [2]. The technology itself will probably become available in the mid of this decade, but its application opportunities are restricted [50].
4. Instead of thinking solely about having all data stored and processed in central cloud data centers, much data processing is better done on decentralized edge units in combination with 5 g communication [40]. This will reduce the energy costs of central data center servers and networking but will reduce the opportunities of exploiting economies of scale that are easier to realize via large data centers. Edge computing is also less useful for data transmissions that will have to go via the Internet core, like large scale video services and social media, but will be especially useful for regional smart industry and transportation development [51]. Edge computing also has an energy costs and it is not sure if edge processing is more efficient than data center processing.
5. Heat nets. Instead of focusing on data centers as energy consumers, they also can be seen as energy suppliers to local households and industry. The heat produced by data centers can be sent through a city heating network, and when at the end heat has been given to houses, the cooler liquid is used for cooling the data center again. This may reduce the cooling energy costs to nearly zero for a data

Table 9
Sensitivity analysis for all data center electricity forecasts.

Data center electricity consumption (TWh)	Scenario 0: Baseline	Scenario 1: Moore's law	Scenario 2: IIoT	All scenarios combined
Number of replications	10,000	10,000	10,000	10,000
95% Lower bound	266.75	326.22	322.86	343.25
Median	334.26	445.03	382.00	565.87
95% Upper bound	421.59	712.63	489.17	1031.27

center, although it does not reduce the energy costs of storage, processing and transmission of data [52,53].

7.3. Data center demand side perspective

Besides of more energy efficient data centers also user behavior changes may contribute to handling the energy scarcity problem. The key idea that motivates our selection of mitigation options is the historic awareness that each innovation after a stage of unlimited growth and explored opportunities comes in a stage of large scale adoption, diffusion and maturity (if successful) that generates scarcity of resources [8,54,55]. Given this scarcity concept, we can think of strategies that disallow certain types of behavior, that encourage certain types of behavior, and that allow certain types of consumption if compensating something worse.

Some highly energy intensive data center usages have an unclear contribution and could be considered to be disallowed. One of them is the block chain, which although without a large diffusion yet, is energy-intense in its code generation and verification networks [56–58]

Another, but more radical, option is the disallowance of electronic advertising. Experiments with adblockers show that about 40% of the energy consumption of mobile phones can be avoided with applying adblockers [59]. The impact of adblockers on the profitability of internet content however is large and may be destructive of content business models [60]. Consequently, content owners need to develop attractive alternative business models that may combine public sponsorship with payments. The feasibility of these new business models however is still in need of research.

A growing and high energy consumption industry is the gaming industry, which also can be defined as a variant of interactive video, the probably most IP traffic consuming variant of data center usage [12]. Some limitations to the growth of this industry may be needed.

For realizing more energy saving behavior, positive encouragements may be useful. Such encouragements can be realized by creating users' energy awareness and given positive feedbacks for efficient users. Creating awareness can be realized by public campaigns, but alternatively we also suggest a metering system that can give users an energy usage report of what they were doing, possibly including a recommender or gamification system to avoid high avoidable usage [61]. Positive feedback can also be created as feed forwards like a pricing system that informs the user before actual consumption about the energy and GHGE report or giving information about a monetary price. For the pricing "solution" a price elasticity must be known, or a behavioral impact prediction must be known before such a strategy would be recommended for practice. For GHGE reporting and feedback, more research is needed on the future ecological footprint that data center usage causes.

As a follow up on euro commissioner Vestager's call for "technology with a purpose" [62], we believe that technology is great if it helps to make life better. So, if data centers cause energy and ecological problems, there would not be a problem if following the next two principles:

1. The service gives a strong and demonstrably great contribution to social goods, like the generally well accepted United Nations Challenges.
2. The service's energy consumption by far should reduce energy needs of other activities (e.g., large improvements in traffic mileages or production energy use avoided), resulting in energy neutral or energy reduction applications.

Both these two principles require research for developing a useful decision method.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2021.116798.

References

[1] Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. Science (80-) 2020;367:984–6. https://doi.org/10.1126/science.aba3758.
[2] Shalf J. The future of computing beyond Moore's Law. Philos Trans R Soc A Math Phys Eng Sci 2020;378:20190061. https://doi.org/10.1098/rsta.2019.0061.
[3] Belkhir L, Elmeligi A. Assessing ICT global emissions footprint: trends to 2040 & recommendations. J Clean Prod 2018;177:448–63. https://doi.org/10.1016/j.jclepro.2017.12.239.
[4] Hintemann R, Hinterholzer S. Energy consumption of data centers worldwide: How will the internet become green? CEUR Workshop Proc.; 2019.
[5] Dayarathna M, Wen Y, Fan R. Data center energy consumption modeling: a survey. IEEE Commun Surv Tutorials 2016;18:732–94. https://doi.org/10.1109/COMST.2015.2481183.
[6] Yu L. Coevolution of information ecosystems: a study of the statistical relations among the growth rates of hardware, system software, and application software. ACM SIGSOFT Softw Eng Notes 2011;36:1–5.
[7] Chen C, Watanabe C, Griffy-Brown C. The co-evolution process of technological innovation—an empirical study of mobile phone vendors and telecommunication service operators in Japan. Technol Soc 2007;29:1–22.
[8] Rogers EM. Diffusion of innovations. Simon and Schuster; 2010.
[9] Barlas Y. Formal aspects of model validity and validation in system dynamics. Syst Dyn Rev 1996;12:183–210. https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4.
[10] Hinton K, Baliga J, Feng MZ, Ayre RWA, Tucker RS. Power consumption and energy efficiency in the internet. IEEE Netw 2011;25:6–12. https://doi.org/10.1109/MNET.2011.5730522.
[11] Van Heddeghem W, Lambert S, Lannoo B, Colle D, Pickavet M, Demeester P. Trends in worldwide ICT electricity consumption from 2007 to 2012. Comput Commun 2014;50:64–76. https://doi.org/10.1016/j.comcom.2014.02.008.
[12] Morley J, Widdicks K, Hazas M. Digitalisation, energy and data demand: the impact of Internet traffic on overall and peak electricity consumption. Energy Res Soc Sci 2018;38:128–37. https://doi.org/10.1016/j.erss.2018.01.018.
[13] Andrae A, Edler T. On Global Electricity Usage of Communication Technology: Trends to 2030. Challenges 2015;6:117–57. https://doi.org/10.3390/challe6010117.
[14] Andrae A. Comparison of several simplistic high-level approaches for estimating the global energy and electricity use of ICT networks and data centers. Int J Green Technol 2019;5:50–63.
[15] Andrae A. Hypotheses for primary energy use, electricity use and CO2 emissions of global computing and its share of the total between 2020 and 2030. WSEAS Trans Power Syst 2020;15:50–9.
[16] Markov IL. Limits on fundamental limits to computation. Nature 2014;512:147–54. https://doi.org/10.1038/nature13570.
[17] Shahidi G. Slow-Down in Power Scaling and the End of Moore's Law? 2019 Int. Symp. VLSI Technol. Syst. Appl., IEEE; 2019, p. 1–1. https://doi.org/10.1109/VLSI-TSA.2019.8804705.
[18] Smith G. Data mining fool's gold. J Inf Technol 2020;35:182–94.
[19] Derbyshire J, Giovannetti E. Understanding the failure to understand New Product Development failures: Mitigating the uncertainty associated with innovating new products by combining scenario planning and forecasting. Technol Forecast Soc Change 2017;125:334–44.
[20] van der Heijden K. Scenarios: The Art of Strategic Conversation. 2nd ed. John Wiley & Sons; 2005. https://doi.org/10.1038/sj.jors.2600027.
[21] Isaacs W, Senge P. Overcoming limits to learning in computer-based learning environments. Eur J Oper Res 1992;59:183–96. https://doi.org/10.1016/0377-2217(92)90014-Z.
[22] Sterman JD. All models are wrong: reflections on becoming a systems scientist. Syst Dyn Rev 2002;18:501–31. https://doi.org/10.1002/sdr.261.
[23] Rahim FHA, Hawari NN, Abidin NZ. Supply and demand of rice in Malaysia: A system dynamics approach. Int J Supply Chain Manag 2017;6:234–40.
[24] Leon H, Osman H, Georgy M, Elsaid M. System dynamics approach for forecasting performance of construction projects. J Manag Eng 2018;34:04017049. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000575.
[25] Fortmann-Roe S. Insight Maker: a general-purpose tool for web-based modeling & simulation. Simul Model Pract Theory 2014;47:28–45. https://doi.org/10.1016/j.simpat.2014.03.013.

[26] Hristoski I, Mitrevski P. Evaluation of business-oriented performance metrics in eCommerce using web-based simulation. J Emerg Res Solut ICT 2016;1:1–16.

[27] Cisco Systems. Forecast and Methodology, 2012–2017; 2013.

[28] Aslan J, Mayers K, Koomey JG, France C. Electricity intensity of internet data transmission untangling the estimates. J Ind Ecol 2018;22:785–98. https://doi.org/10.1111/jiec.12630.

[29] Shehabi A, Smith SJ, Masanet E, Koomey J. Data center growth in the United States: decoupling the demand for services from electricity use. Environ Res Lett 2018;13:124030.

[30] Shehabi A, Smith SJ, Sartor DA, Brown RE, Herrlin M, Koomey JG, et al. United States Data Center Energy Usage Report | Energy Technologies Area. Berkeley Lab; 2016:65.

[31] Barroso LA, Clidaras J, Hölzle U. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second edition. vol. 8; 2013. https://doi.org/10.2200/S00516ED2V01Y201306CAC024.

[32] International Energy Agency -IEA. Global Energy and CO2 Status Report 2018. Iea 2019:29.

[33] Eid C, Koliou E, Valles M, Reneses J, Hakvoort R. Time-based pricing and electricity demand response: existing barriers and next steps. Util Policy 2016;40:15–25.

[34] Cisco Systems. Forecast and Methodology, 2013–2018; 2014.

[35] Cisco Systems. Forecast and Methodology, 2014-2019; 2015.

[36] Cisco Systems. Forecast and methodology 2010-2015; 2011.

[37] Cisco Systems. Forecast and Methodology, 2011-2016; 2012.

[38] Andrae A. New perspectives on internet electricity use in 2030. Eng Appl Sci Lett 2020;3:19–31.

[39] Macaulay J, Buckalew L, Chung G. A collaborative report by DHL and Cisco on implications and use cases for the logistics industry. In: Chung G, Buckalew L, editors. Internet Things Logist., Troisdorf, Germany; 2015, p. 1–27.

[40] Ai Y, Peng M, Zhang K. Edge computing technologies for Internet of Things: a primer. Digit Commun Networks 2018;4:77–86.

[41] Cisco Systems. Cisco annual internet report (2018–2023); 2020.

[42] Ballard M. Why Amsterdam halted data center construction. Data Cent Knowl; 2019.

[43] Pieters J. netherlands-power-grid-cant-handle-tech-developments-data-centers-warn. NLTimes; 2018.

[44] Vakilifard N, Bahri PA, Anda M, Ho G. An interactive planning model for sustainable urban water and energy supply. Appl Energy 2019;235:332–45.

[45] Boccardi F, Heath RW, Lozano A, Marzetta TL, Popovski P. Five disruptive technology directions for 5G. IEEE Commun Mag 2014;52:74–80. https://doi.org/10.1109/MCOM.2014.6736746.

[46] Wang LC, Rangapillai S. A survey on green 5G cellular networks. 2012 Int. Conf. Signal Process. Commun. SPCOM 2012; 2012. https://doi.org/10.1109/SPCOM.2012.6290252.

[47] Reinsel D, Gantz J, Rydning J. The Digitization of the World - From Edge to Core; 2018.

[48] Touch J, Badawy A-H, Sorger VJ. Optical computing. Nanophotonics 2017;6:503–5.

[49] Matsuoka M, Matsuda K, Kubo H. Liquid immersion cooling technology with natural convection in data center. In: 2017 IEEE 6th Int. Conf. Cloud Netw., IEEE; 2017. p. 1–7.

[50] Mohseni M, Read P, Neven H, Boixo S, Denchev V, Babbush R, et al. Commercialize quantum technologies in five years. Nature 2017;543:171–4.

[51] Mocnej J, Miškuf M, Papcun P, Zolotová I. Impact of edge computing paradigm on energy consumption in IoT. IFAC-PapersOnLine 2018. https://doi.org/10.1016/j.ifacol.2018.07.147.

[52] Boesten S, Ivens W, Dekker SC, Eijdems H. 5th generation district heating and cooling systems as a solution for renewable urban thermal energy supply. Adv Geosci 2019;49:129–36.

[53] Averfalk H, Werner S. Economic benefits of fourth generation district heating. Energy 2020;193:116727.

[54] Cleveland CJ. An exploration of alternative measures of natural resource scarcity: the case of petroleum resources in the US. Ecol Econ 1993;7:123–57.

[55] da Silveira TA, dos Santos ECA, Colling AV, Moraes CAM, Brehm FA. E-waste Management and the Conservation of Geochemical Scarce Resources. E-waste Recycl. Manag., Springer; 2020. p. 179–200.

[56] Truby J. Decarbonizing Bitcoin: law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies. Energy Res Soc Sci 2018;44:399–410. https://doi.org/10.1016/j.erss.2018.06.009.

[57] Sedlmeir J, Buhl HU, Fridgen G, Keller R. The energy consumption of blockchain technology: beyond myth. Bus Inf Syst Eng 2020:1–10.

[58] Monrat AA, Schelen O, Andersson K. A survey of blockchain from the perspectives of applications, challenges, and opportunities. IEEE Access 2019;7:117134–51. https://doi.org/10.1109/ACCESS.2019.2936094.

[59] Albasir A, Naik K, Plourde B, Goel N. Experimental study of energy and bandwidth costs of web advertisements on smartphones. In: Proc. 6th Int. Conf. Mob. Comput. Appl. Serv., ICST; 2014, p. 90–7. https://doi.org/10.4108/icst.mobicase.2014.257770.

[60] Shiller B, Waldfogel J, Ryan J. The effect of ad blocking on website traffic and quality. RAND J Econ 2018;49:43–63. https://doi.org/10.1111/1756-2171.12218.

[61] Kaptein M, Markopoulos P, de Ruyter B, Aarts E. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. Int J Hum Comput Stud 2015;77:38–51.

[62] Margarethe Vestager. Speech shaping a digital future for Europe. Https://EcEuropaEu/Commission/Commissioners/2019-2024/Vestager/Announcements/Shaping-Digital-Future-Europe_en 2020. https://ec.europa.eu/commission/commissioners/2019-2024/vestager/announcements/shaping-digital-future-europe_en [accessed February 7, 2020].