Review

Strategies for improving the sustainability of data centers via energy mix, energy conservation, and circular energy

Matteo Manganelli 1,*, Alessandro Soldati 2, Luigi Martirano 1, and Seeram Ramakrishna 3

1 Department of Astronautics, Electrical & Energy Engineering, Sapienza University of Rome, Italy; (matteo.manganelli, luigi.martirano@uniroma1.it)
2 Department of Engineering and Architecture, University of Parma, Italy; alessandro.soldati@unipr.it
3 Center for Nanotechnology & Sustainability, Department of Mechanical Engineering, National University of Singapore; seeram@nus.edu.sg

* Correspondence: matteo.manganelli@uniroma1.it

Abstract: Information and communication technologies (ICT) are increasingly permeating our daily life and we ever more commit our data to the cloud. Events like the COVID-19 pandemic put an exceptional burden upon ICT infrastructures. This involves increasing implementation and use of data centers, which increased energy use and environmental impact. The scope of this work is to take stock on data center impact, opportunities, and assessment. First, we estimate impact entity. Then, we review strategies for efficiency and energy conservation in data centers. Energy use pertains to power distribution, IT-equipment, and non-IT equipment (e.g. cooling): Existing and prospected strategies and initiatives in these sectors are identified. Among key elements are innovative cooling techniques, natural resources, automation, low-power electronics, and equipment with extended thermal limits. Research perspectives are identified and estimates of improvement opportunities are presented. Finally, we present an overview on existing metrics, regulatory framework, and bodies concerned.

Keywords: data center; green data center; sustainability; energy efficiency; energy saving; ICT.

1. Introduction

Digital economy is expanding and so is the demand for information and communication technology (ICT), driving the data center industry. Compared to the recent “age of computing”, the present time is regarded as the “age of data” [1]. Drivers for the recent massive expansion of ICT are the fifth generation mobile networks (5G), modern computing paradigms, internet of things (IoT) [2,3], cryptocurrencies, blockchain [4], big data science, artificial intelligence (AI), and emergencies like the ongoing COVID-19 pandemic [5,6]. Key estimates on 2018-2023 digital evolution by Cisco are reported in Table 1 [7].

The fifth generation mobile network, known as 5G, is being implemented to meet increasing service demand [8]. The related energy demand is under investigation [9,10].

Cryptocurrencies (Bitcoin being the first and most famous) are media of exchange, which are digital, encrypted, and distributed. They are not issued by a central authority but rather are based on a distributed ledger, typically blockchain. Mining is the release of new units of cryptocurrencies [11,12]. The energy and environmental costs of cryptocurrency mining is an emerging issue [13-17]. The estimated energy use related to Bitcoin is reported in Figure 1. Sustainable alternatives are under investigation [18].

A blockchain is an open-source distributed database, based on state-of-the-art cryptography, via a distributed ledger [19]. The first application of blockchains has been to support bitcoin transactions; today they are regarded as disruptive in many applications [20,21], including climate change [22], energy [23], and health [24]. A recent application of blockchains is in smart mobility, supporting Internet-of-Vehicles [25]. The energy and en-
The environmental impact of blockchains is investigated [17,26]. Other drivers are modern computing paradigms, e.g., cloud computing, edge computing, fog computing, and the IoT [3].

The COVID-19 pandemic changed the use of ICT. In March 2020, Microsoft Teams use increased by 775% [27] and Facebook group calls increased tenfold in Italy [28]; Zoom exceeded 200 million daily participants [29]. Changes in social media use following COVID-19 are addressed e.g. by J.P. Morgan [30]; Amazon notably profited [31]. This can also lead to beneficial results: Ong et al. [32,33] estimate the impact of videoconferencing, in terms of energy and CO2 costs over the life cycle, compared to face-to-face meetings.

Information traders (e.g., Google/Alphabet, Amazon, Apple, Facebook, Microsoft) are among top companies by market capitalization [34,35]. ICT electricity demand is expected to accelerate 2020-2030, to 8%-21% (based on scenario) of total electricity demand [36]. Power usage of data centers can be as high as about 400 MW [37]. Top data centers by power are presented in Table 2.

| Table 1. Forecast of digital evolution 2018-2023, elaborated from [7] |
|--------------------------|----------------|----------------|
|                          | 2018 | 2023 | Variation |
| Internet users (billions) | 3.9  | 5.3  | +36%       |
| Internet users (% of world population) | 51%  | 66%  | +29%       |
| Average mobile networked devices and connections per person | 1.2  | 1.6  | +33%       |
| Average total networked devices and connections per person | 2.4  | 3.6  | +50%       |
| Average broadband speed (Mbps) | 46   | 110  | +139%      |
| Average Wi-Fi speed (Mbps) | 30   | 92   | +207%      |
| Average mobile speed (Mbps) | 13   | 44   | +238%      |

Data centers use ICT devices (servers, storage drives, and network devices) which are electrically powered. Heat caused by device operation must be removed by cooling equipment also powered by electric energy. Top items of energy consumptions are cooling and servers, estimated as 43% each, followed by storage drives and network devices (11% and 3% respectively) [39]. Other estimations are roughly 52% IT equipment, 38% cooling system, and 10% remaining equipment (e.g., power distribution, UPS) [40]. Power flows in a typical data center are illustrated in Figure 2 and Figure 3. Electricity use estimations 2010-2030 for ICT and data centers are presented in Figure 4. Percent of global GHG emissions by ICT is presented in Figure 5.
On these grounds, energy and environmental sustainability of data centers is a priority in the ICT industry. European strategies push for data center to be carbon-neutral by 2030 [43].

Main actions on energy use and operational carbon of data centers are: high-performance computing (software); energy conservation of computer rooms (hardware); low-power servers (hardware); renewable energy application (hardware) [44] (Figure 6).
In this work, we focus on energy conservation strategies at hardware level, as outlined in Figure 7.

Figure 4. (a) global ICT annual electricity use; (b) ICT share of global electricity use; (c) global annual electricity use for data centers; (d) data centers share of global electricity use; elaborated from [36].

Figure 5. ICT share of global GHG emissions; elaborated from [36,45-49].
2. Power supply and continuity

Electrical power supply plays a key role in the proper and efficient operation of a data center. The loads of a data center can be classified in two main levels according to the type of service requested:

- **IT equipment** – Computers, servers, storage, electronic and telecommunication equipment enclosed in the IT racks need a vital service with a very high continuity of supply, given the vulnerability even for very short voltage dips (milliseconds).
- **Non-IT equipment/data center physical infrastructure (DCPI)** – Cooling, lighting and other auxiliaries need a preferential service, considering system inertia with a tolerable out of service of several seconds.

Vital service is supplied by Uninterruptible Power Systems (UPS). Emergency service is supplied by Emergency Generator Sets (EGS).

Open-source IT systems are redefining how power is distributed within an IT rack by replacing internal server power supplies with a centralized rack-level power supply.

Rackmount servers use a variety of dc voltages ranging from 12 Vdc down to 1 Vdc, to power internal components, e.g., CPU, GPU, hard drives, memory, fans, peripherals.

Generation of these dc levels starting from ac voltages of the power system occurs in several steps through transformers (ac/ac), rectifiers (ac/dc) and converters/regulators (dc/dc). Inside the IT rack, power supply units (PSU) convert ac to dc power (typically to 12 Vdc, and more recently to 48 Vdc). Voltage Regulator Modules (VRM) on the motherboard are dc/dc converters used to reduce the voltage further for final use by the internal components.

PSU can be divided in single-cord or dual-cord supply. Dual-corded equipment is normally fed from separate sources by two PSU below 50% capacity, so that either PSU can feed total load whenever necessary. Conventional internal server PSU architectures...
and centralized rack-level PSU architectures (12 Vdc and 48 Vdc) are the most used architectures in IT rack data centers. With best-in-class components, the consolidated 12 Vdc rack-level PSU architecture provides a small incremental energy efficiency improvement over the conventional architecture. Consolidating at 48 Vdc provides another small incremental energy efficiency improvement over 12 Vdc.

![Figure 8. Conventional server rack and rack with PSUs disaggregated from servers, elaborated from [51]](image)

### 2.1. Tier classification

Service continuity is of paramount importance for data centers. The Tier classification, proposed by The Uptime Institute [52] and included in the ANSI/TIA-942-2005 standard [53], is the international standard on data center performance. Tiers are assigned based on data center reliability. They consequently define criteria on maintenance, power supply, cooling, and fault capabilities. A summary of Tier classification is reported in Table 3. The Uptime Institute illustrates characteristics of Tiers and features and actions on IT equipment, electrical infrastructure, mechanical infrastructure, and facility operations for reliability, availability, and serviceability [52].

<table>
<thead>
<tr>
<th>Tier Classification</th>
<th>Tier II</th>
<th>Tier II</th>
<th>Tier III</th>
<th>Tier IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denomination</td>
<td>Basic</td>
<td>Redundant components</td>
<td>Concurrently maintainable</td>
<td>Fault tolerant</td>
</tr>
<tr>
<td>Site availability</td>
<td>99.671%</td>
<td>99.749%</td>
<td>99.982%</td>
<td>99.995%</td>
</tr>
<tr>
<td>Site IT downtime (h/a)</td>
<td>28.8</td>
<td>22.0</td>
<td>1.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Delivery paths</td>
<td>1</td>
<td>1</td>
<td>1 active</td>
<td>2 active</td>
</tr>
<tr>
<td>Component redundancy</td>
<td>N</td>
<td>N + 1</td>
<td>N + 1</td>
<td>2(N + 1)</td>
</tr>
<tr>
<td>Susceptible to disruption from planned activities</td>
<td>Susceptible</td>
<td>Less susceptible</td>
<td>Not susceptible</td>
<td>Not susceptible</td>
</tr>
<tr>
<td>Susceptible to disruption from unplanned activities</td>
<td>Susceptible</td>
<td>Less susceptible</td>
<td>Susceptible</td>
<td>Not susceptible</td>
</tr>
<tr>
<td>Fault tolerant to worst event</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Redundancy (multiplication of components) is a pursued strategy to reduce probability of failure and improve reliability (ability to perform under stated conditions for a given time) and availability (degree to which a system is operational when required for use) [41,54,55]. Topology and power distribution for reliability is discussed e.g. by Wiboonrat [56]. It should be remarked that redundancy increases costs and decreases energy efficiency [41]. Therefore sustainability should be assessed also based on reliability. This is also reflected in multidimensional metrics.

### 2.2 Power losses

Multiple causes of energy loss exist in a data center, as discussed e.g. by Rasmussen [42]. Ideally, all power should be delivered to IT equipment; in reality, energy is obviously consumed also by non-IT equipment/DCPI. Part of the DCPI is in series with and powers IT while other is in parallel as it supports IT operation. DCPI efficiency is pursued via...
more efficient devices, accurate sizing, and innovative techniques. A more accurate sizing matching IT load is regarded as the most immediate opportunity.

Losses in DCPI components are usually divided among no-load losses, proportional losses, and square-law losses; typical values as a fraction of full load rating are reported [42]. An energy model of data centers is also available, illustrating various items of energy loss (Figure 10) [42]. Data center modeling in literature is discussed by Ahmed et al. and a model of the electrical energy consumption of data center subsystems, considering their interactions and highlighting power losses, is presented [57].
Figure 9. Scheme of electric system based on Tier: (a) Tier I; (b) Tier II; (c) Tier III; (d) Tier IV; elaborated from [58]
Figure 10. Diagram of energy losses in a data center, elaborated from [42]
2.3. UPS

Data centers are mission-critical, and reliability is expected; hence, UPS are key. Even if they operate in emergency conditions, they are part of the infrastructure and taken into account in efficiency measurements. UPS losses can be grouped into no-load losses, proportional losses, and square-law losses (Figure 11a), as follows [59]:

- No-load losses: caused by no-load losses in transformers, capacitors, and auxiliaries;
- Proportional losses: due to switching losses in transistors and conduction losses in semiconductors and rectifiers;
- Square-law losses: Joule losses.

UPS efficiency can be pursued via technology, topology, and modularity. Efficiency typically decreases with reduced load in common UPS, while it is maintained in efficient UPS (Figure 11b). UPS efficiency is discussed e.g. by Milad and Darwish [60].

UPS efficiency values are usually given at 100% load under the most favorable conditions, leading to nearly identical values for different UPS. UPS efficiency depends on the load – increasing redundancy means adding extra spare capacity, hence redundancy can have a deep impact on efficiency. Efficiency at 30% load is proposed to better specify a UPS [59]. An example can be given as follows. Assuming 800 kW load, so that a 1 000 kW UPS operates at 80% load (typical threshold set by operators), the effect of UPS alternative configurations is as per Table 4. The same load represents a different relative load based on UPS configuration. Above 80% load, the energy cost of UPS losses can be regarded as small with respect to IT load.

Table 4. Example on comparison of loading of different UPS configurations

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Modules</th>
<th>Capacity</th>
<th>Percent load</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internally modular</td>
<td>4 × 250 kW</td>
<td>1 000 kW</td>
<td>80%</td>
<td>Small</td>
</tr>
<tr>
<td>Internally modular redundant</td>
<td>5 × 250 kW</td>
<td>1 250 kW</td>
<td>64%</td>
<td>High</td>
</tr>
<tr>
<td>Parallel redundant</td>
<td>3 × 500 kW</td>
<td>1 500 kW</td>
<td>53%</td>
<td>High</td>
</tr>
</tbody>
</table>

Offline UPS provide the load with a bypass mains power supply without conditioning and ensure maximum efficiency of 99% in comparison with online UPS.

Figure 11. (a) total losses in an UPS, elaborated from [59]; (b) Curve of efficiency of UPS, elaborated from [59,61]

Standard IEC 62040-3 [62] applies to movable, stationary and fixed electronic uninterruptible power systems (UPS) that deliver single/ three-phase fixed frequency ac output voltage not exceeding 1 000 V and that incorporate an energy storage system, generally connected via a dc link. The standard is intended to specify performance and test
requirements of a complete UPS system and not of individual UPS functional units. The standard introduces a code of identification of the service of the UPS, based on:

- Output characteristics, in terms of voltage and frequency independent of the output in comparison with the input;
- Output wave form, normal or in by-pass;
- Dynamic performance in output.

3. Energy conservation of computer rooms

Rong et al. [44] review technologies for optimizing energy use in data centers as of 2016, including energy conservation in computer rooms. Nadjahi et al. [40] discuss thermal loads, active cooling (air conditioning) and passive cooling (free cooling, liquid cooling, two-phase cooling, building envelope), as of 2018. Ebrahimi et al. [50] discuss configuration of data centers, thermal loads and thermal management (cooling systems); they also provide discussion on technologies for waste heat recovery.

Energy conservation of computer rooms may rely on:

- New architecture and control of the cooling system [40];
- Possible control of lighting system [63].

3.1. Improvement of the cooling system

As one of major items of energy use in data centers, improvements of cooling systems are investigated.

3.1.1. Location of data centers

Choosing the most appropriate location for a data center is essential. One common driver in this choice is the risk associated with the site itself [64]. Here, “risk” has a general meaning, pertaining not only to natural adverse events, but also to utility costs and relationship with other services and facilities.

On the other hand, electric air conditioning is found to aggravate urban microclimate (heat island effect), in turn exacerbating the need for cooling [44,65], in a vicious circle. Nonetheless, location choice based on environmental conditions can improve the efficiency of the cooling system [66,67]. For example, locations with abundant water or cold climate are considered for natural cooling: BMW has a data center in Iceland [70,71], Facebook has and is investing in data centers in Sweden [68,69,72], and Google has a data center in Finland [73]. Microsoft deployed an underwater data center off Scotland’s Orkney Islands [74]. In some of those locations, inexpensive, renewable energy is also available.

Lei et al. [75] recently investigated achievable Power Usage Effectiveness (PUE) - i.e., practical minimum PUE with given climate conditions and state-of-the-art technologies - based on locations of 17 Facebook and Google hyperscale data centers, via a simulation model of free cooling with different sources. It is found that this can impact up to twice as much. Impact on different items of consumption is also noted. Other studies on quantification of the impact of the location were by Depoorter et al. [76] and by Shehabi et al [77]. Considerations on location of data centers, other than cooling, are addressed by Atkins [78].
3.1. Room configuration

The configuration of computer rooms impacts on airflow; parameters can be e.g. room ceiling or configuration of floor openings in raised-floor data centers [80-82]. The cost for a proper design of room layout and ventilation system in large data centers is estimated as 8%-10% of total cost and it is usually compensated by energy saving in two to three years [44].

3.1.3. Room temperature

Room temperature and humidity values are recommended, for equipment life. However, as better equipment is being released, thermal limits are relaxing accordingly, to cut cooling costs. ASHARE thermal guidelines recommended 20-25 °C dry-bulb air temperature in 2004 and 18-27°C in 2008; in 2011, equipment classes were defined, allowing a range as wide as 5-45 °C. Dell presented servers able to withstand 45 °C air [83]. Google raised room temperature to 26.7 °C [84].

3.1.4. Airflow pattern and active cooling

Unified rack arrangement is now obsolete, implying mixing of hot and cold air; face-to-face or back-to-back (hot aisle/cold aisle) arrangement is now common [40,44,85]. Bedekar et al. [86] investigated optimal computer room air conditioning (CRAC) location via computational fluid dynamics.

Air distribution in computer rooms is discussed in literature [40,44,87,88]. Air distribution can be divided as upward or downward. Floor design can be hard floor or overhead. Air pattern can be open, local pipeline or full pipeline [44]. Air can be conveyed between CRAC and rack via flooded supply/return, locally ducted supply/return, fully ducted supply/return [88]. To avoid mixing of hot and cold air in hot aisle/cold aisle arrangement, cold aisle containment or hot aisle containment are implemented [40]. The latter is found to be the best of the two [87].

The CRAC unit maintains temperature, airflow, and humidity in the computer room. It typically uses the compression refrigeration cooling. Other than energy consumption, downsides of active cooling are noise and reliability [40]. To address energy consumption, passive cooling has been investigated.

3.1.5. Innovative cooling systems

Innovative cooling systems [89-91] can be based on natural air, chilled water, liquid immersion, evaporative cooling, geothermal cooling, passive cooling, pumped two-phase cooling, also with the contribution of smart monitoring and IoT.

3.1.6. Use of natural cold source

The use of a natural cold source can be a direct use or an indirect use. In direct use, outdoor air is directly introduced; humidity control and filtration are required. In indirect
use, heat exchange equipment is used. The crucial point in using a natural cold source is the efficiency of the heat exchange between indoor and outdoor air [44].

3.1.7. Free cooling

In free cooling, natural fluids are used, without mechanical active components [46]. This increases energy efficiency, savings, and reliability. Free cooling exploits airside or waterside economization. In airside economization, cold air is used, which must be at least 5 °C colder than indoor air. In waterside economization, water is used, from a nearby water source. In both cases, free cooling can be direct or indirect. In addition, the heat pipe concept can be combined with free cooling. Free cooling techniques are discussed by Zhang et al. [92] and by Daraghmeh et al. [93]. Free cooling technologies can be classified as per Table 5.

<table>
<thead>
<tr>
<th>Airside</th>
<th>Waterside</th>
<th>Heat pipe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct airside</td>
<td>Direct water-cooled</td>
<td>Independent</td>
</tr>
<tr>
<td>Indirect airside</td>
<td>Air-cooled systems</td>
<td>Integrated</td>
</tr>
<tr>
<td>Multistage evaporative</td>
<td>Cooling tower systems</td>
<td>Cold storage</td>
</tr>
<tr>
<td></td>
<td>Integrated dry cooler-chiller (water-to-air dry cooler)</td>
<td>Pulsating heat pipe</td>
</tr>
</tbody>
</table>

Table 5. Free cooling technologies

Reviewed studies present PUE in the range 1.10-1.16 and energy savings 30%-40%; particular studies declare coefficient of performance (COP) up to 9 or 12, or energy savings up to 47% or 49%.

3.1.8. Liquid cooling

In high power density data centers, technologies other than air cooling are recommended, e.g., liquid cooling. This has a higher heat transfer capacity per unit mass [94], allowing for a lower temperature difference between equipment and coolant, potentially allowing for passive cooling and also for heat reuse. Liquid cooling systems are discussed e.g. by Capozzoli et al. [95]. Liquid cooling systems can be implemented via micro-channels flow and cold-plate heat exchangers in contact with components (e.g., CPU or DIMM). Studies are e.g. by Zimmermann et al. [96,97] (hot water-cooled electronics and heat reuse; energy performance of Aquasar, the first hot water-cooled prototype), Coles et al. [98] (direct liquid cooling), Iyengar et al. [99,100] (experimental on IBM chiller-less test facility). Commercial systems are proposed e.g. by Asetek [101].

Another emerging technique is the fully immersed direct liquid cooling [87]. Commercial systems are proposed e.g. by Iceotope [102]. Chi et al. [94] compare an air-cooled and a fully immersed liquid-cooled system. Temperatures for liquid-cooling systems are discussed in literature [94,96,98,103].

3.2. Improvement of lighting system

Energy saving in lighting is pursued reducing power losses via efficient equipment (passive measures) and regulating power use via control systems (active measures) [104]. The motivation of lighting control is to provide lighting when, where, and in the amount needed.

3.2.1. Lighting control

Many data centers implement a “lights-out” practice, in which light fixtures are switched manually across a (large) space. The drawbacks are that the illuminated area is large compared to the accessed spot, and that lights can then be left on unnecessarily [63]. A proposed approach is the “follow-me” lighting (implemented e.g. in Facebook’s Oregon and North Carolina data centers) in which lighting is operated as a spotlight following the technician. Motion detectors are implemented in each light fixture and connected to a central application which controls on/off state and intensity (dimming) of each fixture [63].

3.2.2. Light sources
The common, inexpensive technology for data center lighting is fluorescent lighting. Drawbacks are as follows: life is shortened by number of starts and by dimming; maintenance is required, which is aggravated by shorter life, multitude of lamps, and disposal; dimming, as a cause of aging, is seldom implemented [63].

LED lighting has surpassed fluorescent lighting in energy efficiency and light quality, and it is recommended on the grounds of lower electricity use, lower heat release (impacting on HVAC), and dimming capability. The higher price of LED fixtures is dropping and compensated by longer life [63].

To further reduce heat release in the data center, LED fixtures are available, which do not implement drivers and are powered via a central supply, providing power conversion and control [63].

3.2.3. Other strategies for lighting improvement

Although black is the most common finish for ICT equipment, white racking could reduce the number of luminaires and lighting energy use by as much as 37% each [105].

3.3. Net zero energy data center

Net zero energy data centers (NZEDC) are defined, as per Deliverable 4.5 of project RenewIT [106], as data centers that «consume net zero non-renewable primary energy from power grid and district energy networks during their lifetime, while generating as much energy (both electric and thermal) as they use during a year.» A road map towards NZEDC is presented in Figure 13. Many technical concepts are investigated in the mentioned deliverable, as presented in Figure 14, and results on energy flows are discussed.

<table>
<thead>
<tr>
<th>I – Load reduction</th>
<th>II – Renewable energy use</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Efficient building</td>
<td>6. Cooling</td>
</tr>
<tr>
<td>2. Efficient IT-equipment</td>
<td>7. Power supply</td>
</tr>
<tr>
<td>3. Efficient power distribution</td>
<td></td>
</tr>
<tr>
<td>4. Efficient cooling distribution</td>
<td></td>
</tr>
<tr>
<td>5. Reuse of IT waste heat</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13. Road map to net zero energy data center, elaborated from [106]
Figure 14. Advanced technical concepts to reduce electric and cooling load, elaborated from [106].
4. Electronics and other strategies

4.1. Low-power servers

The server is regarded as the basic unit of power and heat flow path [107]. In low-power servers, energy usage is reduced via components configuration. Approaches are illustrated by Rong et al. [44], including features, energy efficiency, and constraint of selected technologies. Recent advancements are presented by Jin et al. [107], who compare existing server power consumption models and identify future research trends. It is found that it is possible to handle a heavier workload without increasing energy use. While accurate power consumption models of servers result in accurate cooling load calculation and avoid excessive redundancy, energy- and thermal-aware managements based on the model results in the best overall energy-savings. Meisner et al. [108] investigated high-power versus low-power server design, questioning the perceived energy efficiency of the latter.

4.2. The Little Box Challenge

Shrinking magnets, capacitance, and heat extraction are the main challenges in the design of high-power density converters [109]. The Little Box Challenge (LBC) was a competition, sponsored by Google and the IEEE Power Electronics Society [110,111], to build a prototype of a power inverter with given specifications (e.g. size approximately 1/10 of that of contemporary state of the art, efficiency above 95%, fitting in a 40 in³ casing) [112]. Improving Google’s data center efficiency was among the scopes [110]. Main design challenges are discussed by Neumayr et al. [113]. The outcomes ignited attention from the consumer electronics community and technology advancements. Design challenges and proposed solutions are examined [109,114] (Table 6).

<table>
<thead>
<tr>
<th>Team</th>
<th>Efficiency (%)</th>
<th>Power density (W/cm³)</th>
<th>Dimensions (cm)</th>
<th>Topology</th>
<th>FOM*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE+T RED</td>
<td>n/a</td>
<td>8.8</td>
<td>6.4 × 4.1 × 8.7</td>
<td>Parallel full-bridge</td>
<td>n/a</td>
</tr>
<tr>
<td>ETH !verter</td>
<td>95.1a</td>
<td>8.2</td>
<td>n/a</td>
<td>Parallel full-bridge</td>
<td>128</td>
</tr>
<tr>
<td>Schneider Electric</td>
<td>n/a</td>
<td>6.1</td>
<td>n/a</td>
<td>Full-bridge</td>
<td>n/a</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
<td>98</td>
<td>3.4</td>
<td>13.5 × 13.2 × 3.3</td>
<td>Full bridge</td>
<td>90</td>
</tr>
<tr>
<td>Taiwan Tech</td>
<td>96.5a</td>
<td>5.6</td>
<td>15.2 × 9.4 × 2.5</td>
<td>Full-bridge</td>
<td>117</td>
</tr>
<tr>
<td>UIUC</td>
<td>97a</td>
<td>13.2</td>
<td>10.2 × 6.2 × 2.4</td>
<td>FCMLI^</td>
<td>316</td>
</tr>
<tr>
<td>Univ. Tennessee</td>
<td>96.9b</td>
<td>6.2</td>
<td>11.1 × 8.8 × 3.3</td>
<td>Full-bridge</td>
<td>143</td>
</tr>
<tr>
<td>Virginia Tech FEEC</td>
<td>98.6b</td>
<td>3.7</td>
<td>n/a</td>
<td>HERIC^</td>
<td>123</td>
</tr>
</tbody>
</table>

Table 6. Main features of LBC finalists [109,114]

<table>
<thead>
<tr>
<th>Team</th>
<th>Ratings (V x A)</th>
<th>Switching frequency (kH z)</th>
<th>Switches</th>
<th>Power decoupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE+T RED</td>
<td>n/a</td>
<td>35-240</td>
<td>GaN</td>
<td>Active synchronous buck to buffer</td>
</tr>
<tr>
<td>ETH !verter</td>
<td>600 × n/a</td>
<td>200-1000</td>
<td>GaN</td>
<td>Active synchronous buck to buffer</td>
</tr>
<tr>
<td>Schneider Electric</td>
<td>n/a</td>
<td>45</td>
<td>SiC</td>
<td>Active ripple filter full bridge to buffer</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
<td>650 × 30</td>
<td>100</td>
<td>GaN</td>
<td>Active half-bridge like three-phase</td>
</tr>
<tr>
<td>Taiwan Tech</td>
<td>650 × 60</td>
<td>25-800^, 200-680^h</td>
<td>GaN</td>
<td>Active synchronous buck to buffer</td>
</tr>
<tr>
<td>UIUC</td>
<td>150 × 48</td>
<td>120v, 720h</td>
<td>GaN</td>
<td>Active series-stacked buffer</td>
</tr>
<tr>
<td>Univ. Tennessee</td>
<td>650 × 30</td>
<td>100</td>
<td>GaN</td>
<td>Passive notch filter</td>
</tr>
<tr>
<td>Virginia Tech FEEC</td>
<td>n/a</td>
<td>60, 400h</td>
<td>GaN</td>
<td>Active interleaved buck 1st power stage</td>
</tr>
</tbody>
</table>

* Via CEC method; ^ peak; a seven-level flying capacitor multilevel inverter; a highly efficient and reliable inverter concept; d dc/ac; * dc/dc; # operated; b obtained; ^ figure of merit, based on efficiency and power density (approximated data).

Approaches to heat management are discussed by Kim et al. [114] (Table 7). Wide-bandgap switches cause lower switching losses than silicon switches. They are decisive in high-power density converters and were widely used in LBC prototypes. It is deduced...
that all teams relied on forced air cooling via fans. The winners (Red Electrical Devils by CE+T) paid much attention to thermal design [115].

Table 7. Thermal management of LBC finalists [114]

<table>
<thead>
<tr>
<th>Team</th>
<th>Thermal management</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE+T RED</td>
<td>Copper enclosure, with gap-pad</td>
</tr>
<tr>
<td>ETH Inverter</td>
<td>Forced air cooling by utilizing high fin-number heat sinks and six ultra-flat blowers</td>
</tr>
<tr>
<td>Schneider Electric</td>
<td>Heat sink over power switches with small fan, two air inlets on case</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
<td>Unspecified cooling system with heat sink</td>
</tr>
<tr>
<td>Taiwan Tech</td>
<td>Six fans, heat sink connected to aluminum case</td>
</tr>
<tr>
<td>UIUC</td>
<td>Copper enclosure, 2 mm tall heat sink fins, 6 radial fans</td>
</tr>
<tr>
<td>Univ. Tennessee</td>
<td>Heat sink over power switches, two small fans, air inlets on top and side</td>
</tr>
<tr>
<td>VT FECC</td>
<td>Copper enclosure, 10 micro-fans on side wall</td>
</tr>
</tbody>
</table>

4.3. Direct-current power supply

The diffusion of dc grids and power supply systems is envisaged in the evolution of the power system and ICT [116,117]. Wong et al. [116] simulate a modular dc power supply unit for servers. Pueschel [117] investigates a 380 Vdc microgrid, serving an office building and the data center of a German company, as an approach to energy efficiency and safety.

4.4. Semiconductors

Until recently, the best efficiency in UPS power stages (96%) was achieved via insulated-gate bipolar transistors (IGBT) with three-level switching topologies. Recently, silicon carbide (SiC) transistors allowed to exceed 98% efficiency and nearly independent of percentage load. This is possible via the properties of wide-bandgap (WBG) semiconductors. SiC devices are proposed by ROHM Semiconductor, Wolfspeed, and ON Semiconductor [118]. As an example, the efficiency of a Mitsubishi SiC-based UPS is reported above 98% for any load above 30% [61].

4.5. Automation, monitoring, and robotization

Integration of robotics in data centers is envisaged for their management and maintenance. While robots cannot completely replace human operators (at least in the near future), they can be used to automate repetitive tasks, relieving operators and increasing productivity. Robotic maintenance can enable the implementation of “lights out” data centers and of a vertical configuration of the space. Challenges and possible benefits are discussed [119]. An unmanned data center has been launched by AOL [120]. In addition, robotics can be used for diagnosis of data centers and energy management [121-123]. Emerging applications of automation, monitoring, and robotization are presented by Levy and Subburaj [124].

4.6. Modular data centers

Modular data centers are mobile data centers, designed for rapid deployment, e.g., for disaster recovery. They feature high energy efficiency and density, and they can be easily scaled. As an example, HP manufactures the Performance Optimized Datacenter (POD). Model 240a, nicknamed “EcoPOD”, has a declared PUE of 1.05 to 1.30.

5. Regulatory environment governing data centers

5.1. Metrics

5.1.1. Indicators

Given the increasing impact of data centers on society under many aspects (energy, security, sustainability), the need for comparing different solutions calls for reliable metrics and figures of merit. That is the reason behind the introduction of the “multidimen-
sional approach” by Levy and Raviv [41,125], who formalized more specific previous attempts, such as [126]. That approach was then specialized on “green” [127,128] and sustainability metrics [129,130].

Concerning data center efficiency, common metrics are the power usage effectiveness (PUE) and the carbon usage effectiveness (CUE). PUE is defined as the ratio of total facility power to ICT equipment power, quantifying extra power required per unit ICT power. The best PUE is ideally 1 and in practice it ranges from 1.2 (very efficient) to 3 (very inefficient). The data center infrastructure efficiency (DCIE) is sometimes used, equivalent to the inverse of PUE. CUE is defined as the product of CO₂ emission factor (CEF) and PUE [131].

Other performance indicators are reported [44,131,132]. Specific indicators are proposed, to quantify effectiveness of on-site renewable generation (on-site energy fraction, OEF, and on-site energy matching, OEM), energy reuse (energy reuse factor, ERF), and water usage (water usage effectiveness, WUE) [131]. Concerning data center sustainability, specific environmental metrics beyond renewable energy and efficiency can be introduced, such as those related to lifecycle assessment, decommissioning costs, use of recycled materials and possibility of second-life reuse of some parts [133,134]. Also, indicators exist to correlate energy to processed information, e.g., joules per bit [135]. Levy and Raviv present a discussion on metrics and sub-metrics and propose a new metric approach, the “data center multidimensional scorecard”, illustrated in Figure 15 [41]. Also Lykou et al. discuss existing metrics and propose a new, sustainability-oriented methodology [136]. A comprehensive taxonomy on data center metrics is presented by Reddy et al. [137].

![Data center scorecard](image)

**Figure 15.** Illustration of the data center multidimensional scorecard metric, elaborated from [41].

### 5.1.2. Trends

A global survey by the Uptime Institute reported average annual PUE to have decreased 2007-2013 and then stalled (Figure 16). Improvements are due to major steps in energy efficiency (hot/cold air separation, increased thermal limits, enhanced control, free cooling). Speculations on the recent stall include exceptional outdoor temperatures, shift of workloads to public cloud services – resulting in data centers operating inefficiently, or diffusion of high power density data center.

![PUE chart](chart)

**Figure 16.** Evolution of industry-average PUE, elaborated from [138]
5.2. Regulations

Main requirements on data centers [139] are on: temperature and humidity control – see ASHRAE specifications [140], static electricity monitoring; fire suppression; physical security; continuity and quality of supply, availability, physical protection, information security (protection of personal and financial data), etc. The regulatory framework concerning data centers is constantly evolving; main institutions which have contributed are listed by Levy and Raviv [41]. An overview on the matter is reported in Table 8.

Table 8. Main best practices, guidelines, and standards on data centers

<table>
<thead>
<tr>
<th>Body</th>
<th>Document</th>
<th>Scope</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICPA</td>
<td>SAS 70</td>
<td>Assurance controls</td>
<td>U.S.</td>
</tr>
<tr>
<td>AICPA</td>
<td>SSAE 16</td>
<td>Assurance controls</td>
<td>U.S.</td>
</tr>
<tr>
<td>AMS-IX</td>
<td>AMS-IX</td>
<td>Data center business continuity standard</td>
<td>International</td>
</tr>
<tr>
<td>ANSI/BICSI</td>
<td>ANSI/BICSI 002</td>
<td>Data center design and implementation</td>
<td>U.S.</td>
</tr>
<tr>
<td>ANSI/ASHRAE</td>
<td>ANSI/ASHRAE 90.4</td>
<td>Data center energy standard</td>
<td>U.S.</td>
</tr>
<tr>
<td>ASHRAE</td>
<td>TC 9.9 guidelines</td>
<td>Data center equipment – thermohygrometric limits</td>
<td>U.S.</td>
</tr>
<tr>
<td>BICSI</td>
<td>BICSI-009</td>
<td>Data center operations and maintenance best practices</td>
<td>U.S.</td>
</tr>
<tr>
<td>CENELEC</td>
<td>EN 50541-1</td>
<td>Power supply – distribution transformers</td>
<td>European</td>
</tr>
<tr>
<td>CENELEC</td>
<td>EN 50160</td>
<td>Power supply – voltage of distribution system</td>
<td>European</td>
</tr>
<tr>
<td>CENELEC</td>
<td>EN 50173</td>
<td>Information technology – cabling</td>
<td>European</td>
</tr>
<tr>
<td>CENELEC</td>
<td>EN 50174</td>
<td>Information technology – cabling</td>
<td>European</td>
</tr>
<tr>
<td>CENELEC</td>
<td>EN 30600</td>
<td>Information technology – data center certification</td>
<td>European</td>
</tr>
<tr>
<td>IAASB</td>
<td>IASAE 3402</td>
<td>Assurance controls</td>
<td>International</td>
</tr>
<tr>
<td>IEC</td>
<td>IEC 62040</td>
<td>Power supply – uninterruptible power systems (ups)</td>
<td>International</td>
</tr>
<tr>
<td>IEC</td>
<td>IEC 60076</td>
<td>Power supply – power transformers</td>
<td>International</td>
</tr>
<tr>
<td>IEC</td>
<td>IEC 60831-1 and 2</td>
<td>Power supply – capacitors</td>
<td>International</td>
</tr>
<tr>
<td>IEC</td>
<td>IEC 61439</td>
<td>Power supply – low voltage switchgear</td>
<td>International</td>
</tr>
<tr>
<td>ISO</td>
<td>ISO 14000</td>
<td>Environmental management system</td>
<td>International</td>
</tr>
<tr>
<td>ISO</td>
<td>ISO 27000</td>
<td>Information security</td>
<td>International</td>
</tr>
<tr>
<td>ISO</td>
<td>ISO 30134</td>
<td>Data centers — key performance indicators</td>
<td>International</td>
</tr>
<tr>
<td>ISO</td>
<td>ISO 45001</td>
<td>Occupational health and safety management systems</td>
<td>International</td>
</tr>
<tr>
<td>ISO</td>
<td>ISO 9001</td>
<td>Quality management system</td>
<td>International</td>
</tr>
<tr>
<td>ISO/IEC</td>
<td>ISO/IEC 11801</td>
<td>Information technology – cabling</td>
<td>International</td>
</tr>
<tr>
<td>ISO/IEC</td>
<td>ISO/IEC 27001</td>
<td>Information technology – information security</td>
<td>International</td>
</tr>
<tr>
<td>ISO/IEC</td>
<td>ISO/IEC 22237</td>
<td>Data centers – facilities and infrastructures</td>
<td>International</td>
</tr>
<tr>
<td>PCI SSC</td>
<td>PCI DSS</td>
<td>Payment card industry data security standard</td>
<td>International</td>
</tr>
<tr>
<td>Singapore Standard</td>
<td>SS 564</td>
<td>Green data centers</td>
<td>Singapore</td>
</tr>
<tr>
<td>TIA</td>
<td>ANSI/TIA-568</td>
<td>Information technology – cabling</td>
<td>U.S.</td>
</tr>
<tr>
<td>TIA</td>
<td>ANSI/TIA-942</td>
<td>Information technology – data center certification</td>
<td>U.S.</td>
</tr>
<tr>
<td>Uptime Institute</td>
<td>Tier classification</td>
<td>Information technology – data center certification</td>
<td>International</td>
</tr>
</tbody>
</table>

5.3. Certifications and initiatives

Data centers fall within certifications or initiatives on sustainable ICT or buildings.

In the United States, possible certifications for green data centers are the Leadership in Energy and Environmental Design (LEED) by the U.S. Green Building Council [141] and the U.S. National Data Center Energy Efficiency Information Program within the ENERGY STAR program [142]. Other than advanced cooling and reduced energy use, features of a LEED compliant data center are a clean backup system, use of renewable energy, green construction, and intelligent design [143].

A number of companies and other stakeholders of data center efficiency are part of The Green Grid consortium [144]. The Green500 list biannually ranks supercomputers, in the TOP500 list, for energy efficiency – the NVIDIA DGX SuperPOD (2.356 Piflops) ranked...
first in November 2020 with 26.195 Gflops/W [145]. Other pertaining initiatives in the U.S. are the Energy Efficiency Improvement Act of 2014 (H.R. 2126), the Data Center Optimization Initiative, and the Federal Data Center Consolidation Initiative [41].

6. Conclusions

Energy use for ICT is ever more increasing and so are concerns on sustainability of data centers. In this paper we review approaches to reduce energy consumption and resource depletion caused by operation of data centers, highlighting promising strategies and future research directions. Main actions are on software (HPC) and hardware (energy conservation of computer rooms – cooling and lighting, energy conservation in electronic equipment, integration of renewable energy). Metrics and the regulatory environment are a useful framework to support actions. Several indicators have been introduced to assess the state of the art and future targets of single aspects of efficiency (energy efficiency, carbon impact, use of resources). As a general concept, the definition of NZEDC was proposed in literature and it can be regarded as a useful benchmark. To reduce cooling load, several concepts have been proposed, taking advantage of favorable environmental conditions (location), natural cold sources, and passive cooling. Also the electronics is evolving to reduce IT load, via energy-aware IT management and new architectures. Also, a balance must be achieved between energy conservation and performances (continuity and quality). The extension of efficiency initiatives to data centers and the investigation of new technologies are desirable. As our life ever more relies on data and thus on the data center industry, in light of the ongoing digital evolution and rising environmental concerns, sustainability of data centers must be pursued.

Funding: This research received no external funding.

Acknowledgments: The authors acknowledge Matteo Dalboni and Ludovica Gentili (Department of Engineering and Architecture, University of Parma, Italy) for the technical support on the figures. Author Seeram Ramakrishna acknowledges the IAF and future research directions. Several indicators have been introduced to assess the state of the art and future targets of single aspects of efficiency (energy efficiency, carbon impact, use of resources). As a general concept, the definition of NZEDC was proposed in literature and it can be regarded as a useful benchmark. To reduce cooling load, several concepts have been proposed, taking advantage of favorable environmental conditions (location), natural cold sources, and passive cooling. Also the electronics is evolving to reduce IT load, via energy-aware IT management and new architectures. Also, a balance must be achieved between energy conservation and performances (continuity and quality). The extension of efficiency initiatives to data centers and the investigation of new technologies are desirable. As our life ever more relies on data and thus on the data center industry, in light of the ongoing digital evolution and rising environmental concerns, sustainability of data centers must be pursued.

Funding: This research received no external funding.

Acknowledgments: The authors acknowledge Matteo Dalboni and Ludovica Gentili (Department of Engineering and Architecture, University of Parma, Italy) for the technical support on the figures. Author Seeram Ramakrishna acknowledges the IAF–PP project “Sustainable Tropical Data Centre Test Bed” recently awarded by the National Research Foundation of Singapore.

References


The Ridiculous Amount of Energy It Takes to Run Bitcoin


42. Rasmussen, N. Electrical Efficiency Measurement for Data Centers - White Paper #154; APC.


68. Harding, L. The node pole: inside Facebook’s Swedish hub near the Arctic Circle. *2015.*


83. Fitch, J. Dell’s Next Generation Servers: Pushing the Limits of Data Center Cooling Cost Savings; 2012.


