Green Data Centers

Robert Basmadjian, Pascal Bouvry, Georges da Costa, László Gyarmati, Dzmitry Kliazovich, Sébastien Lafond, Laurent Lefèvre, Hermann de Meer, Jean-Marc Pierson, Rastin Pries, et al.

CHAPTER 7

GREEN DATA CENTERS

Robert Basmadjian, Pascal Bouvry, Georges Da Costa, László Gyarmati, Dzmitry Kliazovich, Sébastien Lafond, Laurent Lefèvre, Hermann De Meer, Jean-Marc Pierson, Rastin Pries, Jordi Torres, Tuan Anh Trinh, Samee Ullah Khan

COSTIC0804 partners

7.1 Introduction

The widely adoption of the novel Internet services of the last decade, e.g., web 2.0 services, cloud services, and cloud computing, modified the structure of the whole Internet ecosystem. Contrary to the earlier disperse structure, where each service had its own server to be operated on, the infrastructures of the current cloud services are highly centralized; numerous services are run by a single infrastructure. These facilities are commonly known as data centers.

The operators' profit-awareness causes the recent golden age of the data centers. Due to the economics of scale principle, the expenditures (both capital and operational) can be reduced with these highly concentrated architectures. The energy consumption of the data centers is accounted for 15 percent of the total expenditures of a data center [1] while data centers have a non-negligible share of the total energy consumption of the society. Based on the study of J. Kroomey [2], the average power dissipated by data centers was 6.4, 4.7, and 1.8 GW in US, Western Europe, and Japan, respectively in 2005. The energy consumption of data centers was as high as 1.5% of the total energy consumption in the US in 2006. Moreover, these ratios are increasing resulting from the recent data center

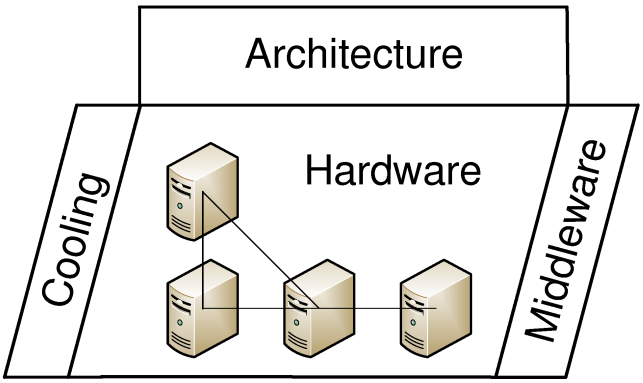

Figure 7.1 Areas where the energy consumption of data centers can be reduced

deployments. In addition the *Efficient Servers* project [3] evaluated the increase of electric energy consumption of servers in Western Europe at 37% between 2003 and 2006 [4]. In 2007 the energy consumed in data centers in Western Europe was 56 TWh and is projected to increase to over 100 TWh per year by 2020 [5]. This will represent about 7 times the capacity of the currently under construction new EPR nuclear reactor in Olkiluoto, Finland.

As the price of electricity is continuously augmenting [1] and the environment aware operation of the companies is becoming more and more desirable by the customers, the energy efficient operation of data centers is required both from a financial and a social viewpoint. Therefore, the operators of the data centers are interested in more energy-efficient data center infrastructures and operations. This is justified by the press releases of leading IT companies; there is a new statement in almost every week.

Data centers have become prevalent in the literature in the recent years; tremendous works have been made towards reducing the energy consumption of the data centers. Albeit this fact, a comprehensive survey of the energy efficiency of the data centers has not been published yet. Therefore, in this chapter we summarize the proposals dealing with the energy consumption and its reduction possibilities. The achievements are presented by the following areas. The energy consumption of data centers' hardware infrastructure is reviewed in Section 7.2. Section 7.3 discusses middleware proposals, which optimize the energy consumption of data centers. Cooling and heat control play a crucial role in the data center facilities: they are necessary to precede hardware failures; however, significant amount of energy is utilized by these equipment. Thus, energy efficient cooling solutions are summarized in Section 7.5. Finally, the properties of data centers' network infrastructures are overviewed because the energy consumption of the switches and routers is not negligible (Section 7.4). The relation of the areas reviewed in this paper is illustrated in Figure 7.1. We hope this survey will serve as a ground that will help the research community to address the open issues of the topic of data centers' energy efficiency.

[1] `http://www.eia.doe.gov/cneaf/electricity/epa/epat7p4.html`

7.2 Overview of energy consumption of hardware infrastructure in data center

7.2.1 Energy consumption rankings and metrics

Data centers are composed of several distributed equipment and infrastructures which contribute to their energy consumption [6]. In most of the papers, we can figure out a portion for chillers, ACDV, power supply, fans and servers. Servers' consumption depends on embedded devices and components [7] like the CPU (37% of power dissipation), memory (17%), PCI slots (23%), motherboard (12%), disk (6%) and fans (5%).

While European Commission has launched the EU code of conduct for data centers [2]; several initiatives propose to evaluate and rank data centers depending on their performance and power usage :

- The TOP500 list that lists the 500 biggest supercomputers in the world. The power used for entire system is also listed in kW. On the June 2010 TOP500 list, the first rank occupied by the Jaguar center from Oakridge National Laboratory uses a power of 6950 kW [3].
- The Green500 [8] ranks supercomputers and also provides the whole consumption of the entire system in kW. The centers are ranked by a metric based on Mflops per watt. Additional lists like the Little List and the HPCC list have recently been added. On the June 2010 Green500 list, the first rank is occupied by the machines from Forschungszentrum Juelich which uses 57.54 kW of power and has an energy efficiency of 773 MFlops per watt [4].
- Through ENERGY STAR Data Center Energy Efficiency Initiatives, the US Environmental Protection Agency proposes to rank data centers depending on their energy efficiency [5]. EPA selected the Power Usage Effectiveness (PUE) as the metric to evaluate data center energy performance. The PUE is a standard industry metric, equal to the total energy consumption of a data center (for all fuels) divided by the energy consumption used for the IT equipment. The PUE generally ranges from 1.25 to 3.0 for most data centers.
- The Green Grid consortium proposes some metrics [9] and tools to evaluate the power efficiency of data centers. [9] re-affirms the use of PUE but redefines its reciprocal as datacenter infrastructure efficiency (DCiE).

Even when not performing any application or services, a data center consumes energy. By correlating usage and energy consumption, we can observe the impact of applications of electrical consumption. As an example, Figure 7.2 presents the energy usage of the Grid5000 [6] [10] site of Lyon (135 computing nodes) on a 6 months period [11]. This

[2] `http://re.jrc.ec.europa.eu/energyefficiency/html/standby_initiative_data_centers.htm`
[3] `http://top500.org`
[4] `http://www.green500.org`
[5] `http://www.energystar.gov/ia/partners/prod_development/downloads/DataCenterRatingGeneral.pdf`
[6] Some experiments of this paper were performed on the Grid5000 platform, an initiative from the French Ministry of Research through the ACI GRID incentive action, INRIA, CNRS, and RENATER and other contributing partners (`http://www.grid5000.fr`)

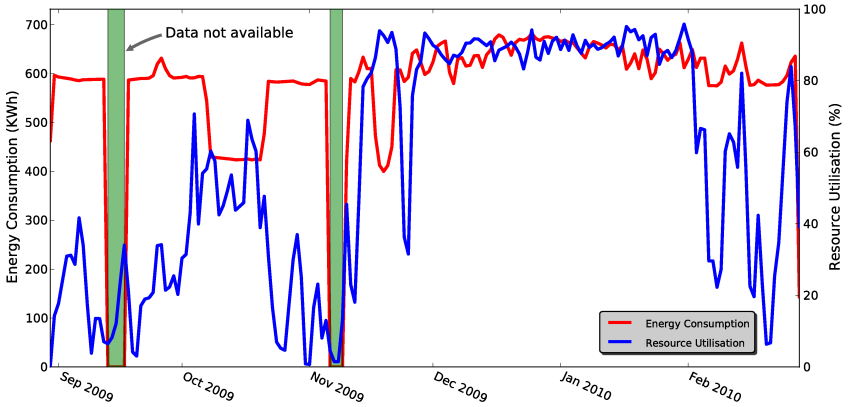

Figure 7.2 Grid5000 Lyon site energy consumption and usage of nodes over six months

figure also presents the resource utilization according to the reservation log obtained from the Resource Management System; the utilization indicates the percentage of reserved nodes, and hence does not imply that CPUs, storage or network resources were used by reservations at the same rate.

Next sections will focus on three main components of the servers: processing, storage, and communicating elements.

7.2.2 Processing: CPU, GPU, and memory

As seen in Section 7.2.1, the processing elements are the main consumers in data centers. Processors (CPU) and memory account together for about 54% of the total consumption, with a rough 37% share for CPU and 17% for memory. When GPUs are present, they can represent up to a tremendous 50% share of the total consumption.

Processors are nowadays all multi-cores, and many-cores are becoming more and more present in data centers. The next generation of powerful machines will embark up to 256 cores in one CPU. Nowadays, most of the data centers rely on four to eight cores per processors. The actual power dissipation processors ranges from 80-100 Watts when idle to 200-250 Watts when loaded, where the consumption of each core is more or less the total divided by the number of cores (but no mechanisms exist for actually measuring it in the processors).

The range of energy consumption of processors can be estimated in two manners: by actual measurements under different circumstances (different loads)—sometimes directly on the main board like in [12]—or indirectly by measuring the total consumption of a node at the plug and deriving from these observations the individual consumptions (see [13] for a model comparison). In the upcoming norm ACPI 4.0 [14], it is possible to get the current power consumptions of individual elements, such as CPU for instance. But no data center is today functioning with such ACPI 4.0 compliant components.

The current ACPI norm embedded in processors found in data centers allows for turning the processor in different operating C-states (C0, C1, C3). C0 is the normal operating state, where the maximum power is consumed. In C1 (halt), the processor is not executing anything, and only a small number of cycles are needed to go back to C0. C3 is the deepest sleep mode, in which more time is necessary to come back to C0, but also where less power

is used for the processor. The worst case wake-up time is provided by the ACPI firmware so that optimized savings can be arranged as a function of further use of the processor. Vendors typically go beyond these states (for instance the C6 in I7 from Intel) but these mechanisms are internal to the processors and can be activated by software. While higher C-states allow for more energy savings, they have to be activated with care, since the energy required to get back to C0 is high: typically, if the time in C6 is too small, the system would actually lose energy. Also, the thermal control zone is used internally to warn operating system when the temperature is getting too high and the state should be changed or even the processor should be stopped immediately. Additionally, P-states indicate at what speed the processor should run: P0 indicates the maximum frequency per voltage combination, but the system can define several Pn with decreasing frequency per voltage values. As the power consumption is a factor of the frequency and of the square of the voltage, these possibilities allow for potential large energy savings [15, 16]. These P-states are controlled by the operating system that reduces the frequency per voltage state under low utilization (and vice-versa). This mechanism, known as DVFS (Dynamic Voltage Frequency Scaling), is the most used one in current data centers middleware, as it will be denoted in Section 7.3.

While traditional data centers are not using GPU or Cells, a current trend for the most powerful computers is to use such alternative hybrid architectures (combining CPU with Cells/GPU) to deliver even more processing power. In the Top500 list, the Chinese Nebulae ranks second (June 2010); it is composed of Intel CPU and Nvidia GPU. In the Green500 list [8], we can find such data centers in the first eight positions. Indeed, from an energy point of view, it can be competitive, since the scheduled jobs finish earlier, energy (which is power $\times$ time) is spent for a shorter time and the Flops/Watt metric reaches 773 MFlops/Watt. Nvidia ships the Tesla GPU Computing Systems, consisting of 1U servers embedding 4 GPUs (for instance the S2050 is delivering 2 TFlops in double precision at the cost of 900 W). Each GPU individually can consume as much as 250 W, for instance the Tesla C2050. The main problem with such infrastructure is when it is idle, since it is not possible to deactivate a GPU card: when installed, it will anyway consume an important minimal amount of power (not less than 50-60 W), and there is no such mechanism to completely switch off GPU elements.

In previous generation multicore processors—still in use in many data centers—it was not possible to manage individually the cores. All cores had to be in the same C- and P-states. Recently, AMD and Intel are producing multicores that allow for a differentiated policy for the different cores (AMD Turbo Core and Intel Throttling in I7 family). Hence, a core can be switched off completely if not needed. New processors even allow a core to increase its frequency above the official maximal frequency (the P0-state) when the overall temperature and power envelop is not exceeded (Turbo Boost Technology from Intel for instance). To the best of our knowledge, no energy consumption comparison has been done with these innovative operating modes.

As already stated, memory banks consume about 17% of the node's consumption. Most of the nodes in a data center have nowadays DRAM DDR3 memory units, composed of memory cells. A memory unit consumes power of basically two different types. First, it always consumes a background power to enable receiving commands (like input/output) and to refresh the data by recharging the capacitors that lose charge over time. Second, it consumes more power when it has to go to the active state (so that it can actually perform data retrieval and communication with outward drivers). On an energy saving point of view, the DVFS and the operating states that we mentioned for the CPU hold also true; hence, a memory can be in different states, each one differentiates with others with the time to come back to operation and the power consumed. Common DDR3 runs at 1.5

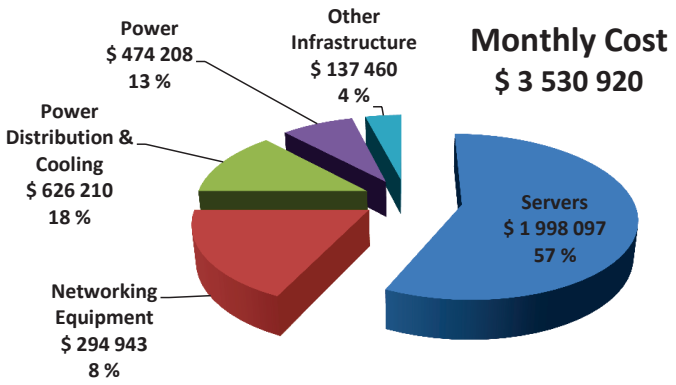

Figure 7.3 Monthly costs for a data center [23, 24]

Volts while Kingston manufacturer has a DDR3 that operates at 1.25 Volts for 1600 MHz, until DDR4 is actually produced in mass, requiring less power (1 Volt) and up to 3200 MHz.

7.2.2.1 Cost and Energy Reduction Evaluation for ARM Based Data Centers Because the processor architectures used in embedded systems have been designed with strong energy efficiency requirements from the beginning, the possibility to use mobile device processors in servers and data centers has lately sparked interest among researchers and the industry. In particular, the feasibility to use ARM based processors has been recently analyzed [17, 18, 19, 20] and few commercial solutions were pushed on the market.

ARM processors are based on Restricted Instruction Set Computer (RISC) CPUs and are therefore designed to operate based on a simplified, highly-optimized and fixed-length set of instructions. Because of the main characteristics of the set of instructions used by ARM CPUs, that are a) one instruction per cycle, b) register to register operations, c) simple address mode and d) simple instruction formats [21], the design of the CPU control unit is considerably simplified and dissipates less power compared to other types of architectures. On the other hand a x86 processor, the over-leading architecture currently found in data centers, is based on Complex Instruction Set Computer (CISC) CPUs using a set of complex instructions of variable length and features multiple addressing modes and multiple instruction formats [22].

In order to evaluate the potential cost savings when using ARM based CPUs in a data center, the overall cost of data centers must be taken into account. Hamilton presents in [23, 24] a cost model for a hypothetical data center and gives a cost comparison between different elements such as infrastructure, networking equipment, servers and power. The model assumes a data center with around 50 000 servers, an overall 10 years infrastructure amortization time, a 4 years amortization time for the networking equipment and a 3 years amortization time for the servers. The model takes into account a five percent yearly interest for the capital used to fund the data center and assumes an energy cost of $0.07 per kWh. An 80% average critical load usage is assumed and a server is assumed to dissipate 165 Watts. The resulting monthly cost of the different cost elements in the data center is shown in Figure 7.3.

Figure 7.3 shows that the direct cost contribution of power accounts for 13% of the total data center cost. However, the power has also an indirect impact on the infrastructure cost as the cooling and power distribution infrastructures are based on the maximal power

Machine	Request / s	Requests / J
Quad Core Intel Xeon E5430 (2.66 GHz)	33000	413
Pentium 4 (2.8GHz)	7100	80
Dual Core Cortex-A9 MPCore (1 GHz)	4600	4600
Quad Core Cortex-A9 MPCore (400 MHz)	3400	2833
Cortex-A8 (600 MHz)	760	760

Table 7.1 Ability of Apache 2.2 to serve a 10 byte static files using different hardware

dissipated by the servers. Therefore, improving the energy consumption of the servers will overall affect 31% of the total data center cost.

Using Hamilton's model, the evaluation of the potential cost reduction when using ARMv7 based CPUs is presented in [17]. In this work a quad-core ARM cortex A9 processor, using a Versatile Express development platform, and a dual-core ARM cortex A9 processor, using a Tegra 200 series developer kit, are evaluated. The versatile express consists of a V2M-P1 motherboard and a CoreTile V2P-CA9 Express A9 MPCore daughter board. The daughter board has 1GB of DDR2 memory and a Cortex A9 NEC CPU clocked at 400MHz. The Tegra 200 series developer kit has a Tegra 250 processor with 1 GB DDR2 memory and is clocked at 1 GHz. Three benchmarks representing typical applications found in data centers and server farms were evaluated on these two platforms: a) Autobench to evaluate the performance of the Apache 2.2 HTTP server, b) SPECweb2005 and c) Erlang run time system.

Table 7.1 shows how the performance and energy efficiency of the Cortex-A9 based platforms for traditional server tasks compared to x86 machines. These results are obtained for Apache 2.2 serving a 10 Bytes static file. From Table 7.1 although the quad-core Intel Xeon platform can handle 7 times more request per second than the dual-core Cortex A9, we notice that the ARM based processor provides a 10 fold better energy efficiency. The SPECweb2005 benchmark was used to evaluate the performance of the Tegra 250 processor with more demanding web services. SPECweb2005 consists of a set of three different workloads: support, ecommerce and banking. The support simulates the workload of a hypothetical customer support web service, the ecommerce workload emulates a web based shopping system and the banking an online banking system. Table 7.2 presents the performance of SPECweb2005 on two x86 machines and the Tegra 250 while Table 7.3 gives the corresponding energy efficiency of the platforms. Two Xeon X3360 machines are used as references in this comparison, the second one having an optimized disk architecture to serve the data requested by the benchmarks. The optimized Xeon X3360 machine can sustain around 33 times more sessions compared to the Tegra 250 platform, but provides a 3 times lower power efficiency.

Finally, an Erlang based SIP proxy is used as benchmark to evaluate the performance and energy efficiency of the studied processors on a typical telecom application found in data centers. The performance of the proxy was measured based on the maximal number of calls per second the platforms were able to handle, and the corresponding energy efficiency is expressed in number of calls per consumed Joule. The reference x86 machine has two quad-core Intel Xeon L5430 processors clocked at 2.66GHz. The performance results are presented in Table 7.4 and the corresponding energy efficiencies are given by Table 7.5. The reference x86 machine was able to handle 400 calls per second while the quad-core

Machine	Ecommerce	Banking	Support
Quad Core Intel Xeon X3360 (1)	3600	2700	4200
Quad Core Intel Xeon X3360 (2)	7360	6240	7840
Dual Core Cortex-A9 MPCore (1 GHz)	230	180	220

Table 7.2 Number of simultaneous sessions using different hardware

Machine	Ecommerce	Banking	Support
Quad Core Intel Xeon X3360 (1)	38	28	44
Quad Core Intel Xeon X3360 (2)	77	66	83
Dual Core Cortex-A9 MPCore	230	180	220

Table 7.3 Number of simultaneous sessions per dissipated watt

SMP	Intel Xeon L5430 (2.66GHz)	Quad Core Cortex-A9 (400 MHz)	Dual Core Cortex-A9 (1 GHz)
1	130	5	5
2	240	12	13
4	350	30	13
8	400	30	13

Table 7.4 Maximum number of calls per second handled by the Erland SIP-Proxy

SMP	Intel Xeon L5430 (2.66GHz)	Quad Core Cortex-A9 (400 MHz)	Dual Core Cortex-A9 (1 GHz)
1	2,6	4	5
2	4,8	10	13
4	7	25	13
8	8	25	13

Table 7.5 Energy efficiency in number of calls per Joule

cortex A9 was able to handle 30 calls per second with 8 schedulers (SMP) as using more schedulers than the number of available physical CPUs does not bring any performance increase. This leads to an energy efficiency of 25 calls per Joule for the quad-core Cortex A9 versus 8 calls per Joule for the Xeon machine.

Using the energy efficiency results of Tables 7.3 and 7.5 in the cost model proposed by Hamilton, we can evaluate the cost saving potential of using ARM cortex A9 processors over the overall data center cost at around 10% for the Erlang SIP proxy and 12,7% for web services represented by the SPECweb2005 benchmarks. In terms of financial cost benefits,

this leads to respectively a $ 350 000 and $ 448 000 monthly cost reduction or a $12,6M and $16,1M cost reduction over the 3 years amortization time for the servers.

Following the demonstration of the cost and energy reduction potential of ARM based data centers, a set of commercial solutions appeared on the market. In 2011 Sandia National Laboratories demonstrated a mini supercomputer based on 196 Cortex-A8 CPUs using the Texas Instrument OMAP3530 chip. The company Calxeda is currently shipping the EnergyCore ECX-1000 chip containing a quad-core Cortex A9 processor as well as a Quad-Node EnergyCard embedding four EnergyCore ECX-1000 chips. The EnergyCore and EnergyCard from Calxeda are directly targeting the data center market.

Recently, a joint initiative in the European funded project EuroCloud pushes ARM Cortex A9 processors, linked with 3D DRAM to create 3D server on chip, serving as a basis for compact and energy efficient data centers [25]. Also within the European funded Mont-Blanc project [26] the Barcelona Supercomputer Center is currently evaluating ARM based supercomputers consisting of prototype boards using Nvidia's Tegra 3 (quad-core Cortex-A9 CPUs) and Samsung Exynos 5 (Dual-core ARM Cortex-A15 CPUs) processors. The Mont-Blanc project aims at designing a new type of computer architecture capable of setting future global High Performance Computing (HPC) standards that will deliver Exascale performance while using 15 to 30 time less energy.

Although a few ARM based commercial solutions targeting data centers and server farms were lately pushed on the market, much expectation is put on the future ARMv8 architectures. The industry is already working on the design of 64 bit 3D many core processors based on the ARMv8 architectures and predict energy efficient cloud data centers of several hundreds of server-in-a-single chip achieving thousands of cores on a single board [19].

7.2.3 Storage

Storage is an important feature of a data center. With estimates foreseeing a growth of 50% of data centers requirements in terms of storage in the next years, it shall continue to draw attention.

Different technologies co-exist for storing data in data centers. Most of the time, a NAS (Network Attached Storage) is present, in order to concentrate the data outside the working nodes, while these nodes keep temporary data and their operating systems. Another possibility is to use a SAN (Storage Area Network) that allows to share and coordinate distributed disks. The difference lies in the access pattern: in a SAN, the devices are directly addressed by blocks by the file system of the nodes, acting as if the distant disk is present locally, while in a NAS, an explicit communication protocol has to be set up over IP, like NFS for instance. Since NAS and SAN involve technologies related to networks, we will let the communication part to the next section and we will focus here on the storage devices themselves.

Traditional hard disk drives (HDD) constitute the most prominent technologies, while solid state drive (SSD) based on flash memory is becoming more and more attracting, together with Hybrid HDDs technologies (magnetic rotating drives like HDD combined with SSD for a part of application/data often used). This later technology is not yet applied in data centers; thus, we will not detail it here.

Despite its 50 years age, HDDs are still widely in use in today businesses. This technology is based on continuous rotating disk platters and a disk head that is positioned dynamically beneath the disk at the right location to read the bytes of data. The disk controls the platters spin and the communication channel with the host. Together these functions

consume 2/3 of the power consumption, even when the activity is high. For reducing energy, three possibilities exist. The first one consists in reducing the spin speed in terms of RPM (Revolutions per Minute). Typically, disks can reduce from 7200 RPM to 5400 RPM when idle. As the power dissipation is quadratic in the speed of rotation, the saving can be as much as half the regular idle consumption. A second mean is the smart control of disk head positioning, reducing the speed of heads so that they arrive just in time when bytes to read are beneath them (for instance with Seagate Just In Time mode). To do so, the power current is reduced; inducing a reduction in energy consumed. The SATA specification [27] includes a Automatic Acoustic Management that allows to reduce the speed of seek operations (resulting in reduced power dissipation). Third, several operating modes exist for disks, including predetermined modes with reduced speed. The SATA specification on Advanced Power Management allows the disk to move from one mode to another automatically after a predefined idle-time or after host and operating system decisions. Typical HDDs consume (numbers for a 2 TB Seagate Constellation ES at 7200 RPM, SATA, 140 MB/s transfer rate, 3”5) about 7 Watts when idle, and 10-11 watts when busy (read operations being more power consuming). Lower disk capacities consume less power, down to 4.6 / 9.4 / 8.2 Watts (idle / read / write) for a 500 GB HDD. On these disks for instance, a PowerChoice mode (a proprietary implementation of T10 and T13 Standards [28, 27]) makes the disk power consumption drop down to 0.53 Watts. Smaller disks (2”5) formally only in laptops, are now getting much interest from data centers despite their more limited capacities at comparable performances. They run at about half the power of 3”5 disks and takes less space in the racks. For instance, the Savvio (15K.2, 15000 RPM, SAS) offers 146 GB only but consumes 4.1 watts when idle.

SDDs are garnering much interests in the last years. Their most important feature is the improved access time. A multilevel cell (MLC) SSD has an access time of 0.5ms compared to an access time of 15.7 ms for a 7,200 RPM HDD. Please note that the highest performances coming for SDD access rate can be limited from an application point of view in some cases (see study on the comparison metrics[29]). As no mechanics exist in a SSD drive, the power consumption is only a fraction of the one of a HDD. A typical Seagate SSD drive (Pulsar, 200 GB, SATA, 300 MB/s) consumes only 0.75 watt when idle and 1.3 watt in operation. This improved energy performance comes with a higher price and limited capacities, making them not really sustainable in big data centers that host Tera or Peta Bytes of data.

7.2.4 Communicating elements

While Section 7.4 presents Data Center Network architectures and their relevant costs; this section focuses on associated network equipment costs in terms of energy. While networks are not main energy consumer equipment in data centers [7], this infrastructure is part of the whole consumption of the system (Network Interface cards, switches, routers, wired links).

Data centers mainly use Ethernet technology as the basic block for communicating equipment. Through the IEEE P802.3az Energy Efficient Ethernet Task Force [7], a consortium mixing academic and industries is proposing new solutions for obtaining Energy Efficient Ethernet solutions. Today, energy consumption of Ethernet networks is not greatly linked with bandwidth utilization. So even in low or no usage context, networks equipment

[7] `http://www.ieee802.org/3/az/`

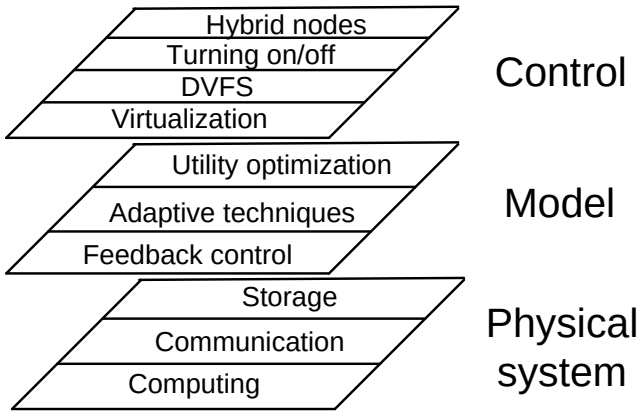

Figure 7.4 Areas where the power consumption of data centers can be reduced

consume energy at high level. As a first approach, by proposing Adaptive Link Rate solutions, energy savings can be obtained by quickly changing the speed of network links in response to the amount of data that is being transmitted. Now, for high speed Ethernet networks (1 and 10 Gbits) used in data centers, the Energy Efficient Ethernet Task Force is proposing low power idle modes which should allow to power down and quickly wake up specific components of Ethernet products.

7.3 Middleware solutions that regulate and optimize the energy consumption in data centers

7.3.1 An overview of the middleware

For many years, research in middleware mainly focused on performance management [30]. However, middlewares are currently challenged to rethink the resource/service management strategies to add energy efficiency to the list of critical operating parameters to control, already including availability, reliability, and performance. The energy parameter has been included in a decisive way shifting the paradigm from 'time to solution' to 'kWh to solution'.

However, considering power management at middleware level is not a new issue in the resource management arena. There are several works from some years ago proposing energy management for servers that focus on applying energy optimization techniques in multiprocessor environments [31, 32]. The proposals range from load balancing for power and performance optimization [33] to economical approaches for managing shared server resources, e.g. [34], where Chase et al. use a greedy resource allocation distributing a web workload among different servers assigned to each service.

Trying to tidy up the research work done until now in this area, we could consider two important aspects in order to classify the existing literature: (1) the system modeling used for making decisions; and (2) the set of control mechanisms required to make decisions effective. Figure 7.4 shows the relation of the physical system, the models, and the controlling mechanisms.

A number of companies, such as Symantec [35], Aperture [36], RackWise [37], iTracs [38], CISCO [39], nlyte [40], Intel [41], HP [41], BMC [42], egenera [41], and Specorp [35] offer commercial software to manage and optimize data centers. Because none of the aforementioned companies have made their solutions transparently available, it is an extremely difficult and perhaps impossible exercise to quantitatively compare middleware systems. The fact that current data centers (and data center middleware) are always designed to and operated at peak performance, such a practice entails promising extensions and exploitation of fundamental interdisciplinary concepts that would further reduce the overall energy consumption of a data center. From scientific literature, we find an isolated problem solving approach, i.e., energy, power, or thermal aspects related to a data center (or a large-scale computing system) are tackled separately in the context of computing, storage, and communications. Therefore, below we give an overview of the state-of-the-art in the three aforementioned categories.

7.3.1.1 Computing Both independent [43] and precedence [44] task models have been considered on uniprocessor [45] or multiprocessor [46] systems using static [47] and dynamic scheduling [37]. The aforementioned models have been treated to present works in the domain of energy-efficient high performance computing [48], web servers [49], computational grids [50], data centers [51], cluster computing [46], and cloud computing [52]. Mapping methodologies [53] based on a given application [54] and machine [55] load have been considered. Energy efficiency also has been the focus for application placement [56], task duplication [35], and task migration [57] models. A majority of the aforementioned works have either used dynamic voltage (or frequency) scaling (DV/FS) [37] as a medium to exploit the complex relationship between processor speed, power dissipation, and energy consumption or dynamic power management (DPM) [58] to completely shutdown processing units.

7.3.1.2 Communications Achieving energy or power efficiency in communication medium is difficult because accurate knowledge [52] about the communications or prediction mechanisms [40] to project communications must be in place [35]. Earlier works opted to treat the entire communication fabric as a uniform medium [48]. Thereafter, DV/FS [59] and dynamic network shutdown (DNS) [60] that is analogous to the DPM technique were introduced to effectively regulate power consumptions [61]. The critical drawback of the aforementioned methodologies is the required additional complex hardware modifications. In contrast, on-off links require much simpler hardware [62] and have been reported to have comparable performance with previous counterparts but reduced switching overhead. The major challenges in energy or power efficient communication fabric include (but not limited to) connectivity, potential network deadlocks, and rerouting when links are asleep [41]. More recently, inspired by the work in fault tolerant routing protocols, incurious researchers have focused on steering network traffic and providing network connectivity as links shut down to save power [54, 63, 64]. However, fault tolerant approaches are reactive and merely performance oriented. The aforementioned methodologies have thus far not been utilized in the context of large-scale distributed computing systems and in particular data centers.

7.3.1.3 Storage The DPM scheme, being the only applicable mechanism in the storage domain [65], covers three levels cache, memory, and disk, which use hardware power management features, such as multiple power states, DRAMs [51], and multiple spin speed hard disks [66]. Moreover, RDRAM chip [67] (if used) in a memory system can be set to an appropriate power state independently. Thereby, enabling dynamic switching of RDRAM

with power-aware page allocation in the operating system [68] becomes a feasible solution. However, page misses may hamper the ability to successfully make the system energy-efficient [38]. It has been noted that a large portion of the power budget goes into disk assessing [68, 69, 64].

7.3.2 System modeling

Middleware requires models that capture the most important factors of the systems while allowing abstract reasoning. The models will allow formalizing behaviors and interactions that help the use of optimization techniques (from simple heuristics to complex techniques) based on what-if predicting techniques. It is important to remark that optimizations at different system levels interfere between each other. This makes the behavior of the current systems unmanageable at execution time. This requires novel optimization techniques that implements self-* properties at run time. These autonomic techniques have been developed to manage workload fluctuations and to determine optimal trade-offs between performance and energy costs.

Taking into account the techniques involved in decision-making, we can group the relevant related work in three main groups (although there can be orthogonalities among them): feedback control theory, adaptive techniques, and utility-based optimization techniques.

7.3.2.1 Feedback control theory Feedback control theory, where a controller manipulates the inputs of a system to obtain the desired effect on the output of the system. The main advantage of this approach is that it guarantees system stability. Furthermore, upon a change in workloads, these mechanisms can accurately model transient behavior and can adjust the system configuration within the time frame of a transitory. Most control theoretic approaches adopt system identification techniques to build linear time invariant models and then apply classical proportional integral differential control. Kusic et al. [70] implement a limited lookahead controller to determine the servers in active state, the operating frequency, and the placement of virtual machines on physical servers. Kalyvianaki et al. [71] propose the use of Kalman filters to track and control the CPU utilization in virtualized environments to guide capacity allocation. Analogously, Raghavendra et al. [72] propose a control-oriented framework to coordinate different kinds of power managers.

7.3.2.2 Adaptive techniques Adaptive techniques, where the learning process is based on the live systems, do not require an analytical model of the system. For example, Tesauro et al. [73] present a capacity allocation technique that determines the assignment of physical servers to applications that maximizes the fulfillment of SLAs. Kephart et al. [74] apply machine learning to coordinate multiple autonomic managers with different goals. A recognized advantage of machine learning techniques is that they accurately capture system behavior with little built-in system-specific knowledge. Reinforcement Learning approaches have also been used to reduce power consumption in clusters. Tesauro, Kephart et al. [75, 74] present a reinforcement learning approach to simultaneous online management of both performance and power consumption. These approaches look at learning what policies should be applied given a system status.

7.3.2.3 Utility-based optimization techniques Utility-based optimization techniques, introduced to optimize users' satisfaction by expressing their goals in terms of user-level performance metrics. For example, a server consolidation project on blade servers based on a power budget mechanism is presented by Ranganathan et al. [76], while Choi et

al. [77] provide power budget policies for virtualized environments and an accurate model to predict server system average power consumption. Yiyu et al. [78] integrate utility-based and control oriented techniques for energy management in hosting centers.

7.3.3 Control Mechanisms

Middleware requires new advanced management mechanisms to provide the necessary control knobs to successfully manage the resources in order to add energy efficiency as an operating parameter. Today most common techniques used in the research literature of the area can be summarized as virtualization, turning on/off servers, dynamic voltage, frequency scaling, and hybrid nodes/hybrid DC.

7.3.3.1 Virtualization Virtualization is a key strategy to reduce power consumption. With virtualization, multiple virtual servers can be hosted on a smaller number of more powerful physical servers, using less electricity.

Virtualization is a mechanism currently used for consolidation. Petrucci et al. [79] propose a dynamic configuration approach for power optimization in virtualized server clusters and outlines an algorithm to dynamically manage the virtualized server cluster. Following the same idea, Liu et al. [80] aim to reduce virtualized data center power consumption by supporting VM migration and VM placement optimization while reducing the human intervention, but no evaluation is provided. Other work of Verma et al. [81] also proposes a virtualization aware adaptive consolidation approach, measuring energy costs executing a given set of applications.

7.3.3.2 Dynamic voltage and frequency scaling Dynamic Voltage and Frequency Scaling (DVFS) allows the reduction of voltage and frequency providing substantial saving in power at the cost of slower program execution. Current microprocessors allow power management by DVFS. DVFS offers dynamic adjustment of supply voltage to the minimum level required for processing elements to operate at a desired clock frequency. Voltage scaling has been widely acknowledged as a very powerful, flexible, and feasible technique for trading off power consumption for execution time.

Depending on the type of tasks executed DVFS approach can be classified into different categories.

The work reported in [82] was the first to characterize a convex function that optimized energy consumption of a set of independent tasks. The work was further extended by Hong et al. [83] that provided a heuristic (for a similar problem) for a fixed priority static scheduling. In continuum, an energy-aware resource allocation heuristic for non-preemptive scheduling was proposed by Quan and Hu [84]. Manzak and Chakrabarti [85] pointed out that extreme variations in power consumption and tasks invalidate the conclusion provided in [37] that uniform voltage scaling was the optimal procedure. To circumvent such an anomaly, the work in [86] reported an iterative slack allocation algorithm based on the Lagrange multiplier method. It is worth addressing the DVS based techniques for soft real-time systems. In such systems, it is not required to fulfill deadlines; therefore, negating the purpose of using deadlines as a criterion for optimization. The DVS techniques for soft real-time systems need to trade-off power savings for average response times for tasks. Therefore, one possible application of such an academic problem could be the conception of energy-efficient Web services.

For scheduling tasks with precedence relationships, Bambha et al. [87] use a combined global/local search strategy. It uses a genetic algorithm combined with simulated annealing for global search, and hill-climbing coupled with Monte Carlo techniques for local

search. Zhang et al. [88] formulate the problem as a Linear Programming (LP) for continuous voltage levels, which can be solved in polynomial time. The work of Gruian and Kuchcinski [89] proposes a scheduling heuristic with a special priority function to trade-off energy reduction for processing delay. The schedule is constructed step-by-step. At each step, a ready task is selected based on an assigned priority and scheduled in the timestamp at which the partial schedule can achieve a maximal probabilistic energy reduction. The complexity of this approach is high due to the number of discrete time steps that must be evaluated in scheduling a task. Moreover, probabilistic evaluation of energy reduction of a partial schedule does not necessarily yield the best decision for the final schedule. Seredynski et al. [90] use a genetic algorithm to optimize task assignment, scheduling the task execution order, and infatuated slack allocation scheme that advocates a small time unit to the task that leads to the most energy reduction in each step. The work reported in [91] alters communication speed selection for communication paths and DVS on processors to achieve a trade-off between communication and computation power. This is the only work that tries to combine the two necessary computing elements (processing elements and communication paths).

The general facility to reduce energy consumption using hardware supporting multiple operating states is introduced in [92]. Ge et al. [93] classified the impact of using DVFS for different application types. This feature could be used by the middleware, e.g. [74], where the authors use frequency scaling in a scheme that trades off web application performance and power usage while coordinating multiple autonomic managers.

7.3.3.3 Turning on/off Turning on/off servers allows that the overall consumption can be reduced through consolidation. Khargharia et al. [94] introduce a theoretical methodology for autonomic power and performance management in e-business data centers. They optimize the performance per Watt at each level of the hierarchy while maintaining scalability. The authors opt for a mathematically-rigorous optimization approach that minimizes wasted power while meeting performance constraints. Petrucci et al. [95] developed a mixed integer programming formulation to dynamically configure the consolidation of multiple services/applications in a virtualized server cluster focused on Web workloads. The approach is power efficiency centered and takes into account the cost of turning on/off the servers. Berral et al. [96] propose a framework that provides an intelligent consolidation methodology using different techniques such as turning on/off machines, power-aware consolidation algorithms, and machine learning techniques to deal with uncertain information while maximizing performance. Other approaches dealing with uncertainty are [97], where statistic methods based on correlation are used to predict usage and so to consolidate works.

7.3.3.4 Hybrid nodes/hybrid Data Centers The Hybrid nodes/hybrid Data Centers mixes low power systems and high performance ones in the same node/data center, offering more control to the management middleware. Today a good approach for energy saving is to have a middleware that can manage a hybrid data center architecture that mixes low power systems and high performance ones in the same data center [98, 99]. Filani et al. [100] offer a solution that includes a platform resident Policy Manager, which monitors power and thermal sensors and enforces platform power and thermal policies. They explain and propose how the PM can be used as the basis of a data center power management solution.

7.3.4 A use case of leveraging energy efficiency in data centers

In this section, we present a middleware solution that takes into account the aforementioned modeling techniques as well as controlling mechanisms. More precisely, the corresponding middleware was realized within the context of EU FP7 FIT4Green[8] project. It has as an aim of reducing the CO_2 emissions as well as the energy consumption of data centers' ICT resources by 20% which will have an indirect impact on the energy use of cooling systems.

7.3.4.1 Concept As mentioned above, the proposed approach tackles the problem by reducing the carbon footprint of data centers through the deployment of ICT technology. There are various approaches of increasing the energy efficiency in data centers. Most of them are hardware oriented through investing in energy-efficient IT equipment or HVAC (heat, ventilation, air condition). Success in these areas, however, can only be incremental, as the capital cost of replacing old equipment is high. Therefore, the proposed solution is based on a different perspective: independently of the current IT and HVAC infrastructure, an energy-aware middleware is proposed that re-arranges the workload in a data center and among a federation of data centers according to the optimal energy and/or CO_2 emissions efficiency. The middleware is designed agnostic of the existing data center automation and management frameworks and takes into account not only transferring workload to the most efficient clusters in a data center, but also re-allocating workload within a federation of data centers with the ultimate objective of reducing the global energy and/or CO_2 emissions. It is worth pointing out that the devised plug-in is suitable for any computing style being traditional, supercomputing or cloud computing.

7.3.4.2 Implementation The cornerstone of the proposed approach is a set of energy optimization algorithms (e.g. policies) that reallocates the workload (e.g. virtual machines, jobs, etc.) by taking into account technical Service Level Agreements (SLAs) and other restrictions, to optimize energy and/or CO_2 emissions through two basic procedures: With the so-called textquotedblleft global optimization",the algorithms check in regular intervals (e.g. every 5 minutes) the state of the system from energy and ICT load point of view and reorganize the workload in case they calculate a potential energy reduction. Additionally an optimization is carried through every time a new workload enters the system, be it the execution of a batch job in the case of supercomputing data center or the creation of a new virtual machine in the case of a cloud computing data center. Those optimization algorithms are based on Constraint Programming (CP) paradigm. To this end, an innovative architecture was designed, in order to cope with the complexity of the various SLAs and data centerrequirements, as well as the different algorithms available.

However, in order that these optimization algorithms can take the most suitable energy- and/or CO_2-saving decisions, the existence of accurate power prediction models becomes primordial. To this end, power consumption estimation models for ICT resources such as servers, storage devices and networking equipment were devised.

Since both the optimization algorithms as well as power estimation models periodically check the state of the data center, a detailed description of data centers' ICT resources is provided with their relevant energy-related attributes and interconnections. The identified energy-related attributes is classified into two classes: Dynamic and static. The former denotes the fact that the value of the attribute changes dynamically and it needs to be kept up-to-date through the data centers' monitoring framework. On the other hand, static

[8] http://www.fit4green.eu/

attributes are those whose value remains constant; most of the times, the values of static attributes can be obtained from the manufacturersdata sheet.

7.3.4.3 Obtained results In order to evaluate the impact of the proposed energy-aware middleware and demonstrate that it works agnostic of the existing data center framework, the choice of three testbeds representing three different computing styles suggested itself. One testbed is a traditional data center that provides business services to internal customers, the other is a scientific supercomputing center and the final one is a cloud computing data center offering IaaS platform. For each testbed, different scenarios were created by taking into account two cases: single-site data center and federation of data centers. In the traditional data center, which is provided by ENI in Italy, the workload is characterized by two peaks occurring at the beginning (between 8-10 AM) of the simulated working day and at the end (between 4-6 PM), with a dip at lunchtime. Even in the single-site case, the proposed middleware managed to reduce by 30% the average consumption by semi-automatically shutting down servers during the times of low-utilization. In the federation case, more savings were achieved in terms of energy consumption then in the single-site case which ranged between 28% – 50%. In the supercomputing testbed – the Forschungszentrum Jülich in Germany – the utilization rate of the resources is regularly much higher than in the traditional data center. Thus the potential for shutting down servers and consolidating workload on fewer servers is much reduced. Therefore, savings were 4% to 27% in single site depending on the utilization of the data center, and 30% to 42%, even 52% in the federated site. These savings were based on setting the unused servers to low-power standby mode and by allocating the new jobs to the different data centers in an energy-efficient manner. The cloud computing scenario is represented by HPIS, a laboratory for cloud computing in Milan. As the laboratory does not offer real services, the workload for the IaaS platform was generated synthetically, through the monitoring of real customer activities. The major load generator of this testbed was the allocation of virtual machines which was done based on the identified workload profile. Through the deployment of the proposed middleware, the energy consumption of the testbed was reduced by 10% to 24% in the single site case – with the middleware itself consuming not more than an additional of 3.5% of energy. The ability to exploit the federation as a unique pool of resources at allocation time allows achieved saving to range from 17% to 22%. These energy savings were achieved by allocating the new virtual machines in an energy-efficient manner and by turning off the unused servers. The number of servers was also optimized by using live migration of virtual machines from server to another energy-efficient one.

7.3.4.4 Conclusion and future perspective In the end, it was shown that through optimization algorithms it is possible to reduce energy consumption of ICT sector. Hence, the proposed approach has the following three-dimensional benefits:

1. For the environment: reduction of CO_2 emissions.
2. For the data center businesses:
   - Reduction of costs and therefore prices.
   - Marketing options for green services.
   - Provision of potential energy legislation.
3. For the data center end users: reduction of cost for services.

In order to go one step further from what was achieved by the proposed middleware solution for data centers, the EU FP7 All4Green project takes into account the ecosystem comprising of the following three entities: Energy provider, data centers, and IT customers. More precisely, during power shortage/surplus situations, the energy provider asks for power adaption collaboration (e.g. decrease/increase) from data centers. This can be achieved either through:

- Local flexibilities such as heating up/ cooling down the datacenter (e.g. air conditioner) or discharging/charging the battery (e.g.UPS).
- External flexibilities involving IT customers who willingly cooperate with data centers by accepting reduced QoS related metrics of their services (e.g. workload shedding or shifting).

All these are realized by means of introducing three novel contracts:

1. GreenSupplyDemandAgreement (GreenSDA): It comprises of contractual terms between an energy provider and data centers. For instance, such terms specify:
   - The minimum and maximum power (in kW) to increase and decrease. Also for each power adaption capability, the minimum and maximum duration (in min) is defined.
   - The number of requests an energy provider can send to a data center per month. Also, the number of rejects a data center can send to energy provider per month.
2. GreenServiceLevelAgreement (GreenSLA): It consists of contractual terms between a data center and its IT customers. For instance, such terms specify the flexibilities based on a time period such as: “High availability and performance in working days” and “low availability and performance during nights and weekends”.
3. GreenWorkloadServicesOutsourcingAgreement (GreenWSOA): It comprises of agreements between two data centers that intend to collaborate in improving each other’s (green) performance/efficiency by exchanging workload. By committing to a GreenWSOA, the collaborating datacenters thus become a federation.

The obtained preliminary results were encouraging that show high potential of data centers to participate in Demand Response programs, such that the data centers can reduce their energy consumption by means of energy-aware middlewares as the one presented in this section.

7.4 Data Center Network Architectures

Although the main power consumers in a data center are the servers, the network, including network interface cards and layer 2/3 switches, consumes about 15 % of the total power consumption [1]. Therefore, we take a closer look at the impact of different data center network architectures on the power consumption.

7.4.1 Architectures

Several different network architectures have been proposed for data centers ranging from switch-centric approaches such as butterfly, Clos network, and VL2 to server-centric approaches such as mesh, torus, ring, Hypercube, DCell, and BCube. In this section, we only

highlight the most promising and well-known approaches and evaluate their impact on the total power consumption.

7.4.1.1 Hierarchical Network Architecture Several small and medium data centers today consists of a two-tier or a three-tier network tree topology. An example of a three-tier topology is shown in Figure 7.5. According to [101, 102], a two-tier design supports up to 5000 hosts and a three-tier topology scales up to several ten thousands hosts. A two-tier data center architecture consists of a core tier as root and an access tier with the servers. A three-tier architecture has an additional middle-tier, the aggregation tier. The servers itself are connected via Gigabit Ethernet while 10 Gigabit Ethernet is used for the core and aggregation network. Within the next years, the 10 Gigabit Ethernet connections will be exchanged by 40 Gbps or 100 Gbps links. This reduces the number of core switches or helps to reduce the oversubscription factor. According to [103] paths through the highest levels of the tree are oversubscribed by factors of 1:80 to 1:240. This high oversubscription rate is used to reduce the number of switches in the core and aggregation layer whose costs are about $700,000 for a 128-port 10 Gigabit Ethernet switch.

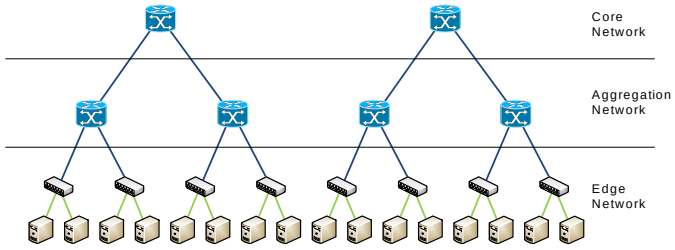

Figure 7.5 Hierarchical data center architecture (three-tier topology)

7.4.1.2 Clos Networks (Fat-tree and VL2) In contrast to the general three-tier topology, a fat-tree topology uses commodity Ethernet switches. The fat-tree architecture was developed to reduce the oversubscription ratio and to remove the single point of failure of the hierarchical architecture. An example of a fat-tree data center architecture is shown in Figure 7.6. Thereby, hosts connected to the same edge switch form their own subnet. Thus, all traffic to the same lower layer switch is switched, whereas all other traffic is routed.

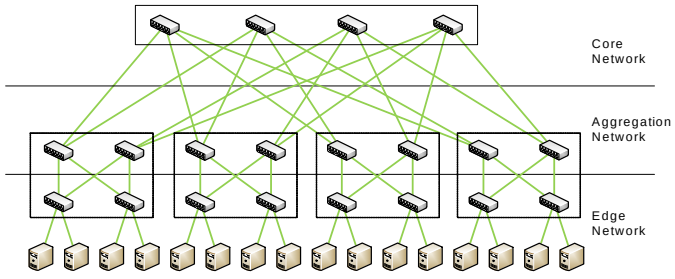

Figure 7.6 Fat-tree data center architecture.

The example in Figure 7.6 shows that fat-tree is a switch-centric structure where the switches are concatenated. The VL2 architecture is quite similar to fat-tree except that fewer cabling is needed. Greenberg et al. [103] claim that switch-to-switch links are faster

than server-to-switch links and therefore use 1 Gbps links between server and switch and 10 Gbps links between the switches. By this, they reduce the number of cables required to implement the Clos topology. However, high end intermediate switches are needed and thus, the trade-off made is the cost of those high-end switches.

7.4.1.3 DCell The DCell data center architecture was developed to provide a scalable infrastructure and to be robust against server failures, link outages, or server-rack failures [104]. A DCell physical structure is a recursively defined architecture whose servers have to be equipped with multiple network ports. Each server is connected to other servers and to a mini switch, cf. Figure 7.7. In the example, $n = 4$ servers are connected to a switch, forming a level-0 DCell. According to [104], n should be chosen ≤ 8 to be able to use commodity 8-port switches with 1 Gbps or 10 Gbps per port. A level-1 DCell is constructed using $n + 1$ level-0 DCells, in our example 5 level-0 DCell form the level-1 DCell. In order to connect the level-0 DCells, each DCell is connected to all other DCells with one link. A level-2 DCell and the level-k DCell are constructed the same way.

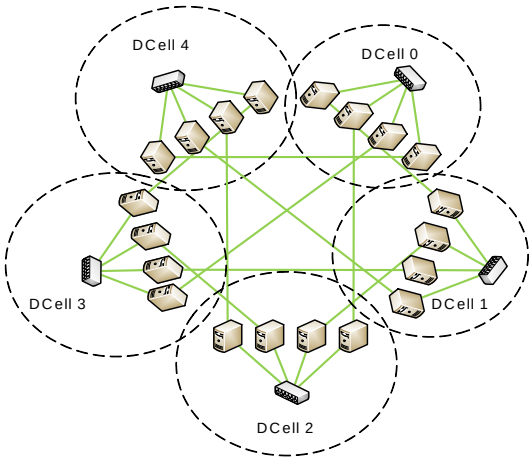

Figure 7.7 DCell data center architecture.

Thus, the DCell architecture is a server-centric structure which uses commodity switches and the fewest number of switches of all presented data center architectures. However, the cabling complexity might prevent large deployments.

7.4.1.4 BCube BCube is similar to the DCell structure, just that the server-to-server connections are replaced by server-to-switch connections for faster processing [105]. Figure 7.8 shows a BCube$_k$ ($k = 1$) architecture with $n = 4$ servers per switch. From the figure we can see that the total number of servers is $N = n^{k+1}$ and each server has to be equipped with $k + 1$ ports. Each level has n^k switches and the total number of levels is $k + 1$.

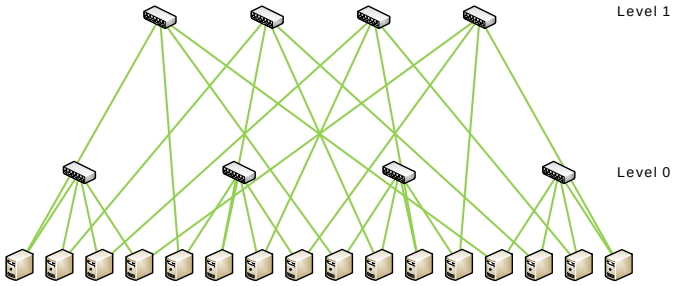

Figure 7.8 BCube data center architecture.

Similar to DCell and in contrast to the fat-tree architecture, BCube is server-oriented and can use existing commodity Ethernet switches.

7.4.1.5 MDCube BCube was designed for intra-container networking with about 2500 servers. In order to connect several containers together, Wu et al. [106] proposed the Modularized Data Center Cube (MDCube). MDCube connects all containers using optical fibers without extra high-end switches or routers. Compared to DCell it reduces the cabling complexity and in comparison to fat-tree, the approach can be built directly with commodity switches without needing any switch upgrades. More details about the construction of an MDCube can be found in [106].

7.4.1.6 High-level properties of the topologies Table 7.6 shows a comparison of the last four presented architectures in terms of performance and costs. Looking at the server-to-server communication, fat-tree achieves the lowest throughput, because each server is only equipped with one port. However, for all-to-all communication, fat-tree performs best. Considering the costs in terms of intra- and inter-container cabling and number of switches, DCell uses the lowest number of switches, while fat-tree uses the largest number. While the cabling costs inside a container are quite similar, MDCube uses the lowest number of cables for inter-container connections.

7.4.2 Power Consumption of Data Center Architectures

Gyarmati and Trinh [107] analyzed four different data center architectures in terms of power consumption. The total power consumption consists of the power requirements of the switches, and the power consumed at the servers that have multiple ports. Thereby, the power consumption of the servers as well as the power consumption of additional devices such as cooling is not taken into account. Table 7.7 shows the power consumption and the

Table 7.6 Performance and cost comparison of different data center architectures [106].

	Fat-tree	DCell	BCube	MDCube
Server-to-server	1	$k'+1$	$k+1$	$log_n t$
All-to-all	N	$\frac{N}{2^{k'}}$	$\frac{n(N-1)}{n-1}$	$\frac{N}{1.5+0.75D}$
Traffic balance	Yes	No	Yes	Yes
Graceful degradation	fair	good	good	good [a]
Switch upgrade	Yes	No	No	No
Inner- cable NO.	$Nlog_{\frac{n}{2}}\frac{t}{2}$	$N\left(\frac{k''}{2}+1\right)$	$Nlog_n t$	$Nlog_n t$
Inter- cable NO.	$Nlog_{\frac{n}{2}}\frac{N}{t}$	$N\left(\frac{k'-k''}{2}\right)$	$Nlog_n\frac{N}{t}$	$\frac{N}{2n}log_n t$
Switches NO.	$\frac{N}{n}log_{\frac{n}{2}}\frac{N}{2}$	$\frac{N}{n}$	$\frac{N}{n}log_n N$	$\frac{N}{n}log_n t$

[a] n is port number of switches. t is the number of servers in a container, while N is the number of all. DCell has $k' = log_2 log_n N$ and $k'' = log_2 log_n t$. MDCube has $log_n t = k+1$.

Table 7.7 Power consumption and diameter of data center architectures [107].

Architecture	Power consumption	Diameter
Tree	$En^k + E_{sw}\sum_{i=0}^{k-1} n^i$	$2k$
Fat-tree	$En^3/4 + E_{sw}[(n/2)^2 + n^2]$	6
DCell	$\approx (E + E_{sw}/n)(n+1)^{2k}$	$s^{k+1}-1$
BCube	$En^{k+1} + E_{sw}\sum_{i=1}^{k+1} n^i$	$k+1$

diameter of four different architectures. The power consumption of a single server and a switch is denoted as E and E_{sw}. It is obvious that the power consumption strongly depends on the number of used ports, denoted by n, and the number of structural levels, denoted by k.

Using these equations from [107], we can see that in small-size data centers, BCell and DCell have roughly the same energy requirements. However, when increasing the number of servers, DCell consumes less power than BCube. The power consumption of the fat-tree architectures is between DCell and BCube. The tree structure of course consumes the fewest power, but is also not robust against link, switch, or port failures.

According to Mahadevan et al. [108], the power consumption of a switch can be further subclassified. The power consumed by a switch depends on the power of the chassis, the power consumption of the linecard as well as the power consumption of different link rates. Looking at Table 7.8, we can see that a 1 Gbps port rack switch consumes almost 5 times more power than a 100 Mbps port.

In the paper, three schemes are presented to reduce the power consumption in a data center. The first scheme is called Link State Adaptation (LSA). In this scheme, the power controller monitors the links and dynamically adapts the line speed to the states *disabled*,

Table 7.8 Switch power consumption [108].

Configuration	Rack switch (in Watts)	Tier-2 switch (in Watts)
$Power_{chassis}$	146	54
$Power_{linecard}$	0 (included in chassis power)	39
$Power_{10Mbps}$ (per port)	0.12	0.42
$Power_{100Mbps}$ (per port)	0.18	0.48
$Power_{1Gbps}$ (per port)	0.87	0.9

10 Mbps, 100 Mbps, or 1 Gbps. However, this line speed adaptation cannot be performed immediately and thus, the delay of the switching has to be taken into account. The second scheme is called Network Traffic Consolidation (NTC). This scheme is also known as Traffic Aggregation Scheme (TAS). Thereby, the traffic in a low-loaded data center is aggregated on a few links while the other links and switches are disabled. Considering a fat-tree, BCube, or DCell architecture, redundant links can also be disabled when not needed. This scheme can reduce the power consumption significantly, while taking into account the trade-off between power savings and availability. The last scheme presented in [108] is the Server Load Consolidation (SLC). Here, server jobs are migrated to fewer servers using virtualization techniques. This is also an indirect way to consolidate network traffic on fewer links and allows a controller to turn off non-utilized ports or switches. However, the energy savings achieved with these three schemes always come along with lower availability and less reliability.

7.4.3 Additional Proposals For Energy-Efficient Data Centers

Finally, we review three proposals dealing with the architecture of the data centers. Albeit these methods are diverse; they all intend to reduce the power consumption of the data center networks. To be more energy efficient, the first one powers off unutilized switches, the second applies residential access gateways to form a data center, while the third introduces a highly scalable and flexible network topology generation method.

7.4.3.1 Elastic Tree All the abovementioned mesh-like approaches (fat-tree, BCube, DCell) except the hierarchical network architecture help to be robust against failures by using more components and more paths. However, as shown in [107], this also increases the power consumption, with the BCube architecture as the largest power consumer. However, although the number of traffic fluctuates during the day, the power consumption is fixed, see e.g. Google production data center [109]. Thus, Heller et al. [109] propose to reduce the power consumption by dynamically turning off switches and links that are not needed. The approach is called Elastic Tree whose underlying topology is a fat-tree. Using a testbed based on OpenFlow, it is shown that in the data center network, up to 60 % power can be saved, depending on the traffic matrices. Safety margins are used to become robust against highly varying traffic fluctuations.

7.4.3.2 Nano Data Centers Nano data centers can be made out of ISP-controlled home gateways to form a distributed, peer-to-peer data center structure [110]. The first order

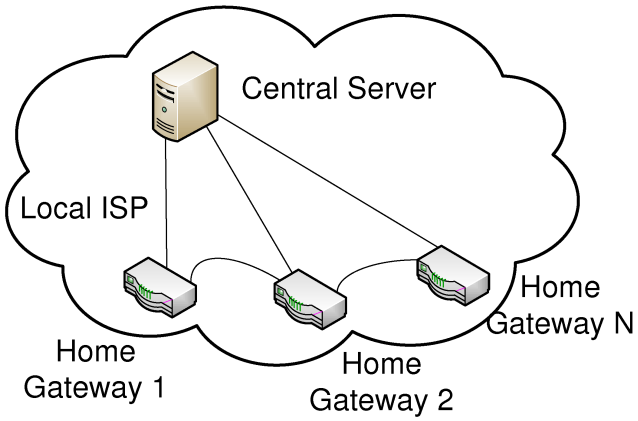

Figure 7.9 The nano data center architecture utilizes the resources of the home gateways of end-users

goal of nano data centers is to form an energy-efficient content delivering data center. To exploit the advantage of the peer-to-peer structure, the users' requests are served from home gateways whenever it is possible; thus, the load of the content servers, located in the facilities of the operator, is decreased. Figure 7.9 illustrates the architecture of the nano data centers.

The architecture shares storage and computational resources among the participants; the solution uses the underutilized resources and the already committed power consumption of the equipment. The energy efficiency of the structure arises from two properties: as the gateways are located in the residence of the subscribers, the heat dissipation is solved without extra cooling facilities; the demand and the services are co-located that reduces the intra-network traffic. Valancius et al. claim that the power consumption can be decreased by at least 20% compared to traditional data center architectures.

7.4.3.3 Scafida A recently proposed data center network generation method [111], called Scafida, offers a highly scalable and flexible design. Scafida is inspired by biological networks, namely scale-free networks, which are energy efficient as they survived the evolutionary competition. The Scafida algorithm generates the data center topology iteratively, i.e., the nodes are added one-by-one to the network. The algorithm's input parameters are the number of servers, the number and type of the switches, and the number of servers' ports; these parameters cause the high scalability and flexibility of Scafida. Due to this, the Scafida algorithm is capable to create data centers out of any set of network switches; accordingly, the operator of the system is able to specify in advance the consumable power of the Scafida data center.

The power consumption of several Scafida topologies is shown in Figure 7.10 by scaling the number of servers within the structure. Topologies are generated with the 5-, 8-, 24-, and 48-port commodity switches; the servers are attached to the network with only one link. The power consumption of Scafida data centers is proportional to the size of the system regardless of the type of the switches; the steps of the plots are only due to the scaling of the simulation parameters. Thus, if Scafida topologies would be generated for all the possible number of servers, the curves of Figure 7.10 would be linear without any

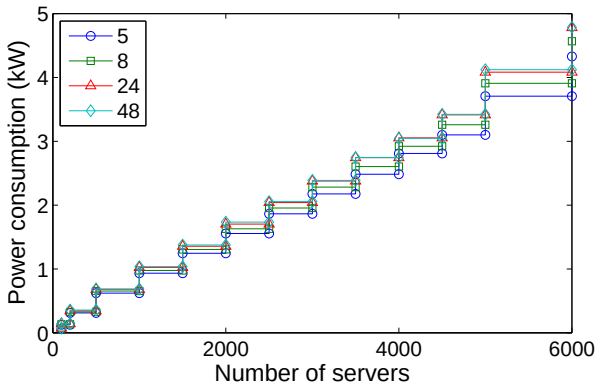

Figure 7.10 Energy proportionality of the Scafida data center networks

significant jumps. This implies that the Scafida data center structure is energy proportional design.

7.5 Solutions for cooling and heat control in data center

Computing equipment dissipate a large amount of heat that is exhausted in the facility. The facility must be maintained to an acceptable level of temperature and humidity. The facility also must be ridiculed of air-born particles and contaminants. We find a rich literature related to climate control dating back to 1970s. However, in this section, we only survey the latest state-of-the-art. Prior to 1980, data centers had many of the same characteristics as those of the modern day facilities with the exception that heat loads were much less. The design considerations have since then not changed drastically as reliability [51], redundancy, maintenance [59], cost, and space remain the primary concerns. Constantly increasing costs and energy consumption of modern cooling systems urge the need for energy-efficient cooling solutions. On the other side overheating of data center components reduces their reliability considerably [51]. Air flow direction has a major effect on the cooling of facilities; therefore, in [112] Obler illustrates various cooling concepts with a few different air flow directions. Air delivery also has been the focus of several works. These works consider whether air should be delivered from overhead or from underneath a raised (false) floor [54] ceiling height requirements that may reduce air stratification [61] raised floor height [112], and proper layout of computing equipment that would reduce hot spots [113].

Data center thermal control solutions can be broadly divided into mechanical- and software-based according to the approach they adapt. Mechanical-based approaches focus on the air flow dynamics for efficient cooling while software-based approaches, being aware of the the thermal properties of the data center infrastructure, optimize the process of the workload scheduling.

7.5.1 Mechanical-based approaches

Mechanical-based approaches can be further divided into: (a) non-raised floor facilities and (b) raised floor facilities. The non-raised floors were the offshoot of the earlier computer

room design. However, quickly it was realized that such a design was not economically feasible for large-scale data centers. Some of the notable works in non-raised floorings include a thorough computational fluid dynamic analysis to determine optimal air flow, and air flow distribution [59]. The raised floors maintain a near layout to all interconnection cabling. Due to the squared dependency between the air flow pressure and its velocity, the airflow patterns remain independent of the flow rate. Air supply through the raised floor along the walls and under the computer equipment with exhaust through a false-ceiling is considered by Grande [114].

The effect of air flow, volume, and tile openings on heat load was considered by Khoshhala et al. [115]. Innovative self-contained air conditioning systems, liquid cooled systems, and chillers with integral air handling and refrigeration mechanisms installed within the facility are discussed in [51, 54, 115, 44]. [51] proposes a framework for the throughput optimization and load balancing of the available power with a focus on systems constrained by the number of power circuits available or having non-uniform power footprint due to the heterogeneous nature of user workloads. Khoshhala et al. [115] propose a system for local cooling demonstrating that due to the high heat transfer rate the inlet air temperature has no significant effect on cooling in certain setups. Wireless sensor networks initially considered for the greenhouse monitoring scenarios can be easily adapted to operate in a data center facility delivering temperature measurements to the main coordination module. Their indoor characteristics are addressed in [44].

In all of the aforementioned methodologies, the key drawbacks are: (a) ”bulk cooling”, which is wasteful, (b) ”threshold cooling”, which is an untimely cooling, and (c) ”uni-methodology cooling”, which does not allow embarrassing state-of-the-art cooling mechanisms.

The efficiency of the heat removal process in a facility is proportional to the available climatic information. Therefore, a low-cost, non-destructive, and readily deployable climatic information gathering wireless sensor network should be developed. The core idea being that in lieu of a central thermostat, distributed temperature sensors are utilized to accurately measure temperature at different locations of the data center that can be as fine-grained as possible. The prior work on wireless sensor network deployment in green houses uses IEEE 802.15.4/ZigBee for (a) measuring substrate water, electrical conductivity, photosynthetic radiation, and leaf wetness [116], (b) regulating climate for rapid melon and cabbage growth [117], (c) measuring soil moisture [44], (d) multi-spectral imaging for cabbages [118], and (e) affect of lighting conditions on ambient temperature [119].

Next generation systems will include a combination of a wireless sensor network and an event-based control system that can effectively and efficiently offer a fine-grained control of the data center atmosphere. For the above, the following issues must be addressed: (a) issues posed to a multivariable, interacting control system by possibly faulty communications, (b) location of sensors to correctly represent, for the purpose of control and spatially distributed quantities, (c) efficient use of actuators to minimize wear, and (d) effects of event-based sampling. The climatic information acquisition coupled with event-based sampling is typically subjected to machine learning based techniques that can proactively control the data center environment. Thermal management of data centers includes: (a) air movement or ventilation, (b) heat rejection, and (c) humidity control. Because Computational Fluid Dynamics (CFD) addresses heat transfer, air movement, and humidification in a unified approach, CFD is excellently suited to address the aforementioned outstanding issues. Thus, applying CFD tools, such as StarCD [120], Fluent [121], or OpenFoam [122] can predict both flow of hot and cold air, and heat transfer in data centers. For the air movement or ventilation, the approach is to predict the distribution of the air velocity in

terms of magnitude and direction within a data center. Because cooling is largely achieved by convection, a sufficient velocity of the air affects the cooling rate to a large extent. The predicted velocity field has to be also analyzed to identify: (a) recirculation zones where no exchange of hot and cold air takes place, (b) dead corners where effectively no or only little air movement occurs, and thus, leads to very reduced cooling rates, (c) bypass air that streams without contact to the equipment, and therefore, does not contribute to cooling, and (d) cold air contamination of coolers takes place if a cooler does not receive the warmest possible air, and thus, operates less efficiently. Based on the aforementioned analysis, measures such as ideal placement of perforated tiles throughout the room can be derived that provide ventilation complying with the requirements of the data centers. This procedure may lead to a scenario of different measures of which all of them are to be evaluated by CFD. Within this process CFD offers an additional advantage because predicting the effect of a proposal involves only a fraction of costs than changing the hardware [116].

Heat rejection is proportional to the temperature difference between the cold air stream and the surface temperature of the equipment. Therefore, the heat rejected can only be assessed by a spatial temperature distribution of the cooling air [112]. Expanding CFD predictions by a prediction of the spatial temperature distribution in conjunction with the spatial distribution of the air velocity will provide a detailed view of local heat transfer rates of the equipment; thereby, identifying hotspots. Once identified, these hotspots can be avoided by (a) reducing the cooling temperature of the air stream that increases the temperature difference, and thus, improves the cooling efficiency and (b) increasing the air velocity in the region of the hot spot that leads also to an improved cooling efficiency. Either measure or a combination of both may be evaluated by a simultaneous prediction of the spatial distribution of both air velocity and temperature by CFD for a better cooling of the equipment [115]. Such a methodology is cost effective because by rating the predicted results of different measures in terms of cooling efficiency allows identifying a solution that perhaps requires minor changes in operation or set-up only or at least involves minimal costs. For humidity control, similar to a CFD prediction that includes air velocity and temperature, the spatial distribution of humidity can be obtained. These results will indicate whether the humidity levels fall within a given recommended range for a data center or otherwise.

7.5.2 Software-based approaches

Software-based approaches aim at minimizing costs associated with data center cooling expenses by intelligent scheduling of incoming jobs. Typically, the policy of the software-based approaches focuses on (a) preventing a server temperature crossing a predefined threshold and (b) increasing efficiency of CRAC units by maximizing their temperature [123]. Raising the temperature coming from CRAC units will minimize the energy consumed by a CRAC unit used to remove a unit of heat contributing into CRAC efficiency. However, the temperature increase should be performed only when inlet server temperatures are within a ”safe” range. In [123] task scheduling is performed according to the power budget of each server which is defined as the product of server power and the deviation of its outlet energy from the reference desired value. Such cooling optimization approaches may nearly half the costs associated with cooling. Thermal-aware scheduling algorithms presented by Tang et al. [124] distribute the jobs spatially preventing excessive heat conditions. Such method will trade the reduction in energy consumption of cooling equipment with a moderate increase in servers’ consumption as idle or under-loaded servers consume more energy per executed task than those highly loaded. Mukherjee et

al. [125] take a step further and extend spatial distribution of jobs adding a temporal dimension. Temporal thermal-aware job scheduling tries to allocate the jobs at energy-efficient equipment extending the task execution time up to the allowed threshold. In a scenario with heterogeneous nature of jobs and data center infrastructure it becomes useful to track thermal footprint of the executed jobs [126]. The availability of such thermal profiles allows distribution on jobs favoring computing resources with minimum levels of heat emission for a certain type of jobs. Current approaches for data center thermal management adapt either mechanical-based or software-based techniques independently. Mechanical-based approaches are simple and can be implemented in a distributed fashion. But, the software-based approaches, being centralized, can deliver better level of optimization in terms of individual jobs and system performance. It is obvious that future thermal management systems for data centers will be complex and include both mechanical-based and software-based techniques.

7.6 Current Practices in Data Centers

7.6.1 Evaluation

The increased pressure of energy consumption awareness leaded to the creation of new tools to evaluate and monitor whole data centers power consumption. The GREEN-GRID consortium established a number of useful documents[9] for designing data centers, measuring, adjusting and so on. Self-feedback on data centers can be achieved using several tools, like the one designed by the CoolEmAll[127] project.

7.6.1.1 Buildings As the environmental pressure rise, news buildings are designed with the energy management as a priority. By instance, EnergyStar helps evaluating the energy impact of a building[10]. It provides the EnergyStar label to buildings that achieve a 75 out of 100 points after evaluation. IBM provides a tool to evaluate energy efficiency of IT infrastructure[11]

- Metrics: To evaluate the quality of a data center in relation to energy several metrics exists:
  - Perf/Watt. This metric is mainly used to evaluate only the computing nodes. By instance Green500[12] uses it to ranks the most powerful supercomputers (mainly clusters). It does not encompass the whole energy consumption of the room (such as AC) but only the consumption of the computing nodes themselves.
  - DCiE (Data center infrastructure efficiency) is the ratio between the ICT equipment power and the total data center power, expressed in percentage. For example, a DCiE value of 50% means that half of the total data center power is spent for the ICT equipment, whilst the other half is spent for maintaining it, such as for the HVAC system.

[9] http://www.thegreengrid.org/library-and-tools
[10] http://www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfoliomanager_benchmarking
[11] http://ibmgreen.bathwick.com/
[12] http://www.green500.org/

– PUE (Power usage effectiveness). This value is complementary to the previous one. It evaluates the ratio between the total energy consumed by the data center facility and the energy provided to the ICT equipment[13]. In 2006, a classical PUE was about 2.0 [128], meaning that half of the energy consumed was to be used for cooling, lightning, and air conditioning (exactly equivalent to a DCiE of 50%). The newest Yahoo center[14] constructed near the Niagara falls uses circulating exterior air to cool the servers, and is able to achieve a PUE of around 1.1. However, several controversial arises on how to fairly calculate the PUE/DCiE values of data centers. In general, an important improvement comes from feedback. More data are available about power usage, the easier it is to optimize a data center consumption[15].

7.6.2 Context aware building

First of all, lightning is not necessary for servers to work; it can be reduced as possible. As self-evident this statement seems, it is common to see a full lightning in data centers. Occupancy sensors and/or economic bulbs can save a lot of energy without a extensive cost[16].

A common believed idea is that a data center in Greenland will consume less than a data center in Sahara, since the external temperature is on average lower. But it has been shown (for instance in the Energy Star study[17], slide 23) that the external temperature has little impact on the overall electricity consumption of data centers. This study does not explicit exactly the infrastructure of the building and the cooling of the server rooms. Indeed, if air circulation coming from outside is in the game, the difference will be significant while if traditional air conditioning is the rule then outside temperature has little influence.

More and more data centers are built so that they are using renewable energy. Solar panels (AISO[18], Phoenix[19], Intel[20], Sun[21], Google[22], ...), wind mills (Google[23], OWC[24], Green House Data[25], Baryonyx[26]) are producing part of the electricity needed by the data

[13] http://www.google.com/corporate/green/datacenters/measuring.html
[14] http://green.yahoo.com/blog/ecogeek/1125/yahoo-data-center-will-be-powered-by-niagara-falls.html
[15] http://technet.microsoft.com/en-us/magazine/2009.gr.datacenter.aspx
[16] http://hightech.lbl.gov/DCTraining/strategies/light.html
[17] http://www.thegreengrid.org/~/media/TechForumPresentations2010/ENERGYSTARforDataCenters.ashx?lang=en
[18] http://www.aiso.net/technology-network-sun.html
[19] http://www.datacenterknowledge.com/archives/2009/06/16/solar-power-at-data-center-scale/
[20] http://www.datacenterknowledge.com/archives/2009/01/19/intel-testing-solar-power-for-data-centers/
[21] http://www.datacenterknowledge.com/archives/2008/05/22/the-solar-powered-blackbox/
[22] http://www.google.com/corporate/green/clean-energy.html
[23] http://www.datacenterknowledge.com/archives/2007/11/29/googles-data-center-windmill-farm/
[24] http://www.datacenterknowledge.com/archives/2009/12/21/data-center-powered-entirely-by-the-wind/
[25] http://www.datacenterknowledge.com/archives/2007/11/29/wind-powered-data-center-in-wyoming/
[26] http://www.datacenterknowledge.com/archives/2009/07/20/wind-powered-data-center-planned/

centers (in one case, all the electricity[27]). Most of the experiences are small size experiences, mainly due to the fact that the cost of these energy productions are still higher than normal electricity for the consumer.

Solutions are also developed to consume renewable electricity in data centers when the cost of electricity is high (typically during daytime) and use chillers during nights. Doing so, the cold that was produced and kept during night can be additionally used with the "free" electricity during day time[28].

This difference of electricity generation and usage can also reflect on the data centers usage itself, offloading the data centers whether during day time (when classical electricity is the rule) or during nights (when solar panels are in the game).

Another trend are the movable data centers. For instance, IBM with portable modular data center (PMDC)[29]. It is advertised that PMDCs have a power usage effectiveness (PUE) of 1.3, including the IT components and physical infrastructure such as chillers, UPS and other components. That compares to a PUE of 2 or higher for most existing data centers, and a PUE of 1.5-1.7 for some of the newer ground-based data centers. Interestingly, Sun proposes a portable solution powered by solar panels[30].

7.6.3 Cooling

In modern data centers, the HVAC (heating, ventilation and air conditioning) consumes approximately half of the total power drawn by the data center [129]. However, an in depth analysis on the cooling infrastructure falls outside the scope of this chapter. Here we will only limit to briefly analyze fans and say some words about common practices on cooling, which represents a not negligible part of the energy consumption of computing elements.

7.6.3.1 Common practices An important part of the data centers energy consumption is wasted for cooling the running components. As explained above, the typical PUE of a data center was about 2.0 in 2006, meaning that one watt for the infrastructure is wasted for each watt used to compute. Among this waste, part of it is due to the cooling.

The first aspect on this is to determine the optimal operational temperature for a data centers. Recent studies tend to exhibit that data centers are often too cold [31] and could operate at higher temperature (with some limits). A consensus is agreed by the industry to maintain an ambient temperature range of 20° to 24°C, while the limit is set to 30°C. A study jointly published by Intel, IBM, HP and Lieberth[32] shows that most data centers are cooled at 20°C while they could operate at 26°C [130].

Several techniques exist and often coexist to cool down the server rooms. Traditionally, air conditioning has been used ever and ever for cooling the infrastructure. Problems arise when the air circulation has not been optimally studied between the racks in the rooms. Some hot spots can exist, and a full investigation taking into account CFD models, cold

[27] http://www.datacenterknowledge.com/archives/2009/06/16/solar-power-at-data-center-scale/
[28] http://www.datacenterknowledge.com/archives/2009/06/16/solar-power-at-data-center-scale/
[29] http://www.environmentalleader.com/2009/12/03/ibm-advances-data-center-efficiencies/
[30] http://www.datacenterknowledge.com/archives/2008/05/22/the-solar-powered-blackbox/
[31] http://www.greenbiz.com/blog/2009/09/01/your-data-center-much-too-cold
[32] http://download.intel.com/pressroom/archive/reference/IPACK2009.pdf

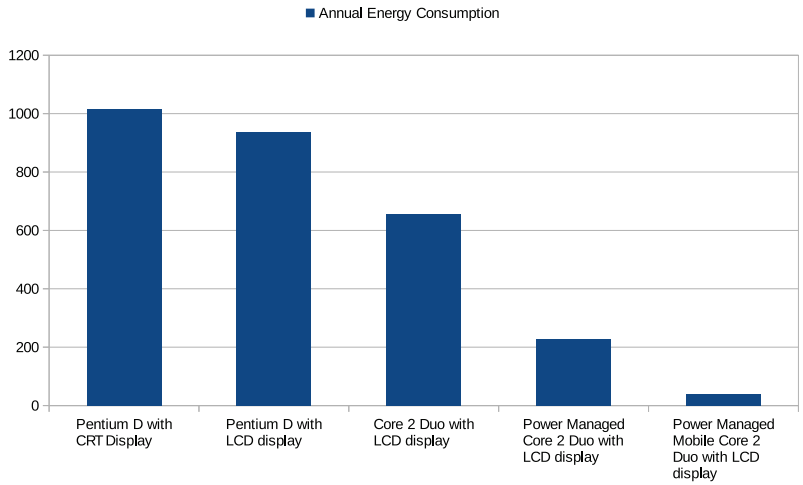

Figure 7.11 Estimated Annual Energy Consumption, data from Intel, Nov 2008

and hot aisle locations, must be done. Some vendors (HP with Dynamic Smart Cooling[33], DegreeC with AdaptivCool[34]) are offering tools to monitor and adjust cooling according to heat dispersion and air circulation.

Another way witnessed is to use cold air column, where the heated air is directed from behind the racks to ease the air circulation. Such an approach can be seen at the Barcelona Marenostrum for instance.

Water cooling is being more and more used, since the efficiency of heat dispersion with water is much higher than with air. In these solutions, water circulates behind the racks and capture the heat and direct it away from the server, before being chilled again and sent back colder. For instance, the CALMIP machine in Toulouse is working with this system.

7.6.4 Upgrades

Citing chip makers, for the same amount of work, energy usage can be vastly reduced by updating hardware.

As an example from Intel (Figure 7.11 Nov 2008), replacing 184 mono-processors from 2005 with the equivalent 21 quad-core from 2009 reduce energy consumption by 92

7.6.5 Uses cases and example of current practices

As energy awareness gains momentum, several uses cases have been fully documented:

- An US best practices repository [131], after an extensive benchmarking of 22 data centers[35]:

[33] `http://www.hp.com/hpinfo/newsroom/press/2006/061129xa.html`
[34] `http://www.adaptivcool.com/`
[35] `http://hightech.lbl.gov/DCTraining/Best-Practices.html`

- In[36] the US Department of Energy shows a joint study with LucasFilm and Verizon;
- In[37] IBM provides information about uses cases where its technology improved energy efficiency;
- In[38] Microsoft shows cases where its technology helped reduce carbon footprint;
- In[39] Accenture and major leaders are forecasting the future (July 2008).

Acknowledgments

This work was partially supported by the COST (European Cooperation in Science and Technology) framework, under Action IC0804. The authors would like to thank particularly G. Da Costa, D. Careglio for their valuable co-ordination of the Working Group 1 deliverable of the Action, entitled ”Hardware leverages to reduce energy consumption in large scale distributed systems” and available for download at www.cost804.org. The authors want to thank A.-C. Orgerie and M. Dias de Assuncao for their contributions to this paper.

[36] http://www1.eere.energy.gov/industry/saveenergynow/pdfs/doe_data_centers_presentation.pdf
[37] http://www-01.ibm.com/software/success/cssdb.nsf/solutionareaL2VW?OpenView&Count=30&RestrictToCategory=corp_Energyefficiency&cty=en_us
[38] http://www.microsoft.com/environment/news_resources/case_studies.aspx
[39] https://microsite.accenture.com/svlgreport/Documents/pdf/SVLG_Report.pdf

REFERENCES

1. A. Greenberg, J. Hamilton, D. A. Maltz, P. Patel, The cost of a cloud: research problems in data center networks, SIGCOMM Comput. Commun. Rev. 39 (1) (2009) 89–73.
2. J. Koomey, Worldwide electricity used in data centers, Environmental Research Letters 3 (2008) 034008.
3. Efficient servers, a project conducted within the eu-programme intelligent energy europe. URL `http://www.efficient-server.eu`
4. B. Schäppi, F. Bellosa, B. Przywara, T. Bogner, S. Weeren, A. Anglade, Energy efficient servers in europe. energy consumption, saving potentials, market barriers and measures. part 1: Energy consumption and saving potentials, Tech. rep., The Efficient Servers Consortium (November 2007).
5. Code of conduct on data centres energy efficiency, version 2.0, Tech. rep., European Commission. Institute for Energy, Renewable Energies Unit (November 2009).
6. P. Jean-Marc, H. Hlavacs (Eds.), Proceedings of the COST Action IC804 on Energy Efficiency in Large Scale Distributed Systems - 1st Year, IRIT, 2010.
7. X. Fan, W.-D. Weber, L. A. Barroso, Power provisioning for a warehouse-sized computer, in: ISCA '07: Proceedings of the 34th annual international symposium on Computer architecture, ACM, New York, NY, USA, 2007, pp. 13–23. doi:http://doi.acm.org/10.1145/1250662.1250665.
8. W. chun Feng, T. Scogland, The green500 list: Year one, in: 23rd IEEE International Parallel and Distributed Processing Symposium (IPDPS) - Workshop on High-Performance, Power-Aware Computing (HP-PAC), Rome, Italy, 2009.
9. A. Rawson, J. Pfleuger, T. Cader, Green grid data center power efficiency metrics: Pue and dcie, in: The Green Grid, 2008.

Please enter `\offprintinfo{(Title, Edition)}{(Author)}`
at the beginning of your document.

10. F. Cappello et al, Grid5000: A large scale, reconfigurable, controlable and monitorable grid platform, in: 6th IEEE/ACM International Workshop on Grid Computing, Grid'2005, Seattle, Washington, USA, 2005.

11. M. Dias de Assuncao, J.-P. Gelas, L. Lefèvre, A.-C. Orgerie, The green grid5000: Instrumenting a grid with energy sensors, in: 5th International Workshop on Distributed Cooperative Laboratories: Instrumenting the Grid (INGRID 2010), Poznan, Poland, 2010.

12. R. Joseph, M. Martonosi, Run-time power estimation in high performance microprocessors, in: ISLPED '01: Proceedings of the 2001 international symposium on Low power electronics and design, ACM, New York, NY, USA, 2001, pp. 135–140. doi:http://doi.acm.org/10.1145/383082.383119.

13. S. Rivoire, P. Ranganathan, C. Kozyrakis, A comparison of high-level full-system power models., in: F. Zhao (Ed.), HotPower, USENIX Association, 2008.
URL http://dblp.uni-trier.de/db/conf/osdi/hotpower2008.html#RivoireRK08

14. A. Specification, www.acpi.info/spec.htm (Apr. 2010).

15. N. Weste, K. Eshragian, Principles of CMOS VLSI Design:A Systems Perspective, Addison-Wesley, 1988.

16. A. P. Chandrakasan, R. W. Brodersen, Minimizing power consumption in digital cmos circuits, in: Proceedings of the IEEE, Vol. 83, IEEE, 1995.

17. O. Svanfeldt-Winter, S. Lafond, J. Lilius, Cost and energy reduction evaluation for arm based web servers, in: Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on, 2011, pp. 480–487. doi:10.1109/DASC.2011.93.

18. R. V. Aroca, L. M. G. GonçAlves, Towards green data centers: A comparison of x86 and arm architectures power efficiency, J. Parallel Distrib. Comput. 72 (12) (2012) 1770–1780. doi:10.1016/j.jpdc.2012.08.005.
URL http://dx.doi.org/10.1016/j.jpdc.2012.08.005

19. S. Saponara, L. Fanucci, M. Coppola, Many-core platform with noc interconnect for low cost and energy sustainable cloud server-on-chip, in: Sustainable Internet and ICT for Sustainability (SustainIT), 2012, 2012, pp. 1–5.

20. K. Fürlinger, C. Klausecker, D. Kranzlmüller, Towards energy efficient parallel computing on consumer electronic devices, in: Proceedings of the First international conference on Information and communication on technology for the fight against global warming, ICT-GLOW'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 1–9.
URL http://dl.acm.org/citation.cfm?id=2035539.2035541

21. W. Stallings, Reduced instruction set computer architecture, Proceedings of the IEEE 76 (1) (1988) 38–55. doi:10.1109/5.3287.

22. T. Jamil, Risc versus cisc, Potentials, IEEE 14 (3) (1995) 13–16. doi:10.1109/45.464688.

23. J. Hamilton, Cooperative expendable micro-slice servers (cems): Low cost, low power servers for internet-scale services, in: 4th Biennial Conference on Innovative Data Systems Research (CIDR), 2009.

24. J. Hamilton, Overall data center costs, http://perspectives.mvdirona.com/2010/09/18/OverallDataCenterCosts.aspx, accessed: 2013-06-19.

25. EuroCloud, www.eurocloudserver.com.

26. M.-B. E. A. T. E. E. H. Performance, http://www.montblanc-project.eu.

27. SATA Specification, INCITS Technical Committee T13 subcommittee ATA.
URL www.t13.org/

28. SCSI Specification, INCITS Technical Committee T10 subcommittee SCSI.
URL www.t13.org/

29. J. Rydning, D. Reinsel, J. Janukowicz, White paper: The need to standardize storage device performance metrics (Sep. 2008).

30. J. Guitart, J. Torres, E. Ayguadé, A survey on performance management for internet applications, Concurrency and Computation: Practice and Experience 22 (1) (2010) 68–106.

31. C. Lefurgy, K. Rajamani, F. Rawson, W. Felter, M. Kistler, T. Keller, Energy management for commercial servers, Computer 36 (12) (2003) 39 – 48. doi:10.1109/MC.2003.1250880.

32. R. Bianchini, R. Rajamony, Power and energy management for server systems, Computer 37 (11) (2004) 68 – 76. doi:10.1109/MC.2004.217.

33. E. Pinheiro, R. Bianchini, E. Carrera, T. Heath, Load balancing and unbalancing for power and performance in cluster-based systems, in: Workshop on Compilers and Operating Systems for Low Power, Vol. 180, Citeseer, 2001, pp. 182–195.

34. J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, R. P. Doyle, Managing energy and server resources in hosting centers, in: SOSP '01: 18th ACM symposium on Operating systems principles, ACM, New York, NY, USA, 2001, pp. 103–116. doi:http://doi.acm.org/10.1145/502034.502045.

35. R. C, Humidity control: Systems meet varied demands, Consulting Specifying Engineering 11 (1992) 50–60.

36. Aperture Data Management, http://www.aperture.com/.

37. J. Pouwelse, K. Langendoen, H. Sips, Energy priority scheduling for variable voltage processors, in: Proceedings of the 2001 international symposium on Low power electronics and design, ISLPED '01, ACM, New York, NY, USA, 2001, pp. 28–33. doi:http://doi.acm.org/10.1145/383082.383089.
URL `http://doi.acm.org/10.1145/383082.383089`

38. R. Banginwar, E. Gorbatov, Gibraltar: Application and network aware adaptive power management for ieee 802.11, in: Wireless On-demand Network Systems and Services, 2005. WONS 2005. Second Annual Conference on, 2005, pp. 98 – 108. doi:10.1109/WONS.2005.20.

39. Cisco, Data center management software, http://www.cisco.com/go/spdatacenter.

40. Microsoft, Top 10 business practices for environmentally sustainable data centers, `http://www.microsoft.com/environment/our_commitment/articles/datacenter_bp.aspx`.

41. Cisco, Data center networking best practices, http://www.cisco.com/en/US/solutions/collateral/ns340/ns414/ns742/ns743/.

42. D. Bell, Distributed Database Systems, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.

43. L. B. N. Laboratory, Data center energy management best practices, http://hightech.lbl.gov/dctraining/best-practices.html.

44. H. Liu, Z. Meng, S. Cui, A wireless sensor network prototype for environmental monitoring in greenhouses, in: Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on, 2007, pp. 2344 –2347. doi:10.1109/WICOM.2007.584.

45. C. Lefurgy, X. Wang, M. Ware, Server-level power control, in: Autonomic Computing, 2007. ICAC '07. Fourth International Conference on, 2007, pp. 4 –4. doi:10.1109/ICAC.2007.35.

46. X. Wang, M. Chen, Cluster-level feedback power control for performance optimization, in: High Performance Computer Architecture, 2008. HPCA 2008. IEEE 14th International Symposium on, 2008, pp. 101 –110. doi:10.1109/HPCA.2008.4658631.

47. E. Pinheiro, R. Bianchini, E. Carrera, T. Heath, Dynamic cluster reconfiguration for power and performance, Compilers and operating systems for low power (2001) 75–93.

48. N. Boden, D. Cohen, R. Felderman, A. Kulawik, C. Seitz, J. Seizovic, W.-K. Su, Myrinet: a gigabit-per-second local area network, Micro, IEEE 15 (1) (1995) 29 –36. doi:10.1109/40.342015.

49. M. Elnozahy, M. Kistler, R. Rajamony, Energy conservation policies for web servers, in: Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems - Volume 4, USITS'03, USENIX Association, Berkeley, CA, USA, 2003, pp. 8–8. URL http://portal.acm.org/citation.cfm?id=1251460.1251468

50. M. Vasic, O. Garcia, J. Oliver, P. Alou, J. Cobos, A dvs system based on the trade-off between energy savings and execution time, in: Control and Modeling for Power Electronics, 2008. COMPEL 2008. 11th Workshop on, 2008, pp. 1 –6. doi:10.1109/COMPEL.2008.4634665.

51. M. E. Femal, V. W. Freeh, Boosting data center performance through non-uniform power allocation, Autonomic Computing, International Conference on 0 (2005) 250–261. doi:http://doi.ieeecomputersociety.org/10.1109/ICAC.2005.17.

52. W. P. Jones, Computer rooms, Air Conditioning Applications and Design 18 (1977) 181–185.

53. S. Ceri, G. Pelagatti, Distributed databases: principles and systems, McGraw-Hill New York, 1984.

54. E. L. Watford, One engineering solution for temperature and humidity control when computers are added, Air Conditioning, Heating Ventilation 56 (1959) 89–90.

55. G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, F. Zhao, Energy-aware server provisioning and load dispatching for connection-intensive internet services, in: Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, NSDI'08, USENIX Association, Berkeley, CA, USA, 2008, pp. 337–350. URL http://portal.acm.org/citation.cfm?id=1387589.1387613

56. Q. Wu, P. Juang, M. Martonosi, L.-S. Peh, D. Clark, Formal control techniques for power-performance management, Micro, IEEE 25 (5) (2005) 52 – 62. doi:10.1109/MM.2005.87.

57. X. Zhou, Y. Cai, G. Godavari, C. Chow, An adaptive process allocation strategy for proportional responsiveness differentiation on web servers, in: Web Services, 2004. Proceedings. IEEE International Conference on, 2004, pp. 142 – 149. doi:10.1109/ICWS.2004.1314733.

58. Epa report on server and data center energy efficiency, http://www.energystar.gov/index.cfm?c=prod_development.server_efficiency_study.

59. R. Schmidt, Thermal management of office data processing centers, in: Advances in Electronic PackagingProceedings of the Pacific Rim/ASME International Electronic Packaging Technical Conference (INTERpack97), 1997, pp. 15–19.

60. R. Arularasan, R. Velraj, Cfd analysis in a heat sink for cooling of electronic devices, International Journal of the Computer, the Internet and Management 16 (3) (2008) 1–11.

61. S. Murugesan, Harnessing green it: Principles and practices, IT Professional 10 (1) (2008) 24 –33. doi:10.1109/MITP.2008.10.

62. P. J. Solly, Air-conditioning for a computer department, Consulting Engineering 20 (5) (1966) 72–76.

63. H. Zhu, H. Tang, T. Yang, Demand-driven service differentiation in cluster-based network servers, in: INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol. 2, 2001, pp. 679 –688 vol.2. doi:10.1109/INFCOM.2001.916256.

64. X. Zhou, C.-Z. Xu, Harmonic proportional bandwidth allocation and scheduling for service differentiation on streaming servers, Parallel and Distributed Systems, IEEE Transactions on 15 (9) (2004) 835 – 848. doi:10.1109/TPDS.2004.43.

65. P. Hutchins, J. Wade, G. Sparts, Energy Savings in Computer/Data Centers, Energy and Pollution Control Opportunities To The Year 2000 (1994) 339–342.

66. M. Song, Energy-aware data prefetching for multi-speed disks in video servers, in: Proceedings of the 15th international conference on Multimedia, MULTIMEDIA '07, ACM, New York, NY, USA, 2007, pp. 755–758. doi:http://doi.acm.org/10.1145/1291233.1291403.
URL http://doi.acm.org/10.1145/1291233.1291403

67. T. Newhall, D. Amato, A. Pshenichkin, Reliable adaptable network ram, in: Cluster Computing, 2008 IEEE International Conference on, 2008, pp. 2 –12. doi:10.1109/CLUSTR.2008.4663750.

68. W. G. Brown, Equipment cooling for modernization, British Telecommunication Engineering 2 (1984) 246–250.

69. R. W. Haines, Keeping cool, Datamation 32 (1986) 83–84.
URL http://portal.acm.org/citation.cfm?id=12273.13307

70. D. Kusic, J. Kephart, J. Hanson, N. Kandasamy, G. Jiang, Power and performance management of virtualized computing environments via lookahead control, in: Autonomic Computing, 2008. ICAC '08. International Conference on, 2008, pp. 3 –12. doi:10.1109/ICAC.2008.31.

71. E. Kalyvianaki, T. Charalambous, S. Hand, Self-adaptive and self-configured cpu resource provisioning for virtualized servers using kalman filters, in: Proceedings of the 6th international conference on Autonomic computing, ICAC '09, ACM, New York, NY, USA, 2009, pp. 117–126. doi:http://doi.acm.org/10.1145/1555228.1555261.
URL http://doi.acm.org/10.1145/1555228.1555261

72. R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, No "power" struggles: coordinated multi-level power management for the data center, SIGARCH Comput. Archit. News 36 (2008) 48–59. doi:http://doi.acm.org/10.1145/1353534.1346289.
URL http://doi.acm.org/10.1145/1353534.1346289

73. G. Tesauro, N. Jong, R. Das, M. Bennani, A hybrid reinforcement learning approach to autonomic resource allocation, Autonomic Computing, International Conference on 0 (2006) 65–73. doi:http://doi.ieeecomputersociety.org/10.1109/ICAC.2006.1662383.

74. J. Kephart, H. Chan, R. Das, D. Levine, G. Tesauro, F. Rawson, C. Lefurgy, Coordinating multiple autonomic managers to achieve specified power-performance tradeoffs, in: Autonomic Computing, 2007. ICAC '07. Fourth International Conference on, 2007, pp. 24 –24. doi:10.1109/ICAC.2007.12.

75. G. Tesauro, R. Das, H. Chan, J. Kephart, D. Levine, F. Rawson, C. Lefurgy, Managing power consumption and performance of computing systems using reinforcement learning, Advances in Neural Information Processing Systems 20 (2007) 1–8.

76. P. Ranganathan, P. Leech, D. Irwin, J. Chase, Ensemble-level power management for dense blade servers, SIGARCH Comput. Archit. News 34 (2006) 66–77. doi:http://doi.acm.org/10.1145/1150019.1136492.
URL http://doi.acm.org/10.1145/1150019.1136492

77. J. Choi, S. Govindan, B. Urgaonkar, A. Sivasubramaniam, Profiling, prediction, and capping of power consumption in consolidated environments, in: Modeling, Analysis and Simulation of Computers and Telecommunication Systems, 2008. MASCOTS 2008. IEEE International Symposium on, 2008, pp. 1 –10. doi:10.1109/MASCOT.2008.4770558.

78. Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, N. Gautam, Managing server energy and operational costs in hosting centers, in: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, SIGMETRICS '05, ACM, New York, NY, USA, 2005, pp. 303–314. doi:http://doi.acm.org/10.1145/1064212.1064253.
URL http://doi.acm.org/10.1145/1064212.1064253

79. V. Petrucci, O. Loques, B. Niteroi, D. Mossé, Dynamic configuration support for power-aware virtualized server clusters, in: WiP Session of the 21th Euromicro Conference on Real-Time Systems. Dublin, Ireland, Citeseer, 2009, pp. 1–4.

80. L. Liu, H. Wang, X. Liu, X. Jin, W. B. He, Q. B. Wang, Y. Chen, Greencloud: a new architecture for green data center, in: Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session, ICAC-INDST '09, ACM, New York, NY, USA, 2009, pp. 29–38. doi:http://doi.acm.org/10.1145/1555312.1555319.
URL http://doi.acm.org/10.1145/1555312.1555319

81. A. Verma, P. Ahuja, A. Neogi, Power-aware dynamic placement of hpc applications, in: Proceedings of the 22nd annual international conference on Supercomputing, ICS '08, ACM, New York, NY, USA, 2008, pp. 175–184. doi:http://doi.acm.org/10.1145/1375527.1375555.
URL http://doi.acm.org/10.1145/1375527.1375555

82. F. Yao, A. Demers, F. Shenker, A scheduling model for reduced CPU energy, in: Proceedings of the 36th Annual Symposium on Foundations of Computer Science, Citeseer, 1995, pp. 374–382.

83. I. Hong, D. Kirovski, G. Qu, M. Potkonjak, M. Srivastava, Power optimization of variable-voltage core-based systems, Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on 18 (12) (1999) 1702 –1714. doi:10.1109/43.811318.

84. G. Quan, X. Hu, Energy efficient fixed-priority scheduling for real-time systems on variable voltage processors, in: Design Automation Conference, 2001. Proceedings, 2001, pp. 828 – 833. doi:10.1109/DAC.2001.156251.

85. A. Manzak, C. Chakrabarti, Variable voltage task scheduling algorithms for minimizing energy, in: Proceedings of the 2001 international symposium on Low power electronics and design, ISLPED '01, ACM, New York, NY, USA, 2001, pp. 279–282. doi:http://doi.acm.org/10.1145/383082.383168.
URL http://doi.acm.org/10.1145/383082.383168

86. N. K. Jha, Low power system scheduling and synthesis, in: Proceedings of the 2001 IEEE/ACM international conference on Computer-aided design, ICCAD '01, IEEE Press, Piscataway, NJ, USA, 2001, pp. 259–263.
URL http://portal.acm.org/citation.cfm?id=603095.603147

87. N. Bambha, S. Bhattacharyya, J. Teich, E. Zitzler, Hybrid global/local search strategies for dynamic voltage scaling in embedded multiprocessors, in: Hardware/Software Codesign, 2001. CODES 2001. Proceedings of the Ninth International Symposium on, 2001, pp. 243 –248. doi:10.1109/HSC.2001.924683.

88. Y. Zhang, X. S. Hu, D. Z. Chen, Task scheduling and voltage selection for energy minimization, in: Proceedings of the 39th annual Design Automation Conference, DAC '02, ACM, New York, NY, USA, 2002, pp. 183–188. doi:http://doi.acm.org/10.1145/513918.513966.
URL http://doi.acm.org/10.1145/513918.513966

89. F. Gruian, K. Kuchcinski, Lenes: task scheduling for low-energy systems using variable supply voltage processors, in: Proceedings of the 2001 Asia and South Pacific Design Automation Conference, ASP-DAC '01, ACM, New York, NY, USA, 2001, pp. 449–455. doi:http://doi.acm.org/10.1145/370155.370511.
URL http://doi.acm.org/10.1145/370155.370511

90. M. Seredynski, P. Bouvry, M. Klopotek, Performance of a strategy based packets forwarding in ad hoc networks, in: Availability, Reliability and Security, 2008. ARES 08. Third International Conference on, 2008, pp. 1036 –1043. doi:10.1109/ARES.2008.181.

91. J. Liu, P. H. Chou, N. Bagherzadeh, Communication speed selection for embedded systems with networked voltage-scalable processors, in: Proceedings of the tenth international symposium on Hardware/software codesign, CODES '02, ACM, New York, NY, USA, 2002, pp.

169–174. doi:http://doi.acm.org/10.1145/774789.774824.
URL http://doi.acm.org/10.1145/774789.774824

92. Y.-H. Lu, L. Benini, G. De Micheli, Operating-system directed power reduction, in: Proceedings of the 2000 international symposium on Low power electronics and design, ISLPED '00, ACM, New York, NY, USA, 2000, pp. 37–42. doi:http://doi.acm.org/10.1145/344166.344189.
URL http://doi.acm.org/10.1145/344166.344189

93. R. Ge, X. Feng, K. W. Cameron, Performance-constrained distributed dvs scheduling for scientific applications on power-aware clusters, in: Proceedings of the 2005 ACM/IEEE conference on Supercomputing, SC '05, IEEE Computer Society, Washington, DC, USA, 2005, pp. 34–. doi:http://dx.doi.org/10.1109/SC.2005.57.
URL http://dx.doi.org/10.1109/SC.2005.57

94. B. Khargharia, S. Hariri, M. Yousif, Autonomic power and performance management for computing systems, Cluster Computing 11 (2008) 167–181, 10.1007/s10586-007-0043-6.
URL http://dx.doi.org/10.1007/s10586-007-0043-6

95. V. Petrucci, O. Loques, D. Mossé, A dynamic configuration model for power-efficient virtualized server clusters, in: 11th Brazillian Workshop on Real-Time and Embedded Systems (WTR), Vol. 2, Citeseer, 2009, pp. 35–44.

96. J. L. Berral, I. n. Goiri, R. Nou, F. Julià, J. Guitart, R. Gavaldà, J. Torres, Towards energy-aware scheduling in data centers using machine learning, in: Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, e-Energy '10, ACM, New York, NY, USA, 2010, pp. 215–224. doi:http://doi.acm.org/10.1145/1791314.1791349.
URL http://doi.acm.org/10.1145/1791314.1791349

97. A. Verma, G. Dasgupta, T. K. Nayak, P. De, R. Kothari, Server workload analysis for power minimization using consolidation, in: Proceedings of the 2009 conference on USENIX Annual technical conference, USENIX'09, USENIX Association, Berkeley, CA, USA, 2009, pp. 28–28.
URL http://portal.acm.org/citation.cfm?id=1855807.1855835

98. B.-G. Chun, G. Iannaccone, G. Iannaccone, R. Katz, G. Lee, L. Niccolini, An energy case for hybrid datacenters, SIGOPS Oper. Syst. Rev. 44 (2010) 76–80. doi:http://doi.acm.org/10.1145/1740390.1740408.
URL http://doi.acm.org/10.1145/1740390.1740408

99. R. Nathuji, C. Isci, E. Gorbatov, Exploiting platform heterogeneity for power efficient data centers, in: Autonomic Computing, 2007. ICAC '07. Fourth International Conference on, 2007, pp. 5 –5. doi:10.1109/ICAC.2007.16.

100. D. Filani, J. He, S. Gao, M. Rajappa, A. Kumar, R. Shah, R. Nagappan, Dynamic Data Center Power Management: Trends, Issues and Solutions, Intel Technology Journal 12 (01).

101. M. Al-Fares, A. Loukissas, A. Vahdat, A scalable, commodity data center network architecture, in: SIGCOMM '08: Proceedings of the ACM SIGCOMM 2008 conference on Data communication, Seattle, WA, USA, 2008, pp. 63–74.

102. D. Kliazovich, P. Bounvry, Y. Audzevich, S. U. Khan, GreenCloud: A Packet-level Simulator of Energy-aware Cloud Computing Data Centers, in: IEEE Globecom, Miami, FL, USA, 2010.

103. A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, S. Sengupta, VL2: A scalable and flexible data center network, SIGCOMM Comput. Commun. Rev. 39 (4) (2009) 51–62.

104. C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, S. Lu, DCell: A Scalable and Fault-Tolerant Network Structure for Data Centers, SIGCOMM Comput. Commun. Rev. 38 (4) (2008) 75–86.

105. C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, S. Lu, BCube: a high performance, server-centric network architecture for modular data centers, in: SIGCOMM '09:

Proceedings of the ACM SIGCOMM 2009 conference on Data communication, Barcelona, Spain, 2009, pp. 63–74.

106. H. Wu, G. Lu, D. Li, C. Guo, Y. Zhang, MDCube: A High Performance Network Structure for Modular Data Center Interconnection, in: Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT), Rome, Italy, 2009, pp. 25–36.

107. L. Gyarmati, T. A. Trinh, How can Architecture Help to Reduce Energy Consumption in Data Center Networking, in: e-Energy '10: Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, Passau, Germany, 2010, pp. 183–186.

108. P. Mahadevan, P. Sharma, S. Banerjee, P. Ranganathan, Energy aware network operations, in: INFOCOM'09: Proceedings of the 28th IEEE international conference on Computer Communications Workshops, Rio de Janeiro, Brazil, 2009, pp. 25–30.

109. B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, N. McKeown, Elastic Tree: Saving Energy in Data Center Networks, in: 7th USENIX Symposium on Networked System Design and Implementation (NSDI), San Jose, CA, USA, 2010, pp. 249–264.

110. V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, P. Rodriguez, Greening the internet with nano data centers, in: CoNEXT '09: Proceedings of the 5th international conference on Emerging networking experiments and technologies, ACM, New York, NY, USA, 2009, pp. 37–48. doi:http://doi.acm.org/10.1145/1658939.1658944.

111. L. Gyarmati, A. T. Trinh, Scafida: A Scale-Free Network Inspired Data Center Architecture, SIGCOMM Comput. Commun. Rev. 40 (2010) 5–12.

112. H. Obler, Energy efficient computer cooling, Heating, Piping, Air Conditioning Engineering 54 (1) (1982) 107–111.

113. P. A. Green, A one-system, 'equipment first' cooling plan for computer installations, Heating, Piping, Air Conditioning Engineering 33 (12) (1961) 96–99.

114. F. J. Grande, How to select and integrate equipment for computer room air conditioning, Heating, Piping, Air Conditioning Engineering 35 (7) (1963) 96–98.

115. A. Khoshhala, M. Rahimia, A. Alsairafib, Cfd investigation on the effect of air temperature on air blowing cooling system for preventing tube rupture, International Communications in Heat and Mass Transfer 36 (7) (2009) 750–756.

116. S. Yoo, J. Kim, T. Kim, S. Ahn, J. Sung, D. Kim, A2s: Automated agriculture system based on wsn, IEEE International Symposium on Consumer Electronics.

117. J. D. Lea-Cox, G. Kantor, J. Anhalt, A. G. Ristvey, D. S. Ross, A wireless sensor network for the nursery and greenhouse industry, Southern Nursery Association Research Conference (2007) 454–458.

118. I.-C. Yang, S. Chen, Y.-I. Huang, K.-W. Hsieh, C.-T. Chen, H.-C. Lu, C.-L. Chang, H.-M. Lin, Y.-L. Chen, C.-C.Chen, Y. Lo, Rfid-integrated multi-functional remote sensing system for seedling production management, ASABE Annual International Meeting.

119. J. M. Ayers, Air conditioning needs of computers pose problems for new office building, Heating, Piping, Air Conditioning Engineering 34 (8) (1962) 107–112.

120. CD-Adapco, http://www.cd-adapco.com/.

121. F. Inc., http://www.fluent.com/.

122. OpenCFD, http://www.openfoam.com/.

123. J. Moore, J. Chase, P. Ranganathan, R. Sharma, Making scheduling "cool": temperature-aware workload placement in data centers, USENIX Annual Technical Conference.

124. Q. Tang, S. K. S. Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach, IEEE Transactions on Parallel and Distributed Systems archive 19 (11).

125. T. Mukherjee, A. Banerjee, G. Varsamopoulos, S. K. S. Gupta, S. Rungta, Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers, Computer Networks: The International Journal of Computer and Telecommunications Networking 53 (7).

126. L. Wang, G. von Laszewskiand J. Dayal, X. He, A. J. Younge, T. R. Furlani, Towards thermal aware workload scheduling in a data center, Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks.

127. M. V. D. Berge, G. D. Costa, M. Jarus, A. Oleksiak, W. PiaTek, E. Volk, Modeling data center building blocks for energy-efficiency and thermal simulations, in: International Workshop on Energy-Efficient Data Centres, 2013.

128. C. Malone, C. Belady, Metrics to characterize data center & it equipment energy use, in: Proceedings of 2006 Digital Power Forum, 2006.

129. T. G. Grid, The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE, Technical Committee White Paper, 2008.

130. 2008 ashrae environmental guidelines for datacom equipment, expanding the recommended environmental envelope - ashrae tc 9.9.

131. S. Greenberg, E. Mills, B. Tschudi, P. Rumsey, B. . Myatt, August), Best practices for data centers: Lessons learned from benchmarking.