# Survey of Techniques and Architectures for Designing Energy-Efficient Data Centers

Junaid Shuja, Kashif Bilal, Sajjad A. Madani, Mazliza Othman,
Rajiv Ranjan, Pavan Balaji, and Samee U. Khan

***Abstract*—Cloud computing has emerged as the leading paradigm for information technology businesses. Cloud computing provides a platform to manage and deliver computing services around the world over the Internet. Cloud services have helped businesses utilize computing services on demand with no upfront investments. The cloud computing paradigm has sustained its growth, which has led to increase in size and number of data centers. Data centers with thousands of computing devices are deployed as back end to provide cloud services. Computing devices are deployed redundantly in data centers to ensure 24/7 availability. However, many studies have pointed out that data centers consume large amount of electricity, thus calling for energy-efficiency measures. In this survey, we discuss research issues related to conflicting requirements of maximizing quality of services (QoSs) (availability, reliability, etc.) delivered by the cloud services while minimizing energy consumption of the data center resources. In this paper, we present the concept of inception of data center energy-efficiency controller that can consolidate data center resources with minimal effect on QoS requirements. We discuss software- and hardware-based techniques and architectures for data center resources such as server, memory, and network devices that can be manipulated by the data center controller to achieve energy efficiency.**

***Index Terms*—Controller design, data centers, energy efficiency.**

## I. Introduction

INFORMATION TECHNOLOGY (IT) industry has evolved from its birth in the last century to one of the most prominent industries in today's world. Along with its rapid growth, IT has changed our lifestyle and has become a technology enabler for many veteran industries and businesses [1]. The IT growth has been generally fueled by cheaper and more powerful computing resources. However, recently, the on-demand availability of computing resources in the form of cloud has led to a major technological revolution. The cloud facilities deploy large number of computing resources networked together in the form of warehouse-size data centers. Clients utilize these computing resources on-demand through managed cloud applications at flexible prices. The on-demand utility-based cloud model has characteristics such as lower operating costs, small investments, high scalability, and multitenancy that are ideal for small industries and businesses [2]. IT technologies, including cloud and mobile cloud computing, have become a major consumer of energy due to rapid growth and integration with many industries [3], [4]. Energy efficiency is a global challenge for today's world, where conventional energy resources are being consumed at a very rapid pace, and call for alternative energy resources is growing. Therefore, many IT technologies have focused on saving energy by developing energy-efficient protocols, architectures, and techniques [1]. Although cloud computing is termed as an inherently energy-efficient platform due to scalable nature of its resources and multitenant capability, densely populated data centers are usually over provisioned and consume large amounts of energy in a competitive race among the service providers to ensure 24/7 availability of services to the clients [5].

Recently, many studies have focused on the energy consumption levels of data centers [6], [7]. The Environmental Protection Agency (EPA) study estimated that the data center energy consumption would double from 2006 (61 billion kWh) to 2011 [6]. In 2009, data centers accounted for 2% of worldwide electricity consumption with an economic impact of US $30 billion [8]. Gartner Group forecasted data center hardware expenditure for 2012 to be at US $106.4 billion, a 12.7% increase from 2011, whereas cloud computing revenue is forecasted to jump from US $163 billion in 2011 to US $240 billion in 2016 [9]. Another drawback of energy consumption by the IT sector is the emission of greenhouse gases. As the electricity production process emits large amount of carbon dioxide depending on the type of fuel used, the IT sector contributes indirectly to carbon dioxide emissions [7]. The IT sector was responsible for 2% of carbon dioxide worldwide in 2005, a figure that is estimated to grow by 6% per year [6]. Aggressive energy-efficiency measures for all devices inside the data center can reduce 80% of energy costs and 47 million metric tons of carbon dioxide emissions [6]. The data center efficiency is measured by a metric called power usage effectiveness (PUE), which is the ratio of total data center energy usage to IT equipment energy usage [6]. The average PUE value of data centers in a 2005 survey was found to be 2 [10]. It was estimated that the average PUE value will fall to 1.9 by 2011, although aggressive energy-efficiency measures can result in ideal PUE value of 1.2 [6]. Higher PUE value indicates that most of data center energy

Manuscript received July 16, 2013; revised November 15, 2013 and February 14, 2014; accepted March 24, 2014.

J. Shuja and M. Othman are with the Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia (e-mail: junaidshuja@siswa.um.edu.my; mazliza@um.edu.my).

K. Bilal and S. U. Khan are with the Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND 58108-6050 USA (e-mail: kashif.bilal@my.ndsu.edu; samee.khan@ndsu.edu).

S. A. Madani is with COMSATS Institute of Information Technology, 22060 Abbottabad, Pakistan (e-mail: madani@ciit.net.pk).

R. Ranjan is with the Information Engineering Laboratory, CSIRO ICT Centre, Canberra, A.C.T. 0200, Australia (e-mail: raj.ranjan@csiro.au).

P. Balaji is with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439 USA (e-mail: balaji@mcs.anl.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSYST.2014.2315823

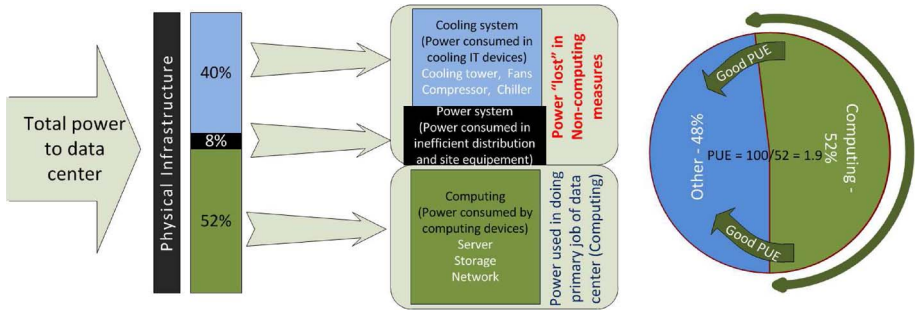



Fig. 1. Power distribution among physical infrastructure.

is consumed in cooling measures instead of computing. An energy-efficient data center with a lower PUE index can lead to several benefits such as: **a)** reduced energy consumption, hence, lesser operational expenses; **b)** lesser greenhouse gas emissions; and **c)** lesser device costs [11]. Data center operators are also interested in the total cost of ownership: the sum of capital expenses (capex) required in setting up the data center and the operational expenses (opex) required in running the data center [6]. Fig. 1 illustrates the distribution of data center power among cooling, power distribution, and computing units.

### A. Research Problem Discussion

The fact that the energy consumption is distributed among many data center resources and their components, an ideal energy-efficiency approach requires consideration of all data center resources, including the server (i.e., motherboard, fan, and peripherals), network devices (i.e., chassis, line card, and port transceivers), storage devices (i.e., hard disks, RAM, and ROM chips), and cooling devices (i.e., fans and chillers) [5]. While achieving lower PUE index, cloud providers have to provide quality services in an ever-increasing competitive cloud market. Cloud providers hosting diverse applications need to maintain service-level agreements; (SLAs), achieve low access latencies; meet task deadlines; and provide secure, reliable, and efficient data management. Cloud provider business objectives often conflict with low-cost hardware designs and resource optimization techniques deployed in back-end data centers for energy efficiency. Data center energy optimization is a hard problem due to online consideration of dynamic factors such as workload placement, resource mapping, cooling schedule, interprocess communication, and traffic patterns. However, cloud providers have been forced to consider energy optimization techniques for back-end data centers due to escalating energy bills and competitive price market for cloud services [12].

The average workload on the data center usually remains at 30% and does not require functioning of all computing resources [11]. Therefore, some of the underutilized resources can be powered off to achieve energy efficiency while fulfilling the data center workload demands. However, scheduling of data center resources requires careful considerations of data center traffic patterns, client SLAs [13], latency and performance issues [14], network congestion [11], and data replication [15]. Data center energy-efficiency controllers for servers [16], memory devices [15], and network topologies [13] have been proposed separately in the literature. Moreover, security cannot be ignored while optimizing data center resources for energy efficiency [17], [18]. Researchers have proposed a resource optimization strategy composed of confidentiality, integrity, and authentication measures considering real-time computation and deadline constraints [19]. The practice of data center resource optimization was considered taboo previously due to business continuity objective. Google recently published a data center resource management patent that shows that large IT companies are now focusing on the issue of energy efficiency [12]. Cloud providers deploy redundant resources to ensure 99.9% availability of services. On the other hand, energy efficiency advocates minimizing the data center resources while eliminating redundancy. In order to achieve energy efficiency without compromising quality of service (QoS) and SLAs, data center resources need to be managed carefully according to the workload profile of cloud service.

In this paper, we present the concept of a central high-level energy-efficiency controller (see Fig. 2) for data centers. The data center controller interacts with resource-specific controllers (i.e., server, memory, and network) in a coordinated manner to make decisions regarding workload consolidation and device state transitions (sleep, idle, etc.) for energy efficiency. User SLAs enforce cloud providers to ensure a required performance level to meet business objectives. The data center controller is required to make resource consolidation and state change decisions based on workload profile to avoid SLA violations. The data center controller divides control among various resource controllers that manage a single data center resource domain. However, the resource controllers often work in opposing manner to counter each other's effects [20], [21]. Therefore, to manage data center energy efficiency, the resource controllers' coordinate control mechanism between data center resources through the central controller. Resource controllers' coordinate workload conditions within their domain such as network congestion, server overload, and thermal hotspots to the central controller to avoid further workload placement. The central controller takes decision at a global level and forwards feedback to resource controllers in order to mitigate adverse workload conditions. The feedback helps resource controllers make decisions regarding device state transitions that result in energy efficiency while meeting SLA requirements.

To optimize energy, the central controller employs both hardware- and software-based techniques for all data center resources. This survey has been organized considering three main resources in the data centers, i.e., server, storage, and network devices. In each section, we discuss the energy-efficiency

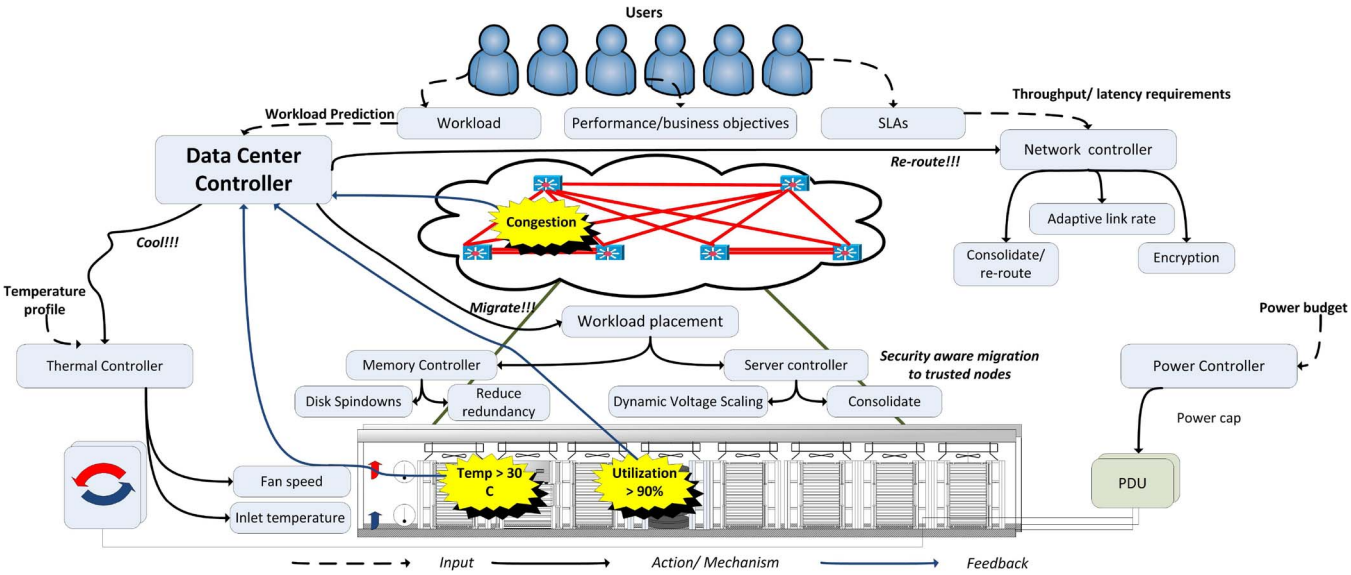



Fig. 2. Central controller for energy-efficient data center.

TABLE I
SERVER POWER STATES

| CPU state | Range | Mechanism |
|---|---|---|
| Global states | S1,...,S3 | Multiple core management |
| Performance states | P0,...,P15 | Dynamic voltage frequency scaling (DVFS). The power consumption of a server is proportional to CPU frequency and square of supplied voltage |
| Throttling states | T0,...,T3 | Induces STPCLK signal in CPU duty cycle to reduce CPU frequency |
| Core states | C0,...,C3 | % STPCLK signal and DVFS |

techniques and architectures for the corresponding data center resource. Section II elaborates energy-efficient server devices and their architecture. Cooling and power distribution devices are also discussed in the sections. Section III highlights energy-efficient storage devices and architectures. Section IV details the data center communication architectures, devices, and protocols.

## II. COMPUTING SYSTEMS

Servers are the most prominent and energy-hungry element of the data centers. Servers account for 50%–90% of IT electricity consumption in a data center of different space types [6]. Data centers are mainly composed of three kinds of servers, i.e., volume servers, mid-range servers, and high-end servers. The EPA study estimated in 2007 that the volume servers were the major server class used in data centers and consumed 34% of total data center electricity [6]. As the same study pointed out the escalated electricity consumption values in the data centers, data center operators started search for energy-efficient data center building blocks, specifically servers. A server node can be managed in different operational states according to the current workload to achieve temporal energy proportional computing [22]. Dynamic voltage and frequency scaling (DVFS) of processor results in lower power states with lower computation cycles. Lesser computation cycles may affect QoS of the hosted cloud services. The Advanced Configuration and Power Interface (ACPI), available on most of the operating systems, provides a standard interface for managing server power states [23]. ACPI standard provides different server states defined in Table I. The power consumption of a server is proportional to CPU frequency and square of supplied voltage. Many researchers have proposed DVFS techniques for power efficiency in data centers [5], [24]. However, next-generation semiconductor technology would operate on optimal voltage that would not be further minimized [8]. Hence, such hardware will have lesser advantage of DVFS techniques. On the other hand, researchers have proposed "race to halt" techniques for energy efficiency, i.e., execute task at highest operational frequency and sleep at completion [22]. We will discuss energy-efficient server architectures, server power, and cooling requirements in the following sections.

### A. Server Architectures

Reduced instruction set computer (RISC) architectures are inherently energy efficient than their counterpart Complex Instruction Set Computer (CISC) architectures. Therefore, they are predominantly used in mobile and embedded systems. RISC-based processors have lesser number of transistors and gates but require more instruction cycles for a task than a CISC-based processor. Aroca and Gonçalves [25] have made comparison of several x86 and ARM processors for power efficiency on a hardware testbed. Typical server applications were run on the processors, and analysis of server temperature, CPU utilization, latency, and power consumption for varying workloads was made. The results show that ARM-based processors are three to four times more energy efficient than the x86-based processors while comparing requests per second per Watt relation. Li *et al.* [26] carried out study on ARM cores over a power and area modeling framework. The simulation results show 23% and 35% reductions in capex and power costs, respectively.

TABLE II
COMPARISON OF SERVER DESIGNS IN DATA CENTER SPACE

| Server design | Energy efficiency | Thermal efficiency | Space efficiency |
|---|---|---|---|
| CISC | 1 | 1 | none |
| RISC | 300-400% less than CISC | 100-150% more than CISC | none |
| SoC | 35% less than RISC | 50% more than RISC | 300-500% more |

System-on-Chip (SoC) designs have emerged from the ARM-based processor class. A SoC design integrates several processor cores, memory units, network interfaces, graphics processor, and input/output (I/O) controllers on a single die to reduce the number of interconnects, die area, and power consumption [26]. SoC designs reduce pin crossings, which lead to lesser power consumption than discrete and physically separate designs. Table II shows energy, cooling, and space efficiency tradeoffs between various server designs [25], [27].

The server technology is becoming increasingly out of balance from the memory technology in terms of bandwidth and speed. Moreover, server designs emphasize on performance rather than work done per unit cost or energy. Hamilton [28] proposed Collaborative Expandable Micro-slice Servers (CEMS) design for modular data center built on a customized rackable server chassis consisting of x86 cores sharing a single power supply, double data rate (DDR) 2 memory units, and I/O interfaces. Performance metrics, such as request per server (RPS), price, power consumption, and RPS per dollar were evaluated. Results show that CEMS design leads to 50% energy efficiency without affecting system performance. Three-dimensional stacked design similar to SoC design is another promising technique for space compaction and energy efficiency in data centers. Researchers proposed a 3-D component stacking server architecture named PicoServer [29]. PicoServer architecture consists of stacked cores and multiple DRAMs connected through low-latency buses, thus eliminating the need of L2 caches. Multiple cores allow clock frequency to be lowered without affecting throughput of the system and also satisfying thermal constraints. The essence of 3-D stacking architectures is the core-to-DRAM interfaces that provide low latency and high throughput and consume lesser energy. A typical SoC design is illustrated in Fig. 3.

Multiple cores integrated on single-chip compaction designs (e.g., blade, SoC, and CEMS) increase power and heat density of the die. Therefore, designing densely populated computing modules requires careful consideration of power density, clock frequency, and system throughput [30]. Furthermore, SoC designs have to consider application and thread level parallelism for performance optimization. Critics to such designs have pointed out several challenges that need to be addressed before the benefits of power efficiency can be achieved. Performance of such server designs is bounded by Amdahl's law for multicore systems. According to Amdahl's law, a balanced system should: **a)** need a bit of sequential I/O per second per instruction per second; and **b)** have a memory with megabyte per millions of instructions per second (MB/MIPS) ratio close to 1 [31]. Processing capacity, memory, I/O, and network bandwidth need to be increased in balance so that the performance of one

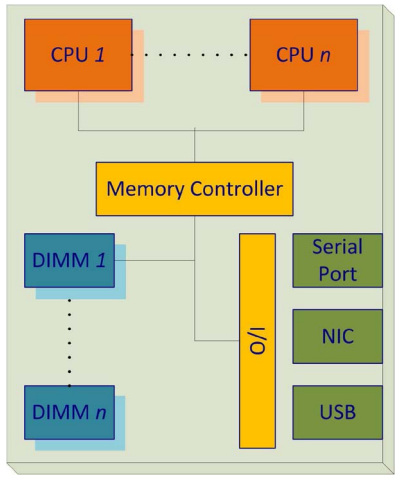



Fig. 3. SoC 3-D stack design of server.

component does not bottleneck the performance of the whole system [32]. Moreover, multiprocessor SoC parallelization can impose significant overhead of interprocessor communications [33], [34]. Wang *et al.* [35] proposed a task-scheduling technique that jointly reschedules computation and interprocess communication of each task for multiprocessor SoC designs to minimize both intercore communication overhead and task memory usage.

Data center operators have predominantly adopted high-end server designs in data center buildings as high-end servers are suitable for high-performance computing (HPC) applications [7]. High-end servers are similar in design to rack-mounted systems with multiple cores, dedicated storage, and I/O bandwidth all at exascale computing level. High-end servers provide: **a)** high performance for data- and compute-intensive applications, **b)** better power distribution among computing resources through efficient power distribution unit (PDU) design, and **c)** higher utilization levels for HPC applications due to exascale design. Blade servers also offer a similar design of a stripped-down server with modular components and several cores wired to a die. Several thermal and virtual resource management techniques are based on the blade server model of data centers [36]. However, researchers have advocated the usage of commodity low-cost low-power cores for scale-out applications for increased performance per dollar rather than increased performance/server metric. Lim *et al.* [27] present the intuition behind the usefulness of commodity and embedded processors: Volume servers drive higher costs, whereas commodity and embedded servers target cheaper markets, therefore providing better performance per dollar. The authors evaluated six server designs from volume, mobile, and embedded space. Results showed that mobile and embedded processors provided better performance per dollar than the volume servers for different data center applications. A similar study [37] conducted on seven different server class designs showed that mobile class systems are 80% and 300% energy efficient than the embedded class systems and volume servers, respectively. However, I/O interface limitations of mobile class systems limit their usability for data-intensive workloads. Graphic processing units (GPUs) have also become essential part of data centers hosting scientific

applications due to their efficiency in floating-point functions. Moreover, GPUs are highly energy efficient compared with CPUs. The top two energy-efficient supercomputing machines are GPU based.[1] GPUs provide nearly 85 times lesser power consumption for a floating-point calculation than x86-based processors [38].

Employing DVFS techniques and using commodity server designs lead to energy efficiency at the cost of performance. The scale-out approach (3-D stacking, SoC, and commodity servers) can lead to better performance per unit energy with customized I/O and memory designs for data-intensive computations [39]. Moreover, SoC designs compact computing units in a small space and require dedicated thermal fail over management [30]. Furthermore, scale-up design optimizes system throughput rather than work done per unit cost or energy. Due to escalating data center energy costs, systems designs aimed at lower energy per unit work need to be investigated. RISC-based architectures are inherently energy efficient and can be utilized in scale-out designs [26]. However, high-end systems optimally support virtualization techniques that are making their way in low-power RISC designs [25].

### B. Server Power Distribution

A medium-size data center with 288 racks can have a power rating of 4MW [40]. The data center power infrastructure is complex, and power losses occur due to multistage voltage conversions. Power loads of the medium-size data center require high-voltage power supply of up to 33KV. This high-voltage supply is stepped down to 280–480-V range by automated transfer switches. The stepped-down voltage is supplied to an uninterrupted power supply (UPS) system. The UPS converts the ac to dc to charge the batteries and converts dc back to ac before supplying power to the PDUs. The PDU steps down voltage according to specification of each computing unit. The multiple power conversion phases have efficiency in the range of 85%–99%, which results in power losses [22]. Therefore, along the lines of *smart grid*, the data center power distribution and conversion need to be intelligently and efficiently designed [20]. The PDUs are arranged across the data center in wrapped topology; each PDU serves two racks, and each rack draws current from two PDUs. The power is provided to each server through connectors called whips that split 480-V three-phase ac to 120-V single-phase ac. Fig. 4 depicts the conventional power distribution system of a data center.

Power capping and multilevel coordination is required to limit data center power budget across the cyberphysical space. Researchers have proposed a power capping scheme for data centers that require coordination among various data center resources, such as virtual machines, blade servers, racks, and containers [21]. The server controller caps power budget by varying P-states according to reference utilization and actual utilization levels of the server. Modern servers are equipped with dynamic power management solutions to cap local power budget [41]. Controllers at higher level react to the changes of lower level controllers in the same manner as they react to

[1]http://www.green500.org/

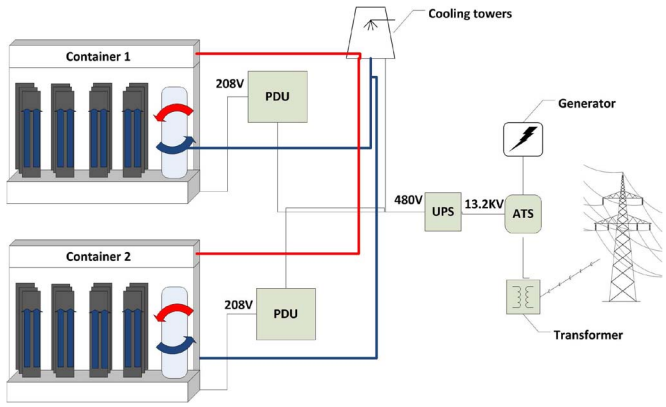



Fig. 4. Data center power distribution.

varying workloads. In modern rack-based data centers, modular power distribution is adopted with redundant power distribution elements. Pelley *et al.* [40] adopt the technique of shuffled power topologies, where a server draws power from multiple power feeds. A central controller manages switching of servers to power feeds to: **a)** enforce power budgets through control loops by modulating processor frequency and voltage; **b)** distribute load over feeds in case of feed failure; and **c)** balance power draws across the three ac phases to avoid voltage and current spikes that reduce device reliability. The power-scheduling technique, *power routing*, sets power budget for each server at the PDU. If the server power demand increases its budget cap, the power routing seeks additional power on the power feeds linked to the server. For underutilized servers, the power routing controller reduces their budget caps and creates power slacks that might be required elsewhere in the power distribution system. Estimating server power usage and provisioning data center power accordingly requires thorough investigation of: **a)** nameplate versus actual peak power values; **b)** varying data center workloads; and **c)** application-specific server resource utilization characteristics [42]. The gap between the theoretical and actual peak power utilization of cluster servers allows the deployment of additional computing devices under the same power budget. Meisner *et al.* [43] argue that the PDUs' efficiency remains in *green zone* above 80% for server utilization. At lower server utilization levels, the PDU efficiency drops below 70%. PDU scheduling scheme, PowerNap, is devised to achieve higher efficiency. PowerNap deploys redundant array of inexpensive load sharing (RAILS) commodity PDUs. PDUs operating at less than 80% are shifted to nap mode, and their load is shifted to adjacent PDUs.

DC power distribution architectures achieve better efficiency than the ac power distribution due to lesser ohmic power loss. However, dc power distribution is prone to voltage spikes that can cause device failure. Voltage spike protection devices need to be added to the dc power distribution architecture for safe operations. Researchers have investigated the role of dc/dc converter that acts as a voltage stabilizer in a 400-V dc power distribution system [44]. The dc/dc converter consists of hybrid pair silicon super junction and silicon carbide Schottky barrier diode that result in low conduction and low power loss. The dc/dc converter provides 97% efficiency and power density

greater than 10 W/cm$^3$ that is ideal for blade servers and SoC server designs.

### C. Server Cooling

A large portion of data center power is used in cooling measures for large number of heat dissipating servers and power distribution systems. The cooling measures account for 38% of total data center power, whereas 6% of power is lost in power distribution [6]. As this power is not utilized in computing work, it is considered wasted and degrades the data center performance metric, PUE. Heat removal has been the key for reliable and efficient operation of computer systems from early days. The increasing heat dissipation in computing chips is contributed to both scale-up and scale-out trends in data centers [8], with current power density estimated at 3000 W/m$^2$ [45]. Scale-out trends have led to more densely populated data centers, thus requiring dynamic thermal management techniques. Scale-up trends are pushing more transistors on small chip space. In such scale-up designs, the shrinking dimensions have led to higher power consumption by leakage currents. The transistor switching power has not reduced fast enough to keep pace with increasing transistor densities; hence, chip power density has increased [8]. As the root of heat dissipation in data centers, the transistors and their constituent materials require consideration for parameters, such as leakage power per unit area. For now, chip heat dissipation continues to rise with power and transistor density and requires integrated fluid-cooling mechanisms at data center level [46]. ASHRAE has suggested maximum server inlet temperature of 27 °C for data center environments [47]. Every 10 °C rise in temperature over 21 °C can result in 50% decrease in hardware reliability [48]. A network of temperature sensors is usually deployed to monitor thermal dynamics of data centers [49]. Thermal resource management has been modeled with computational fluid dynamic techniques. There are two thermal management techniques in data centers: **a)** air-flow-based solution; and **b)** fluid-based solution [8].

Liquids are more efficient in heat removal than air due to higher heat capacities. However, fluid cooling comes with increased mechanical complexity [8]. IBM supercomputing node 575 with racked core design comes with modular water-cooling units (MWUs) for a heat exchange system [50]. Each node in the rack-mounted supercomputer connects to two fluid couplers that supply water to cold plates. Cold plates are coupled to chips by conduction through a thermal interface material. An air-to-liquid heat exchanger, rear-door heat exchanger (RDHX), acts as exhaust and cools the air exiting the racks. Water is supplied to the RDHX and the fluid couplers by the MWU. In the hybrid fluid–air cooling system, 80% of heat dissipated is rejected to the MWU, whereas the rest of the heat is rejected to room air. The MWU consumes 83% less power than the conventional air cooling systems [50]. Fluid cooling comes with the advantage of waste heat utilization. The thermal energy transferred to the coolant can be reused for heating in cold climates [51], [52]. Moreover, fluid-based cooling is emission free and further waste utilization can reduce carbon footprint by a large factor

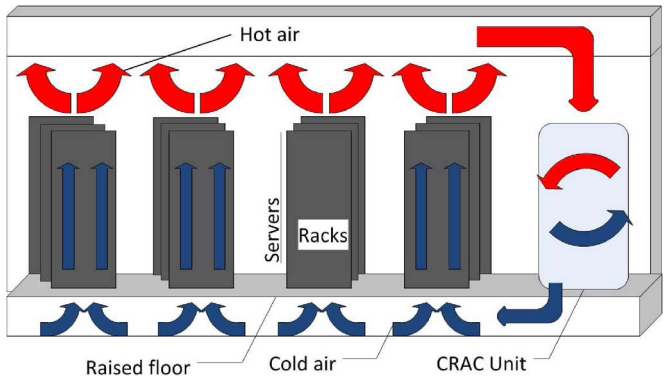



Fig. 5. CRAC.

[53]. Researchers have also proposed green energy sources such as solar and wind energy for data centers and computing clusters [54]. The amount of energy generated by the renewable sources varies due to weather conditions. However, the data center requires a fixed input current to operate the cluster of servers. Li *et al.* [55] proposed SolarCore, a hybrid power utilization architecture that switches between solar and grid power. SolarCore maximizes solar energy utilization by matching supply power with utilized power and handles power variations using DVFS to processor cores. *iSwitch* architecture [56], utilizing wind energy, switches the computing load between energy sources to balance the supply load variations. Compute load switching is enabled by the virtual machine migration in data center environments.

Computer room air conditioning (CRAC) units utilize air ventilation and conditioning techniques to dissipate heat produced by data center resources. Air-side economizer can be utilized for free cooling in regions of low temperature and humidity [57]. Higher server utilization and CRAC unit failures can lead to nonuniform heat distribution and hot spots in specific racks. Dynamic provisioning of cooling resources is required for the nonuniform heat profiles of data centers [45]. Dynamic thermal management of data centers results in benefits such as: **a)** uniform temperature distribution with reduced hot spots; **b)** reduced cooling costs; and **c)** improved device reliability [49]. Dynamic thermal management techniques have varying approaches to cooling data centers such as workload placement in cool racks [48], combined workload placement and cooling management techniques, and virtual machine placement techniques [36]. A typical CRAC configuration is depicted in Fig. 5.

## III. Storage Systems

Memory devices account for 5% of electricity consumption within the data center. Storage volumes continue to increase 50%–70% per year. Cloud data centers often provide Data as a Service (DaaS) facilities. DaaS facilities maintaining user data have to meet the often conflicting requirements of high availability, redundancy, and energy efficiency [2]. There are two basic techniques to achieve energy efficiency in data center storage medium: **a)** making storage hardware energy efficient; and **b)** reducing data redundancy [58]. To reduce data redundancy and achieve energy efficiency, careful consideration of

data replication, mapping, and consolidation is required. Storage systems traditionally utilize redundant array of independent disks (RAID) technology for performance and availability. Several proposals have targeted redundant disk shutdowns to achieve energy efficiency [59]. Verma *et al.* [15] proposed Sample-Replicate-Consolidate Mapping (SRCMap) technique that utilizes virtualization to map a workload on minimum number of physical nodes. The integrated virtualization manager consists of subcontrollers: **a)** replica placement controller that samples the active set of virtual disks to be replicated, **b)** active disk controller that consolidates active virtual disks to physical disks, and **c)** consistency manager that makes sure replicas are consistent across physical disks. However, storage migration over the network is expensive and prohibitive as it affects network performance. Moreover, disk spin downs have been also proposed for energy-efficient RAID storage [60]. Disk spin downs have severe effect on access latencies. However, such techniques can be applied to secondary disks during light workloads. In the following sections, we will discuss the storage architectures, emerging energy-efficient storage technologies, and DRAM architectures in the context of data centers.

### A. Storage Architectures

Storage in data centers can be provided in many different ways. On-chip DRAM units are energy efficient. However, only a small fraction of application data can be stored in DRAMs due to higher costs and lower capacities [61]. High volume storage is provided in storage towers composed of hard disk drives (HDDs) connected to the server racks through optical fiber or twisted pair copper wires. High-speed storage is provided by on-chip memory units such as DRAMs and caches. Data center storage can be categorized into three forms: **a)** direct attached storage connected directly to the server; **b)** storage area network (SAN) residing across the network; and **c)** network attached storage (NAS) accessible at higher level of abstraction such as files [22]. The shareable storage is stored on HDDs' storage towers accessed through NAS and SAN. Although fiber optic cabling provides high-speed and energy-efficient storage networking technology, it is cost prohibitive [62]. Fiber channel is mostly used as the networking technology for SAN and NAS access. However, Ethernet and infiniband solution have been also proposed [22]. Fig. 6 depicts the data center storage architecture.

### B. Storage Technologies

The main focus of energy-efficient storage hardware has been the solid-state drives (SSDs). SSDs utilize Flash memory, i.e., nonvolatile memory with characteristics similar to electronically erasable programmable read-only memory. To write a Flash memory, an erase operation is required that can be performed on contiguous block of pages. The erase operation is very slow compared with the read operation and also wears memory cells, which degrades the memory lifetime [22]. NAND Flash is most popular Flash memory, but other types of Flash devices have been also proposed for data center operations [65]. A Flash translation layer (FTL) performs: **a)** logical-to-physical

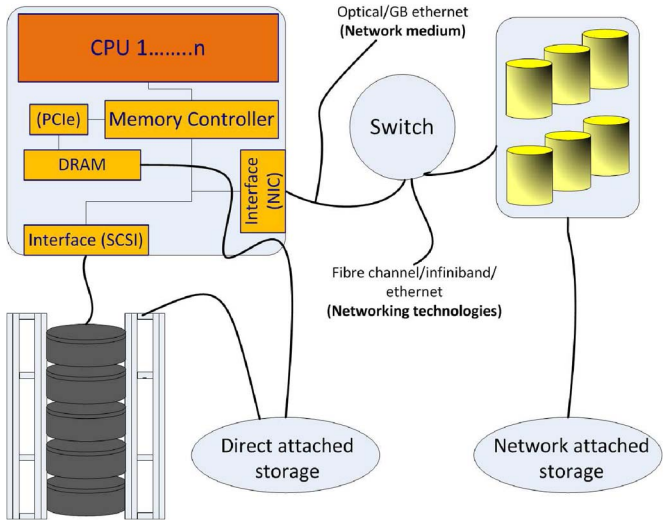



Fig. 6. Data center storage architecture.

address translation; **b)** out-of-order updates to avoid erase operation for writes; and **c)** wear leveling policies to distribute erase operations equally over memory cells [22]. The SSDs have energy proportional characteristics, i.e., the energy used is proportional to I/O operations per second. On the other hand, HDDs consume 85% of energy when idle [61]. SSDs are the future of energy-efficient primary storage in data centers. For now, the price gap between HDDs and SSDs is decreasing but is large enough to prohibit data center operators to perform a complete HDD to SSD migration. Narayanan *et al.* [61] analyzed several data centers workload traces on HDDs, SSDs, and two-tier hybrid architectures. Least cost storage configurations were measured to support performance, fault tolerance, and capacity requirements. The hybrid architecture consisting of SSDs performing the role of a mid-tier fast cache between server storage and SAN provides better performance and is lesser cost prohibitive than SSD-only approach. Ferroelectric-NAND Flash memory design with nonvolatile page buffer has been also proposed for high-speed low-power data center applications [65]. The proposed design consumes 300% less power than conventional NAND Flash memory designs due to lower erase/program voltages. High write/erase endurance is achieved by inserting buffer layer between Flash and substrate layers that reduces stress-induced leakages. A batch write algorithm is proposed, which speeds up write operations by writing data on available free pages and updating the OS addresses of the pages through flash controller. Researchers [63] proposed a flash controller to manage an array of Flash drives and implement FTL. The FTL provides interface to the processors for logical block addresses of the Flash array and is tailored to meet requirements of data-intensive workloads. Each processor has access to the Flash array that reduces the ratio of processing power to memory bandwidth. FTL can be designed in three mapping architectures, i.e., **a)** page-level FTL that allocates sequential pages in a block; **b)** block-level FTL that allocates random pages within a block and uses block offset to address pages; and **c)** hybrid-level mapping that stores data in data blocks while changes to data are stored in log blocks [67]. Liu *et al.* [68] proposed a reuse-aware FTL that avoids block

TABLE III
COMPARISON OF STORAGE ARCHITECTURES IN DATA CENTER SPACE

| Storage type | Storage technology | Energy efficiency | Cost per capacity | Energy Efficiency techniques |
|---|---|---|---|---|
| SAN and NAS | HDD, magnetic, optical storage | 10W for active, 5W for idle state | 0.2-0.3$Gb for small to large capacity | HDDs are mostly used in hybrid storage schemes [61] |
| DAS | SSD, flash memory | 200-300% less for SSD, 3-50% less for flash memory | 1000-1500% more for SDD, 200-300% more for flash memory | DRAM [63], [64], NAND flash [65], PCM [66] |

erase operation in case the block has a threshold of free pages. The reuse-aware strategy reduces the number of erases by selecting blocks with lesser free pages for merge operations. Researchers have also proposed a hybrid-level FTL that utilizes both data and log blocks for reuse strategy [67]. FTL garbage collection technique collects invalid pages scattered over many blocks. Qin *et al.* [69] proposed a hybrid FTL that stores data and update log in the same physical block to concentrate address mapping. Garbage collection is delayed until there is no free block for memory allocation. Victim block is selected on the basis of least number of valid pages. Researchers have proposed Meta-Cure, a technique for FTL metadata fault detection and avoidance [70]. Meta-Cure replicates Flash metadata and requests error correction codes before update to prevent error accumulation.

Emerging main memory technologies such as phase-change memory (PCM) [66], [71], CMOS, spin-torque transfer RAM [72], and memristor [73] have been also proposed as energy-efficient main memory alternatives. These technologies present attractive features such as high density, nonvolatility, and lower leakage currents for data center environments. However, these technologies suffer in performance due to high latency of the write operation that also reduces the memory endurance [74]. Researchers have proposed several techniques to increase endurance of PCM. Redundant writes are removed and balanced equally to all memory cells through row shifting and memory segment swapping techniques. Write updates can be delayed and stored in buffer for frequently updated records to increase PCM endurance [71].

In spite of energy-efficient characteristics, there are still many issues that need to be addressed for adoption of SSDs in data centers. Due to decreasing cell areas, Flash memory devices are becoming less reliable and less durable [66]. Therefore, the FTL needs to be redefined to mask media errors and provide object abstraction for application-specific patterns. Wear leveling techniques need to be devised to increase Flash endurance. SSDs have slow write operations, as compared with DRAM. Data-intensive tasks usually faced in data center environments require large number of write operations that degrade SSDs' performance [63]. Table III compares storage architectures and technologies for energy efficiency and cost estimates [75].

### C. DRAM Architectures

Several research works have also focused on increasing DRAM efficiency for data center operations. Sudan *et al.* [64] proposed a DRAM colocation scheme of frequently accessed pages in row buffers to increase hit ratios. As the row buffer hit ratio increases, the system efficiency increases in terms of performance, latency, and energy consumption. Frequently accessed pages are grouped together and migrated to the same row buffer. The hardware-assisted migration scheme introduces a translation layer between physical addresses and the memory controller. The translation layer keeps track of the new physical addresses of the migrated pages without changing page table entries. The DRAM page migration schemes increase performance by 9% and reduce memory energy consumption by 15%. Researchers proposed high-performance storage, RAMCloud, based entirely on DRAM components from commodity servers [76]. As RAM storage is volatile, disk-level durability is achieved by keeping two copies of each object on different servers and logging write updates on a disk. The RAMCloud architecture reduces the access latencies and increases throughput ten times, as compared with Flash-based storage. However, RAMCloud architecture consumes five times more energy than Flash-based storage. Andersen *et al.* [75] proposed fast array of wimpy nodes (FAWN) architecture based on RAM and Flash devices to reduce hot spots in primary-key storage applications. HDDs, Flash, and DRAM components were analyzed for FAWN architecture on the basis of four parameters: cost per gigabyte, access time, throughput, and power consumption. For small storage capacities (up to 10 GB), Flash drives provide better cost per gigabyte, whereas large storage capacities can be supported by HDDs at better cost per gigabyte. DRAM provides throughput on the order of gigabytes per second, whereas HDDs and Flash devices lack behind in this regard. DRAM access latencies are on the order of nanoseconds, whereas HDD and Flash access latencies are on the order of milliseconds. DRAM and HDDs consume more energy due to their capacitor charge and mechanical nature, respectively. Whereas, Flash memory consumes ten times less power than HDDs and DRAM.

Next-generation memory devices will be energy efficient with newer DDR technologies operating at lower voltages [22]. DDR memory can be also operated with DVFS techniques to achieve energy efficiency. Like server power states, DDR power states introduce state transition latency that increases with lower power modes. DDR power modes are managed across various components of DDR RAM, such as DRAM array, I/O ports, and registers [77].

## IV. COMMUNICATION SYSTEMS

Network devices are another energy-hungry resource of data centers. Data center networks often extend connectivity to non-IT equipment for monitoring of environment factors such as humidity and temperature [49]. An ideal data center network would connect every node with every other node within the data center so that cloud applications can be independent of resource location. However, such scale interconnects are not possible in data centers with tens of thousands of nodes. Therefore, data

center networks are usually hierarchical in nature with multiple subhierarchies connected to each other through switches [78]. Although the number of routers and switches inside the data center is much less than the number of servers, still the network devices consume 5% of the data center electricity as the router peak power can be 90 times greater than that of a server [5]. Cluster switches tailor made for warehouse scale data centers can connect up to 3456 servers at a time [62]. Most of routers are modular in nature, i.e., modules comprising multiple ports can be added to the router chassis [78]. Power consumption in network devices can be managed through adaptive link rate (ALR) techniques [79], [80]. The ALR techniques achieve energy efficiency by either: **a)** lower data rates; or **b)** lower power/idle (LPI) state transitions that are managed according to the data loads [81]. The standardization of ALR techniques has resulted in IEEE 802.3az Protocol for Energy Efficient Ethernet [82]. An ALR technique consists of: **a)** a mechanism that defines link rate synchronization and corresponding negotiations; and **b)** a policy that defines controlling of link data rate. The mechanisms to implement an ALR technique can be: **a)** medium access control (MAC) frame handshake protocol in which the sender and receiver adjust data rates through MAC frames [83], **b)** ALR autonegotiation protocol that permits communicating devices to set data rate and flow control parameters [84]; and **c)** IEEE Energy Efficient Ethernet 802.3az [82]. The LPI state in the 802.3az protocol consumes only 10% power, as compared with the active state [81].

Most of cloud applications, such as MapReduce and web search, reach hundreds of servers within the data center for a single query. Therefore, intra-data center communications consume 70% of the network bandwidth [81]. Server-to-server communication requires very low latencies to meet task deadlines. However, the data center networks are usually oversubscribed, which affects network performance and reduces full bisection bandwidth. Multiple servers with higher aggregate uplinks share a switch with a lesser aggregate uplink that causes oversubscription [3]. To overcome oversubscription ratio and remove bandwidth bottleneck: **a)** server-centric [85] and hybrid [86] network architectures have been adopted; and **b)** 100-Gb Ethernet solutions have been proposed [62]. Scalability is another issue with data center networks. The network architecture should scale with the increase in the number of servers [87]. In the sections below, we discuss network architectures, optical interconnects, and routing aspects of data centers.

### A. Data Center Network Architectures

Data center network topologies (often cited as architectures in the literature) can be classified into two broad categories: **a)** conventional tiered topologies, such as three tier (3T) built from high-end devices [11]; and **b)** Clos, fat-tree, and hybrid topologies built from commodity off-the-shelf resources [86], [87]. The tiered topology is mostly used in data centers networks. The tiered topologies use high-end switches that have high cost, enhanced functionality, and higher energy consumption. On the other hand, unconventional topologies such as BCube [86] and DCell [88] advocate the use of commodity resources that are low cost, provide low functionality, and consume lesser energy. The unconventional data center topologies also take advantage of economy of scale of consumer hardware and follow scale-out designs. The scale-out designs deploy large number of simple and inexpensive switches. On the contrary, conventional topologies follow scale-up designs, which enforce higher port densities and more functionality per device.

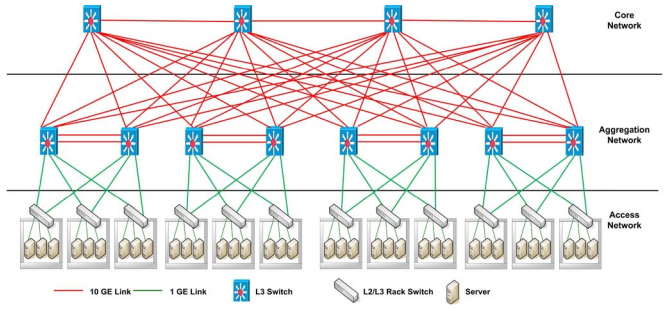



Fig. 7. 3T data center topology.

The 3T topology fulfills the requirements of HPC applications. However, there are serious issues regarding tiered data center topologies: **a)** High-end switches result in oversubscribed links in data center networks, which results in limited server-to-server bandwidth [89]; **b)** the routing designs in tiered data center architectures create fragmentation of server resources and addresses that limit the utility of virtualization techniques; and **c)** tiered data center topologies are not scalable [90]. Moreover, high-end switches in tiered data center topologies use variants of equal cost multipath (ECMP) routing protocols that assign paths to flows based on their IP header hash values. The ECMP routing results in unbalanced flow scheduling where there are few hot spots inside the data center while other switches remain idle. Fig. 7 presents the conventional tiered topology adopted in data centers [78].

Several unconventional topologies have been proposed that use commodity hardware and custom routing protocols for energy proportional data center operations. Al-Fares *et al.* [87] proposed one such data center topology built from commodity resources. The proposed $k$-ary fat-tree topology consists of $k$ pods, each containing two layers of $k/2$ switches having $k$ ports each. Each switch is connected to $k/2$ pod hosts with remaining $k/2$ ports connecting to aggregate layer pods. The proposed fat-tree architecture consumed 56% less energy than tiered architectures. Hybrid network architectures utilize servers and switches as packet forwarding devices [86], [88]. BCube [88] utilizes servers having multiple ports with forwarding characteristics connected to mini switches. ServerSwitch is a server cum switch design that has the capability to process and forward data packets [91]. The usage of commodity devices in hybrid architectures leads to lower energy consumption. However, as servers work as packet forwarding entities, hybrid architectures limit the scope of server sleep and off modes for energy efficiency. Researchers proposed a network traffic and server load consolidation scheme that limits data center workload to a subset of servers and route traffic through limited number of switches [92]. Results showed 74% energy saving in the data center network. Abts *et al.* [93] proposed an energy proportional flattened butterfly (FBFLY) topology

that is inherently energy efficient due to reduced number of redundant network devices. The FBFLY topology utilizes ALR techniques according to workload characteristics to achieve energy proportional data center network.

### B. Optical Data Centers

Optical networks have been proposed to overcome some of the problems faced by conventional data center networks. As most of the Internet service providers are providing high bandwidth to end users through fiber-to-the-home technologies, data center networks are becoming a bottleneck when subjected to high-speed access. However, optical technology used in telecommunication networks requires redesign for data center networks. For intra-data center communication, the size and power consumption of small form-factor pluggable (SFP) transceivers at switches are high. Quad SFP transceivers providing four times port density can be used instead of SFP transceivers, whereas vertical-cavity-surface-emitting lasers have been suggested instead of horizontal cavity lasers for low-space low-power data center operations [62]. For inter-data center communications, high distances require use of optical amplifier that nonlinearly respond to different wavelength-division multiplexing (WDM) wavelengths. Digital signal processing enabled optical receivers provide better tolerance to amplitude spontaneous emission noise and signal dispersion. Several electrical–optical hybrid data center network architectures have been also proposed to provide energy-efficient network operations in data centers [94]. The idea is to replace some of the electrical packet switches (EPSs) with optical circuit switching (OCS) and the corresponding cables with optical cables. The advantages of such an approach are: **a)** Per-port power consumption of an optical switch is 50 times lesser than that of EPS; **b)** optical data rates are much higher; **c)** WDM can be used to switch multiple channels through a single port; and **d)** reduced cabling complexity. The Helios [94] design of a hybrid optical/electrical data center network fulfills the packet switching and circuit switching demands of the data center network by a controlled topology configuration of EPS and OCS. The Helios topology manager classifies flows into high-bandwidth point-to-point flows and bursty many-to-many flows and assigns them to OCS and EPS, respectively. The OCS and fiber interconnects have several barriers to data center integration, such as: **a)** higher capex of optical instruments; **b)** lower port densities in optical switches prohibitive of large-scale data centers; and **c)** higher switching times than electrical switches [89].

### C. Data Center Routing

Efficient routing protocols play an important role in utilizing network capacity of the underlying topology. Shang *et al.* [95] put forward the idea of energy-aware routing where few network devices provide the basic routing and throughput according to a performance threshold while the rest of devices are powered off. The NP-hard problem of finding the energy-aware routing subset is reduced to 0–1 knapsack problem. Heller *et al.* [13] proposed ElasticTree, a data center network comprising OpenFlow-enabled switches [96] for energy-efficient data center operations. ElasticTree works as a network controller that calculates a subset network topology that satisfies current workload. OpenFlow-enabled switches help divert flows to the energy-efficient subset of nodes and present a standard API to manage flow entries in the switch flow table. OpenFlow switches are essential to energy-efficient operations of data centers [13].

## V. Conclusion

The rapid rise of cloud computing paradigm has changed the landscape of IT. Cloud computing has emerged as an innovative way of delivering IT services and is perceived to deliver computing as the fifth human resource utility. However, cloud data centers consume large amount of electricity, which has led to call for energy and resource optimization. The main business objective of cloud providers is to provide 99.99% available cloud services. Energy-efficiency techniques conflict with the business objectives and call for reduced redundancy of cloud resources that may lead to violation of user SLAs. Moreover, energy-efficient hardware devices such as SSDs and optical interconnects do not fit into current cloud paradigm due to their high cost. Furthermore, energy-efficient hardware devices such as ARM processors and PCM drives do not provide comparable performance to prevalent hardware technology. The DVFS and ALR techniques achieve energy efficiency but compromise performance of devices. An energy optimization technique needs to consider data center workload profile and user SLAs to balance the energy efficiency and performance requirements. In case the data center workload is low, devices can be transitioned to lower power states or turned off. In this survey, we have analyzed mechanisms to control and coordinate data center resources for energy-efficient operations. We have also presented a central controller design and formulated coordination among resource controllers. Energy-efficient hardware designs for data center resources were also discussed in detail.

## References

[1] S. Zeadally, S. Khan, and N. Chilamkurti, "Energy-efficient networking: Past, present, and future," *J. Supercomput.*, vol. 62, no. 3, pp. 1093–1118, May 2011.
[2] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: State-of-the-art and research challenges," *J. Internet Serv. Appl.*, vol. 1, no. 1, pp. 7–18, Apr. 2010.
[3] J. Shuja, S. Madani, K. Bilal, K. Hayat, S. Khan, and S. Sarwar, "Energy-efficient data centers," *Computing*, vol. 94, no. 12, pp. 973–994, Dec. 2012.
[4] H. Qi and A. Gani, "Research on mobile cloud computing: Review, trend and perspectives," in *Proc. 2nd Int. Conf. Digit. Inf. Commun. Technol. Appl.*, May 2012, pp. 195–202.
[5] D. Kliazovich, P. Bouvry, Y. Audzevich, and S. Khan, "GreenCloud: A packet-level simulator of energy-aware cloud computing data centers," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–5.
[6] R. Brown, "Report to congress on server and data center energy efficiency public law 109-431," U.S. Environ. Protection Agency, Washington, DC, USA, 2007.
[7] J. Koomey, *Growth in Data Center Electricity Use 2005 to 2010*. Oakland, CA, USA: Analytics Press, Aug. 2011.
[8] G. Meijer, "Cooling energy-hungry data centers," *Science*, vol. 328, no. 5976, pp. 318–319, Apr. 2010.
[9] G. Group, *Forecast: Data centers, worldwide, 2010–2015*, Accessed March 2013. [Online]. Available: http://www.gartner.com

[10] S. Greenberg, E. Mills, B. Tschudi, P. Rumsey, and B. Myatt, "Best practices for data centers: Lessons learned from benchmarking 22 data centers," in *Proc. ACEEE Summer Study Energy Efficiency Build. Asilomar*, Pacific Grove, CA, USA, Aug. 2006, vol. 3, pp. 76–87.

[11] D. Kliazovich, P. Bouvry, and S. Khan, "DENS: Data center energy-efficient network-aware scheduling," in *Proc. IEEE/ACM GreenCom*, Dec. 2010, pp. 69–75.

[12] A. Trossman, G. Iszlai, M. Mihaescu, M. Scarth, P. Vytas, M. Li, and D. Hill, "Method and system for managing resources in a data center," Patent 8 122 453, Feb. 21, 2012.

[13] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "ElasticTree: Saving energy in data center networks," in *Proc. NSDI*, Apr. 2010, vol. 10, pp. 249–264.

[14] B. Khargharia, S. Hariri, F. Szidarovszky, M. Houri, H. El-Rewini, S. Khan, I. Ahmad, and M. Yousif, "Autonomic power performance management for large-scale data centers," in *Proc. IEEE Int. Parallel Distrib. Process. Symp.*, Mar. 2007, pp. 1–8.

[15] A. Verma, R. Koller, L. Useche, and R. Rangaswami, "SRCMap: Energy proportional storage using dynamic consolidation," in *Proc. FAST*, Feb. 2010, pp. 267–280.

[16] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-aware server provisioning and load dispatching for connection-intensive Internet services," in *Proc. 5th USENIX Symp. Netw. Syst. Des. Implementation*, Berkeley, CA, USA, Apr. 2008, pp. 337–350.

[17] M. Qiu, L. Zhang, Z. Ming, Z. Chen, X. Qin, and L. T. Yang, "Security-aware optimization for ubiquitous computing systems with seat graph approach," *J. Comput. Syst. Sci.*, vol. 79, no. 5, pp. 518–529, Aug. 2013.

[18] H. Ning, H. Liu, and L. Yang, "Cyber-entity security in the Internet of things," *Computer*, vol. 46, no. 4, pp. 46–53, Apr. 2013.

[19] W. Qiang, D. Zou, S. Wang, L. T. Yang, H. Jin, and L. Shi, "CloudAC: A cloud-oriented multilayer access control system for logic virtual domain," *IET Inf. Security*, vol. 7, no. 1, pp. 51–59, Mar. 2013.

[20] Z. Wang, N. Tolia, and C. Bash, "Opportunities and challenges to unify workload, power, and cooling management in data centers," in *Proc. 5th Int. Workshop Feedback Control Implementation Des. Comput. Syst. Netw.*, New York, NY, USA, Apr. 2010, pp. 1–6.

[21] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No power struggles: Coordinated multi-level power management for the data center," *ACM SIGARCH Comput. Archit. News*, vol. 36, no. 1, pp. 48–59, Mar. 2008.

[22] K. Kant, "Data center evolution: A tutorial on state of the art, issues, and challenges," *Comput. Netw.*, vol. 53, no. 17, pp. 2939–2965, Dec. 2009.

[23] R. A. Bergamaschi, L. Piga, S. Rigo, R. Azevedo, and G. Araújo, "Data center power and performance optimization through global selection of p-states and utilization rates," *Sustainable Comput., Inf. Syst.*, vol. 2, no. 4, pp. 198–208, Dec. 2012.

[24] X. Zhu, C. He, K. Li, and X. Qin, "Adaptive energy-efficient scheduling for real-time tasks on DVS-enabled heterogeneous clusters," *J. Parallel Distrib. Comput.*, vol. 72, no. 6, pp. 751–763, Jun. 2012.

[25] R. V. Aroca and L. M. G. Gonçalves, "Towards green data centers: A comparison of x86 and arm architectures power efficiency," *J. Parallel Distrib. Comput.*, vol. 72, no. 12, pp. 1770–1780, Dec. 2012.

[26] S. Li, K. Lim, P. Faraboschi, J. Chang, P. Ranganathan, and N. Jouppi, "System-level integrated server architectures for scale-out datacenters," in *Proc. 44th Annu. IEEE/ACM Int. Symp. Microarchitect.*, Dec. 2011, pp. 260–271.

[27] K. Lim, P. Ranganathan, J. Chang, C. Patel, T. Mudge, and S. Reinhardt, "Understanding and designing new server architectures for emerging warehouse-computing environments," in *Proc. 35th Int. Symp. Comput. Archit.*, Jun. 2008, pp. 315–326.

[28] J. Hamilton, "Cooperative Expendable Micro-slice Servers (CEMS): Low cost, low power servers for Internet-scale services," in *Proc. Conf. Innov. Data Syst. Res.*, Pacific Grove, CA, USA, Jan. 2009, pp. 1–8.

[29] T. Kgil, A. Saidi, N. Binkert, S. Reinhardt, K. Flautner, and T. Mudge, "PicoServer: Using 3D stacking technology to build energy efficient servers," *ACM J. Emerging Technol. Comput. Syst.*, vol. 4, no. 4, pp. 1–34, Oct. 2008.

[30] P. Ranganathan, P. Leech, D. Irwin, and J. Chase, "Ensemble-level power management for dense blade servers," *ACM SIGARCH Comput. Archit. News*, vol. 34, no. 2, pp. 66–77, May 2006.

[31] T. Mudge and U. Holzle, "Challenges and opportunities for extremely energy-efficient processors," *IEEE Micro*, vol. 30, no. 4, pp. 20–24, Jul./Aug. 2010.

[32] A. Szalay, G. Bell, H. Huang, A. Terzis, and A. White, "Low-power Amdahl-balanced blades for data intensive computing," *ACM SIGOPS Oper. Syst. Rev.*, vol. 44, no. 1, pp. 71–75, Jan. 2010.

[33] Y. Wang, H. Liu, D. Liu, Z. Qin, Z. Shao, and E. H.-M. Sha, "Overhead-aware energy optimization for real-time streaming applications on multiprocessor system-on-chip," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 16, no. 2, p. 14, Mar. 2011.

[34] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "Memory-aware optimal scheduling with communication overhead minimization for streaming applications on chip multiprocessors," in *Proc. IEEE RTSS*, Dec. 2010, pp. 350–359.

[35] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "Optimally removing inter-core communication overhead for streaming applications on MPSOCS," *IEEE Trans. Comput.*, vol. 62, no. 2, pp. 336–350, Feb. 2013.

[36] J. Xu and J. Fortes, "Multi-objective virtual machine placement in virtualized data center environments," in *Proc. IEEE/ACM Int. Conf. CPSCom*, Dec. 2010, pp. 179–188.

[37] L. Keys, S. Rivoire, and J. Davis, "The search for energy-efficient building blocks for the data center," in *Proc. Comput. Architect.*, 2012, pp. 172–182.

[38] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, "GPUS and the future of parallel computing," *IEEE Micro*, vol. 31, no. 5, pp. 7–17, Sep./Oct. 2011.

[39] P. Lotfi-Kamran, B. Grot, M. Ferdman, S. Volos, O. Kocberber, J. Picorel, A. Adileh, D. Jevdjic, S. Idgunji, E. Ozer, and B. Falsafi, "Scale-out processors," in *Proc. 39th Int. Symp. Comput. Archit.*, Jun. 2012, pp. 500–511.

[40] S. Pelley, D. Meisner, P. Zandevakili, T. Wenisch, and J. Underwood, "Power routing: Dynamic power provisioning in the data center," *ACM Sigplan Notices*, vol. 45, no. 3, pp. 231–242, Mar. 2010.

[41] M. Floyd, S. Ghiasi, T. Keller, K. Rajamani, F. Rawson, J. Rubio, and M. Ware, "System power management support in the IBM POWER6 microprocessor," *IBM J. Res. Develop.*, vol. 51, no. 6, pp. 733–746, Nov. 2007.

[42] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," *ACM SIGARCH Comput. Archit. News*, vol. 35, no. 2, pp. 13–23, May 2007.

[43] D. Meisner, B. T. Gold, and T. F. Wenisch, "PowerNap: Eliminating server idle power," *SIGPLAN Notices*, vol. 44, no. 3, pp. 205–216, Mar. 2009.

[44] R. Simanjorang, H. Yamaguchi, H. Ohashi, K. Nakao, T. Ninomiya, S. Abe, M. Kaga, and A. Fukui, "High-efficiency high-power dc–dc converter for energy and space saving of power-supply system in a data center," in *Proc. IEEE APEC Expo.*, Mar. 2011, pp. 600–605.

[45] A. H. Beitelmal and C. D. Patel, "Thermo-fluids provisioning of a high performance high density data center," *Distrib. Parallel Databases*, vol. 21, no. 2/3, pp. 227–238, Jun. 2007.

[46] S. V. Garimella, L.-T. Yeh, and T. Persoons, "Thermal management challenges in telecommunication systems and data centers," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 2, no. 8, pp. 1307–1316, Aug. 2012.

[47] D. Quirk and M. Patterson, "Ab-10-c021 the "right" temperature in datacom environments," *ASHRAE Trans.*, vol. 116, no. 2, p. 192, 2010.

[48] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, "Making scheduling "cool": Temperature-aware workload placement in data centers," in *Proc. Annu. Conf. USENIX Annu. Tech. Conf.*, Apr. 2005, pp. 61–75.

[49] N. Jiang and M. Parashar, "Enabling autonomic power-aware management of instrumented data centers," in *Proc. IEEE Int. Symp. Parallel Distrib. Process.*, May 2009, pp. 1–8.

[50] M. Ellsworth, L. Campbell, R. Simons, and R. Iyengar, "The evolution of water cooling for IBM large server systems: Back to the future," in *Proc. 11th Intersoc. Conf. Thermal Thermomech. Phenom. Electron. Syst.*, May 2008, pp. 266–274.

[51] S. Zimmermann, I. Meijer, M. K. Tiwari, S. Paredes, B. Michel, and D. Poulikakos, "Aquasar: A hot water cooled data center with direct energy reuse," *Energy*, vol. 43, no. 1, pp. 237–245, Jul. 2012.

[52] *Treat your data centre as an energy source*, Accessed Nov. 13, 2013. [Online]. Available: http://www.future-tech.co.uk/treat-your-data-centre-as-an-energy-source/

[53] T. Brunschwiler, B. Smith, E. Ruetsche, and B. Michel, "Toward zero-emission data centers through direct reuse of thermal energy," *IBM J. Res. Develop.*, vol. 53, no. 3, pp. 11:1–11:13, May 2009.

[54] Y. Wang, R. Chen, Z. Shao, and T. Li, "SolarTune: Real-time scheduling with load tuning for solar energy powered multicore systems," in *Proc. 19th IEEE Int. Conf. Embedded RTCSA*, Aug. 19–21, 2013, pp. 101–110.

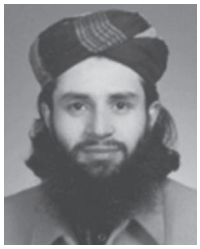
[55] C. Li, W. Zhang, C.-B. Cho, and T. Li, "SolarCore: Solar energy driven multi-core architecture power management," in *Proc. IEEE 17th Int. Symp. HPCA*, Feb. 2011, pp. 205–216.

[56] C. Li, A. Qouneh, and T. Li, "iSwitch: Coordinating and optimizing renewable energy powered server clusters," in *Proc. 39th Annu. ISCA*, Jun. 2012, pp. 512–523.

[57] J. Siriwardana, S. Jayasekara, and S. K. Halgamuge, "Potential of air-side economizers for data center cooling: A case study for key Australian cities," *Appl. Energy*, vol. 104, pp. 207–219, Apr. 2013.
[58] A. Yoder, "Energy efficient storage technologies for data centers," in *Proc. Workshop Energy-Efficient Des.*, Saint Malo, France, Jun. 2010, pp. 13–18.
[59] Q. Zhu, Z. Chen, L. Tan, Y. Zhou, K. Keeton, and J. Wilkes, "Hibernator: Helping disk arrays sleep through the winter," *ACM SIGOPS Oper. Syst. Rev.*, vol. 39, no. 5, pp. 177–190, Dec. 2005.
[60] C. Weddle, M. Oldham, J. Qian, A. Wang, P. Reiher, and G. Kuenning, "PARAID: A gear-shifting power-aware raid," *ACM Trans. Storage*, vol. 3, no. 3, p. 13, Oct. 2007.
[61] D. Narayanan, E. Thereska, A. Donnelly, S. Elnikety, and A. Rowstron, "Migrating server storage to SSDs: Analysis of tradeoffs," in *Proc. 4th ACM Eur. Conf. Comput. Syst.*, Apr. 2009, pp. 145–158.
[62] C. Lam, H. Liu, B. Koley, X. Zhao, V. Kamalov, and V. Gill, "Fiber optic communication technologies: What's needed for datacenter network operations," *IEEE Commun. Mag.*, vol. 48, no. 7, pp. 32–39, Jul. 2010.
[63] A. Caulfield, L. Grupp, and S. Swanson, "Gordon: Using flash memory to build fast, power-efficient clusters for data-intensive applications," *ACM Sigplan Notices*, vol. 44, no. 3, pp. 217–228, Mar. 2009.
[64] K. Sudan, N. Chatterjee, D. Nellans, M. Awasthi, R. Balasubramonian, and A. Davis, "Micro-pages: Increasing dram efficiency with locality-aware data placement," *ACM Sigplan Notices*, vol. 45, no. 3, pp. 219–230, Mar. 2010.
[65] T. Hatanaka, R. Yajima, T. Horiuchi, S. Wang, X. Zhang, M. Takahashi, S. Sakai, and K. Takeuchi, "Ferroelectric (Fe)—NAND flash memory with non-volatile page buffer for data center application enterprise solid-state drives (SSD)," in *Proc. VLSI Circuits Symp.*, Jun. 2009, pp. 78–79.
[66] P. Zhou, B. Zhao, J. Yang, and Y. Zhang, "A durable and energy efficient main memory using phase change memory technology," *ACM SIGARCH Comput. Archit. News*, vol. 37, no. 3, p. 14, Jun. 2009.
[67] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "An endurance-enhanced flash translation layer via reuse for NAND flash memory storage systems," in *Proc. DATE*, Mar. 2011, vol. 11, pp. 14–20.
[68] D. Liu, Y. Wang, Z. Qin, Z. Shao, and Y. Guan, "A space reuse strategy for flash translation layers in SLC NAND flash memory storage systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 6, pp. 1094–1107, Jun. 2012.
[69] Z. Qin, Y. Wang, D. Liu, Z. Shao, and Y. Guan, "MNFTL: An efficient flash translation layer for MLC NAND flash memory storage systems," in *Proc. 48th ACM/EDAC/IEEE Des. Autom. Conf.*, Jun. 2011, pp. 17–22.
[70] Y. Wang, L. A. D. Bathen, N. D. Dutt, and Z. Shao, "Meta-cure: A reliability enhancement strategy for metadata in NAND flash memory storage systems," in *Proc. 49th ACM/EDAC/IEEE DAC*, Jun. 2012, pp. 214–219.
[71] D. Liu, T. Wang, Y. Wang, Z. Qin, and Z. Shao, "A block-level flash memory management scheme for reducing write activities in PCM-based embedded systems," in *Proc. Conf. Des. Autom. Test Eur.*, Mar. 2012, pp. 1447–1450.
[72] X. Wang, Y. Chen, H. Li, D. Dimitrov, and H. Liu, "Spin torque random access memory down to 22 nm technology," *IEEE Trans. Magn.*, vol. 44, no. 11, pp. 2479–2482, Nov. 2008.
[73] D. Niu, Y. Chen, C. Xu, and Y. Xie, "Impact of process variations on emerging memristor," in *Proc. 47th ACM/IEEE DAC*, Jun. 2010, pp. 877–882.
[74] A. K. Mishra, X. Dong, G. Sun, Y. Xie, N. Vijaykrishnan, and C. R. Das, "Architecting on-chip interconnects for stacked 3D STT-RAM caches in CMPS," *ACM SIGARCH Comput. Archit. News*, vol. 39, no. 3, pp. 69–80, Jun. 2011.
[75] D. Andersen, J. Franklin, M. Kaminsky, A. Phanishayee, L. Tan, and V. Vasudevan, "FAWN: A fast array of wimpy nodes," in *Proc. ACM SIGOPS 22nd Symp. Oper. Syst. Principles*, Oct. 2009, pp. 1–14.
[76] J. Ousterhout, P. Agrawal, D. Erickson, C. Kozyrakis, J. Leverich, D. Mazières, S. Mitra, A. Narayanan, G. Parulkar, M. Rosenblum, S. M. Rumble, E. Stratmann, and R. Stutsman, "The case for RAMClouds: Scalable high-performance storage entirely in DRAM," *ACM SIGOPS Oper. Syst. Rev.*, vol. 43, no. 4, pp. 92–105, Jan. 2010.
[77] H. David, C. Fallin, E. Gorbatov, U. R. Hanebutte, and O. Mutlu, "Memory power management via dynamic voltage/frequency scaling," in *Proc. 8th ACM Int. Conf. Auton. Comput.*, Jun. 2011, pp. 31–40.
[78] *Cisco Data Center Infrastructure 2.5 Design Guide*, Cisco Press, Indianapolis, IN, USA, 2007.
[79] K. Bilal, S. U. Khan, L. Zhang, H. Li, K. Hayat, S. A. Madani, N. Min-Allah, L. Wang, D. Chen, M. Iqbal, C.-Z. Xu, and A. Y. Zomaya, "Quantitative comparisons of the state-of-the-art data center architectures," *Concurr. Comput., Pract. Exp.*, vol. 25, no. 12, pp. 1771–1783, Aug. 2013.
[80] K. Bilal, S. U. R. Malik, O. Khalid, A. Hameed, E. Alvarez, V. Wijaysekara, R. Irfan, S. Shrestha, D. Dwivedy, M. Ali, U. S. Khan, A. Abbas, N. Jalil, and S. U. Khan, "A taxonomy and survey on green data center networks," *Future Gen. Comput. Syst.*, vol. 36, pp. 189–208, Jul. 2014, to be published.
[81] K. Bilal, S. Khan, S. Madani, K. Hayat, M. Khan, N. Min-Allah, J. Kolodziej, L. Wang, S. Zeadally, and D. Chen, "A survey on green communications using adaptive link rate," *Cluster Comput.*, vol. 16, no. 3, pp. 575–589, Oct. 2013.
[82] K. Christensen, P. Reviriego, B. Nordman, M. Bennett, M. Mostowfi, and J. A. Maestro, "IEEE 802.3az: The road to energy efficient Ethernet," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 50–56, Nov. 2010.
[83] H. Anand, C. Reardon, R. Subramaniyan, and A. George, "Ethernet Adaptive Link Rate (ALR): Analysis of a MAC handshake protocol," in *Proc. 31st IEEE Conf. Local Comput. Netw.*, Nov. 2006, pp. 533–534.
[84] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen, "Reducing the energy consumption of Ethernet with adaptive link rate (ALR)," *IEEE Trans. Comput.*, vol. 57, no. 4, pp. 448–461, Apr. 2008.
[85] H. Abu-Libdeh, P. Costa, A. Rowstron, G. O'Shea, and A. Donnelly, "Symbiotic routing in future data centers," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 51–62, Oct. 2010.
[86] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, "DCell: A scalable and fault-tolerant network structure for data centers," in *Proc. ACM SIGCOMM Conf. Data Commun.*, New York, NY, USA, Oct. 2008, pp. 75–86.
[87] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 63–74, Oct. 2008.
[88] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, and S. Lu, "BCube: A high performance, server-centric network architecture for modular data centers," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 4, pp. 63–74, Aug. 2009.
[89] A. Vahdat, H. Liu, X. Zhao, and C. Johnson, "The emerging optical data center," presented at the Optical Fiber Communication Conference, Los Angeles, CA, USA, Mar. 2011 Paper OTuH2.
[90] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Vl2: A scalable and flexible data center network," in *Proc. ACM SIGCOMM Conf. Data Commun.*, New York, NY, USA, Oct. 2009, pp. 51–62.
[91] G. Lu, C. Guo, Y. Li, Z. Zhou, T. Yuan, H. Wu, Y. Xiong, R. Gao, and Y. Zhang, "ServerSwitch: A programmable and high performance platform for data center networks," in *Proc. NSDI*, Apr. 2011, pp. 1–14.
[92] P. Mahadevan, S. Banerjee, P. Sharma, A. Shah, and P. Ranganathan, "On energy efficiency for enterprise and data center networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 94–100, Aug. 2011.
[93] D. Abts, M. Marty, P. Wells, P. Klausler, and H. Liu, "Energy proportional datacenter networks," in *ACM SIGARCH Comput. Archit. News*, Jun. 2010, vol. 38, no. 3, pp. 338–347.
[94] N. Farrington, G. Porter, S. Radhakrishnan, H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: A hybrid electrical/optical switch architecture for modular data centers," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 339–350, Oct. 2011.
[95] Y. Shang, D. Li, and M. Xu, "Energy-aware routing in data center network," in *Proc. 1st ACM SIGCOMM Workshop Green Netw.*, New Delhi, India, Aug. 2010, pp. 1–8.
[96] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling innovation in campus networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, Mar. 2008.

**Junaid Shuja** received the B.S. degree in computer and information science from Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan, in 2009 and the M.S. degree in computer science from COMSATS Institute of Information Technology, Abbottabad, Pakistan, in 2012. He is currently working toward the Ph.D. degree in the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia.

His research interests include mobile virtualization, ARM-based servers, and energy-efficient data centers networks, and BCS.



**Kashif Bilal** is a doctoral student at North Dakota State University. His research interests include cloud computing, data center networks, green computing, and distributed systems. He is a student member of IEEE.

**Sajjad A. Madani** is currently an Associate Professor with and the Chairman of the Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad, Pakistan. He has published more than 60 research articles in peer-reviewed journals and conferences. His research interests include wireless sensor networks, application of IT to electrical energy networks, and large-scale systems.

**Mazliza Othman** received the M.Sc. and Ph.D. degrees from the University of London, London, U.K.

She is currently a Senior Lecturer with the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. She is the author of "Principles of Mobile Computing & Communications".

**Rajiv Ranjan** is currently a Senior Research Scientist and Julius Fellow with CSIRO Computational Informatics, Canberra, Australia, where he is working on projects related to cloud and big data computing. He has been conducting leading research in the area of cloud and big data computing developing techniques for quality-of-service-based management and processing of multimedia and big data analytic applications across multiple cloud data centers.

**Pavan Balaji** holds appointments as a Computer Scientist and Group Lead with Argonne National Laboratory, Lemont, IL, USA; as a Research Fellow with the Computation Institute, The University of Chicago, Chicago, IL; and as an Institute Fellow with the Northwestern Argonne Institute of Science and Engineering, Northwestern University, Evanston, IL. His research interests include parallel programming models and runtime systems for communication and I/O, modern system architecture, cloud computing systems, and resource management.

**Samee U. Khan** is currently an Assistant Professor with North Dakota State University, Fargo, ND, USA. His work appears in more than 225 publications. His research interests include optimization, robustness, and security of cloud, grid, cluster, and big data computing, social networks, wired and wireless networks, power systems, smart grids, and optical networks.

Mr. Khan is a Fellow of The Institution of Engineering and Technology.