Energy Efficiency (2019) 12:1399–1428
https://doi.org/10.1007/s12053-018-9753-2

REVIEW ARTICLE

A review on energy efficiency and demand response with focus on small and medium data centers

Thiago Lara Vasques · Pedro Moura · Aníbal de Almeida

Received: 22 June 2017 / Accepted: 13 November 2018 / Published online: 29 November 2018

Abstract Data centers are the backbone of a growing number of activities in modern economies. However, the large increase of digital content, big data, e-commerce, and Internet traffic is also making data centers one of the fastest-growing users of electricity. The total energy consumption of data centers corresponded to almost 1.5% of the global electricity consumption and has an approximated annual growth rate of 4.3%. Therefore, it is very important to increase the energy efficiency in data centers with actions such as power usage management, server consolidation, energy-efficient components and systems, as well as demand response programs and renewable energy sources. Small and medium data centers account for more than 50% of the total electricity consumption in this sector. In fact, surveys indicate that this data center profile waste more energy than larger facilities. Nevertheless, existing studies tend to be focused on the energy-related issues for large data centers rather than small and medium data centers. Therefore, through a meticulous state-of-the-art literature review of data centers energy efficiency and demand response perspectives, this paper aims to present how an intensive energy consumer, such as small and medium data centers, can become more efficient from the energy point of view and how they can take advantage of demand response programs to decrease costs and to cooperate with the grid to ensure higher reliability and sustainable development goals.

Keywords Data centers · Information and communication technologies · Energy efficiency · Demand response

Introduction

The availability of Internet and the relevance that Information and Communication Technology (ICT) has in modern society is changing the way in which computing resources are typically provisioned and allocated, where the computing infrastructure itself is provided as a service to its users. In a not very distant past, data were generated and communicated primarily among ICT systems—albeit of diminishing size. In the future, data-producing systems will increasingly involve small, low-power sensors and actuators embedded in the physical world—a network of cyber-physical systems, also referred to as the Internet of Things (SIA 2015). However, the increasing demand of computing resources has brought an inevitable growth in the energy consumption associated with these infrastructure, fostering a set of ICT to reduce the environmental impacts called Green ICT (Craig-wood et al. 2010; Uddin and Rahman 2012).

T. L. Vasques (✉) · P. Moura · A. de Almeida
Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal
e-mail: tlvasques@isr.uc.pt

P. Moura
e-mail: pmoura@isr.uc.pt

A. de Almeida
e-mail: adealmeida@isr.uc.pt

Jiang et al. (2015) globally conceptualize a data center as a facility used to house enterprise's ICT equipment such as servers, telecommunications, and storage systems, including also supporting infrastructures of high-quality power delivery and cooling systems. Nowadays, data centers are the backbone of contemporary economies, having different profiles, such as server rooms that power small- to medium-sized organizations, enterprise data centers that support large corporations, and server farms that run cloud computing services hosted by major market players.

On the other hand, the increase of digital content, big data, e-commerce, and Internet traffic is also making data centers one of the fastest-growing users of electricity (Josh and Delforge 2014). The total energy consumption of data centers in 2012 was about 270 TWh, which corresponds to almost 1.5% of the global electricity consumption, and has an approximated annual growth rate of 4.3% (Van Heddeghem et al. 2014). Just in 2014, US data centers consumed 70 TWh of electricity and such consumption is projected to increase to roughly 73 TWh by 2020 (Shehabi et al. 2016), costing US$13 billion annually in electricity bills and emitting nearly 150 million metric tons of greenhouse gas (GHG) emissions per year. If worldwide data centers were a country, they would be the globe's 12th largest consumer of electricity, ranking somewhere between Spain and Italy (Josh and Delforge 2014). Thus, an understanding of data center energy use, disaggregated energy efficiency options, as well as the metrics used to characterize data center energy performance are fundamental to address this large load in the most sustainable way.

Connected with the above reality, conventional power systems have been facing a noticeable transition from a centralized supply side management to a decentralized supply and demand side management, as a result of the inclusion of distributed renewable generation, among other factors (Wang et al. 2011). The future electric power grid infrastructure, the so-called smart grid (Aghaei and Alizadeh 2013), should ensure a higher efficiency and reliability through automated control, high-power converters, novel communication infrastructures, sensing and metering technologies, sophisticated energy management techniques, renewable energy, and network availability (Wiboonrat 2012; Cecati et al. 2010; Panajotovic et al. 2011; Fang et al. 2012; Güngör et al. 2011). Data centers are completely immerse in this grid context as a player and a typical architecture of them under smart grid environment is presented in Fig. 1, where the flow of relationships between service requests and power supply can be identified. On the one hand, data centers can take advantage of the flexibility of their loads to implement load management strategies aiming at reducing the operation costs and, on the other hand, play an important role to ensure the efficient and reliable operation of electrical grids by providing demand response (DR) services.

This possibility is corroborated by the National Institute of Standards and Technology (NIST) and the Department of Energy (DoE) in the USA, which have both identified demand response as one of the precedence areas for the future smart grid and furthermore, the National Assessment of Demand Response Potential report has identified DR as the potential to reduce up to 20% of the total peak electricity demand in the USA (Liu et al. 2013). The European model of smart grid (SGAM) addresses demand response focused on aggregators and applied to flexible loads, such as data centers (Andrén et al. 2016). The European Network of Transmission System Operators for Electricity (ENTSO-E) has also quantified a load reduction potential of about 11 GW available throughout continental Europe (Gils 2014).

Nevertheless, the considerable majority of studies about energy-related issues, i.e., power usage management, server consolidation, load management, or demand response programs, does not consider the dimensions of data centers and among those which consider it, nearly all are focused on large, and just a few on small and medium ones, which account for more than 50% of the total electricity consumption (Josh and Delforge 2014). Many organizations, such as laboratories, research institutes, universities, industries, and enterprises have multiple small and medium data centers scattered around their facilities. In fact, surveys indicate that this data center profile waste more energy than larger facilities, whereby the power consumption is often overlooked, because the energy cost of an individual data center usually accounts for just a small portion of total spending (Josh and Delforge 2014; Delforge 2014; Bennett and Delforge 2012).

Based on this wide and currently heterogeneous scenario, the main purpose of this paper is bringing light to the above-mentioned issues through a comprehensive review in energy efficiency and demand response, providing the state-of-the-art perspective and interconnection to small and medium data centers.

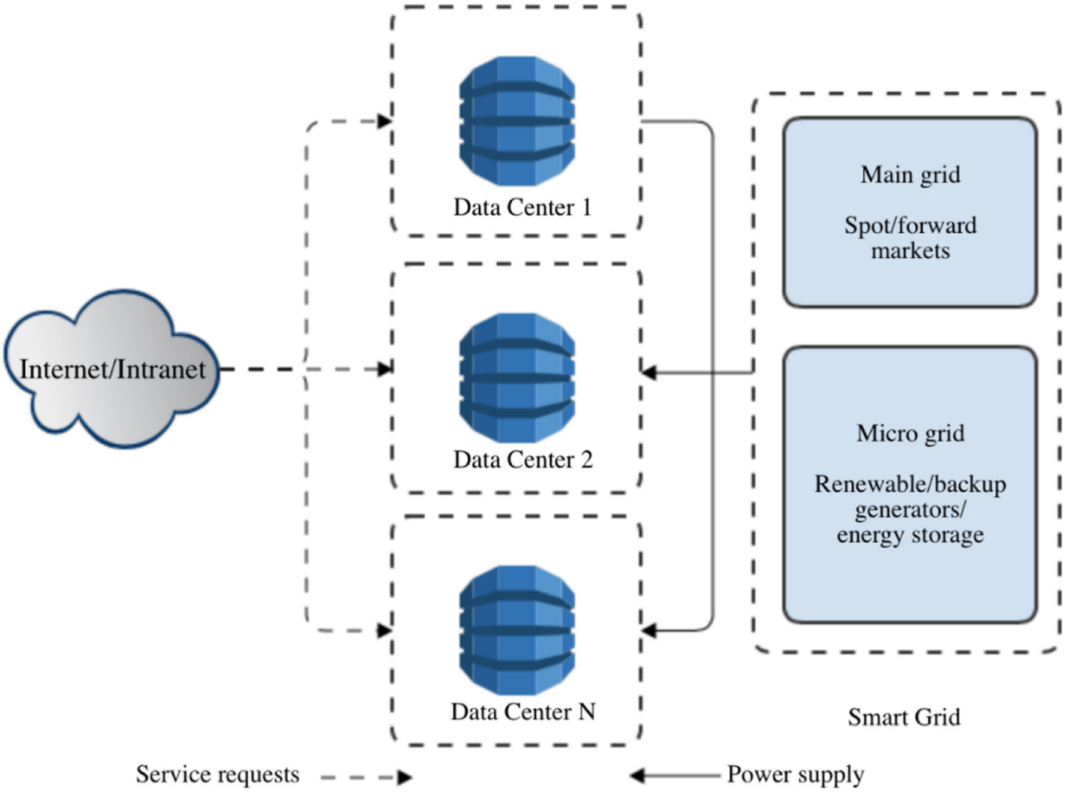


Fig. 1 A typical architecture of multiple data centers

Several studies have conducted extensive literature reviews whose focus was specifically on data centers, energy efficiency, demand response, or related issues. Ebrahimi et al. (2014), Fulpagare and Bhargav (2015) and Zhang et al. (2014) reviewed aspects related to the main cooling solutions used in data centers, as well aspects related with waste heat recovery and advances in thermal management. Energy efficiency in networks, telecom systems, power-efficient algorithms, and server consolidation were addressed, respectively, in Hammadi and Mhamdi (2014), Garimella et al. (2013), Uddin et al. (2015), and Ahmad et al. (2015), nevertheless in a cloud data center context. Concerning demand response, there are a few broad reviews such as Oconnell et al. (2014) and (Paterakis et al. 2017); however, they are not always focused on data centers specifically, mainly dealing with specific strategies.

Therefore, the present literature review arises from the initiative to relevantly increment and interconnect all the assumptions discussed hitherto separately, understanding that the panorama of this paper should be aligned with the future of data centers, where energy efficiency and demand response should go hand in hand taking advantage of all joint potential from a technical (grid services and reliability) and economic (costs minimization) point of view.

The approach used in this paper to achieve this goal is dismembering server components generically and analyzing their energy consumption profile and the energy efficiency strategies applied nowadays. By the same criterion, virtualization, cooling, UPS, and energy management will be analyzed, as well as their demand response strategies and related renewable sources integration. Firstly, this systematic analysis will be performed globally, considering data centers as a whole and finally, a small and medium data center perspective will be given based on recent studies, implementations, and respective adaptions.

The remainder of the paper is structured as follows: "Data centers overview" presents an overview on data centers. A review of energy efficiency strategies, as well as the associated energy savings is presented in "Energy efficiency strategies." In "Demand response in data centers," a perspective on demand response, as well as its challenges and overcoming are analyzed. Then, the specific case of small and medium data centers, as well as their impact in the previous areas are presented in "Small and medium data centers perspective." Finally, "Conclusions" summarizes the paper, emphasizing its main conclusions.

Data centers overview

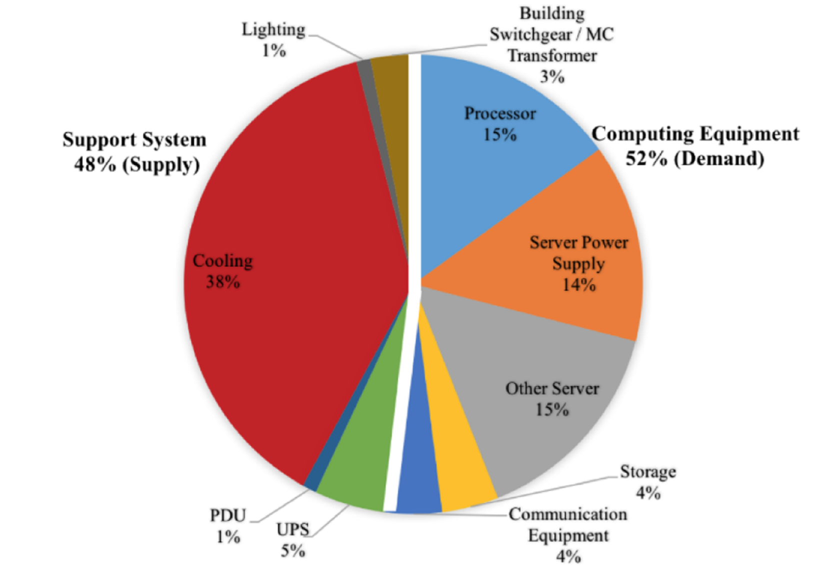
Pierson (2015) address data centers as being structures, or group of structures, dedicated to the centralized accommodation, interconnection, and operation of ICT and network telecommunications equipment, providing data storage, processing, and transport services, along with all the support facilities for high-quality power supply and environmental control with the levels of resilience and security required to provide the desired service availability.

This particular infrastructure is divided in three spaces: ICT room, data center support area, and ancillary spaces (Rong et al. 2016). The ICT room is an environmentally controlled space that houses equipment and cabling directly related to computer and telecommunications systems which generate considerable amounts of heat. Moreover, the ICT equipment is highly sensitive to temperature and humidity fluctuations, so a data center must keep restricted environmental conditions for assuring the integrity and functionality of its hosted equipment. Data centers support areas are all those where different systems, such as the uninterruptible power supply (UPS) systems, cooling control system, and switch boards are located. Finally, the ancillary spaces include mainly offices, lobby, and restrooms (Grice et al. 2013; Oró et al. 2015).

Based on Ghatikar et al. (2010) data centers can include more than 100,000 hardware devices and the electrical load can range from about 1 kW to about 100 MW with different sizes and profiles. A typical example in terms of demand and supply, as well as the power draw unbundling of a medium data center is categorized in Fig. 2.

In the context of this huge load spectrum and some particularities, according to Sheppy et al. (2011), data centers load profile is almost steady, normally, and up to 76% of existing facilities are oversized and therefore inefficient. In addition, it is estimated that up to 58% of energy is wasted in unnecessary and inefficient components such as chips, slots, fans, voltage regulators, power supplies, and many servers (International Energy Agency 2014).

Data centers fall into two general categories: internal and external. Internal data centers, very often also called enterprise data centers, are dedicated to the needs of the organization that operates them and typically serve one of two main functions: production or research and development. External data centers provide services to companies that have outsourced some or all of their ICT functions. Both can still be subdivided into small-,



Category	Power Draw
Computing	588 kW
Lighting	10 kW
UPS and distribution losses	72 kW
Cooling power draw for computing and UPS losses	429 kW
Building switchgear/MV Transformer/other losses	28 kW
Total	1,127 kW

Fig. 2 Analysis of a typical 465 m^2 data center (Emerson 2015)

medium-, and large-sized profiles, depending on their purpose, mission, and financial resources. Salom et al. (2017) present a way to express the dimension of data centers by using ICT power capacity, with the following breakdown; however, Whitney and Delforge (2014) also claim that the small- and medium-sized organization category comprises four data center types: server closet, server room, localized, and mid-tier as summarized in Table 1. In this paper, a combination of these two approaches will be considered, where server closet, server room, very small, small, and localized are examples of small data centers. The medium data center is common in the two classification methodologies.

Energy efficiency strategies

The computational management with emphasis on energy efficiency was initially applied in the context of mobile devices powered by batteries. In such devices, energy consumption should be minimized in order to increase the battery lifetime (Beloglazov and Buyya 2013). Although servers with their different generalizations (e.g., blade, tower, and rackable) and data centers can use techniques developed for mobile devices, these systems require specific methods. Thus, the power consumption in data centers is affected by two main factors. The first one, from the hardware point of view, is divided in two points: one is caused by the amount of active computational resources and the other is the energy efficiency of the physical components. A way of dealing with the energy efficiency of physical components is the use of a power management system to keep their operation in proportion to the demand for use by applications. This has been done through solutions in hardware and firmware, which will be analyzed in the following sections.

Another factor in the problem of energy consumption, handled from a software point of view, is the inefficiency in the use of computational resources. A study of more than 5000 production servers over a 6-month period showed that even when they are not idle, most of the time, the server utilization is between 10 and 50% of its work capacity, generating heat and unnecessarily consuming energy (Barroso and Hölzle 2007). The existence of a set of computational resources much higher than the average use is justified by the need to deal with peak loads. Although this peak scenario occurs with a low frequency, it is necessary to ensure that performance is not adversely affected, which would happen if an application was executed on an overloaded server.

Judge et al. (2008) found that, even when standing idle, servers consume about 70% of the energy consumed during peak working hours. However, according to the data obtained from the SPEC power benchmark (SPEC 2016), the server configurations designed in the end of 2015 consume about 15 to 34% of the energy when idle. Despite the significant reduction of consumption, primarily due to the development of more efficient architectures, maintaining a server connected with a low level of usage is still highly inefficient from the energy consumption point of view.

Based on the highlighted context, it is fundamental to analyze the role of the servers and every supportive technological environment that surrounds it under the energy efficiency panorama in a detailed way, dismembering each component, and pointing out the contemporary energy efficiency strategies used in each of them, as well as potential savings divided by CPU, memory, disk, network interface card, and the impact of virtualization, cooling, and UPS.

CPU

The processor is the central part of servers and as claimed by Wang et al. (2017a), typically, CPU has been the largest, yet not prevailing, contributor to the power consumption, as characterized in Fig. 3. In order to provide extra performance when necessary, traditional CPUs are equipped with additional procedures, whose purpose is to minimize the active and static power consumption in an energy efficiency procedure (Varrette et al. 2015).

Dynamic voltage and frequency scaling (DVFS) is conceptualized by Sueur and Heiser (2010) as being a commonly used power-management technique where the clock frequency of a processor is reduced to allow a related decreasing in the supply voltage intending to establish an energy efficiency relationship, where the reduction of energy consumption is befitting with the processed workload. The reduction of power consumption leads to a meaningful decreasing in the energy requested for computation, specifically for memory-bound workloads. However, Lu et al. (2016) emphasizes that the lowest operating frequency is constrained

Table 1 Typical characteristics of data center space types (Masanet et al. 2011)

Space type	Typical size (m^2)	ICT power capacity	Typical ICT features	Typical infrastructure system characteristics
Server closet	< 19	< 50 kW	1–2 servers No external storage	Typically conditioned through an office heating, ventilation, and air-conditioning (HVAC) system. Environmental conditions are not as tightly maintained as for other data center types. HVAC energy efficiency associated with server closets is probably similar to the efficiency of office HVAC systems
Server room	< 47	50–250 kW	A few to dozens of servers No external storage	Typically conditioned through an office HVAC system, with additional cooling capacity, probably in the form of a split system specifically designed to condition the room. The cooling system and power backup equipment are typically of average or low efficiency because there is no economy of scale to make efficient systems more first-cost competitive
Localized data center	< 93	250–1000 kW	Dozens to hundreds of servers Moderate external storage	Typically use under-floor or overhead air distribution systems and a few in-room air-conditioning (AC) units. AC units in localized data centers are more likely to be air cooled and have constant-speed fans and are thus relatively low efficiency. Operational staff is likely to be minimal, which makes it likely that equipment orientation and airflow management are not optimized. Air temperature and humidity are tightly monitored. However, power and cooling redundancy may reduce overall system efficiency
Mid-tier data center	< 465	1–2 MW	Hundreds of servers Extensive external storage	Typically use under-floor air distribution and in-room AC units. The larger size of the center relative to those listed above increases the probability that efficient cooling, e.g., a central chilled water plant and central air handling units with variable speed fans, is used. Staff at this size data center may be aware of equipment orientation and airflow management best practices. However, power and cooling redundancy may reduce overall system efficiency
Enterprise-class data center	> 465	> 2 MW	Hundreds to thousands of servers Extensive external storage	The most efficient equipment is expected to be found in these large data centers. Along with efficient cooling, these data centers may have energy management systems. Equipment orientation and airflow management best practices are most likely implemented. However, enterprise-class data centers are designed with maximum redundancy, which can reduce the benefits gained from the operational and technological efficiency measures

by the stable voltage conditions of the circuit. The power consumption in watts by DVFS at a frequency f is given by Eq. (1).

$$P(f) = CN_{\mathrm{sw}}V_{\mathrm{dd}}^{2}f \tag{1}$$

where C is the capacitance of the circuit, a significant percentage of which is wire-associated, N_{sw} is the average number of circuit switches per clock cycle, and V_{dd} is the supply voltage of the CPU (Zhuravlev et al. 2013). As the maximum frequency is linearly subjected to the supply voltage, DVFS has a cubic effect on the power savings in watts, as given by Eq. (2) (Elnozahy et al. 2003).

$$P(f) = c_0 + c_1f^3 \tag{2}$$

where, c_0 is a constant that includes the power consumption of all components except the CPU, plus the base power consumption of the CPU. DVFS minimizes the power requirements, but maximizes the application execution time or requests response time, whereas the DVFS total energy efficiency scheme concerns the fact that power reduction is a cubic effect of frequency at the cost of rise in the execution time. When considering the delay of sensitive applications (e.g., Internet services), it is fundamental to maintain the response time within certain thresholds. Furthermore, DVFS increases the overall execution time of the task in delay-tolerant applications. As an outcome, there is a penalty of using the growth in the execution time even when instantaneous power savings are achieved. Nevertheless, to increase the energy savings without exceeding the execution deadline the energy delay trade-off can be used (Arianyan et al. 2017).

In this context, there is a linear relationship between frequency and execution time. The execution time of a task instance, denoted by $C_{\mathrm{i,\,j}}$, is inversely proportional to the frequency at which the processor executes. If X is the processor frequency, then its execution time $C_{\mathrm{i,\,j}}$ is given by Eq. (3).

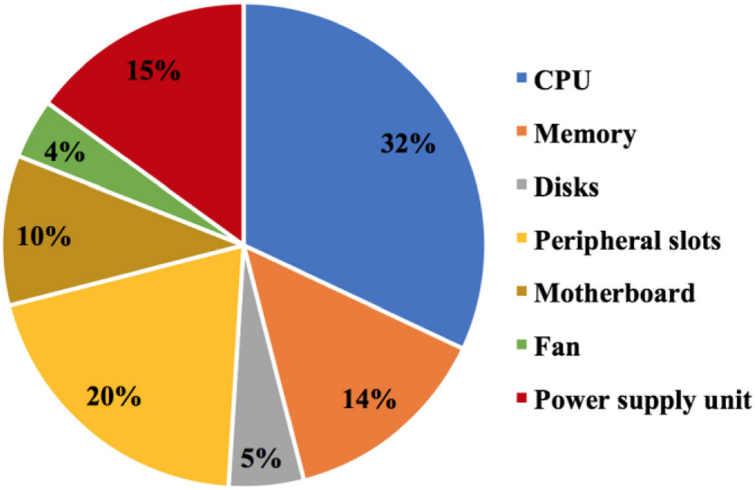
$$C_{\mathrm{i,j}} = \frac{C_{\mathrm{i,j}}^{\max}}{X} \times F_{\max} \tag{3}$$

where, $C_{\mathrm{i,j}}^{\max}$ is the execution time obtained at $F_{\max}$ (Thekkilakattil et al. 2010).

Other CPU procedures to provide energy efficiency are the low-power sleep modes and dynamic power switching. The former, also known by core power gating (Johannah et al. 2017), is utilized when the CPU is idle. By cutting the clock signal and power from idle units, the CPU might be commanded to enter in a low-power mode providing energy savings. Dynamic power switching technique enables the power management processing in all the domains and constantly monitoring to switch its state to a lower power mode when required.

A different approach was conducted by Krzywda et al. (2017) in which several actuators were analyzed jointly to optimize data center servers: DVFS and CPU pinning, which defines the set of CPU cores that each thread can run, were tested and results show that DVFS rarely reduces the power consumption of underloaded servers by more than 5%, but it can be used to limit the maximal power consumption of a saturated server by up to 20% (at a cost of performance degradation). CPU pinning reduces the power consumption of underloaded

Fig. 3 Breakdown of power consumption in servers (Emerson 2015)



server (by up to 7%) at the cost of performance degradation, which can be limited by choosing an appropriate CPU pinning scheme.

Memory

Another important component is the local memory, which exists at different levels. The main memory (dynamic random access memory (DRAM)) is responsible for a significant fraction of a server's power consumption. However, memory with various power states have been developed by a few major organizations. Therefore, any memory power management should assure the performance of memory if DRAM low power states are present.

The main memory contains *static* and *dynamic* energy consumption. In the decoding of addresses and fetching the data from the memory, dynamic energy is consumed. The static energy is consumed during the active period repaid over the number of data transfers. If E_{rw} is the energy per read or write, $\mathrm{BW}_{\mathrm{rw}}$ is the read or write bandwidth, D are the total DRAM channels, E_{AP} is the energy required to activate and pre-charge, and f_{AP} is their frequency then the energy consumption for each DRAM channel, in kilowatt hour, is given by Eq. (4) (Ahn et al. 2009).

$$E_{DRAM} = StaticEnergy + E_{rw}BW_{rw} + DE_{AP}f_{AP} \tag{4}$$

The time needed to regain data from the main memory impacts the memory-based applications performance. To collect data from the main memory, the probability of hit and miss of the previous level of memory, i.e., caches, can be perceived. If the hit probability is p_{hit}, the miss probability is given by Eq. (5). The time, in seconds, needed to obtain data from main memory with one level of cache, taking into account the access time t_{access} and miss time t_{miss}, is given by Eq. (6).

$$p_{miss} = 1 - p_{hit} \tag{5}$$

$$t_{DRAM} = p_{hit}t_{access} + p_{miss}t_{miss} \tag{6}$$

Based on several studies, presented by Pore et al. (2015), the main technologies used in memory to promote a more efficient use of energy are:

- Memory architecture modifications. Dividing the memory into ranks and making use of smaller subsets of memory rather than the entire memory, results into activation savings and pre-charge energy associated with the rank subsets that are not accessed. Nevertheless, the prompt impact of this technique is the data path for each access becoming longer. The design of memory schemes incorporates different factors such as load balancing across memory ranks, number of memory ranks impacting the effective bandwidth, as well as the application features. Several other methods of power savings comprise managing the refresh rates of memory, use of memory buffer, etc.
- Memory low power modes. Currently, new memory sorts have more power states, i.e., RDRAM (Rambus DRAM) establishes four different power states: active, standby, nap, and shutdown. Power management patterns for the memory utilize these states to minimize the energy consumption.

– Static power management. The memory is assigned to a low-power state and when the memory access takes place, the chip has to resume to the active power state.

– Dynamic power management. The low-power state time interval is varied according to the access pattern. The limit time interval after which the memory is in low power state is a crucial design aspect of the power management. This limit is given for improving energy savings; nonetheless, the delay is within the time limitation of the application.

Taking into account the SDRAM (synchronous DRAM) memory, meaning numerous types of DRAM synchronized with the clock speed optimized by the microprocessor and increasing the number of instructions that the processor can perform in a given time, the current generation, Double Data Rate 4 (DDR4) differs from its predecessor, DDR3, due to a 20% decrease in energy consumption (Kim 2016). Whereas DDR3 normally demands 1.5 V of electrical power, DDR4 demands as little as 1.2 V. In data centers implementing servers running as much as a terabyte of memory on a 24/7 profile, associated with onboard fans and external ventilation systems as cooling solutions, upgrading to DDR4 means big return of investment in the form of energy savings (Islam et al. 2015). Furthermore, development of specifications for the new next generation, DDR5 SDRAM, has started, which will be two times faster than DDR4, having double the density, twice the gigabyte capacity, and also more power efficiency (JEDEC 2017).

Disk

According to Dayarathna et al. (2016) and Tang et al. (2017), the majority of the storage disk have transition to on-off power states capability. They are either in idle, standby, or off states when the disks are not in use. Taking into account that d_n is the total number of memory fetched in the storage disk, the active state power, P_{active}, is proportional to d_n by a constant factor d, P_{standby} is the power during disk I/O in the low-power state, t_{active} is time spent in active state and t_{standby} is time spent in the low-power state, then the energy consumption in kilowatt-hour is given by Eq. (7).

$$E_{disk} = dP_{active}t_{active} + P_{standby}t_{standby} \tag{7}$$

The time requested in seconds to obtain data from disk is essential in the disk design of power management patterns. If t_{seek} is seek time, t_{RL} is the rotational latency, and t_{tt} is the transfer time from disk to higher-level cache, then the time requested to obtain data from disk is given by Eq. (8).

$$t_{disk} = d_n(t_{seek} + t_{RL} + t_{tt}) \tag{8}$$

Based on several studies demonstrated by Pore et al. (2015), the main technologies used in disk to promote more efficient use of energy are:

- Disk spinning down. Spinning down is the best-known procedure of power management in disks (i.e., switching the power off) when not in utilization (Verma et al. 2010). Nevertheless, to avoid the delay-sensitive applications performance degradation, since the time to re-establish the disk to the active state takes a few seconds with sudden fluctuations in the data center workload, prediction-based techniques are used. In addition, to schedule the disk spinning down in the idle timeframe in the workloads, it is possible the performance breakdown of the applications due to power state transitions.
- Consolidation, aggregation, and compression. Moving data to a specific number of storage devices using a scalable architecture (Tsirogiannis et al. 2010), maximizing the idle times among the operations by postponing read or writes (Gupta and Singh 2007) and to use data compression in some workload cases (Gupta and Singh 2007) are jointly an effective way to provide more opportunities for energy savings.
- Managing data storage and replication. There is often a large data set stored in multiple storage disks in data center applications, involving popular data, which are more frequently accessed than the remaining data. Identifying the most requested data, storing on fewer disks, and replicating them for performance, whereas the rest of data is stored on remaining disks is the role of this technique. The disks with most requested data are always in active state while more power management patterns are used to the remaining disks. Other power management techniques involve the usage of hybrid disk types such as a combination of solid-state drives (SSD) (Tsirogiannis et al. 2010), flash storage devices (Härder et al. 2011), and dynamic random access memory (DRAM) (Deng 2011) to control the data storage based on the combination of their power, performance features, and costs. The most requested data is migrated to more energy-efficient devices, but moving the data frequently might exceed the savings obtained by a spinning down of disks (Amur et al. 2010).

Thus, to unite power-saving techniques and identify several conflicts among their use, a wide literature review was performed by Bostoen et al. (2013) and several studies related with these technologies have been done (Zakarya and Gillam 2017) concluding that an ideally proportional system has an energy reduction potential of 40–75% involving disks and storages appliances.

Networking Interface

Chen et al. (2016a) highlights that the average use in the data centers is very low and idle network devices such as ports, line cards, and switches, are one of the considerable consumers of energy in low-utilization periods (Gupta and Singh 2007). Therefore, a power management procedure that is frequently used for network devices is to turn off the network components during idle timeframes (Gupta and Singh 2003). When the network components are not in use, they are either in idle or standby states. If P_{active} is the active state power and P_{standby} is the power in the standby state, t_{active} is the time spent in the active state and t_{standby} is the idle period, E_s is the energy needed for switching between the power states and n_s is the number of switching that occurred, then the energy consumption in kilowatt-hour is given by Eq. (9).

$$E_{net} = P_{active}t_{active} + P_{standby}t_{standby} + n_S \tag{9}$$

The time, in seconds, needed to transfer the data through a network component in terms of the transfer time, t_{transfer}, and switching time, $t_{\text{switching}}$, is given by Eq. (10).

$$t_{net} = t_{transfer} + t_{switching} \tag{10}$$

Network cards are generally off-board in the main data center servers and therefore are included in the estimation of 20% of the peripheral slots. The main techniques to promote an efficiency use of energy associated with networking interfaces are (Pore et al. 2015):

- Switching off the network component. Use of a reactive scheme that switches off the network component for a certain time after observing that there is no workload for a few seconds. Some procedures imply proactive patterns where network interfaces are continually monitored for learning the inter-arrival time between packets in a window-based method.
- Managing the workload. Data can be aggregated, stored in buffers for some period, and sent if the application deadlines are not stringent, enabling the network components to be turned off during the idle period.
- Sleep. The network components, such as switches and routers, are in sleeping mode or turned off in the idle timeframe between the workload arrivals, reaching energy savings between 10 and 20% (Gupta and Singh 2007; Nedevschi et al. 2008).
- Aggregation. The network topology is modified to consolidate the network flow on fewest possible routes, such that the data is sent on minimum active series of network devices. Bonetto et al. (2014) present that even simple policies allows to save from 30 to 50%.
- Rate adaptation. With this technique, the workload rate is adjusted such that traffic is serviced within the required time constraints, achieving energy savings between 10 and 90% (Gupta and Singh 2007).
- Traffic shaping. The traffic is divided into bursts, in the elastic tree procedure. This traffic to same destinations is buffered before it is routed. This scheme increases the idle periods between the traffic bursts applied to transition the network devices into low-power states and can save up to 50% of network energy, while maintaining the ability to handle traffic surges (Heller et al. 2010).

Virtualization framework

Virtualization is an increasing and leading technology to mutualize the energy required by a single server operating multiple virtual machines (VMs) instances. Nevertheless, short consensus has been produced about the capacity overhead in energy consumption and the throughput minimization for virtualized servers and/or computing components. Other way to address this topic is conceptualized by Mazumdar and Pranzo (2017), where virtualization helps to reduce the power consumption within a cloud infrastructure by enabling consolidation of heterogeneous applications in few active servers.

The main structure of any virtualization platform and thereby Cloud middleware, remains the hypervisor or virtual machine manager. Thereafter, a VM operation managed by a hypervisor is called a guest machine. The main two categories of hypervisors are native and hosted, but only the former (also called bare-metal) contains an interest in data center context. This category of hypervisor operates straight on the host's hardware to take over the hardware and to supervise guest operating systems. Nevertheless, from the energy efficiency point of view, to implement efficient virtualization mechanisms, it is required to define the allocation of a VM to a physical machine and live VM migration during overburdening situations.

It is important to take into account a virtualized data center with servers that host a sub-assembly of applications by supplying a virtual machine for each application hosted upon it. Hence, an application might have numerous tiers and there might be multiple instances of each tier operating across different VMs in the data center, as it can be seen in Fig. 4.

The VM management determines how many VMs are necessary for each tier of an application, how the workload should be propagated among multiple instances of the application for each tier, how the

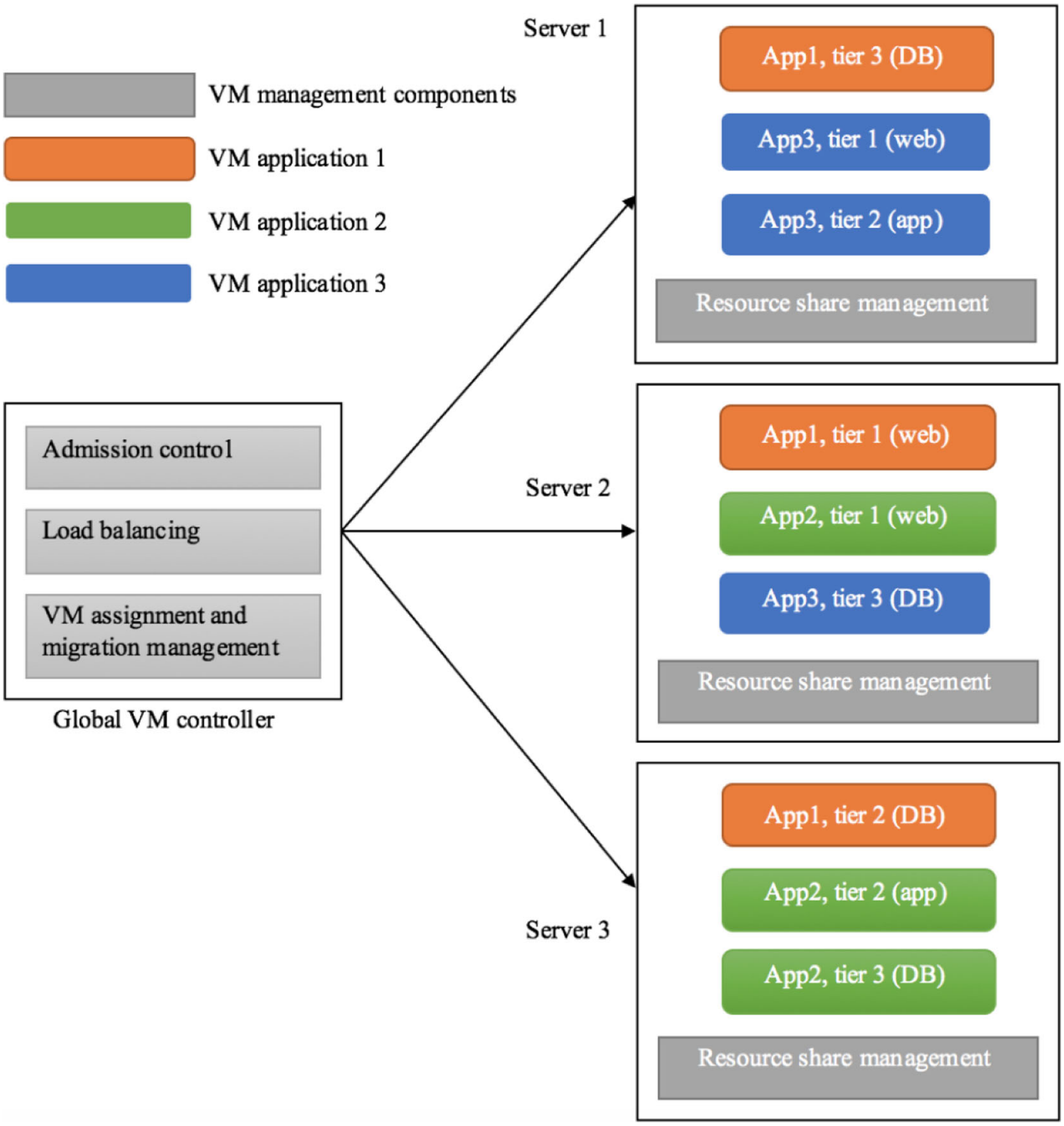


Fig. 4 A pictorial example of VM management components (Pore et al. 2015)

resources should be held in common among VMs that are collocated in a VM, when and how VM migration should be conducted, and how the resources should be held in common among collocated VMs.

Several works have been addressing the problem of VM migration. Dhiman et al. (2010) highlights the variations in the power consumption for CPU- and I/O-intensive applications through trials, where different applications have different characteristic specifications at different times and thereby VM allocation in physical server can be reached when power peak does not occur in parallel. It is stated that placing a hybrid of CPU- and I/O-intensive application in a physical server produces less power consumption compared to any other arrangement. An implementation of the system on a state-of-the-art testbed of server machines called vGreen, an open control loop to manage the application assignments (VMs) to the physical server, is presented maximizing average performance and system-level energy savings by around 40%, using benchmarks with different specificities (Dhiman et al. 2010).

From the *VM Dynamic Resource Allocation* point of view, to make the services in the form of VM applications available, the cloud presents autonomous management of the available physical resources. Two techniques are available for autonomic resource allocation to VMs: Static resource allocation (Hermenier et al. 2009), and dynamic resource allocation (Wang and Wang 2011; Padala et al. 2009; Nathuji and Schwan 2007). The static resource allocation model assumes that the resource demand of VMs are known in advance and the VMs' resource demands does not modify significantly during their life and the VM allocation is carried out in accordance with the peak resource requirements of the application. The pattern discussed in the literature for dynamic resource allocation interval analyzes how to optimize a utility model that catch service-level agreement (SLA) revenue cost as well as energy cost (Ardagna et al. 2012; Urgaonkar et al. 2008; Padala et al. 2009), machine learning techniques for learning resource requirement of applications (Tesauro et al. 2006), and control theory methodologies (Wang and Wang 2011).

Specifically on energy efficiency measures, Castro et al. (2013) propose three new approaches for dynamic consolidation of VMs that take into account both CPU and RAM usage. A heuristic called CPU and RAM Energy Aware (CREW), which uses an energy model that jointly considers the consumption of CPU usage and RAM was proposed to define the allocation of VMs, ensuring the lowest possible power consumption. The implementation and evaluation of tenders done in the CloudSim (Calheiros et al. 2009) simulator used real workload VMs from the PlanetLab (Spring et al. 2006) and an enterprise cluster Google (Reiss et al. 2012). The results showed a reduction on the energy consumption in up to 33% and an increase on the QoS guarantee.

Based on this context, there are still demands in the VM management, given that VM migration overhead deteriorate not only the performance of the migrated VM; nevertheless, also the performance of the VMs collocated in the source and destination physical machine (Lim et al. 2011). Constant cost per each migration is the base for many studies (Sanders et al. 2004), using migration time, which affects many elements as follows: (1) the memory update rate and the memory content of each virtual machine, (2) the virtual machines migrated total number, (3) network bandwidth availability for migration, and (4) the destination servers at the time of migration and the workload of the source (Dargie 2014). Lastly, virtualization approach simplifies dynamic power management and minimize power consumption; however, the applicability of the virtualization under various situations such as real time is not well researched. This is essential, considering that the VM overhead, the delay demands of some applications in an underused VM might not be satisfied.

Uninterrupted power supply

UPS are key components of ICT systems, ensuring reliability by maintaining the continuity and quality of the systems' power supply. A UPS is understood to be a short-duration (minutes to hours) power supply system that maintains the functions of the connected load when the main continuous power source has failed or has significant disturbances (IEC 2013). Therefore, the primary purpose of a UPS is to bridge an unexpected power gap and/or to provide the amount of power needed to safely power down the connected load. A UPS may also be used to continuously maintain the quality (e.g., harmonic content) and stability (voltage and frequency) of the power to the connected load. In data centers, the UPS systems are used to ensure the service continuity of ICT, to protect it from risk of halts in data processing, contributing to 7% of the total energy consumption (Moura et al. 2016).

The energy consumption of UPS should be an important consideration due to its high impact on the lifecycle costs; nevertheless, in most applications of UPS, energy efficiency is not the most important issue since the operational reliability of the ICT systems and the related security of data processing and storage are the major concerns. However, the conversion efficiency of UPS systems has been improving in recent years and large energy savings can be achieved with the adoption of new technologies without a reduction of the reliability levels.

Regarding UPS life cycle, Khan and Khan (2015) present the amount of power that can be stored or retrieved from the batteries, at a given time *t*, of the UPS, which are limited by their maximum amounts. The lifetime of the UPS is constrained by the number of cycles of UPS charging and discharging (Wang et al. 2012) and therefore, the operating cost of the UPS also depends upon UPS charging and discharging cycles.

Moura et al. (2016) analyzes the UPS efficiency clamming that product performance reliability and system configurations with high redundancy often conflict with optimized life cycle costs. At a given level of supply security, these costs are an important consideration for the user.

In UPSs, a decrease in the consumed energy of the product and their installation system architecture leads not only to a direct decrease in the UPS energy costs but also produces cooler operating conditions within the installation environment. This leads to a reduction in ventilating and air-conditioning energy and infrastructure installation costs, an extension of the service life of UPS key components (e.g., such as energy storage batteries and capacitors) and an increase in the overall lifetime reliability of the UPS system.

The key factors that must be considered regarding energy efficiency are the size of the UPS, load type, and load level. Larger UPS modules typically have higher energy efficiency than smaller ones because the power required for control electronics and auxiliary components becomes a smaller portion of the total capacity of the UPS system. The efficiency of a UPS depends on the load level, achieving the highest efficiency with a 100% load, as depicted in Fig. 5. However, the curve is relatively flat with load levels higher than 50%. A UPS operating with a low load level will have significant losses when compared with the same UPS operating at full load. In a realistic scenario, the load level is typically between 10 and 30%, which leads to a 4–17% reduction of efficiency (Moura et al. 2016).

The load type also has a strong influence on the achieved efficiency. UPS efficiency is usually tested with resistive or linear loads, but several UPSs are used with non-linear loads, with low-power factor, and high total harmonic distortion (Pier 2008). The low-power factor will require a higher peak current from the UPS, decreasing its efficiency (Fig. 6).

Moura et al. (2016) assess the potential savings considering several technologies at UPS' component and product level, modeling the main design options and considering policy options focused on minimum efficiency performance standards and energy labelling. The results show a potential for energy savings in the European Union in 2025 of 11.4 TWh (65% energy saving relative to predicted energy requirement of EU ICT system UPS based on current practice).

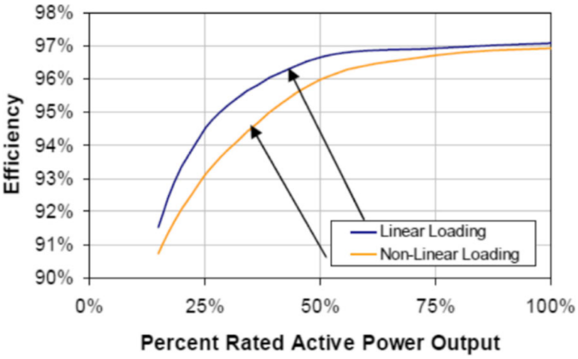


Fig. 6 UPS efficiency with linear and non-linear loads (Pier 2008)

Another alternative that can be considered concerning energy efficiency related to UPS in accordance with Boulos et al. (2014) is the direct current (DC) distribution, since all data centers ICT equipment such as servers and storage devices are essentially DC-based loads. The backup supply system commonly required for critical facilities consists of batteries, which are also based on DC. Thus, by deploying a DC distribution rather than conventional AC, several conversion steps in the power delivery system can be eliminated, reducing distribution losses. By using DC entirely throughout, a data center will save 10 to 20% in power costs and improve reliability (Sithimolada and Sauer 2010).

As seen in Table 2, DC-DC converters can reach an efficiency of 90–96% as compared to AC-DC power supplies which provide an efficiency of 65–75%. Even comparing best-in-class AC-DC to DC-DC, a 2–5% advantage to the DC-DC solution can be reached. The efficiency findings show that 380 V DC provides the highest efficiency DC option, particularly when compared with the 48 V DC system; however, it requires a

Fig. 5 Typical UPS efficiency curve (Moura et al. 2016)

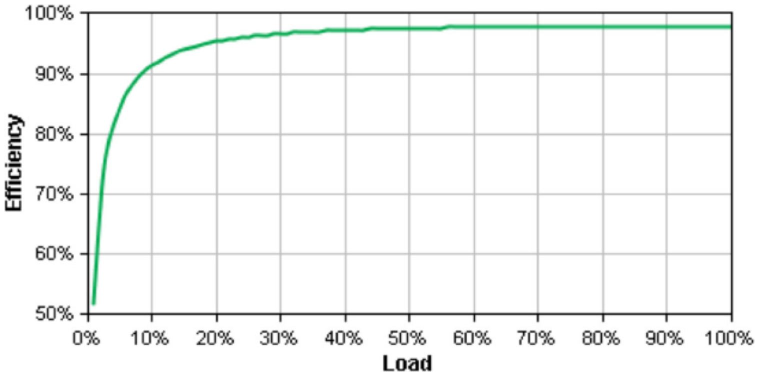


Table 2 Power distribution efficiency comparing AC and different DC distribution methods (Moreno-Munoz et al. 2011)

	UPS (%)	Distribution wiring + power distribution unit (PDU) (%)	Power supply unit (PSU) (%)	Load converter 12 V–1 V (%)	Total efficiency (%)
Facility AC-UPS	92.00	99.00	75.00	88.00	60.00
Facility DC-UPS 48 V/24 V	92.86	99.00	91.54	88.00	74.00
Facility DC-UPS 380 V	96.00	99.00	91.75	88.00	76.73
Distributed DC-UPS	92.00	99.00	94.00	88.00	75.34

critical mass of 380 V DC commercial equipment to exist in the building before any user could decide for this option (Moreno-Munoz et al. 2011).

Cooling

One of the important elements to manage a data center reaching the expected performance is the cooling system. Depending on their power consumption, server systems need a specific amount of cool air at the intake and exhausts the same volume of heated air at the outtake. In such situation that the room is not able to supply this amount of air, the server will draw in its own exhaust air, overheating the device. Thus, a proper cooling approach is inevitable for an uninterruptable server functioning, however Ni and Bai (2017) claim that more than half of the data centers' air-conditioning systems are inefficient. Thereby, in order to increase the energy efficiency, it is fundamental that many factors must be taking into account before adopting a cooling solution, such as energy use, facility location, density per rack, power density, and other user specificities. One practical example of this heterogeneity it that the power used by single racks might vary dramatically, with an average of around 1.7 up to 20 kW in high-density servers (Varrette et al. 2015), directly affecting the adopted cooling solution.

There are three main cooling architectures used in data centers, such as room-, row- and rack-oriented. In the first scenario, computer room air conditioners (CRAC) units are linked with the room (Fulpagare and Bhargav 2015) and cool air might be unrestricted or partly limited by ducts or vents when provided by the conditioners. Because the air supply uniformity is poor due to specific room designs, such as their shape or obstructions, the full rated capacity of the CRAC unit cannot be used in most situations. In row-oriented solutions, CRAC units are linked with a row, being their performance higher, as the airflow paths are shorter. Consequently, the requested CRAC fan power supply is smaller, decreasing the energy cost. In the last scenario, rack-oriented, CRAC units are linked with the rack allowing the cooling to be accurately adapted to the constraints of servers. On the other hand, the disadvantage is requiring a large amount of air-conditioning equipment (Sohel Murshed and Nieto de Castro 2017).

Cooling systems mainly utilize air or a refrigerant and depending on the kind of cooling systems, different operation temperatures will be required. In addition, depending on the different amounts of heat exchanger processes, different elements will compose the cooling system, with different efficiency or losses (vor dem Berge et al. 2015). In order to reach energy-efficient data center cooling, numerous solutions have been presented, being the three most broadly adopted and effective: hot- and cold-aisle isolation, closed-coupled cooling, and free cooling.

Considering the hot- and cold-aisle isolation, through a raised floor, a steel grid resting on stanchions installed 60–120 cm on the concrete floor, the cold air is supplied by CRAC units. By perforated tiles, the cold air will flow into racks and then, hot air will be exhausted through a rear side of rack after absorbing heat produced by servers in the rack. One strategy to maximize the cooling efficiency is to avoid mixing the cold air supplied by CRAC and hot air exhausted by servers. This is guaranteed by a solution named hot- and cold-aisle isolation, arranging server racks such that the intakes of cold air in server racks are faced each other. The hot air is eventually drawn by the CRAC and the cold air is once more supplied to cold aisles by exchanging the heat with cold air (or water) delivered from chillers. This solution can achieve up to 40% energy savings according to Kim et al. (2015).

On the other hand, according to Alkharabsheh et al. (2015) and Capozzoli et al. (2015), there are inefficiencies associated with the cooling scheme in data centers, being air mixing is one of the most important. The cold aisle–hot aisle implementation does not completely isolate the cold air streams in the cold aisle and the hot air streams in the rack exhaust for two causes: hot air recirculation and cold air bypass. The former indicates that the hot air enters into the cold aisle from the top of the racks and the front end of the cold aisle closest to the cooling units. The recirculating warm air mixes with the cold air and increases the inlet temperature in a difficult way to foresee. Thereby, air mixing affects the reliability of the ICT equipment. Avoiding this would require the cooling system to have a complicated control system. The latter, cold air bypass, happens by the time the cold air from the perforated tiles overshoots the racks and returns to the cooling unit at a lower temperature. This can also happen due to floor leakage between the cold aisle and the cooling unit. The resulting low cooling unit extract temperature decrease the efficiency of the cooling system by narrowing the temperature difference between the extract and the supply.

With the aim of decreasing the losses incurred throughout the supply of the cooling medium and quickly reacting to spatial temperature distribution, closed-couple cooling solutions place cooling units nearer to computing units. There are mainly two categories according to the granularity of computing cluster covered by single-cooling unit, i.e., in-row and in-rack coolings. An in-row cooling adapts the cooling requirements at every row in accordance with the corresponding conditions, while an in-row cooling adjusts its cooling settings in accordance with operating condition at each rack, achieving energy savings up to 40% (Kim et al. 2015). Nevertheless, the capital expenses for the installation is very high.

Free cooling is a cooling design fundament, converging a broad spread of applications in the utilization of cooling from natural resources (Oró et al. 2015). The adoption of free cooling schemes is currently one of the most utilized techniques to turn data center more efficient. Zhang et al. (2014) and Ebrahimi et al. (2014) reviewed the improvements of data center free cooling schemes mainly with attention on setting characteristics and performances which can be roughly separated as follows:

- Airside free cooling. Use outside air for cooling data centers.

  – Direct airside free cooling. Drawing the cold outside air straight, after filtering into the data center.
  – Indirect airside free cooling. Running through air to air heat exchangers.

- Waterside free cooling. Use natural cold source by cooling water infrastructure.

  – Direct water cooled system. Natural cold water is used directly to cool the infrastructure through a heat exchange between the warm air and sea, river, ground water.
  – Air cooled system. Air cooler is utilized to cool the water circulating to CRACs when wet-bulb temperature of the outside air is low enough.
  – Cooling tower system. A cooling tower is adopted to cool the water circulating in CRACs and heat exchangers. Two water loops are required; a cooling (external) water loop and a chilled (internal) water loop.

Siriwardana et al. (2013) researched the inclusion of outside air with desired supply air for data center cooling in the Australian climate conditions. It was found there is a significant potential for using this scheme in some states that could lead to significant savings on cooling costs. Subsequently, Lee and Chen (2013) using a dynamic building energy simulation program (eQUEST) have found some energy-saving potential of airside free cooling for data centers in worldwide climate zones, highlighting that sizable direct air free cooling potential was reached in data centers positioned in humid climate zones. In addition, in dry climate conditions, substantial humidification is needed using techniques such as evaporative cooling, where raising the humidity of air lowers the temperature and thereby the water consumption must be considered, since 1 MW in a data center can consume 68 m^3/day of water for cooling, as presented by Ristic et al. (2015).

However, there is a potential risk to damage ICT equipment using direct airside free cooling due to the risk of particulate contaminants entering data centers. In this context, Dai et al. (2012a, b, 2013) studied different strategies to minimize the risks for ICT and telecommunication devices under this cooling policy. Nevertheless, even though Shehabi et al. (2007, 2008) claimed that with an appropriated filtration, the ASHRAE (2011) suggests particulate contaminant concentration for data centers is accomplished and significant economic saving can be achieved.

Some data center operators are leveraging to utilize seawater and geothermal energy to produce cooling for their infrastructure and pursuing green practices and minimize energy costs. After implementing a unique seawater cooling system, a 1600 m^2 data center located in Stockholm reduced energy costs by 80% (Oró et al. 2015). Furthermore, it reused the seawater to heat local offices and residential buildings before returning it to the sea. Consequently, the data center has lowered its metric power usage effectiveness (PUE) value to 1.09 minimizing its ICT load enough to enable additional customers to distribute in the facility, reinforcing its economic benefit. Similarly, data centers in Iowa and Nebraska are cooled by a geothermal bore field utilizing the cool temperatures underground to cool down the servers (Oró et al. 2015).

Demand response in data centers

A data center perspective on programs relies on the intersection of two important social issues. First, as ICT becomes increasingly pivotal to society, the associated energy demand are way up, being the growth in electricity demand for ICT ten times larger than the overall growth of electricity demand (Dreibholz et al. 2007; Koomey 2011; Ghatikar et al. 2012). Second, the integration of renewable energy into the power grid is essential for enhancing sustainability however causes significant challenges for management of the grid (Zhu et al. 2012). The focus behind demand response and energy efficiency in data centers is that these two challenges are in fact cooperative to reduce carbon emissions from electric power generation and to combat the effects of global climate change. Therefore, it is important not only to assess how an intensive energy consumer, such as data centers, can decrease costs by increasing their energy efficiency but also how they can take advantage of demand response programs to decrease costs and cooperate with the grid.

It is important to highlight that demand response requires the flexibility of electricity demand, being not only implemented with a temporary load reduction but also with a temporary load increase to stimulate demand in periods with renewable energy generation surplus. Usually, most studies adopt the subdivision given by ICT-flexible workload, where a set of delay-tolerant loads will be run in a given time and non-ICT workloads, comprise cooling solutions through set point adjustment and UPS strategy by charge and discharge. In this sense, all available technologies can be used both to decrease and to increase demand, such as promoting a load increase by charging a UPS, pre-cooling a building, or anticipating a specific ICT workload. These actions are achieved through flexibility and not by wasting energy, so even in these situations, there is no contradiction in the use of demand response to cause the temporarily load raise with energy efficiency.

Data centers represent very large loads that can reach up to 100 MW and are particularly well-suited for participation in demand response programs (Wierman et al. 2014), since they have flexible loads (Ghatikar et al. 2010) and are greatly automated and monitored, since the power load, the state of ICT equipment, and cooling facilities are usually monitored and adjusted. For example, recent empirical studies by Lawrence Berkeley National Laboratory (LBNL) quantified the flexibility in power usage of four data centers under different management approaches with no impact to operations or service-level agreements (Ghatikar et al. 2012).

The survey developed by Liu et al. (2014) indicates that data centers can use 40 times more energy than conventional office facilities, and 5% of the load can typically be shed in 5 min and 10% in 15 min, with no impact on the ICT procedures. Moreover, if workload management approaches are exploited, the flexibility level can be even larger, without additional time needed to shed the load.

Ghatikar et al. (2010, 2012) concluded that data centers present significant load-reduction potential; nonetheless, not all data centers can take advantage of all approaches because of different operational profiles. Some strategies are appropriated for energy efficiency; however, additional incremental benefits can be reached by temporarily decreasing service levels for a few hours a day and a few days a year for the implementation of demand response. Consequently, data center demand response strategies generally are divided between load shedding (reduction or interruption of the load) and load shifting (moving load from peak to off-peak periods).

A US Federal Energy Regulatory Commission (FERC) assessment lists demand response programs presented by Tang et al. (2012, 2014) as dynamic pricing without enabling technology, dynamic pricing with enabling technology, direct load control (DLC), interruptible tariffs, and other programs, such as capacity/

Table 3 Price-based programs (Tang et al. 2014)

Price-based programs	
Types	Description
Time-of-use (TOU)	TOU rates differ in different blocks of time. The rate reflects the average cost of electricity during different periods
Critical peaking price (CPP)	CPP benefits the participants by reducing their energy usage during CPP events
Real-time pricing (RTP)	The price signal of RTP is released a day or an hour or even shorter ahead of the time for which it applies

demand bidding and wholesale programs. These programs are categorized into:

- Price-based, market-led, or stability-based programs offer participants time-varying rates that reflect the value and cost of electricity in different time periods, as presented in Table 3.
- Incentive-based, system-led, reliability-based, or economic-based programs offer participants discount rates or rebates for their participation or load reduction performance on demand response signals, as presented in Table 4.

The progress of advanced metering technologies will enable all types of customers to participate in automated demand response programs and, in the data center case, taking advantage of the flexible features highlighted by (Irwin et al. 2011), as follows:

- Servers are equipped with programmable power management procedures, settling their power consumption by commands from selected interfaces.
- Many workloads are tolerant to delays or performance degrading, allowing data centers to suit the power consumption in response to price fluctuations.
- Data centers consume a massive amount of energy with a substantial impact on grid requirement.

Tang et al. (2012, 2014) argue the achievement of demand control approaches relies on several factors, including: frequency, duration, local weather patterns,

Table 4 Incentive-based programs (Tang et al. 2014)

Incentive-based programs	
Types	Description
Direct load control (DLC)	DLC program operators offer a participant an incentive, usually financial, in the form of credits on the utility bill
Interruptible/curtailable service (CS) programs	Participants of these programs receive a rate discount or bill credit in return for agreeing to reduce load during certain time periods
Demand bidding/buy back (DB)	Participants offer their most cost-beneficial bids, price and reducible load, to an electricity market
Emergency demandresponse (EDR)	Participants receive incentives for measured load reductions during emergency conditions; however, curtailment is voluntary
Capacity market (CM)	Participants who commit to providing contracted load reductions when necessary receive incentives
Ancillary service market (ASM)	Participants must adjust huge amounts of load quickly when an event occurs. The response duration is typically in minutes rather than hours

or electric grid conditions. However, to undertake a proper demand response control strategy, it is needed to assemble enough power consumption information of the participating facilities.

Flexibility and optimization in data centers

For the purpose of reducing the energy cost and performing a gradual inclusion of data centers in demand response programs, recent and relevant works have been conducted proposing theoretical frameworks of robust optimization and low computational complexity to obtain close-to-optimal solutions.

In this context, Cioara et al. (2016) propose an electronic marketplace designed for trading energy flexibility and ancillary services, enacting data centers to shape their energy demand to buy additional energy when prices are low and sell energy surplus when prices are high. A new pricing mechanism, which extracts load reductions from tenants in colocation data centers during emergence demand response events, was proposed by Chen et al. (2015). Their results present benefits to the environment and data center operators (by decreasing the need for backup diesel generation), while also aiding tenants (by providing payments for load reductions). The dynamic interactions between smart grid and data centers as a two-stage price optimization problem were presented in (Wang et al. 2016), using a heuristic algorithm and simulations to achieve a win-win solution for both the utility and data centers. Fridgen et al. (2017) present an economic analysis of spatial load migration as an alternative form of demand side flexibility compared to load shifting and load shedding using virtualization, finding that spatially migrating load provides an interesting alternative to economically balancing a grid which has previously only been attributed to transmission lines.

A multi-objective energy-efficient task scheduling problem on a green data center partially powered by renewable energy, where the computing nodes are DVFS-enabled is highlighted in Lei et al. (2016). The solution is provided by a multi-objective co-evolutionary algorithm that searches the suitable computing node, supply voltage, and clock frequency for the task computation, and the smart time scheduling strategy is employed to determine the start and finish time of the task on the chosen node. Wang et al. (2017b) propose scheduling algorithms and adjust their scheduling policies for the incoming jobs according to the performance target and the behavior of other competitors based on the game theory. The results show the capacity to reduce the conflict and hence improve the scheduling performance in the data centers deployed with clusters and distributed schedulers. A novel method based on demand response was proposed to control the cooling supply related to ICT dynamic load in Zhu et al. (2017) and the assessment pointed out reductions of 7.9%, 14.2%, 15.6%, and 17.9% at room, row, rack, and server levels, respectively. Determining the cooling demand according to the ICT load at server level, the reduction of the electricity consumption of cooling systems was by 0.9% and considering the dynamic energy efficiency of cooling units, ICT load shifting could be optimized in 1.2%.

As well as there are energy efficiency strategies exploiting workloads with direct impact on CPU, memory, disk and network, virtualization, cooling, and UPS, there are data centers demand response approaches in terms of flexibility. The technique addressed in Cioara et al. (2016) provides flexibility mechanisms defined for hardware components, such as load time shifting, alternative usage of non-electrical cooling devices, and charging/discharging UPS, evidencing potential to shape and modify the data center baseline energy profile to meet energy network levels goals and to provide several types of energy and balancing services.

Tran et al. (2014) and Li et al. (2015) define optimization problems, where processing, virtualization, quality of service, and cooling solutions are analyzed together in a demand response scenario, resulting in a significant energy consumption and electricity cost reduction. In Ghatikar et al. (2012), an initial set of control and load migration strategies by virtualization and economic feasibility for four data centers was evaluated. The findings show that with minimal or no impact to data center operations demand savings of 25% at the data center level or 10 to 12% at the whole building level can be achieved with demand response strategies, such as server and CRAC unit shutdown, load shifting, or queuing ICT jobs while server are idle, temperature set point adjustment, and load migration between homogeneous and heterogeneous cluster systems.

In the same context, it is important to consider the adaptation to data centers of measures already used in other sectors with a higher tradition of participation in DR programs. In this way, Paterakis et al. (2017) perform an extensive overview on the theme covering different sectors and international experiences.

Regarding industrial customers, they claim that the demand can be decreased by on-site generation, energy storage, consumption shifting, non-critical load curtailment, and temporary shut down of several processes. Temporarily interrupting one or more processes may result in significant load reductions. Nevertheless, several constraints such as the criticality of a process, the number of available production lines, the required production target, inventory restrictions, etc., may have longer term impacts on the process line, in a very similar way that occurs in the case of data centers, respecting the differences of equipment and business profile. Analyzing commercial and other non-residential customers, the main DR strategy is load reduction, where air conditioner is the most significant load that can be controlled. However, the project of energy-intelligent buildings monitoring energy consumption and managing locally available resources, as well as the energy procurement from the grid, has been introduced (Christantoni et al. 2016) and can also be used in small and medium data centers. Residential customers are suitable for DLC and price-based DR programs and can invest on an automated system, which monitors and controls the consumption of several appliances. Following this same premise, data centers can use algorithms, software, or even appliances to provide this level of supervision in their loads (Paul et al. 2017), taking advantage of specific demand response programs. Nevertheless, the main similarity of the residential sector compared to the small and medium data centers market is the fact that their loads have a low potential to participate in demand response events acting individually, and therefore, there is a need for aggregation, as suggested the review conducted by Carreiro et al. (2017), covering several cases of aggregators in the context of end users, namely in the residential case.

Challenges and overcoming

Even though simulation results of many studies have been demonstrating that it is possible to improve power grid reliability and provide an important source of economic benefits in demand response market, nowadays, data centers are largely non-participants due to four main challenges to overcome (Wierman et al. 2014):

1. Regulation and market maturity. Many of these demand response programs are not yet available to data centers in several markets due to the need for adjustments in the regulatory aspects. As a result, the opportunities for their participation may be limited to simple and traditional programs, such as smart and coincident peak pricing, as demonstrated by Liu et al. (2013) and Brocanelli et al. (2014). The first step to overcome challenges in this aspect is transitioning to deregulated market with an independent energy regulator, providing more specific roles and competition in the whole value chain through, for example, aggregators, which are companies that pool the generation or flexible demand capacities of a number of smaller consumers (Flanagan 2013).
2. Risk management. Camacho et al. (2014) claim that data centers prefer to negotiate long-term energy contracts with fixed usage prices because their main business is the maximizing uptime and performance, and energy issues are certainly secondary when compared with the need to maintain strong guarantees about these primary measures. However, the electric sector needs to provide information about this old mindset, using scientific works such as Basmadjian et al. (2015), in which new and appropriated data centers contracts are proposed and validated ensuring performance and reliability to data center operations.
3. Control. “Direct load control” programs, for which the grid sends a signal to a controller of the program participant are not always acceptable to participants and other extreme alternative is “prices-to-devices,” where real-time prices are conveyed to participants. Nevertheless, such programs typically require huge price variation to extract desired responses (Wang and Groot 2013). This volatility is not acceptable given the risk tolerance of data centers, thus other programs such as the pricing and operation strategy optimized used in (Jin et al. 2017a) must be developed to facilitate their participation.
4. Market complexity. The complexity to automate and incorporate the bidding process into a data center management system, as well as the high regulation have prevented data centers from entering these markets despite the financial opportunities (Wang et al. 2013). Nevertheless, the ICT industry continues to refine technical capabilities in relation to power-capping, load management, and virtualization of workload that will help manage any perceived risk and along with alternatives as microgrid optimal dispatch in (Jin et al. 2017b), changing some complexities.

Therefore, it is fundamental to promote the necessary adjustments with very specific polices in order that the progress made so far can widely also contemplate the emerging reality in the market of small and medium profile data centers.

Small and medium data centers perspective

Energy efficiency measures, demand response, and integration of renewable sources discussed so far were analyzed without considering size or proportion of data centers. However, while most technical sectors and even scientific community attention are focused on the largest data centers, these hyper-scale cloud computing profile represent only a small portion of overall energy consumption in this market. The extensive majority of data center energy is consumed in small, medium, and multi-tenant environments. These profiles have generally made much less advancement than their hyper-scale cloud equivalent due to market barriers, lack of more specific data centers energy efficiency metrics, and also misalignment of inducements according to Delforge (2014). The fact that the cost center is generally disassociated from the data center itself, since it is not the main business, is another key factor that influences the investment capacity in adopting new and more efficient technologies.

However, the above-mentioned realities are also applicable to small and medium data centers, dimensions respected. Specifically, what can change is the amount of financial resources available to implement such measures, because in terms of technological level, what will be changed is the proportion of a data center resources available.

Regarding data center energy efficiency, there has been a significant advance in the last decade, with server farms operated by large companies leading the way. Nevertheless, these hyper-scale cloud computing enterprises account for approximately only 5% of all data center, as depicted by Fig. 7. The corporate-owned enterprises, small- and medium-sized organizations, and multi-tenant data centers are far behind in terms of efficiency, requiring focused actions, such as utility incentive programs to reduce waste in the huge amount of electricity used by data centers of all sizes (Delforge 2014).

Natural Resources Defense Council (NRDC) report states that large, mid-tier, and enterprise-class data centers represent only half of all US servers, as can be seen in Table 5. The other half constitutes small server rooms and closets typically available in small and medium businesses and organizations, as well as in research institutes, departments, and subsidiaries of larger organizations.

Even in the face of such statistics, in terms of energy efficiency, an estimated 20 to 30% of servers (Delforge 2014) in even these large data centers are obsolete or

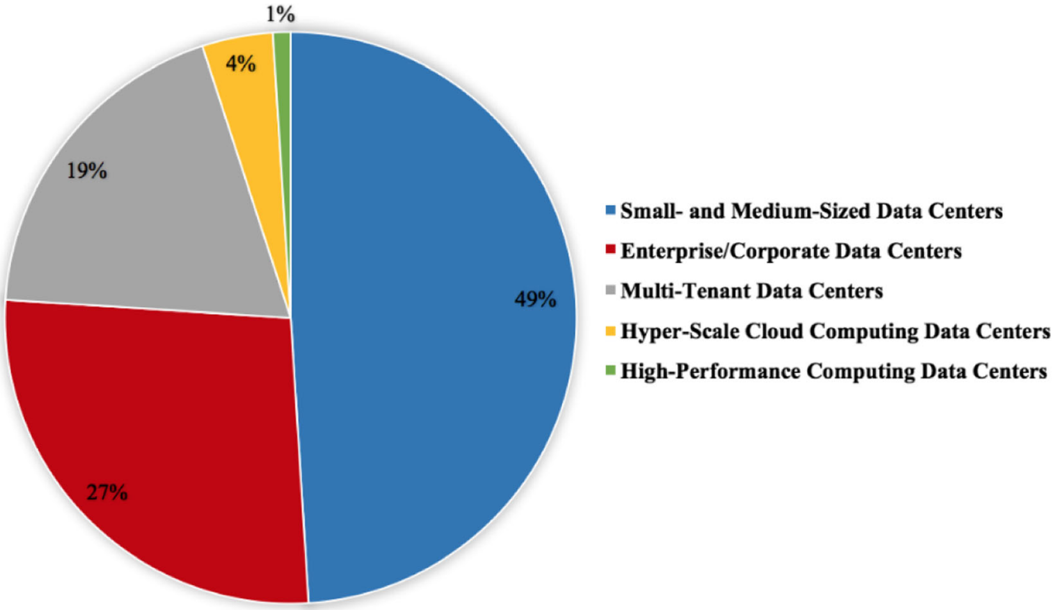


Fig. 7 Estimated US data center electricity consumption by market segment (2011) (Josh and Delforge 2014)

unused because the completion or changes in project profiles; however, they are still in operation and consuming electricity every day for lack of awareness of their real need.

In order to verify this reality, three latest surveys with different approaches are presented. The Green Grid (2016) questioned 150 key European ICT decision makers with data center responsibilities in the UK, France, and Germany. It was found that while most organizations are facing growing pressures to improve the efficiency of their data centers, 43% of those surveyed have no energy efficiency objectives in place, as presented in Fig. 8. Key findings addressed include the following:

- Energy efficiency and operating costs are the most common areas of the data center reported as requiring improvement.
- Two in five respondents reported that their data centers are expensive to run (48%) or upgrade (41%), demonstrating that cost is the most commonly reported impact of data center operations.
- The difficulty in predicting future costs (43%) and the cost of refreshing hardware (37%) are cited as the top challenges to developing resource efficient data centers, along with the difficulty in meeting environmental targets (33%).

Some of the findings presented a positive outlook for future innovations in data center resource efficiency, with nearly all those surveyed clearly seeing areas for improvement and 55% stating that energy efficiency was their highest priority when making (The Green Grid 2016).

The other survey, carried out by NRDC (Bennett and Delforge 2012), surveyed 30 US ICT managers. The focus was on virtualization and server purchasing and replacement, since they are the largest and most cost-effective savings opportunities profile in small and medium data centers. From the energy efficiency point of view, it is possible to reduce operating costs by replacing and using greener equipment, as well as to decrease the amount of hardware in use and idle processing time, improving the utilization of resources and providing energy savings through virtualization. Regarding demand response, one of the used techniques to exploit the flexibility potential is the use of virtualization on servers to perform workloads shifting or shedding,

Table 5 Estimated US data center electricity consumption by market segment (2011) (Whitney and Delforge 2014)

Segment	Percent of stock (based on # of servers)	Average PUE	Average server utilization	Average server age (years)	2011 Electricity use (GWh)	Server power at average utilization level (watts)	Data center market segmentation by electricity consumption (%)
Small and medium server rooms	40	2.0	10%	3	37.5	149	49
Enterprise/corporate data centers	30	1.8	20%	2	20.5	120	27
Multi-tenant data centers	22	1.8	15%	2	14.1	113	19
Hyper-scale cloud computing	7	1.5	40%	1	3.3	101	4
High-performance computing	1	1.8	50%	2	1	169	1
	100				76.4		100

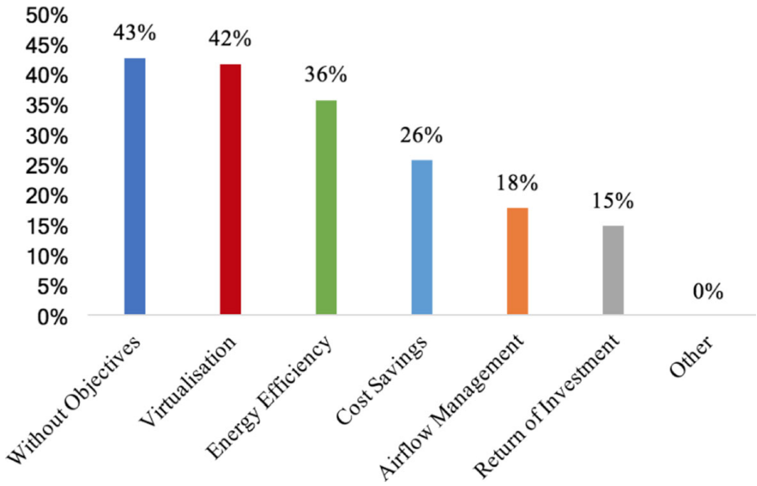


Fig. 8 Organizations' energy efficiency objectives (The Green Grid 2016)

taking advantage of a time window where the energy cost is more profitable. In this context, the questions addressed various issues related to their current server fleet, virtualization practices, cloud computing, obstacles to implementing efficiency improvements, energy use, and billing. The results from this survey compared with results from other survey of large companies' virtualization practices conducted by the market research firm (Vanson-Bourne 2011), as presented in Fig. 9 have shown that:

- Energy-efficient upgrades in small- and medium-sized businesses are particularly challenging due to the enormous variety of ownership configurations.
- Virtualization adoption by smaller businesses are behind large ones.
- Small- and medium-sized organizations tend to implement virtualization usage more broadly than large ones.
- Nearly all large companies have used virtualization and plan an increased use in their operations.
- Currently, 60% of the decision-maker employees regarding server purchasing do not have access to their company's energy bill.
- Replacement of servers coming to the end of their warranty is the most cost-effective way to implant efficiency best practices and half of all organizations surveyed plan on a server room upgrade in the following year.

The survey developed in Vasques et al. (2017) aimed to analyze different contexts in different countries to investigate different energy efficiency patterns and actions taken by 22 small and medium data centers actors. A set of questions was asked to find out the level of energy efficiency of the servers, whereas for each set of 10 servers, 3 are

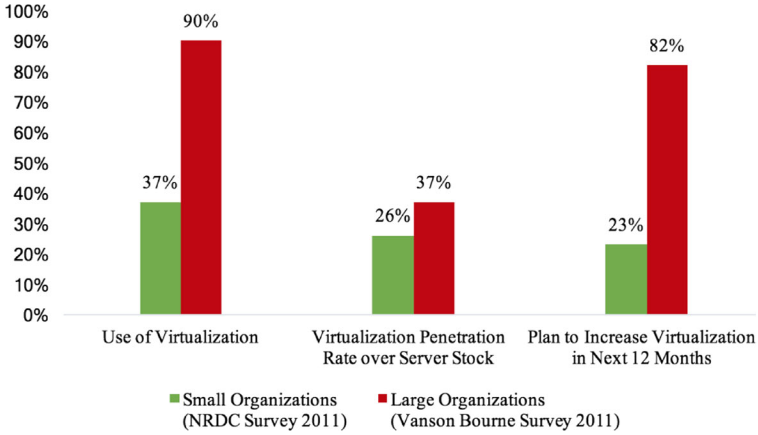


Fig. 9 Comparative survey results (Bennett and Delforge 2012)

legacy, 8 have some energy efficiency label, and only 2 have DVFS enabled. It was concluded that all the actors who have storage appliances in their facilities have already acquired them with some energy efficiency label. However, for each 10 data centers, nearly 3 of them do not have storage appliances.

Figure 10 presents the results regarding the adoption of monitoring actions and they are alarming from the energy efficiency point of view, as the energy consumption of most of the servers, storage appliances, and network devices are not monitored. A high proportion did not calculate annual energy consumption. Moreover, 14% of respondents were unaware of metrics such as PUE and 18% did not know if there was an electrical load diagram in the facility. The results also show that nearly 73% answered that another sector is responsible for managing the overall data center energy consumption and nearly 91% answered that payment of energy bill is the responsibility of the institution that owns the data center and not the data center itself. Such results are therefore good indicator to justify the low adoption of energy efficiency measures.

Despite the reality of energy efficiency terms evidenced in these surveys, the approach presented in Emerson 2015 highlights the best energy efficiency practices applied to a medium 465 m^2 data center based on real-world technologies and operating parameters. The results have shown nearly 50% reduction in data center energy consumption without compromising performance or availability, as presented in Table 6.

Therefore, Josh and Delforge (2014) concluded that several strategical actions can improve data center efficiency; nevertheless, systemic measures are needed to develop the conditions for best-practice efficiency behaviors throughout the small and medium data center industry, including:

- Adoption of a simple server utilization metric, wherever possible to solve the underutilization of servers.
- Alignment of incentives between decision-makers, ensuring that incentives are aligned to provide financial rewards for efficiency best practices.
- Disclosure of data center energy and carbon performance for demonstrating leadership and driving behavior change throughout a whole sector.

It is important to highlight as an important example that Josh and Delforge (2014) claim that if just half of the technical savings potential for data center energy efficiency identified in their paper were realized, electricity consumption in US data centers could be cut by up to 40%.

Moreover, in terms of policies, voluntary approaches such as the European Code of Conduct for Data Centre Energy Efficiency (Avgerinou et al. 2017) are paramount and can be effective in addressing climate and energy issues even in small and medium data centers case, with the average PUE of the participating facilities declining year after year, reaching the value of 1.64 in 2016.

Covering the case of demand response applied to small and medium data center, firstly there is no understanding that it should be a market handled separately, which is a negative reality, as this paper has shown. Secondly, demand response studies and practices applied directly to this dimension profile are incipient in literature. However, Tang et al. (2012) concluded that the aggregation of small and medium data centers has great

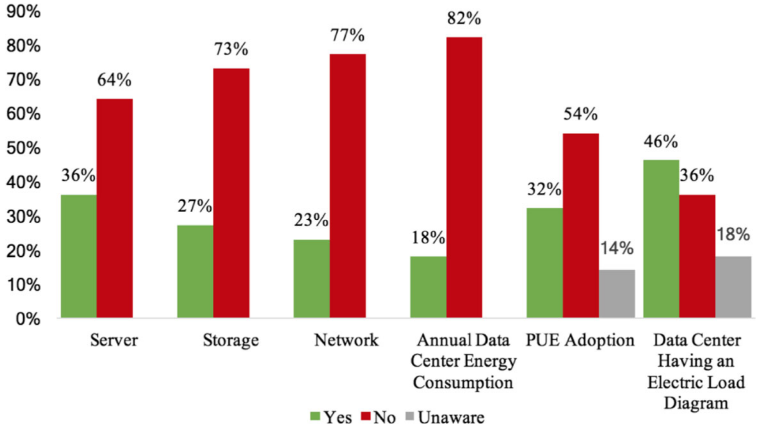


Fig. 10 Energy efficiency monitoring actions (Vasques et al. 2017)

Table 6 Benefits from efficiency improvement actions (Emerson 2015)

Efficiency improvement area	Savings (kW)	
ICT polices	Low-power processors	111
	High-efficiency power supplies	124
	Server power management	86
ICT projects	Blade servers	7
	Virtualization	86
Best practices	Higher voltage AC power distribution	20
	Cooling best practices	15
	Variable-capacity cooling	49
Infrastructure projects	High-density supplemental cooling	72
	Monitoring and optimization	15
Total		585

potential if the power consumption provides satisfactory detail and the control strategies are well planned.

Tang et al. (2012, 2014), through a load control method in a school, performed the required power reduction when a demand response event occurs and claims that major limitation of small data centers participation in this type of programs comes from the minimum capacity requirement set by the demand response program providers. The police suggested the aggregation of small and medium data centers to form a qualifiable unit to overcome this limitation. Contracts run to specific dimensions of data centers, such as small and medium-sized were designed and optimized in Basmadjian et al. (2015) proving their complete adaptability. Besides current optimization problems, an algorithm developed for large data centers can be adapted and applied to match the power demand and specificities of small and medium data centers.

Similarly, it is not very common to find in the literature renewable resources implementation in small and medium data centers. Liu et al. (2013) developed two algorithms for data centers regardless of size by combining workload scheduling and local power generation with photovoltaic, to avoid the coincident peak and reduce the energy costs. The potential of reduction was 35–40% in energy costs and 10–15% in emissions. A systematic framework is proposed in Tianyi Chen et al. (2016b) to integrate renewable energy sources, distributed storage units, cooling facilities, as well as dynamic pricing into the workload and energy management tasks of a data center network. Coping with renewable energy sources uncertainty, the resource allocation task is formulated as a robust optimization problem minimizing the worst-case net cost, offering a new perspective in dealing with the uncertainties involved in this scenario.

Conclusions

The growth of data center demand in recent years has led to an increase in their size and power, and therefore, in their electricity consumption, being the impact of these unique infrastructures on the worldwide energy map is more and more relevant. Data centers designs incorporate oversized and redundant systems, in most cases, running at partial load, and the power and cooling requirements are greatly overstated, leading to extra investment and operational cost expenses. Due to the large savings potential, many researchers have been focused on the development of knowledge, tools, and systematic standards to use efficient technologies to reduce the data center consumption and integrate renewables in their energy portfolio within smart grid context.

This review paper has focused in researching the several available technologies to promote energy efficiency, the most prominent and recently discussed by various authors summarizing the most relevant aspects focusing on dismemberment strategy of computational components of servers, software solutions such as virtualization, cooling technologies, energy storage and management, addressing also the integration of renewable energy sources. All these aspects have been discussed on the framework and the perspective of small and medium data centers, which account for more than 50% of the total electricity consumption around the world, discussing demand response programs as a prominent solution to decrease costs and to cooperate with the grid to ensure higher reliability and sustainable development goals.

On one hand, this work aimed to alert how demand response is still insipid in this sort of small and medium profile environment, as well as the policies for this type of energy consumer has often been neglected. On the other hand, data centers can take advantage of several demand response benefits in the economic, environmental, and market areas. As widely discussed, large data centers have greater prominence in current programs, despite this, adaptations can and should be made to treat small and medium data centers as a single market. It is reasonable to implement this scenario using economic incentives, rather than the application of a blunt regulatory instrument. A solution that offers revenue and savings for small and medium data centers helps ease the challenges and satisfies this new and future market as a whole in the form of lower costs and more secure, sustainable supply within reach.

The discussed surveys aimed to precisely present a worrying scenario in terms of energy efficiency in small and medium data centers, i.e., discussing energy efficiency in large and mid-small profile is the same as discussing two different, antagonistic worlds and this paper warned about this reality. The results of the surveys showed in a practical way how the initially presented efficiency strategies are not yet fully implemented at all in this type of environment.

Hence, based on the reality presented in this paper, it could be concluded that studies to further develop strategies in small and medium data centers are needed, including more case studies, simulation, planning of polices, and overall evaluation to provide a clearer view of how the aforementioned actors can as a matter of fact benefit from each other, thereby promoting effective transformations in the energy market. Therefore, it is a desirable goal to achieve a trade-off between the maximum benefits of data centers and the minimum environmental impact by considering various aspects such as cost, energy consumption, and environment. With the building pace and scale of data centers greatly increasing, various energy-saving technologies should be integrated into data centers to better improve resource utilization and reduce energy consumption, and truly achieve green data centers.

Funding The present work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil).

References

Aghaei, J., & Alizadeh, M. I. (2013). Demand response in smart electricity grids equipped with renewable energy sources: a review. *Renewable and Sustainable Energy Reviews, 18*, 64–72. https://doi.org/10.1016/j.rser.2012.09.019.

Ahmad, R. W., Gani, A., Hamid, S. H. A., Shiraz, M., Yousafzai, A., & Xia, F. (2015). A survey on virtual machine migration and server consolidation frameworks for cloud data centers. *Journal of Network and Computer Applications, pp., 52*, 11–25. https://doi.org/10.1016/j.jnca.2015.02.002.

Ahn, J. H., Jouppi, N. P., Kozyrakis, C., Leverich, J., & Schreiber, R. S. (2009). Future scaling of processor-memory interfaces. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09* (p. 1). New York: ACM Press. https://doi.org/10.1145/1654059.1654102.

Alkharabsheh, S., Fernandes, J., Gebrehiwot, B., Agonafer, D., Ghose, K., Ortega, A., Joshi, Y., & Sammakia, B. (2015). A brief overview of recent developments in thermal management in data centers. *Journal of Electronic Packaging, 137*(4), 040801. https://doi.org/10.1115/1.4031326.

Amur, H., Cipar, J., & Gupta, V. (2010). Robust and flexible power-proportional storage. In *Proceedings of the 1st ACM symposium on Cloud computing - SoCC '10* (pp. 217–228). https://doi.org/10.1145/1807128.1807164.

Andrén, F. P., Strasser, T. and Kastner, W. (2016) 'Applying the SGAM methodology for rapid prototyping of smart grid applications', *IECON Proceedings (Industrial Electronics Conference)*, pp. 3812–3818. https://doi.org/10.1109/IECON.2016.7794057.

Ardagna, D., Panicucci, B., Trubian, M., & Zhang, L. (2012). Energy-aware autonomic resource allocation in multitier virtualized environments. *IEEE Transactions on Services Computing, 5*(1), 2–19. https://doi.org/10.1109/TSC.2010.42.

Arianyan, E., Taheri, H., & Khoshdel, V. (2017). Novel fuzzy multi objective DVFS-aware consolidation heuristics for energy and SLA efficient resource management in cloud data centers. *Journal of Network and Computer Applications, 78*, 43–61. https://doi.org/10.1016/j.jnca.2016.09.016.

ASHRAE (2011) '2011 Gaseous and particulate contamination guidelines for data centers', American Society of Heating, Refrigerating and Air-Conditioning Engineers, pp. 1–22.

Avgerinou, M., Bertoldi, P. and Castellazzi, L. (2017) 'Trends in data centre energy consumption under the European code of conduct for data centre energy efficiency', Energies, 10(10). https://doi.org/10.3390/en10101470.

Barroso, L. A., & Hölzle, U. (2007). The case for energy-proportional computing. *Computer, 40*(12), 33–37. https://doi.org/10.1109/MC.2007.443.

Basmadjian, R., Müller, L., & De Meer, H. (2015). Data centres' power profile selecting policies for demand response: insights of green supply demand agreement. *Ad Hoc Networks, 25*(PB), 581–594. https://doi.org/10.1016/j.adhoc.2014.11.007.

Beloglazov, A., & Buyya, R. (2013). Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints. *IEEE Transactions on Parallel and Distributed Systems, 24*(7), 1366–1379. https://doi.org/10.1109/TPDS.2012.240.

Bennett, D. and Delforge, P. (2012) 'Small server rooms, big energy savings opportunities and barriers to energy efficiency on the small server room market', *NRDC*, (February).

Bonetto, E., Finamore, A., Mellia, M., & Fiandra, R. (2014). Energy efficiency in access and aggregation networks: from current traffic to potential savings. *Computer Networks, 65*, 151–166. https://doi.org/10.1016/j.comnet.2014.03.008.

Bostoen, T., Mullender, S., & Berbers, Y. (2013). Power-reduction techniques for data-center storage systems. *ACM Computing Surveys, 45*(3), 1–38. https://doi.org/10.1145/2480741.2480750.

Boulos, S., Nuttall, C., Harrison, B., Moura, P. and Jehle, C. (2014) ErP lot 27—uninterruptible power supplies: preparatory study—final report. Ricardo-AEA, Intertek, ISR – University of Coimbra.

Brocanelli M, Li S, Wang X, and Zhang W. (2014) 'Maximizing the revenues of data centers in regulation market by coordinating with electric vehicles', Sustainable Computing: Informatics and Systems. Elsevier Inc., pp. 1–13. https://doi.org/10.1016/j.suscom.2014.03.004, 2015.

Calheiros, R. N., Ranjan, R., De Rose, C. A. F. and Buyya, R. (2009) 'CloudSim: a novel framework for modeling and simulation of cloud computing infrastructures and services', *arXiv preprint arXiv:*0903.2525, p. 9.

Camacho, J., Zhang, Y., Chen, M. and Chiu, D. M. (2014) 'Balance your bids before your bits: the economics of geographic load-balancing', *Proc. of the 5th Int. Conf. on Future Energy Systems (ACM e-Energy)*, pp. 75–85. https://doi.org/10.1145/2602044.2602068.

Capozzoli, A., Chinnici, M., Perino, M. and Serale, G. (2015) 'Review on performance metrics for energy efficiency in data center: the role of thermal management', *Energy Efficient Data Centers*, pp. 135–151. https://doi.org/10.1007/978-3-319-15786-3_9.

Carreiro, A. M., Jorge, H. M., & Antunes, C. H. (2017). Energy management systems aggregators: a literature survey. *Renewable and Sustainable Energy Reviews, 73*, 1160–1172. https://doi.org/10.1016/j.rser.2017.01.179.

Castro, P. H. P., Cardoso, K. V and Corrêa, S. (2013) 'Uma Abordagem Baseada no Consumo de CPU e RAM para a Eficiencia Energetica em Centros de Dados para Computação em Nuvem', *Wscad-Ssc 2013*, (February), p. 8. https://doi.org/10.13140/2.1.4681.0242.

Cecati, C., Mokryani, G., Piccolo, A. and Siano, P. (2010) 'An overview on the smart grid concept', *IECON Proceedings (Industrial Electronics Conference)*, pp. 3322–3327. https://doi.org/10.1109/IECON.2010.5675310.

Chen, N., Ren, X., Ren, S. and Wierman, A. (2015) 'Greening multi-tenant data center demand response', Performance Evaluation. Elsevier B.V., 91, pp. 229–254. https://doi.org/10.1016/j.peva.2015.06.014.

Chen, T., Gao, X., & Chen, G. (2016a). The features, hardware, and architectures of data center networks: a survey. *Journal of Parallel and Distributed Computing, 96*, 45–74. https://doi.org/10.1016/j.jpdc.2016.05.009.

Chen, T., Zhang, Y., Wang, X., & Giannakis, G. B. (2016b). Robust workload and energy management for sustainable data centers. *IEEE Journal on Selected Areas in Communications, 34*(3), 651–664. https://doi.org/10.1109/JSAC.2016.2525618.

Christantoni, D., Oxizidis, S., Flynn, D., & Finn, D. P. (2016). Implementation of demand response strategies in a multipurpose commercial building using a whole-building simulation model approach. *Energy and Buildings, 131*, 76–86. https://doi.org/10.1016/j.enbuild.2016.09.017.

Cioara, T., Anghel, I., Bertoncini, M., Salomie, I., Arnone, D., Mammina, M., Velivassaki, T.-H., & Antal, M. (2016). Optimized flexibility management enacting data centres participation in smart demand response programs. *Future Generation Computer Systems., 78*, 330–342. https://doi.org/10.1016/j.future.2016.05.010.

Craig-wood, K., Krause, P. and Mason, A. (2010) 'Green ICT: oxymoron or call to innovation?', *Transport*, (Ict), pp. 978–981. https://doi.org/10.5176/978-981-08-7240-3.

Dai, J., Das, D., & Pecht, M. (2012a). A multiple stage approach to mitigate the risks of telecommunication equipment under free air cooling conditions. *Energy Conversion and Management, 64*, 424–432. https://doi.org/10.1016/j.enconman.2012.06.018.

Dai, J., Das, D., & Pecht, M. (2012b). Prognostics-based risk mitigation for telecom equipment under free air cooling conditions. *Applied Energy, 99*, 423–429. https://doi.org/10.1016/j.apenergy.2012.05.055.

Dai, J., Das, D., Ohadi, M., & Pecht, M. (2013). Reliability risk mitigation of free air cooling through prognostics and health management. *Applied Energy, 111*, 104–112. https://doi.org/10.1016/j.apenergy.2013.04.047.

Dargie, W. (2014) 'Estimation of the cost of VM migration', Proceedings - International Conference on Computer Communications and Networks, ICCCN. https://doi.org/10.1109/ICCCN.2014.6911756.

Dayarathna, M., Wen, Y., & Fan, R. (2016). Data center energy consumption modeling: a survey. *IEEE Communications Surveys & Tutorials, 18*(1), 732–794. https://doi.org/10.1109/COMST.2015.2481183.

Delforge, P. (2014) 'America's data centers are wasting huge amounts of energy', *Natural Resources Defense Council (NRDC)*, IB:14-08-a(august), pp. 1–5.

Deng, Y. (2011). What is the future of disk drives, death or rebirth? *ACM Computing Surveys, 43*(3), 1–27. https://doi.org/10.1145/1922649.1922660.

Dhiman, G., Marchetti, G., & Rosing, T. S. (2010). VGreen: a system for energy-efficient management of virtual machines. *ACM Transactions on Design Automation of Electronic Systems, 16*(1), 1–27. https://doi.org/10.1145/1870109.1870115.

Dreibholz, T., Becke, M. and Adhari, H. (2007) 'Report to congress on server and data center energy efficiency public law 109-431', *tdr.wiwi.uni-due.de*, 109, p. 431.

Ebrahimi, K., Jones, G. F., & Fleischer, A. S. (2014). A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews, 31*, 622–638. https://doi.org/10.1016/j.rser.2013.12.007.

Elnozahy, E. N., Kistler, M. and Rajamony, R. (2003) 'Energy-efficient server clusters', in, pp. 179–197. https://doi.org/10.1007/3-540-36612-1_12.

Emerson (2015) 'Energy logic: reducing data center energy consumption by creating savings that cascade across systems', White Paper, Emerson Network Power, pp. 1–21.

Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart grid—the new and improved power grid: a survey. *IEEE Communications Surveys & Tutorials, 14*(4), 944–980. https://doi.org/10.1109/SURV.2011.101911.00087.

Flanagan, C. (2013) 'A data center perspective on demand response', *Data Centers Dynamics*.

Fridgen, G., Keller, R., Thimmel, M., & Wederhake, L. (2017). Shifting load through space—the economics of spatial demand side management using distributed data centers. *Energy Policy, 109*, 400–413. https://doi.org/10.1016/j.enpol.2017.07.018.

Fulpagare, Y., & Bhargav, A. (2015). Advances in data center thermal management. *Renewable and Sustainable Energy Reviews, 43*, 981–996. https://doi.org/10.1016/j.rser.2014.11.056.

Garimella, S. V., Persoons, T., Weibel, J. and Yeh, L. T. (2013) ‘Technological drivers in data centers and telecom systems: multiscale thermal, electrical, and energy management’, *Applied Energy*, pp. 66–80. https://doi.org/10.1016/j.apenergy.2013.02.047.

Ghatikar, G., Piette, M. A., Fujita, S., McKane, A., Dudley J. H., Radspieler, A., Mares, K. C. and Shroyer, D. (2010) Demand response and open automated demand response opportunities for data centers, Lawrence Berkeley National Laboratory.

Ghatikar, G., Ganti, V. and Matson, N. (2012) Demand response opportunities and enabling technologies for data centers: findings from field studies, *Lawrence Berkeley National Laboratory.*

Gils, H. C. (2014) Assessment of the theoretical demand response potential in Europe. *Energy, 67*, 1–18. https://doi.org/10.1016/j.energy.2014.02.019.

Grice, J. W., Dean, N. and Eddie, S. (2013) ‘Sustainable site selection: the convergence of data center site’, *The Green Grid. Research Report*, (Sustainable Site Selection), p. 20.

Güngör, V. C., Sahin, D., Kocak, T., Ergüt, S., Buccella, C., Cecati, C., & Hancke, G. P. (2011). Smart grid technologies: communication technologies and standards. *IEEE Transactions on Industrial Informatics, 7*(4), 529–539. https://doi.org/10.1109/TII.2011.2166794.

Gupta, M. and Singh, S. (2003) ‘Greening of the internet’, *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications - SIGCOMM ‘03*, pp. 19–26. https://doi.org/10.1145/863956.863959.

Gupta, M. and Singh, S. (2007) ‘Using low-power modes for energy conservation in Ethernet LANs’, In *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, pp. 2451–2455. https://doi.org/10.1109/INFCOM.2007.299.

Hammadi, A., & Mhamdi, L. (2014). A survey on architectures and energy efficiency in data center networks. *Computer Communications, 40*, 1–21. https://doi.org/10.1016/j.comcom.2013.11.005.

Härder, T., Hudlet, V., Ou, Y. and Schall, D. (2011) ‘Energy efficiency is not enough, energy proportionality is needed!’, *DASFAA Workshops*, pp. 226–239. https://doi.org/10.1007/978-3-642-20244-5.

Heller, B., Seetharaman, S., Mahadevan, P., Yiakoumis, Y., Sharma, P., Banerjee, S. and McKeown, N. (2010) ‘ElasticTree: saving energy in data center networks’, *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, pp. 17–17. https://doi.org/10.1021/ci2004779.

Hermenier, F., Lorca, X., Menaud, J.-M., Muller, G. and Lawall, J. (2009) ‘Entropy: a consolidation manager for clusters’, *Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments - VEE ‘09*, p. 41. https://doi.org/10.1145/1508293.1508300.

Iec (2013) 'Uninterruptible power systems (UPS)—part 1: general and safety requirements for UPS.' *International Electrotechnical Commission. Reference number.* IEC 62040–1.

International Energy Agency (2014) ‘Energy supply security 2014 Part 3’, *Energy supply security: the emergency response of IEA countries—2014 Edition*, pp. 1–105.

Irwin, D., Sharma, N. and Shenoy, P. (2011) ‘Towards continuous policy-driven demand response in data centers’, *Proceedings of the 2nd ACM SIGCOMM workshop on Green networking - GreenNets ‘11*, p. 19. https://doi.org/10.1145/2018536.2018541.

Islam, M. A., Arafath, M. Y. and Hasan, M. J. (2015) ‘Design of DDR4 SDRAM controller’, *8th International Conference on Electrical and Computer Engineering: Advancing Technology for a Better Tomorrow, ICECE* 2014, pp. 148–151. https://doi.org/10.1109/ICECE.2014.7026950.

JEDEC (2017) ‘JEDEC DDR5 & NVDIMM-P standards under development’, JEDEC’s Server Forum.

Jiang, T., Yu, L. and Cao, Y. (2015) *Energy management of internet data centers in smart grid*. Springer. https://doi.org/10.1007/978-3-662-45676-7.

Jin, M., Feng, W., Liu, P., Marnay, C., & Spanos, C. (2017a). MOD-DR: microgrid optimal dispatch with demand response. *Applied Energy, 187*, 758–776. https://doi.org/10.1016/j.apenergy.2016.11.093.

Jin, M., Feng, W., Marnay, C., & Spanos, C. (2017b). Microgrid to enable optimal distributed energy retail and end-user demand response. *Applied Energy., 210*, 1321–1335. https://doi.org/10.1016/j.apenergy.2017.05.103.

Johannah, J. J., Korah, R., Kalavathy, M., & Sivanandham. (2017). Standby and dynamic power minimization using enhanced hybrid power gating structure for deep-submicron CMOS VLSI. *Microelectronics Journal, 62*, 137–145. https://doi.org/10.1016/j.mejo.2017.02.003.

Josh, W. and Delforge, P. (2014) Data center efficiency assessment: scaling up energy efficiency across the data center industry: evaluating key drivers and barriers. *NRDC and Anthesis*.

Judge, J., Pouchet, J., Ekbote, A. and Dixit, S. (2008) ‘Reducing data center energy consumption’, ASHRAE Journal, 50(November).

Khan, M. U. S., & Khan, S. U. (2015). Smart data center. In *Handbook on data centers* (pp. 247–262). New York: Springer New York. https://doi.org/10.1007/978-1-4939-2092-1_7.

Kim, J. (2016) Strong, thorough, and efficient memory protection against existing and emerging DRAM errors. University of Texas.

Kim, J., Sabry, M. M., Ruggiero, M., & Atienza, D. (2015). Power-thermal modeling and control of energy-efficient servers and datacenters. In *Handbook on data centers* (pp. 857–913). New York: Springer New York. https://doi.org/10.1007/978-1-4939-2092-1_29.

Koomey. (2011). *Growth in data center electricity use 2005 to 2010* (p. 3). Oakland: Analytics Press. https://doi.org/10.1088/1748-9326/3/3/034008.

Krzywda, J., Ali-Eldin, A., Carlson, T. E., Östberg, P.-O., & Elmroth, E. (2017). Power-performance tradeoffs in data center servers: DVFS, CPU pinning, horizontal, and vertical scaling. *Future Generation Computer Systems., 81*, 114–128. https://doi.org/10.1016/j.future.2017.10.044.

Lee, K. P., & Chen, H. L. (2013). Analysis of energy saving potential of air-side free cooling for data centers in worldwide climate zones. *Energy and Buildings, 64*, 103–112. https://doi.org/10.1016/j.enbuild.2013.04.013.

Lei, H., Wang, R., Zhang, T., Liu, Y., & Zha, Y. (2016). A multi-objective co-evolutionary algorithm for energy-efficient scheduling on a green data center. *Computers and Operations Research, 75*, 103–117. https://doi.org/10.1016/j.cor.2016.05.014.

Li, J., Bao, Z., & Li, Z. (2015). Modeling demand response capability by internet data centers processing batch computing jobs. *IEEE Transactions on Smart Grid, 6*(2), 737–747. https://doi.org/10.1109/TSG.2014.2363583.

Lim, S.-H., Huh, J.-S., Kim, Y., & Das, C. R. (2011). Migration, assignment, and scheduling of jobs in virtualized environment. *HotCloud, 2011*, 1–5.

Liu, Z., Wierman, A., Chen, Y., Razon, B. and Chen, N. (2013) 'Data center demand response: avoiding the coincident peak via workload shifting and local generation', Performance Evaluation. Elsevier B.V., 70(10), pp. 770–791. https://doi.org/10.1016/j.peva.2013.08.014.

Liu, Z., Liu, I., Low, S., & Wierman, A. (2014). Pricing data center demand response. *Acm Sigmetrics, 42*, 111–123. https://doi.org/10.1145/2591971.2592004.

Lu, T., Pande, P. P., & Shirazi, B. (2016). A dynamic, compiler guided DVFS mechanism to achieve energy-efficiency in multi-core processors. *Sustainable Computing: Informatics and Systems, 12*, 1–9. https://doi.org/10.1016/j.suscom.2016.04.003.

Masanet, E. R., Brown, R. E., Shehabi, A., Koomey, J. G., & Nordman, B. (2011). Estimating the energy use and efficiency potential of U.S. data centers. *Proceedings of the IEEE, 99*, 1440–1453. https://doi.org/10.1109/JPROC.2011.2155610.

Mazumdar, S., & Pranzo, M. (2017). Power efficient server consolidation for cloud data center. *Future Generation Computer Systems, 70*, 4–16. https://doi.org/10.1016/j.future.2016.12.022.

Moreno-Munoz, A., De La Rosa, J. J. G., Pallarés-Lopez, V., Real-Calvo, R. J., & Gil-De-Castro, A. (2011). Distributed DC-UPS for energy smart buildings. *Energy and Buildings, 43*(1), 93–100. https://doi.org/10.1016/j.enbuild.2010.08.018.

Moura, P., Nuttall, C., Harrison, B., Jehle, C., & de Almeida, A. (2016). Energy savings potential of uninterruptible power supplies in European Union. *Energy Efficiency, 9*(5), 993–1013. https://doi.org/10.1007/s12053-015-9406-7.

Nathuji, R. and Schwan, K. (2007) 'VirtualPower: coordinated power management in virtualized enterprise systems', *Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles - SOSP '07*, p. 265. https://doi.org/10.1145/1294261.1294287.

Nedevschi, S., Popa, L., Iannaccone, G., Ratnasamy, S., & Wetherall, D. (2008). Reducing network energy consumption via sleeping and rate-adaptation. *Symposium a quarterly journal in modern foreign literatures, 21*(3), 323–336. https://doi.org/10.11143/3471.

Ni, J., & Bai, X. (2017). A review of air conditioning energy performance in data centers. *Renewable and Sustainable Energy Reviews, 67*, 625–640. https://doi.org/10.1016/j.rser.2016.09.050.

Oconnell, N., Pinson, P., Madsen, H., & Omalley, M. (2014). Benefits and challenges of electrical demand response: a critical review. *Renewable and Sustainable Energy Reviews, 39*, 686–699. https://doi.org/10.1016/j.rser.2014.07.098.

Oró, E., Depoorter, V., Garcia, A., & Salom, J. (2015). Energy efficiency and renewable energy integration in data centres. Strategies and modelling review. *Renewable and Sustainable Energy Reviews, 429*, 445–445. https://doi.org/10.1016/j.rser.2014.10.035.

Padala, P., Hou, K.-Y., Shin, K. G., Zhu, X., Uysal, M., Wang, Z., Singhal, S. and Merchant, A. (2009) 'Automated control of multiple virtualized resources', *Proceedings of the 4th ACM European conference on Computer systems*, pp. 13–26. https://doi.org/10.1145/1519065.1519068.

Panajotovic, B., Jankovic, M. and Odadzic, B. (2011) 'ICT and smart grid', *2011 10th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, TELSIKS 2011 - Proceedings of Papers*, pp. 118–121. https://doi.org/10.1109/TELSKS.2011.6112018.

Paterakis, N. G., Erdinç, O. and Catalão, J. P. S. (2017) 'An overview of demand response: key-elements and international experience', *Renewable and Sustainable Energy Reviews*, pp. 871–891. https://doi.org/10.1016/j.rser.2016.11.167.

Paul, D., De Zhong, W., & Bose, S. K. (2017). Demand response in data centers through energy-efficient scheduling and simple incentivization. *IEEE Systems Journal, 11*(2), 613–624. https://doi.org/10.1109/JSYST.2015.2476357.

PIER (2008) 'Uninterruptible power supplies: A data center efficiency opportunity. Technical brief. California Energy Commission's Public Interest Energy Research (PIER) Program'.

Pierson, J.-M. (Ed.). (2015). *Large-scale distributed systems and energy efficiency*. Hoboken, NJ, USA: John Wiley & Sons, Inc.. https://doi.org/10.1002/9781118981122.

Pore, M., Abbasi, Z., Gupta, S. K. S., & Varsamopoulos, G. (2015). Techniques to achieve energy proportionality in data centers: a survey. In *Handbook on data centers* (pp. 109–162). New York: Springer. https://doi.org/10.1007/978-1-4939-2092-1_4.

Reiss, C., Tumanov, A., and Ganger, G. (2012) 'Towards understanding heterogeneous clouds at scale: Google trace analysis', *Center for Cloud*.

Ristic, B., Madani, K., & Makuch, Z. (2015). The water footprint of data centers. *Sustainability, 7*(8), 11260–11284. https://doi.org/10.3390/su70811260.

Rong, H., Zhang, H., Xiao, S., Li, C. and Hu, C. (2016) 'Optimizing energy consumption for data centers', Renewable and Sustainable Energy Reviews. Elsevier, 58, pp. 674–691. https://doi.org/10.1016/j.rser.2015.12.283.

Salom, J., Urbaneck, T. and Oró, E. (2017) *Advanced concepts for renewable energy supply of data centres*. River publishers. https://doi.org/10.13052/rp-9788793519411.

Sanders, P., Sivadasan, N. and Skutella, M. (2004) 'Online scheduling with bounded migration', Automata , Languages and Programming, Proceedings, pp. 1111–1122. https://doi.org/10.1287/moor.1090.0381.

Shehabi, A., Tschudi, W. and Gagdil, A. (2007) Data center economizer contamination and humidity study. *Lawrence Berkeley National Laboratory.*

Shehabi, A., Horvath, A., Tschudi, W., Gadgil, A. J., & Nazaroff, W. W. (2008). Particle concentrations in data centers. *Atmospheric Environment, 42*(24), 5978–5990. https://doi.org/10.1016/j.atmosenv.2008.03.049.

Shehabi, A., Smith, S. J., Sartor, D. A., Brown, R. E., Herrlin, M., Koomey, J. G., Masanet, E. R., Horner, N., Azevedo, I. L. and Lintner, W. (2016) 'United States data center energy usage report', *Ernest Orlando Lawrence Barkeley National Laboratory*, (June)

Sheppy, M., Lobato, C., Van Geet, O., Pless, S., Donovan, K., & Chuck, P. (2011). *Reducing data center loads for a large-scale, low-energy office building: NREL's research support facility.* Colorado: Golden.

SIA (2015) ‘Rebooting the IT revolution: a call to action. Semiconductor Industry Association and the Semiconductor Research Corporation’, p. 40.

Siriwardana, J., Jayasekara, S., & Halgamuge, S. K. (2013). Potential of air-side economizers for data center cooling: a case study for key Australian cities. *Applied Energy, 104*, 207–219. https://doi.org/10.1016/j.apenergy.2012.10.046.

Sithimolada, V. and Sauer, P. W. (2010) ‘Facility-level DC vs. typical AC distribution for data centers: a comparative reliability study’, *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, pp. 2102–2107. https://doi.org/10.1109/TENCON.2010.5686625.

Sohel Murshed, S. M., & Nieto de Castro, C. A. (2017). A critical review of traditional and emerging techniques and fluids for electronics cooling. *Renewable and Sustainable Energy Reviews, 78*, 821–833. https://doi.org/10.1016/j.rser.2017.04.112.

SPEC (2016) *SPECpower_ssj2008 benchmark. http://www.spec.org/power_ssj2008/results/. Accessed 14-01-2016.*

Spring, N., Peterson, L., Bavier, A., & Pai, V. (2006). Using PlanetLab for network research: myths, realities, and best practices. *ACM SIGOPS Operating Systems Review, 40*, 17–24. https://doi.org/10.1145/1113361.1113368.

Sueur, E. L. and Heiser, G. (2010) ‘Dynamic voltage and frequency scaling: the laws of diminishing returns’, Proceedings of the 2010 international conference on power aware computing and systems, pp. 1–8.

Tang, C., Dai, M. and Chuang, C.-C. (2012) ‘Demand response control strategies for on-campus small data centers’, in *Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference on*, pp. 217–224. https://doi.org/10.1109/UIC-ATC.2012.97.

Tang, C.-J., Dai, M.-R., Chuang, C.-C., Chiu, Y.-S. and Lin, W. S. (2014) ‘A load control method for small data centers participating in demand response programs’, Future Generation Computer Systems. Elsevier B.V., 32, pp. 232–245. https://doi.org/10.1016/j.future.2013.07.020.

Tang, Y., Sun, H., Wang, X., & Liu, X. (2017). Achieving convergent causal consistency and high availability for cloud storage. *Future Generation Computer Systems, 74*, 20–31. https://doi.org/10.1016/j.future.2017.04.016.

Tesauro, G., Jong, N. K., Das, R. and Bennani, M. N. (2006) ‘A hybrid reinforcement learning approach to autonomic resource allocation’, In *Proceedings of the IEEE International Conference on Autonomic Computing, 2006. ICAC’06.*, pp. 65–73. https://doi.org/10.1109/ICAC.2006.1662383.

The Green Grid (2016) *‘The green grid’. http://www.thegreengrid.org.*

Thekkilakattil, A., Pillai, A. S., Dobrin, R. and Punnekkat, S. (2010) ‘Preemption control using frequency scaling in fixed priority scheduling’, In *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*. IEEE, pp. 281–288. https://doi.org/10.1109/EUC.2010.47.

Tran, N., Ren, S., Han, Z., Man Jang, S., Il Moon, S. and Seon Hong, C. (2014) ‘Demand response of data centers: a real-time pricing game between utilities in smart grid’, *9th International Workshop on Feedback Computing*.

Tsirogiannis, D., Harizopoulos, S. and Shah, M. a (2010) ‘Analyzing the energy efficiency of a database server’, *the 2010 International Conference*, p. 231. https://doi.org/10.1145/1807167.1807194.

Uddin, M., & Rahman, A. A. (2012). Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. *Renewable and Sustainable Energy Reviews, 16*, 4078–4094. https://doi.org/10.1016/j.rser.2012.03.014.

Uddin, M., Darabidarabkhani, Y., Shah, A., & Memon, J. (2015). Evaluating power efficient algorithms for efficiency and carbon emissions in cloud data centers: a review. *Renewable and Sustainable Energy Reviews, 51*, 1553–1563. https://doi.org/10.1016/j.rser.2015.07.061.

Urgaonkar, B., Shenoy, P., Chandra, A., Goyal, P., & Wood, T. (2008). Agile dynamic provisioning of multi-tier internet applications. *ACM Transactions on Autonomous and Adaptive Systems, 3*(1), 1–39. https://doi.org/10.1145/1342171.1342172.

Van Heddeghem, W., Lambert, S., Lannoo, B., Colle, D., Pickavet, M. and Demeester, P. (2014) ‘Trends in worldwide ICT electricity consumption from 2007 to 2012’, Computer Communications Elsevier B.V., 50, pp. 64–76. https://doi.org/10.1016/j.comcom.2014.02.008.

Vanson-Bourne (2011) V-index: virtualization penetration rate in the enterprise. Available at: http://www.v-index.com/full-report.html.

Varrette, S., Bouvry, P., Jarus, M., & Oleksiak, A. (2015). Energy efficiency in HPC data centers: latest advances to build the path to Exascale. In *Handbook on data centers* (pp. 81–107). New York: Springer. https://doi.org/10.1007/978-1-4939-2092-1_3.

Vasques, T., Moura, P. and Almeida, A. de (2017) ‘Energy efficiency insight into small and medium data centres: a comparative analysis based on a survey’, 13th European Council for an Energy Efficient Economy Summer Study on energy efficiency (ECEEE 2017), Presqu’île de Giens (France), 29th May - 3rd June.

Verma, A., Koller, R., Useche, L. and Rangaswami, R. (2010) ‘SRCMap: Energy proportional storage using dynamic consolidation’, *FAST’10 Proceedings of the 8th USENIX conference on File and storage technologies*, (VM), p. 20.

vor dem Berge, M., Buchholz, J., Cupertino, L., Da Costa, G., Donoghue, A., Gallizo, G., Jarus, M., Lopez, L., Oleksiak, A., Pages, E., Piątek, W., Pierson, J.-M., Piontek, T., Rathgeb, D., Salom, J., Sisó, L., Volk, E., Wössner, U., & Zilio, T. (2015). CoolEmAll: models and tools for planning and operating energy efficient data Centres. In *Handbook on data centers* (pp. 191–245). New York: Springer New York. https://doi.org/10.1007/978-1-4939-2092-1_6.

Wang, C. and De Groot, M. (2013) ‘Enabling demand response in a computer cluster’, *2013 IEEE International Conference on Smart Grid Communications, SmartGridComm* 2013, pp. 181–186. https://doi.org/10.1109/SmartGridComm.2013.6687954.

Wang, X., & Wang, Y. (2011). Coordinating power control and performance management for virtualized server clusters. *IEEE Transactions on Parallel and Distributed Systems, 22*(2), 245–259. https://doi.org/10.1109/TPDS.2010.91.

Wang, P., Huang, J. Y., Ding, Y., Loh, P. and Goel, L. (2011) ‘Demand side load management of smart grids using intelligent trading/metering/billing system’, *2011 IEEE Trondheim PowerTech*, pp. 1–6. https://doi.org/10.1109/PTC.2011.6019420.

Wang, D., Ren, C., Sivasubramaniam, A., Urgaonkar, B., & Fathy, H. (2012). Energy storage in datacenters: what, where and how much. *ACM SIGMETRICS Performance Evaluation Review, 40*(1), 187. https://doi.org/10.1145/2318857.2254780.

Wang, C., Urgaonkar, B., Wang, Q. and Kesidis, G. (2013) 'A hierarchical demand response framework for data center power cost optimization under real-world electricity pricing', pp. 1–10.

Wang, H., Huang, J., Lin, X., & Mohsenian-Rad, H. (2016). Proactive demand response for data centers: a win-win solution. *IEEE Transactions on Smart Grid, 7*(3), 1584–1596. https://doi.org/10.1109/TSG.2015.2501808.

Wang, B., Chen, C., He, L., Gao, B., Ren, J., Fu, Z., Fu, S., Hu, Y., & Li, C. T. (2017a). Modelling and developing conflict-aware scheduling on large-scale data centres. *Future Generation Computer Systems., 86*, 995–1007. https://doi.org/10.1016/j.future.2017.07.043.

Wang, Y., Brun, E., Malvagi, F. and Calvin, C. (2017b) 'Competing energy lookup algorithms in Monte Carlo neutron transport calculations and their optimization on CPU and Intel MIC architectures', Journal of Computational Science. https://doi.org/10.1016/j.jocs.2017.01.006.

Whitney, J. and Delforge, P. (2014) 'Data center efficiency assessment scaling up energy efficiency across the data center industry: evaluating key drivers and barriers', (August).

Wiboonrat, M. (2012) 'Next generation data center design under smart grid', Ubiquitous and Future Networks (ICUFN), 2012 Fourth International Conference on, pp. 103–108. https://doi.org/10.1109/ICUFN.2012.6261673.

Wierman, A., Liu, Z., Liu, I. and Mohsenian-Rad, H. (2014) 'Opportunities and challenges for data center demand response', *International Green Computing Conference*. IEEE, pp. 1–10. https://doi.org/10.1109/IGCC.2014.7039172.

Zakarya, M., & Gillam, L. (2017). Energy efficient computing, clusters, grids and clouds: a taxonomy and survey. *Sustainable Computing: Informatics and Systems, 14*, 13–33. https://doi.org/10.1016/j.suscom.2017.03.002.

Zhang, H., Shao, S., Xu, H., Zou, H., & Tian, C. (2014). Free cooling of data centers: a review. *Renewable and Sustainable Energy Reviews, 35*, 171–182. https://doi.org/10.1016/j.rser.2014.04.017.

Zhu, W., Garrett, D., Butkowski, J., & Wang, Y. (2012). Overview of distributive energy storage systems for residential communities', *2012 IEEE Energytech*. *Energytech, 2012*(1), 1–6. https://doi.org/10.1109/EnergyTech.2012.6304671.

Zhu, K., Cui, Z., Wang, Y., Li, H., Zhang, X., & Franke, C. (2017). Estimating the maximum energy-saving potential based on IT load and IT load shifting. *Energy, 138*, 902–909. https://doi.org/10.1016/j.energy.2017.07.092.

Zhuravlev, S., Saez, J. C., Blagodurov, S., Fedorova, A., & Prieto, M. (2013). Survey of energy-cognizant scheduling techniques. *Parallel and Distributed Systems, IEEE Transactions on, 24*(7), 1447–1464. https://doi.org/10.1109/TPDS.2012.20.