THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le *31/01/2023* **par :**

Igor FONTANA DE NARDIN

On-line scheduling for IT tasks and power source commitment in datacenters only operated with renewable energy

JURY

PREMIER MEMBRE	Professeur d'Université	Rapporteur
SECOND MEMBRE	Professeur d'Université	Rapporteur
TROISIÈME MEMBRE	Professeur d'Université	Examinateur
QUATRIÈME MEMBRE	Professeur d'Université	Examinateur
CINQUIÈME MEMBRE	Professeur d'Université	Examinateur

École doctorale et spécialité :

MITT : Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Unité de Recherche :

Laplace (UMR 5213) et IRIT (UMR 5505)

Directeur(s) de Thèse :

Patricia STOLF et *Stéphane CAUX*

Rapporteurs :

Premier RAPPORTEUR et *Second RAPPORTEUR*

Acknowledgments

Acknowledgments

Abstract

Abstract

Résumé

Résumé

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Global warming is one of the biggest challenges humanity is facing. A recent rapport shows that we are walking toward a global mean temperature increase by 2100 of 2.7°C, well above the 1.5°C defined by the Paris Agreement [1]. The same rapport predicts the rise in mean global temperature will be around 1.8°C even after implementing all announced Paris Agreement goals. Achieving 1.5°C demands an engagement of all sectors to reduce greenhouse gas (GHG) emissions. GHG is generated during the combustion process of fossil fuel, one of the world's main sources of energy production [8].

One significant GHG emitter is the Information and Communications Technology (ICT) sector. It produces around 1.8-2.8% of the world's total GHG [2]. Inside ICT, Data centers and transmission networks are responsible for nearly 1% of global energy-related GHG emissions [9]. The data center sector is one of the most electricity-expensive ICT actors due to its uninterrupted operation. A report revealed that Google data centers consumed the same amount of energy as the entire city of San Francisco in 2015 [10]. In addition, the situation tends to get even worse due to the improvements reduction in processor technologies and the predicted expansion of internet usage [2, 11].

Big cloud providers such as Google and Amazon are trying to reduce energy consumption and increase the power coming from renewable sources (RES) [12]. RES is the most encouraging method to eliminate fossil fuel use [8]. Renewable sources generate energy from clean sources such as biomass, hydropower, geothermal, solar, wind, and marine energies [7, 13, 14, 15, 16]. A significant drawback of RES is the weather conditions dependency, creating power intermittence. These providers smooth this intermittence by not migrating entirely to RES, maintaining a connection to the grid [7]. Therefore, they are not 100% clean. A renewable-only data center must consider this intermittence in its decision-making. Another source of uncertainty comes from the user's demand. Users can send their requests at any time. Providing high availability is a challenge for a renewable-only data center.

A way to reduce the impact of RES power production intermittence is by adding storage elements [7]. Batteries and hydrogen tanks can shift generation and/or consumption over time. A renewable-only data center demands a massive storage capacity [7]. For example, Google plans to use energy from 350 MW solar panels connected to a storage system with 280 MW [17]. While helping to deal with RES intermittence, storage management introduces another level of decision. For example, it can store energy during the day using at night. Nevertheless, the demand during the day could be higher than at night, so maybe it is better to use the energy during the day. This is another big challenge for migrating to a 100% clean data center.

Some works propose ways to deal with both demand and weather uncertainties using predictions [18, 19, 20, 21]. Forecasting the upcoming requests and the weather helps to plan storage usage. They use these predictions to maximize renewable usage but with the grid as backup. All these works are valuable and important to optimize renewable usage. However, the forecast can vary from the actual values. Other works focus on reacting to real events [22, 23, 24, 25]. They try to minimize the data center operational cost, maximize renewable usage, increase the revenue of job execution, or improve the Quality of Service. Usually, they define ways to schedule the jobs, optimizing their objective. However, they focus on short-term decisions without long-term management. Since these works also have the grid as backup, storage management is not a concern. Some works mix predictions with reactive actions. For example, Goiri et al. [21] propose a scheduling algorithm that predicts solar power production and uses it to define the best moment to start new jobs, using brown energy (from the grid) when necessary. Also, Venkataswamy et al. [26] created a job scheduler that defines job placement according to the available machines. The available machines are given by a fixed plan (which can use power from renewable, batteries, or grid), with no modifications.

Few research initiatives are investigating how to design and operate a renewable-only data center. One of them is the ANR Datazero2 project [27]. This project aims to define a feasible architecture to maintain a renewable-only data center. This architecture includes several elements to provide energy to the IT servers, such as Wind turbines, Solar panels, Batteries, and Hydrogen tanks. Considering the decision-making, Datazero2 divides the problem into two parts: offline and online. The offline module predicts power demand and production. Using these predictions and considering long-term constraints, this module creates a power and IT plan for the near future.

The online module schedules the users' jobs, using the offline plan as a guide. Online is the only one that knows exactly the jobs submitted to the data center. So, it needs to place them in the available servers. Online could just apply the offline plan without modifications. However, this behavior would impact the Quality of Service (QoS). Online can improve the QoS, increasing storage usage to turn on more servers (to run more jobs) or speed up the running servers (to finish jobs earlier). Also, online must be renewable production aware. For example, online can identify a lower production that can dry the storage faster, so it must reduce its usage. Finding a good trade-off between QoS and storage management is even harder in online mode, which demands fast decisions. In this thesis, we focus on these online decisions. The goal is to design and prove the efficiency of a novel approach for scheduling users' jobs, finding a good trade-off between QoS and storage management.

1.2 Problem Statement

A data center powered by renewable energy demands several levels of decision. Several works aim to optimize some of these decisions. We can cite demand and production predictions, cost optimization, sizing, shifting demand, battery management, admission control, and job scheduling, to mention a few. Usually, these works introduce a link to the grid, using it as a backup to cope with peak demand. Removing the grid of the context adds several challenges. This context increases the need for predictions to manage weather and workload uncertainties. Another key element in renewable-only data centers is storage. Aligning prediction and storage elements allows it to define the best strategy to handle users' requests. However, actual demand and production can vary from the predictions.

So, the online module must react to the actual values. This reaction can improve the QoS (e.g., when there is more production than expected) or reduce the impact of critical events.

Figure 1.1 illustrates all the elements in the decision process. We consider only renewable sources and storage elements without grid connection. An offline optimization gives an offline plan using production and demand prediction. The offline plan has a limited size named time window (e.g., three days). So, offline suggests actions to online during this time window. Online receives the actual renewable production from wind turbines and solar panels. Online adapts storage usage according to the actual production. Since hydrogen has a longer start-up time, it is difficult to manage it in online mode. Therefore, we let hydrogen usage from the offline optimization, using it to provide energy during periods with low renewable production (e.g., during the winter). So, online decides about battery usage only.

Battery management introduces two new challenges regarding the Battery's State of Charge (SoC). SoC means the level of charge of a battery relative to its capacity. A good practice to extend the battery's lifetime is to avoid drying or overcharging it [28]. So, maintaining the SoC between reasonable levels is the first challenge. Online has the entire time window to make modifications in battery usage. However, it must finish the time window close to the expected SoC (given by the offline plan). This is the second challenge. Since the data center runs continuously, it is not viable to always use more battery than expected for every time window.

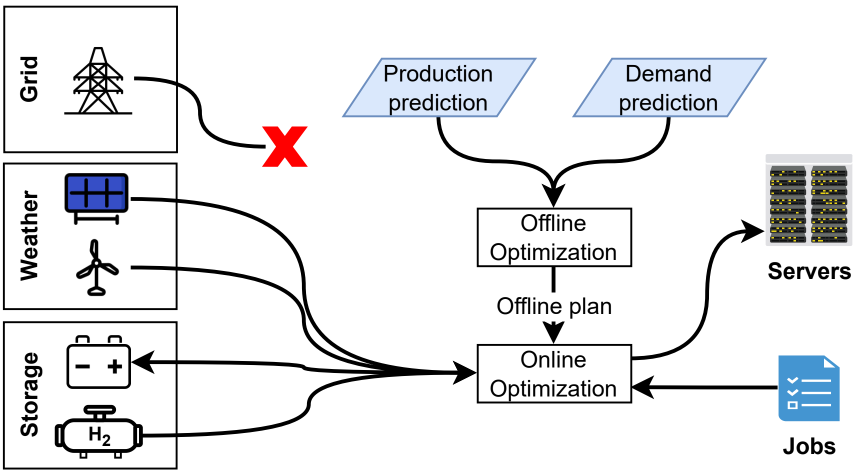

Figure 1.1: Problem overview. Online receives an offline plan, the actual renewable production, and the users' jobs. It must define storage usage, job placement in the servers, and server speed.

On the IT side, online receives the jobs from the users and must schedule them on the available servers. Online receives an offline plan for server configuration (machine on/off and speed). However, it can modify the server configuration to react to incoming events (e.g., more production, demand peak). Changing the speed of a server is possible due to the Dynamic voltage and frequency scaling (DVFS) technique. DVFS allows servers' speed reduction, spending less energy. However, putting a job on a server with a decreased speed can impact the job QoS. To sum up, online must manage the battery (maintaining the SoC between thresholds and finishing the time window with the battery level close to the target), schedule the jobs, and balance the servers' speed.

This thesis' first objective is to make online modifications in the power decisions given by the offline plan, coping with the uncertainty coming from renewable production and workload demand. The second goal is mixing power and scheduling online decisions, turning the scheduling storage aware. This mix allows the scheduling to make better decisions than usual algorithms. The last objective is to add the predictions to the online decision. These goals help to find a better trade-off between QoS and storage management. Different contributions address these questions in this manuscript.

1.3 Main contributions

Proposing a simulation environment

A crucial step to simulate data center management is defining the workload, weather, server configuration, and simulation tool. We detail in Section 3 the simulation environment, providing a framework for future works. Regarding the workload, some traces are used in literature, such as Google [29], Parallel Workloads Archive [30], and Alibaba [31]. We propose a trace from Parallel Workloads Archive named Metacentrum [32]. We detailed the filtering process of this trace. Considering the weather, it is possible to collect data from everywhere in the world. We present the methodology to generate power production from a NASA trace, using the framework Renewables.ninja[1] [33]. The third input is the server configuration. We demonstrate the data collected from a server in GRID5000[2] used in this thesis. Finally, we present the simulation tool named BATSIM[3], based on SIMGRID[4]. We introduced in this simulation tool the modifications needed to manage battery and power production. The ensemble of these data and definitions allows future work inside and outside the Datazero2 project.

Defining offline power and IT decisions

As illustrated in Figure 1.1, an important part is the offline plan. This plan must consider the power and demand predictions to define the actions for the next time window. We demonstrate in Section 3 a model to use both predictions. We separate the problem into two parts. First, we present the optimization problem to define power engagement, giving a power prediction. This optimization problem results in expected renewable power production, storage usage, and expected SoC. The sum of the expected renewable power production and storage usage is named the power envelope. The second part is the IT servers' state (on/off) and speed definition. This optimization problem defines the state and speed according to the power envelope. The objective of this optimization problem is to maximize the servers' speed. The results of both optimizations are the input for the online module.

Reacting to power fluctuations

Given the result of the optimization problem, next, we propose a heuristic to react to the power fluctuations. Since there is no perfect prediction, one source of divergence is the difference between the prediction and actual values. This divergence occurs in both power demand and production. Also, the offline model considers that the servers will maintain constant power usage. However, the server consumption can vary according to

[1]https://www.renewables.ninja/
[2]https://www.grid5000.fr
[3]https://batsim.org/
[4]https://simgrid.frama.io/

the scheduling and/or job. Yet, the scheduling can modify the battery usage to improve the QoS (e.g., avoiding killing jobs). Considering all these sources of power fluctuations, the heuristic must adapt the usage, aiming to approximate the state of charge of the target level at the end of the time window. Since this is an online problem, we can not re-run the offline optimization solution with the actual values. Therefore, we propose four policies to compensate for these divergences in the power envelope. Each one finds a different moment in the future to place the compensation.

Learning the actions to deal with power fluctuations

The four compensation policies apply the same behavior throughout the entire execution. However, different moments inside the time window can demand distinct policies. So, our next goal is to learn when to use each policy. So, we introduce two Reinforcement Learning (RL) algorithms to discover the best mix of policies. Considering each policy as RL's action, we present the RL's state and reward. The premise of applying RL is that optimizing the decisions locally generates a global optimal. In other words, if the algorithm chooses the best action each time, in the end, we will have the best results. We implemented two well-known RL algorithms named Contextual Multi-Armed Bandit and Q-Learning. We present the learning results and a comparison between the RL algorithms and random choices.

Defining storage-aware scheduling using production and demand predictions

Finally, the last contribution is a storage-aware scheduling heuristic. This algorithm is based on the well-known EASY-Backfilling. The algorithm is named BEASY. BEASY uses the predictions given by the offline to predict dangerous moments, where it must be careful in the scheduling. Also, we introduce another level of validation, verifying if the servers allocated to the job would be available during the entire execution. Regarding power compensations, it creates several possible scenarios of production and demand using the forecasts. According to these scenarios, the heuristic finds the best moment to make the compensations. For example, BEASY tries to reduce the usage before the moments when the predictions indicate that the battery could be lower than a critical value. This heuristic mixes all decisions providing a well-balanced answer to the online multi-objective problem.

1.4 Publications and Communication

Submitted Peer Reviewed Conferences:

- I. F. de Nardin, P. Stolf and S. Caux, "Adding Battery Awareness in EASY Backfilling", 2023 IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Porto Alegre, Brazil, 2023.

Accepted Peer Reviewed Conferences:

- I. F. de Nardin, P. Stolf and S. Caux, "Analyzing Power Decisions in Data Center Powered by Renewable Sources", 2022 IEEE 34th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Bordeaux, France, 2022, pp. 305-314;
- I. F. de Nardin, P. Stolf and S. Caux, "Evaluation of Heuristics to Manage a Data Center Under Power Constraints", 2022 IEEE 13th International Green and Sustainable Computing Conference (IGSC), Pittsburgh, PA, USA, 2022, pp. 1-8;

- I. F. de Nardin, P. Stolf and S. Caux, "Mixing Offline and Online Electrical Decisions in Data Centers Powered by Renewable Sources", IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, Brussels, Belgium, 2022, pp. 1-6;

- I. F. de Nardin, P. Stolf and S. Caux, "Smart Heuristics for Power Constraints in Data Centers Powered by Renewable Sources", Conférence francophone d'informatique en Parallélisme, Architecture et Système (COMPAS 2022), Jul 2022, Amiens, France. paper 7.

Others Disseminations:

- Talk: Analyzing Power Decisions in Data Center Powered by Renewable Sources, GreenDays@Lyon, March 2023.

1.5 Dissertation Outline

The remaining dissertation has the following organization:

Chapter 2 - Context and Related Work: This chapter presents the fundamentals to understand this dissertation. Considering the scope of the topic, the context consists of four parts. First, we introduce the context of global and ICT GHG emissions. Then, we describe renewable energy as an alternative to replace brown energy. After, we explain the usage of renewable to power a data center. Then, we define the uncertainties of weather and workload in a renewable-only data center. This last part also clarifies the importance of using predictions but with an online adaptation. After presenting the context, we introduce a list of works that solve part of our problem, highlighting the existing gaps in the state-of-the-art;

Chapter 3 - Modelling, Data, and Simulation: In this chapter, we describe the model to deal with the several elements that compose a renewable-only data center. Datazero2 creates a division between Offline and Online decisions. We present the model to deal with offline decisions using predicted power demand and production. Then, we demonstrate the output of Offline used by the Online. Finally, we define the Online model, which englobes the job scheduling and modifications in the Offline plan. After describing the model, we explain the source of the different data (e.g., workload, weather, servers) applied in the simulations. We present an explanation of the work done in the traces of the literature. Finally, we present the simulation tools used in this work;

Chapter 4 - Introducing Power Compensations: This chapter describes the proposed algorithm to react to power uncertainties. We created four heuristics to find the best place to compensate for battery changes, which aim to reduce the number of killed jobs and the distance between the battery level and the target level. The results presented are related to the publications [34] and [35];

Chapter 5 - Learning Power Compensations: This chapter presents the idea and the results of the introduction of Reinforcement Learning (RL) in the power compensation problem. We propose two RL algorithms (Q-Learning and Contextual Multi-Armed Bandit) to learn the best moment to compensate;

Chapter 6 - Adding Battery Awareness in EASY Backfilling: This chapter explains a heuristic to mix scheduling and power compensation decisions. This heuristic is based on the EASY Backfilling scheduling algorithm but considers the battery's State of Charge to make better decisions;

Chapter 7 - Conclusion and Perspectives: Finally, in this chapter, we summarize the contributions of this work, providing a discussion about future works.

Chapter 2

Context and Related Work

Contents

2.1 Global Warming and ICT Role

Global warming is one of the most critical environmental issues of our day [36]. Global warming is the effect of human activities on the climate, mainly the burning of fossil fuels (coal, oil, and gas) and large-scale deforestation [36]. Both activities have grown immensely since the industrial revolution. The burning of fossil fuels process results in greenhouse gas emissions [8]. Today, fossil fuels are one of the world's main sources of energy production, helping to emit more and more GHG [8]. GHG stays in the atmosphere creating a layer as a blanket over the planet's surface. Without this blanket, the Earth can balance the radiation energy from the sun and the thermal radiation from the Earth to space [36]. However, this human-generated blanket imposes a barrier to the thermal radiation from the Earth, letting it into the atmosphere and heating the planet, working as a greenhouse. All this process works as a greenhouse which is the reason for the name greenhouse gas [36].

This situation brings us to United Nations Climate Change Conference (COP21) in Paris, France, on 12 December 2015. At this conference, 196 signed the Paris Agreement aiming to [37]:

1. Reduce global greenhouse gas emissions substantially, limiting the global temperature increase in this century to 2°C while pursuing measures to limit the growth even further to 1.5°C;

2. Review countries' commitments every five years (through the Nationally Determined Contribution, or NDC);

3. Provide financing to developing countries to mitigate climate change, strengthen resilience, and enhance their abilities to adapt to climate impacts.

These are ambitious but necessary objectives. Since then, countries and organizations have proposed several actions and pledges. However, a recent report indicates that the actual world's effort is not enough [1]. Figure 2.1 shows GHG emission and temperature estimations. We could see that there is a small reduction in emissions increase tendency. Nevertheless, this figure estimates that real-world actions based on current policies will lead to an increase of somewhere between 2.6 and 2.9°C by 2100. This estimation is well above the 1.5°C pursued by the Paris Agreement. Considering the targets proposed by the countries through NDC, the temperature will be around 2.4°C. In a scenario based on NDC targets and submitted and binding long-term targets, the prediction is a temperature of 2°C by 2100, the limit proposed by the Paris Agreement. The report forecasts an optimistic scenario analyzing the effect of net zero emissions targets of about 140 countries that are adopted or under discussion. Even in this optimistic scenario, the estimated temperature would be 1.8°C. The situation tends to be even worst with the gold rush for gas [38]. The report indicates that in 2022 we arrived at 1.2°C warming [1].

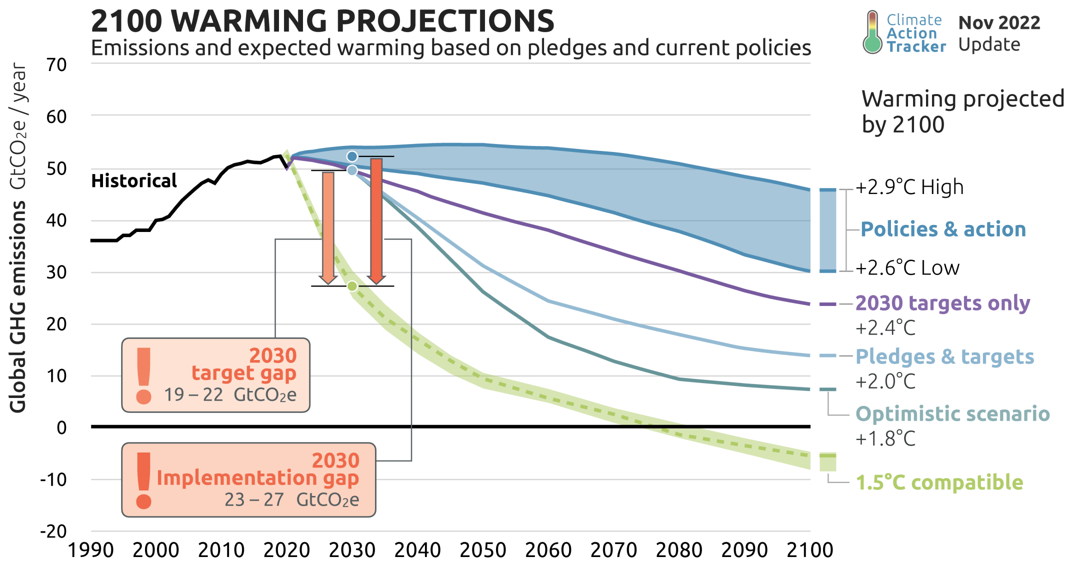

Figure 2.1: Estimated global GHG emissions [1].

We have started to feel the impacts of global warming on humanity, such as heatwaves, droughts, and floods, impacting flora and fauna directly [39, 40]. In a cascade effect, this increases food and water insecurity worldwide [40, 41]. Also, high temperatures increase mortality, impact labor productivity, impair learning, increase adverse pregnancy outcomes possibility, increase conflict, hate speech, migration, and infectious disease spread [42]. Therefore, an increase of the temperature by 2.7°C as forecasted would impact one-third (22–39%) of the world's population by 2100 [42]. Climate change has already impacted around 9% of people (>600 million) [42]. Reducing global warming from 2.7 to 1.5°C results in a ∼5-fold decrease in the population exposed to unprecedented heat (mean annual temperature ≥29°C) [42]. Thus, all sectors must reduce their GHG emissions as much as possible.

Information and Communication Technology is one of these sectors which has accelerated growth in the last 70 years. Unesco defines ICT as [43]:

> "Information and communication technologies (ICT) is defined as a diverse set of technological tools and resources used to transmit, store, create, share or

exchange information. These technological tools and resources include computers, the Internet (websites, blogs, and emails), live broadcasting technologies (radio, television, and webcasting), recorded broadcasting technologies (podcasting, audio and, video players, and storage devices), and telephony (fixed or mobile, satellite, visio/video-conferencing, etc.)."

Regarding the ICT role in GHG emissions, the global share is around 1.8%-2.8%, or 2.1%-3.9% considering the supply chain pathways in 2020 [2]. The situation tends to get even worst, driven by the boom in Internet-connected devices. A Cisco report indicates that the Internet had 3.9 billion users in 2018 [11]. The same report predicts an increase to 5.3 billion in 2023 (66 percent of the global population). Also, they predicted 3.6 networked devices per capita in 2023, up from 2.4 networked devices per capita in 2018. However, International Telecommunication Union (ITU), a United Nations specialized agency for ICTs, indicates that we arrived at 5.3 billion connected users in 2022 due to the COVID-19 pandemic [44]. But will the growth in internet users increase GHG emissions? Andrae and Edler [4] and Belkhir and Elmeligi [3] agree that this growth could lead to an increase in GHG emissions. Figure 2.2 shows the predictions of both works.

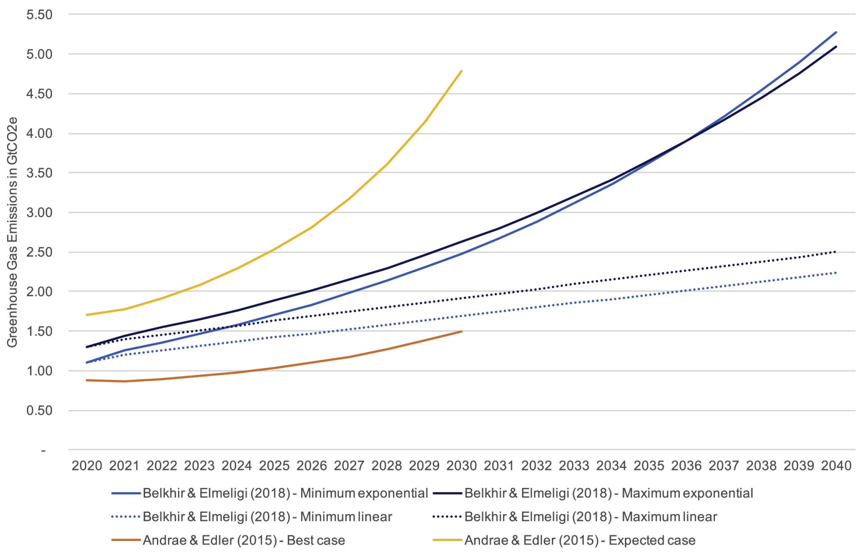

Figure 2.2: Projections of ICT's GHG emissions from 2020 [2].

This figure illustrates the contraction in the Paris Agreement demand and the predictions about usage in the ICT sector. In all forecasts of Figure 2.2, the tendency is emissions growth. However, ICT needs to reduce its emissions drastically. Figure 2.3 illustrates the carbon emission share if the ICT stays at the same level as 2020 and the other sectors decrease their emissions. Without changes, ICT would have 35.1% of global emissions in 2050. So, ICT must move towards reducing its emissions. Figure 2.4 presents the estimations of ICT's GHG emissions for 2015 and 2020 from different authors. This figure breaks down these emissions into different components. One of them, with a good share in some cases, is Data centers. IBM defines the data center as "A data center is a physical room, building or facility that houses IT infrastructure for building, running, and delivering applications and services, and for storing and managing the data associated with those applications and services" [45]. The International Energy Agency (IEA) defines data center as [9]:

"Data centers are facilities used to house networked computer servers that store, process and distribute large amounts of data. They use energy to power both the IT hardware (e.g., servers, drives, and network devices) and the supporting infrastructure (e.g., cooling equipment)."

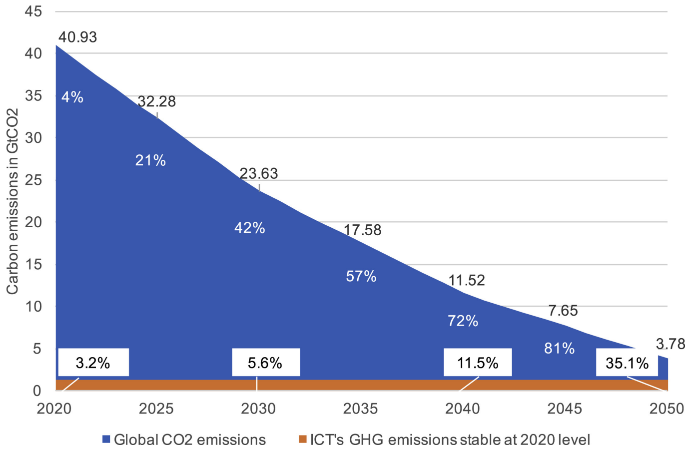

Figure 2.3: ICT's emissions, assuming the 2020 level remains stable until 2050, and global CO2 emissions reduced in line with 1.5°C [2].

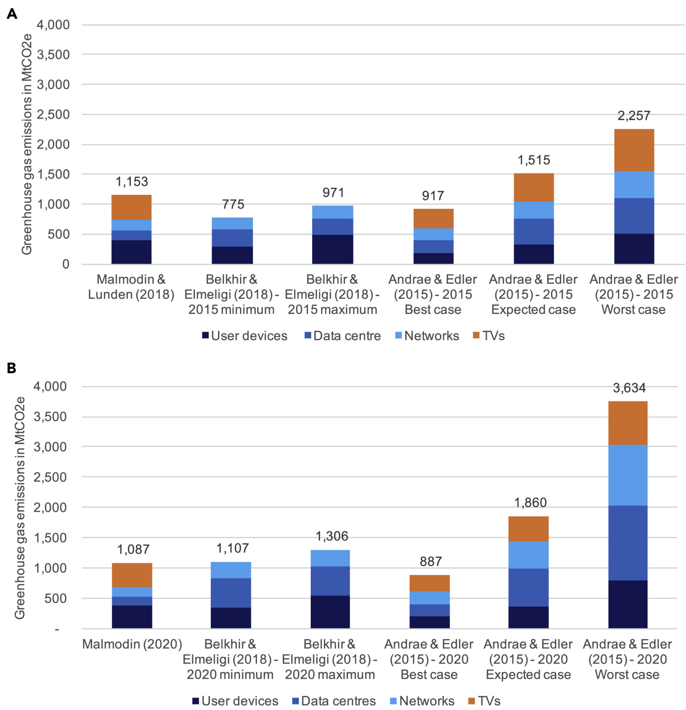

Figure 2.4: Estimations for global ICT's GHG emissions in 2015 and 2020 [2]. The authors consolidated the works from [3, 4, 5, 6].

Data centers are very energy consumers. IEA published an article indicating that data centers and networks were responsible for almost 1% of energy-related GHG emissions in 2020 [9]. Also, Google data centers consumed the same amount of energy as the entire city of San Francisco in 2015 [10]. Global data center electricity use in 2021 was 220-320 TWh, corresponding to 0.9-1.3% of the global demand [9]. For example, the domestic electricity consumption of Italy was 300 TWh in 2021 [46]. In Ireland, electricity consumed by data centers went from 5% of the total electricity consumption in 2015 to 14% in 2021 [47]. Denmark predicts to triple data center consumption, corresponding to 7% of the country's electricity use [48].

Despite the strong growth in demand, data center energy usage has only moderately grown [9]. A reason that explains it is the improvements in IT hardware energy consumption [9]. These improvements allowed a boost in microchips' speed with a reduction in their power consumption, letting big data center companies cope with the peak in demand. Gordon Moore predicted in 1965 (Moore's law) that [49]:

> "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years."

Even if he predicted it just until 1975, it is the case nowadays. However, the future is uncertain, and the community is divided to confirm continuous efficiency improvements [2]. While Andrae and Edler [4] and Belkhir and Elmeligi [3] expected an ending in power-consuming improvements (indicated in Figure 2.2), Malmodin and Lundén [5] are more optimistic. They suggest that ICT's carbon footprint in 2020 could halve by 2030. To achieve that, he considers two key factors. First, the improvements will continue. Second, the migration to renewable sources.

2.2 Renewable Energy Sources

The ICT migration to renewable energy sources (RES) is one of the factors that helped reduce the growth in GHG emissions despite the rapidly growing demand for digital services [9]. RES is one of the principal solutions to decarbonize electrical production [7, 8]. RES is also named green energy, in contrast to brown energy from fossil fuels. Basically, RES generates energy from natural sources, such as solar, wind, geothermal, hydropower, wave and tidal, and biomass [7, 13, 14, 15, 16]. These natural sources have a low impact on GHG emissions. For example, manufacturing is the stage with higher emissions for wind and solar [50]. So, these components could produce energy with no or low GHG emissions. The renewable term comes from the idea that these sources are constantly replenished. On the other hand, fossil fuels are non-renewable because they need hundreds of millions of years to develop. In the Net Zero Emissions by 2050 Scenario, RES is responsible for one-third of the reductions between 2020 and 2030 [51]. Some countries focus on nuclear power plants to produce energy [52]. Even if nuclear power is a low carbon emissions energy source, it introduces the risk of accidents and environmental impacts of radioactive wastes [52].

The biggest challenge of implementing RES is its intermittence [7]. Since RES production comes from nature, it depends on the climate conditions. For example, there is no power production from solar during the night. There are two approaches to reducing brown usage: on-site and off-site generation [53]. On-site generation uses local renewable resources, and off-site takes resources available on the grid. In an off-site generation, it is not possible to guarantee that the incoming energy is from RES since the grid mixes all types of power generation [7]. Giant cloud providers (e.g., Google, Amazon, and Facebook) invest in solar and wind power plants in an off-site approach [12, 17, 54]. So, they could say that, on average, they provide RES to the grid with the same amount that they expend. However, they transfer the RES uncertainty problem to third parties [7]. For example, in a case with a peak in demand, they will use the power from the grid, renewable or not. So, they are still non-renewable-dependent.

2.3 Renewable-only Data center

Since data centers have a controlled infrastructure, they are a good target to migrate to a renewable-only environment [7]. However, creating a non-renewable independent data center imposes several challenges. In this kind of data center, all the generation is on-site without backup from the grid. Nevertheless, the production and demand can not match. Figure 2.5 exemplifies the mismatch between the power demanded by a data center and power generation. This mismatch requires a production (electrical) or a load (IT) shift. We will present both electrical and IT elements needed for a renewable-only data center.

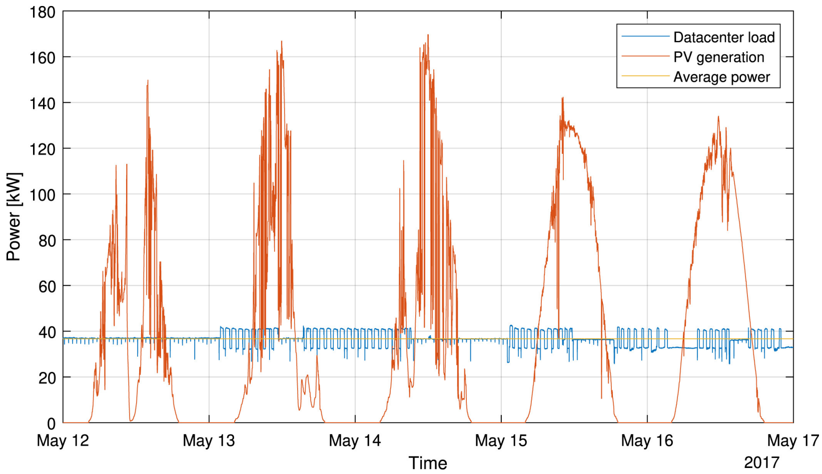

Figure 2.5: Comparison of small data center load and the generation from a theoretical photovoltaic in Belfort, France. Both load and production have the same average value [7].

2.3.1 Electrical elements

As mentioned before, different renewable sources can generate power. We focus on wind and solar since they were the most prominent in the past few years [51]. For wind turbines, the wind speed is crucial. Equation 2.1 gives the power output $P_{WT}(t)$ of a wind turbine, given the wind speed v [55, 56, 57].

$$P_{WT}(t) = \begin{cases} 0 & v \leq v_{in} \text{ or } v(t) > v_{out} \\ P_{WT,rated} \times \frac{v(t)-v_{in}}{v_{rated}-v_{in}} & v_{in} < v(t) \leq v_{rated} \\ P_{WT,rated} & v_{rated} < v(t) \leq v_{out} \end{cases} \tag{2.1}$$

Where:

- $P_{WT}(t)$: Power generated by a wind turbine (kW);
- v: Wind speed (m/s);
- v_{in}: Cut-in wind speed (m/s);

- v_{out}: Cut-out wind speed (m/s);
- v_{rated}: Speed related to wind turbine nominal power (m/s);
- $P_{WT,rated}$: Wind turbine nominal power (kW).

If the wind speed v is lesser or equal to the cut-in v_{in} or greater than the cut-out v_{out}, it does not produce power. It tests the cut-out v_{out} to protect the generator. If the speed v is greater than the cut-in v_{in} and lesser or equal to the rated speed v_{rated}, it generates proportionally to the rated power $P_{WT,rated}$ and rated speed v_{rated}. Finally, if the speed v is greater than the rated speed v_{rated} and lesser or equal to the cut-out v_{out}, it produces constant power $P_{WT,rated}$.

Regarding solar production, the photovoltaic (PV) system uses solar panels to generate power from solar irradiation. Equation 2.2 demonstrates how to calculate the output power of a solar panel $P_{pv}(t)$ [56, 57, 58].

$$P_{pv}(t) = P_{R,PV} \times (R/R_{ref}) \times \eta_{PV} \tag{2.2}$$

Where:

- $P_{pv}(t)$: Power generated by each PV panel (W);
- $P_{R,PV}$: PV panel Nominal power (kW);
- R: Solar radiation (W/m^2);
- R_{ref}: solar radiation at reference conditions. Usually set as 1000 (W/m^2) [56];
- η_{PV}: PV efficiency.

Regarding PV efficiency η_{PV}, it can consider the temperature of the solar panel [57, 58]. However, some works simplify it by applying a constant value [19, 56].

2.3.2 IT elements

- Write about the servers;
- Write about the power consumption (e.g., DVFS, on-off, idle, etc);
- Write about the jobs (e.g., types, resources demanded, etc);

2.4 Sources of Uncertainty

2.4.1 Weather Uncertainties

- Describe wind uncertainty;
- Describe solar irradiation uncertainty;

2.4.2 Workload Uncertainties

- Describe job arrival uncertainty;
- Describe job size uncertainty;

2.5 Optimization Strategies for Dealing with Uncertainties

- Write about weather predictions
- Write about optimization for the weather;
- Write about scheduling algorithms to deal with workload uncertainties;
- Write about mixing both renewable production and workload uncertainties;

2.6 Literature Review

- Present the 20 articles selected;
- Present a table with each article and the following points:
  - Name;
  - Year;
  - Source of power (solar, wind, battery, grid, etc);
  - Level of decision (offline, online, both);
  - Power adaptations (battery compensations, renewable adaptations, etc).

2.6.1 Discussion and Classification of the Literature

Chapter 3

Modelling, Data, and Simulation

Contents

3.1 Model

3.1.1 Offline Decision Modules

Power Decision Module

IT Decision Module

3.1.2 Offline Plan

3.1.3 Online Decision Modules

Job scheduling

Modifying Power Plan

Modifying IT Plan

3.2 Data

3.2.1 Workload Trace

3.2.2 Weather Trace

3.2.3 Platform Configuration

3.3 Simulation

3.3.1 Simulator

3.3.2 Metrics

3.3.3 Datazero2 Middleware

3.4 Conclusion

Chapter 4

Introducing Power Compensations

Contents

4.1 Introduction

4.2 Model

4.3 Heuristics

4.4 Results Evaluation

4.5 Conclusion

Chapter 5

Learning Power Compensations

Contents

5.1 Introduction

5.2 Algorithms

5.2.1 Random

5.2.2 Q-Learning approach

5.2.3 Contextual Multi-Armed Bandit approach

5.3 States

5.4 Actions

5.5 Rewards

5.6 Results Evaluation

5.7 Conclusion

Chapter 6

Adding Battery Awareness in EASY Backfilling

Contents

6.1 Introduction

6.2 Model

6.3 Heuristic

6.3.1 Predictions

6.3.2 Job Scheduling

6.3.3 Power compensation

6.4 Conclusion

Chapter 7

Conclusion and Perspectives

7.1 Conclusion

7.2 Perspectives

Bibliography

[1] Climate Action Tracker, "2100 warming projections: Emissions and expected warming based on pledges and current policies," *Warming Projections Global Update, November*, 2022, Available at: https://climateactiontracker.org/global/temperatures/.

[2] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. Blair, and A. Friday, "The climate impact of ict: A review of estimates, trends and regulations," 2021.

[3] L. Belkhir and A. Elmeligi, "Assessing ict global emissions footprint: Trends to 2040 & recommendations," *Journal of cleaner production*, vol. 177, pp. 448–463, 2018.

[4] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.

[5] J. Malmodin and D. Lundén, "The energy and carbon footprint of the global ict and e&m sectors 2010–2015," *Sustainability*, vol. 10, no. 9, p. 3027, 2018.

[6] J. Malmodin. (2020) The ict sector's carbon footprint. Presentation at the techUK Conference in London Tech Week on 'decarbonising Data'. [Online]. Available: https://spark.adobe.com/page/dey6WTCZ5JKPu/

[7] G. Rostirolla, L. Grange, T. Minh-Thuyen, P. Stolf, J.-M. Pierson, G. Da Costa, G. Baudic, M. Haddad, A. Kassab, J.-M. Nicod *et al.*, "A survey of challenges and solutions for the integration of renewable energy in datacenters," *Renewable and Sustainable Energy Reviews*, vol. 155, p. 111787, 2022.

[8] A. Olabi and M. A. Abdelkareem, "Renewable energy and climate change," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112111, 2022.

[9] G. Kamiya, "Data centres and data transmission networks," International Energy Agency, Paris, Tech. Rep., 2022.

[10] M. A. Khan, A. P. Paplinski, A. M. Khan, M. Murshed, and R. Buyya, "Exploiting user provided information in dynamic consolidation of virtual machines to minimize energy consumption of cloud data centers," in *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*. IEEE, 2018, pp. 105–114.

[11] U. Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.

[12] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020. [Online]. Available: https://science.sciencemag.org/content/367/6481/984

[13] C. Augustine, R. Bain, J. Chapman, P. Denholm, E. Drury, D. G. Hall, E. Lantz, R. Margolis, R. Thresher, D. Sandor *et al.*, "Renewable electricity futures study. volume 2. renewable electricity generation and storage technologies," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2012.

[14] N. L. Panwar, S. C. Kaushik, and S. Kothari, "Role of renewable energy sources in environmental protection: A review," *Renewable and sustainable energy reviews*, vol. 15, no. 3, pp. 1513–1524, 2011.

[15] "What is renewable energy?" https://www.un.org/en/climatechange/what-is-renewable-energy, accessed: 2023-06-08.

[16] R. Gross, M. Leach, and A. Bauen, "Progress in renewable energy," *Environment international*, vol. 29, no. 1, pp. 105–122, 2003.

[17] M. Branscombe, "Google's solar deal for nevada data center would be largest of its kind," *Informa PLC, London*, 2020.

[18] P. Wiesner, D. Scheinert, T. Wittkopp, L. Thamsen, and O. Kao, "Cucumber: Renewable-aware admission control for delay-tolerant cloud and edge workloads," in *Euro-Par 2022: Parallel Processing: 28th International Conference on Parallel and Distributed Computing, Glasgow, UK, August 22–26, 2022, Proceedings.* Springer, 2022, pp. 218–232.

[19] M. Haddad, J. M. Nicod, C. Varnier, and M.-C. Peéra, "Mixed integer linear programming approach to optimize the hybrid renewable energy system management for supplying a stand-alone data center," in *2019 Tenth international green and sustainable computing conference (IGSC).* IEEE, 2019, pp. 1–8.

[20] Y. Lu, R. Wang, P. Wang, Y. Cao, J. Hao, and K. Zhu, "Energy-Efficient Task Scheduling for Data Centers with Unstable Renewable Energy: A Robust Optimization Approach," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Jul. 2018, pp. 455–462.

[21] Í. Goiri, M. E. Haque, K. Le, R. Beauchea, T. D. Nguyen, J. Guitart, J. Torres, and R. Bianchini, "Matching renewable energy supply and demand in green datacenters," *Ad Hoc Networks*, vol. 25, pp. 520–534, 2015.

[22] W. Liu, Y. Yan, Y. Sun, H. Mao, M. Cheng, P. Wang, and Z. Ding, "Online job scheduling scheme for low-carbon data center operation: An information and energy nexus perspective," *Applied Energy*, vol. 338, p. 120918, 2023.

[23] H. He, H. Shen, Q. Hao, and H. Tian, "Online delay-guaranteed workload scheduling to minimize power cost in cloud data centers using renewable energy," *Journal of Parallel and Distributed Computing*, vol. 159, pp. 51–64, 2022.

[24] S. Caux, P. Renaud-Goud, G. Rostirolla, and P. Stolf, "Phase-based tasks scheduling in data centers powered exclusively by renewable energy," in *2019 31st International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD).* IEEE, 2019, pp. 136–143.

[25] N. Sharma, S. Barker, D. Irwin, and P. Shenoy, “Blink: managing server clusters on intermittent power,” in *Proceedings of the sixteenth international conference on Architectural support for programming languages and operating systems*, 2011, pp. 185–198.

[26] V. Venkataswamy, J. Grigsby, A. Grimshaw, and Y. Qi, “Rare: Renewable energy aware resource management in datacenters,” in *Job Scheduling Strategies for Parallel Processing: 25th International Workshop, JSSPP 2022, Virtual Event, June 3, 2022, Revised Selected Papers*. Springer, 2023, pp. 108–130.

[27] J.-M. Pierson, G. Baudic, S. Caux, B. Celik, G. Da Costa, L. Grange, M. Haddad, J. Lecuivre, J.-M. Nicod, L. Philippe, V. Rehn-Sonigo, R. Roche, G. Rostirolla, A. Sayah, P. Stolf, M.-T. Thi, and C. Varnier, “DATAZERO: DATAcenter with Zero Emission and RObust management using renewable energy,” *IEEE Access*, vol. 7, p. (on line), juillet 2019. [Online]. Available: http://doi.org/10.1109/ACCESS.2019.2930368

[28] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen, “Modeling of lithium-ion battery degradation for cell life assessment,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1131–1140, 2016.

[29] C. Reiss, J. Wilkes, and J. L. Hellerstein, “Google cluster-usage traces: format+ schema,” *Google Inc., White Paper*, vol. 1, pp. 1–14, 2011.

[30] D. G. Feitelson, D. Tsafrir, and D. Krakov, “Experience with using the parallel workloads archive,” *Journal of Parallel and Distributed Computing*, vol. 74, no. 10, pp. 2967–2982, 2014.

[31] K. Wang, Y. Li, C. Wang, T. Jia, K. Chow, Y. Wen, Y. Dou, G. Xu, C. Hou, J. Yao *et al.*, “Characterizing job microarchitectural profiles at scale: Dataset and analysis,” in *Proceedings of the 51st International Conference on Parallel Processing*, 2022, pp. 1–11.

[32] D. Klusáček, Š. Tóth, and G. Podolníková, “Real-life experience with major reconfiguration of job scheduling system,” in *Job Scheduling Strategies for Parallel Processing: 19th and 20th International Workshops, JSSPP 2015, Hyderabad, India, May 26, 2015 and JSSPP 2016, Chicago, IL, USA, May 27, 2016, Revised Selected Papers 19*. Springer, 2017, pp. 83–101.

[33] S. Pfenninger and I. Staffell, “Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data,” *Energy*, vol. 114, pp. 1251–1265, 2016.

[34] I. F. de Nardin, P. Stolf, and S. Caux, “Mixing offline and online electrical decisions in data centers powered by renewable sources,” in *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2022, pp. 1–6.

[35] ——, “Analyzing power decisions in data center powered by renewable sources,” in *2022 IEEE 34th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, 2022, pp. 305–314.

[36] J. Houghton, “Global warming,” *Reports on progress in physics*, vol. 68, no. 6, p. 1343, 2005.

[37] U. Nations, "The Paris Agreement," publisher: United Nations. [Online]. Available: https://www.un.org/en/climatechange/paris-agreement

[38] Climate Action Tracker, "Massive gas expansion risks overtaking positive climate policies," *Warming Projections Global Update, November*, 2022.

[39] V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock *et al.*, "Global warming of 1.5 c," *An IPCC Special Report on the impacts of global warming of*, vol. 1, no. 5, pp. 43–50, 2018.

[40] IPCC Climate Change, "A threat to human wellbeing and health of the planet," *Taking Action Now Can Secure our Future*, 2022.

[41] T. Wheeler and J. von Braun, "Climate change impacts on global food security," *Science*, vol. 341, no. 6145, pp. 508–513, 2013. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1239402

[42] T. M. Lenton, C. Xu, J. F. Abrams, A. Ghadiali, S. Loriani, B. Sakschewski, C. Zimm, K. L. Ebi, R. R. Dunn, J.-C. Svenning *et al.*, "Quantifying the human cost of global warming," *Nature Sustainability*, pp. 1–11, 2023.

[43] UNESCO., "Guide to measuring information and communication technologies (ict) in education," 2009.

[44] "Measuring digital development - facts and figures 2022," https://www.itu.int/itu-d/reports/statistics/facts-figures-2022/, accessed: 2023-06-07.

[45] "What is a data center?" https://www.ibm.com/topics/data-centers, accessed: 2023-06-07.

[46] "Electricity domestic consumption," https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html, accessed: 2023-06-07.

[47] "Data centres metered electricity consumption 2021," https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2021/keyfindings/, accessed: 2023-06-07.

[48] "Klimastatus og -fremskrivning 2023," https://ens.dk/service/fremskrivninger-analyser-modeller/klimastatus-og-fremskrivning-2023, accessed: 2023-06-07.

[49] G. E. Moore *et al.*, "Cramming more components onto integrated circuits," 1965.

[50] N. Y. Amponsah, M. Troldborg, B. Kington, I. Aalders, and R. L. Hough, "Greenhouse gas emissions from renewable energy sources: A review of lifecycle considerations," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 461–475, 2014.

[51] P. Bojek, "Renewables - energy system overview," International Energy Agency, Paris, Tech. Rep., 2022.

[52] P. L. Kunsch and J. Friesewinkel, "Nuclear energy policy in belgium after fukushima," *Energy policy*, vol. 66, pp. 462–474, 2014.

[53] C. Ren, D. Wang, B. Urgaonkar, and A. Sivasubramaniam, "Carbon-aware energy capacity planning for datacenters," in *2012 IEEE 20th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*. IEEE, 2012, pp. 391–400.

[54] "Amazon sets a new record for most renewable energy purchased by a single company," https://www.aboutamazon.eu/news/sustainability/amazon-sets-a-new-record-for-most-renewable-energy-purchased-by-a-single-company, 2023, accessed: 2023-06-08.

[55] R. S. Garcia and D. Weisser, "A wind–diesel system with hydrogen storage: Joint optimisation of design and dispatch," *Renewable energy*, vol. 31, no. 14, pp. 2296–2320, 2006.

[56] W. Dong, Y. Li, and J. Xiang, "Optimal sizing of a stand-alone hybrid power system based on battery/hydrogen with an improved ant colony optimization," *Energies*, vol. 9, no. 10, p. 785, 2016.

[57] A. Maleki and F. Pourfayaz, "Optimal sizing of autonomous hybrid photovoltaic/wind/battery power system with lpsp technology by using evolutionary algorithms," *Solar Energy*, vol. 115, pp. 471–483, 2015.

[58] S. Sinha and S. Chandel, "Review of recent trends in optimization techniques for solar photovoltaic–wind based hybrid energy systems," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 755–769, 2015.