2013

Technological Drivers in Data Centers and Telecom Systems: Multiscale Thermal, Electrical, and Energy Management

S V. Garimella
Purdue University, sureshg@purdue.edu

T. Persoons
University of Dublin, Trinity College

J. Weibel
Purdue University

L.-T. Yeh
Huawei Technologies

Follow this and additional works at: https://docs.lib.purdue.edu/coolingpubs

http://dx.doi.org/http://dx.doi.org/10.1016/j.apenergy.2013.02.047

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Technological Drivers in Data Centers and Telecom Systems: Multiscale Thermal, Electrical, and Energy Management *

Suresh V. Garimella a, Tim Persoons b, Justin Weibel c, Lian-Tuu Yeh d

a School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, USA, E-mail: sureshg@purdue.edu (Corresponding author. Phone: (765) 494 5621)

b Department of Mechanical and Manufacturing Engineering, University of Dublin, Trinity College, Dublin, Ireland, E-mail: tim.persoons@tcd.ie

c School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, USA, E-mail: jaweibel@purdue.edu

d Huawei Technologies, North American Headquarters, 2330 Central Expressway, Santa Clara, CA 95050, USA, E-mail: Jjyeh2@aol.com

Abstract – We identify technological drivers for tomorrow’s data centers and telecommunications systems, including thermal, electrical and energy management challenges, based on discussions at the 2nd Workshop on Thermal Management in Telecommunication Systems and Data Centers in Santa Clara, California, on April 25-26, 2012. The relevance of thermal management in electronic systems is reviewed against the background of the energy usage of the information technology (IT) industry, encompassing perspectives of different sectors of the industry. The underlying drivers for progress at the business and technology levels are identified. The technological challenges are reviewed in two main categories – immediate needs and future needs. Enabling cooling techniques that are currently under development are also discussed.

Keywords – Electronics cooling; 3D chip packaging; Power utilization effectiveness; Total cost of ownership; Cloud computing; Alternative air movers; Liquid cooling.

* This paper is based on discussions at the 2nd Workshop on Thermal Management in Telecommunication Systems and Data Centers, held in Santa Clara, CA on April 25-26, 2012, which was dedicated to Richard Chu, IBM Fellow Emeritus, for his 50+ years of distinguished services and significant contributions to the heat transfer community. The authors offer this paper in memory and honor of his towering presence in the community.
1 Background

1.1 Statistics on IT energy usage

The combined worldwide electricity consumption of data centers has increased from 71 billion kWh per year (in 2000) to 152 billion kWh per year (in 2005) [1] to approximately 238 billion kWh per year (in 2010) [2], representing a growth of roughly 11% per year over the last decade (Fig. 1). As a fraction of the worldwide total electricity usage for all sectors [3], the contribution of data centers has increased from 0.53% in 2000 and 0.97% in 2005 to 1.31% in 2010. Based on electricity statistics by country for 2009, the overall electricity consumption of data centers is comparable to the electricity production of countries like Australia, Mexico, Saudi Arabia, and Iran [4,5].

The growing IT demand is outpacing technological developments in sustainable energy management for these systems. Between 2003 and 2008, the total energy consumption of servers has doubled, showing only a temporary slowdown during the global economic crisis [6]. Such growth levels are unsustainable, and are especially worrying because IT equipment already contributes significantly to global energy use and carbon emissions. For Australia in 2005, the equivalent carbon emissions of the IT equipment (comprising data centers and smaller distributed equipment throughout the commercial sector) amounted to approximately 1.5% of the country total [7]. Based on the same study, the carbon footprint of IT equipment is comparable to that of civil aviation and the metal industry (respectively 1% and 2.3% of the country total).

According to a study by The Uptime Institute, the average ratio of three-year site costs (comprising operational and capital costs) to one rack unit (1U) server costs has exceeded 100% as of 2005 and the projected increase is not tenable [8]. Most of the available statistics report on energy consumption as an operational expenditure. The energy cost is the fastest growing expenditure in data centers, currently averaging about 12% of the total operating cost [9]. From a holistic point of view, one should also look at the life-cycle cost comprising energy consumption of manufacturing, mining, materials processing, and recycling. Of particular concern are mobile and personal computing devices, with a rapidly growing market and a typical lifetime of only a few years. For such devices, indirect energy use, viz. in manufacture and end-of-life treatment, may account for over 50% of the total lifetime energy consumption according to a life-cycle assessment performed by the University of California Energy Institute [10].

In an average data center, the IT equipment itself (i.e., rack-mounted servers) uses only about half of the total energy, with the remaining 50% overhead being used for cooling and electrical power delivery [1,2]. The electrical power delivery losses arise from the uninterruptible power supply (UPS), several voltage conversions and electrical transport within the facility. External transport of electrical power and electronic data between the data center and the grid is not accounted for in typical energy usage statistics. The dominating factor in the 50% overhead is the mechanical cooling plant, accounting for roughly 33% of the total energy consumption of a data center facility. Figure 2 shows a Sankey diagram representing the distribution of the incoming power among the three main system components (IT equipment, system cooling and electrical power delivery). A number of sections of this paper that focus
on particular aspects of this energy chart are indicated, with red text representing energy usage metrics (Sect. 1.2) and strategies for energy saving (Sect. 1.3), and green representing the cooling challenges and solutions discussed in Sect. 3 and 4.

1.2 Metrics for IT energy management

1.2.1 Power Usage Effectiveness (PUE)

The most commonly used descriptor of data center energy efficiency is the power usage effectiveness (PUE) proposed by the Green Grid initiative [11]. PUE represents the ratio of total power required to operate a system (including cooling, power distribution and other overheads) to the power used only by the IT equipment. PUE is the reciprocal of the data center infrastructure efficiency (DCiE). Symbolically,

\[
PUE = \frac{P_{\text{tot}}}{P_{IT}} = \frac{P_{IT} + P_T + P_E}{P_{IT}} = \frac{1}{\text{DCiE}}
\]

where \( P_{IT}, P_T \) and \( P_E \) are the power consumption of the IT equipment, thermal management system and electrical power distribution system, respectively.

Between 2000 and 2010, the worldwide average PUE value has favorably decreased only slightly from about 2.0 to a value between 1.83 and 1.92 (see Fig. 3) [1,2]. A study by the U.S. Environmental Protection Agency (EPA) [12] found an average PUE of 1.91 based on a voluntarily submitted survey of 120 data centers, which may represent an optimistic estimate of the actual average.

There is a wide variation in PUE between different data centers, as exemplified by the range of PUE values between 1.25 and 3.75 reported by the aforementioned EPA study [12]. The value depends not only on the type of cooling infrastructure (e.g., free cooling, mechanical cooling) and IT equipment (e.g., type of servers, spatial arrangement, primary cooling medium), but also on the geographical location and local climate around the facility. The lowest attainable PUE value for a given system depends strongly on the average ambient temperature available as the ultimate heat sink. For the majority of systems, the ultimate heat sink is the ambient air although in some cases nearby water streams or reservoirs are used because of their more stable temperature. One of the lowest PUE values on record is for a data center in the National Renewable Energy Laboratory [13] at PUE = 1.06.

From a business point of view, striving for PUE values close to unity is beneficial in terms of reducing the percentage overhead operational cost associated with electricity consumption for non-computing purposes. However, capital investment costs tend to increase in a nonlinear fashion, with diminishing returns as PUE approaches 1.0, resulting in a trade-off which depends on the factors affecting the PUE as mentioned above. In recent years, Google has made significant efforts to decrease its company-wide PUE to about 1.14 (Fig. 3 [2]) but the benefits of further reductions do not warrant the costs, from an economic perspective. Yet, for a dominant presence such as Google to reduce the PUE slightly from 1.14 to 1.13 would mean a reduction in their total energy consumption of approximately 19 GWh annually [2], or a reduction in total power consumption of 2.2 MW.
While economic considerations may prevent further reduction of the PUE values toward 1.0, energy savings of a comparable magnitude may instead be realized through a shift in the governing objective – to reduce the absolute value of energy consumption of the system (e.g., by increasing server utilization, or enhancing the energy efficiency within the server rack).

1.2.2 Changing the definition of PUE

The standard definition of the power usage effectiveness is somewhat simplistic and perhaps too restrictive, using only the most readily measurable quantities. Firstly, the IT electrical power consumption is taken at the input to the server racks. Therefore, power used by fans or other cooling devices inside a server rack is included in the IT equipment contribution, not in the thermal management contribution. This may result in misleading PUE values when racks with more intensive internal cooling are used. Secondly, some non-standard operational expenditures are not accounted for, such as energy expenditure in the treatment of river or sea water used for cooling, which may again result in misleading PUE values. Thirdly, the PUE ignores the embedded energy resulting from a life-cycle assessment of the equipment and infrastructure. Application of this kind of holistic approach is in its infancy, but it might present a very different picture that rewards lean manufacturing processes and operational sustainability.

Moreover, in the original PUE definition, the useful output is identified as the energy consumption of the IT equipment, yet the actual useful output is the computational power, which is commonly expressed in teraflops. Therefore, some authors have proposed a coefficient of performance \[ COP = \frac{P_{IT} - P_{IT,\text{leak}} - P_{IT,fans} - P_{IT,idle}}{P_T + P_E + P_{IT,\text{leak}} + P_{IT,fans} + P_{IT,idle}} \] (2)

The energy consumption due to (i) transistor leakage power \( P_{IT,\text{leak}} \), (ii) rack-internal cooling power \( P_{IT,fans} \), and (iii) idling power \( P_{IT,idle} \) is subtracted from the IT equipment power consumption \( P_{IT} \), leaving only the power consumption which goes to computational purposes in the numerator. The three overhead components are lumped together with the infrastructure power consumption (thermal management and electrical power delivery), so that the numerator represents the overall power overhead required to enable the computation to take place.

Other types of metrics have been proposed by HP to optimize data center operation, such as (i) a temperature-based thermal correlation index used to identify the zone of influence of a computer room air conditioning (CRAC) unit [15], (ii) air delivery efficiency metrics [16,17], and (iii) computational workload placement metrics [18-20].

1.2.3 Energy cost per teraflop

Energy-performance metrics are gaining importance for IT equipment, and are usually expressed in terms of the number of logic operations per unit energy, or number of logic operations per second per unit power. From 2006 to 2012, Intel Xeon® servers have maintained about the same level in terms of
electrical power but the number of operations per second has increased nearly tenfold [21]. The improvement is largely CPU and platform-driven, so one way to achieve better efficiency is to upgrade old systems.

For instance, LRZ-Munich’s high performance computing center uses 40°C water for cooling its servers to achieve free cooling (i.e., direct heat exchange between the primary cooling water and the outside air, without resorting to mechanical refrigeration) throughout the year, resulting in a PUE value of 1.15 [22]. For three successive generations of servers used in this facility between 2000 and 2012, the total power consumption increased six-fold, resulting in a commensurate increase in energy costs. However, the number of logic operations per unit energy increased nearly 400-fold over the three generations of servers. As such, in terms of energy cost per teraflop, the system is currently operating at €7/teraflop compared to €2,300/teraflop in 2000-2006, although the total energy bill is still increasing to levels that are cause for concern [22].

Based on the current projections for supercomputers [23], computational performance approaching 1 exaflops (i.e., 1 million teraflops) is expected by 2019. Even with continued exponential growth of performance per unit energy, further developments are needed to avoid excessive power consumption levels.

The growing awareness about the energy use in IT equipment itself is encouraging. While striving for a PUE ratio close to 1, the absolute value of IT power consumption (i.e., the denominator in Eq. (1)) should not be overlooked.

While these general energy usage statistics and energy efficiency metrics for the IT industry focus on data centers, the conclusions broadly apply to telecommunication systems as well. Strategies for energy savings discussed below are relevant to both data centers and telecommunication systems.

1.3 Strategies for energy savings

1.3.1 Alternative design and operating methodologies

The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) expanded the operating envelope for air-cooled data processing facilities to a broader range of temperatures [24], adding classes with a maximum rack inlet temperature up to 45°C (Class A4) in 2011, whereas the previous maximum allowable was only 35°C (Class A2) in 2008. For an air-cooled data center, operating the server room at elevated temperatures results in (i) the need for increased internal air flow rates within the racks (e.g., higher average fan speeds) to compensate for the higher air inlet temperature, and (ii) a higher heat transfer potential to the outside ambient air. For a typical data center, the overall net effect is a reduction of the infrastructure energy consumption.

The challenge is to engineer platforms with a full range of components capable of performing reliably and consistently at the warmer temperatures. To safeguard reliability and avoid overheating issues, advanced operating methods can take advantage of the instrumentation already available on a typical server platform. Current volume servers (i.e., a computer server packaged in a 1U or 2U high rack-mount
chassis, typically fitted with up to four processor sockets \([25]\)) have over 50 temperature sensors on board, spread out over various components and the motherboard.

A three-fold strategy is recommended by Google for achieving radically stripped down air-cooled data centers: (1) manage the air flow by metering flows and separating cold and hot streams, (2) optimize the efficiency of the mechanical cooling plant (i.e., the largest overhead component in the thermal management system) using advanced economizer design, which enables 80% efficiency (see Sect. 1.3.2), and (3) optimize the UPS system to an efficiency of 99.9% by trickle-charging battery banks.

Additional specific examples are given along with immediate and future challenges (Sect. 3).

### 1.3.2 Alternative ultimate heat sinks

The outside air temperature has a direct effect on the efficiency of the cooling system. However the air temperature in any location exhibits strong fluctuations in time (both day to night variations and average daily temperature variations throughout the year), even in temperate, maritime and colder climates. As such, it may be desirable to use other ultimate heat sinks with a more stable temperature level, such as large water reservoirs (e.g., seas, lakes, and rivers).

For instance, Google uses this approach for some of their data centers: using a river-fed water-side economizer in Saint Ghislain, Belgium; using an economizer with water sprayed over wetted media in Dublin, Ireland; and using sea water from the Bay of Finland in Hamina, Finland. In all cases, the servers are operated at the maximum possible air inlet temperature, to take advantage of the maximum possible temperature differential. This allows free cooling without the need for mechanical chillers.

Use of sea water for cooling poses challenges, although corrosion risks can be limited using plastic piping. Care should be taken to limit the temperature of the discharged water after passing through the heat exchangers so as to avoid adverse effects on the ecosystem.

### 1.3.3 Alternative energy supplies

To replace, back up or supplement traditional (electrical) power supply, several alternative energy sources could be considered. Thermoelectric generators could help to recover some energy from waste heat, or fuel cells could be used to generate power locally from alternative sources; both approaches typically have low overall generation efficiencies, but may utilize otherwise wasted energy sources. Depending on availability, renewable energy sources could be attractive in certain regions, yet most of these sources are inherently intermittent. Integrated thermal storage solutions using CO\(_2\) heat pumps might help to increase the potential for electronics thermal management, resulting in an intermittency-friendly data center \([26]\). From among the myriad potential options, the viability of specific alternative electrical energy sources depends on a number of interconnected and time-varying technical, geographic, economic, and political factors, and is likely to be determined on a case by case basis. Hence, discussion of specific alternative options is outside the scope of this article.
1.3.4 Alternative uses for waste heat

Space heating accounts for about 24% of the worldwide energy consumption [3], so there is significant potential for reusing waste heat from IT equipment to supply part of this heating demand. The main challenge is not so much technological, but rather in the logistics of identifying the value of heat and finding the most suitable customers within reasonable distances from the data center locations.

There are regional differences in the value of heat, because of climate, available infrastructure and population density. In Europe, heat is worth about 16 Eur/GJ (with a total district heating energy supply of 397,000 GWh in 2009) [27] which is roughly half of the average price of electricity in Europe (0.092-0.12 Eur/kWh in 2009) [28]. Because of the abundance of urban district heating systems, finding customers is relatively straightforward. There are no clear cost models for heat, but its value scales with the temperature difference between the transport medium and the local ambient temperature. As such, heat has a higher value in cooler climates for space heating, but it can have alternative uses in hotter climates as well, e.g., desalination of sea water, or for space cooling using adsorption chiller systems. Heat can be efficiently transported over distances of several tens of km, with a linear loss with distance. Losses are generally about ten times higher than for a comparable power level of electricity.

From this perspective, improvements in exergy efficiency can mean savings in operating cost [29]. From a design perspective, recognizing the value of heat can lead to a completely different optimum. The dissipated heat can be taken from the core of the IT equipment using liquid cooling, transported in the same form to nearby customers and sold as heat, bypassing air cooling altogether. By collaborating with urban development authorities, new integrated ecosystems could emerge around data centers.

2 Drivers for progress

The driving forces underlying the further development of data center thermal management are reviewed, at the business (Sect. 2.1) and technology (Sect. 2.2) levels. This summary reflects the topics of discussion at the 2nd Workshop on Thermal Management in Telecommunication Systems and Data Centers in Santa Clara, CA on April 25-26, 2012. The tag cloud in Fig. 4 represents the keywords brought up during the panel discussions.

2.1 Business level: Impact of energy usage on strategic plans

Recent years have seen a shift in engineering focus towards the data center infrastructure, whereas previously the focus was on the IT equipment with cooling provided in an ad hoc manner. In terms of power density, we have reached a crossroads with two opposing directions: (i) increasing power density for high performance computing where the dominant cost is attributed to interconnects and communication between processors and memory, and (ii) lower density systems for cloud computing operations, possibly limited to 8-14 kW/rack to enable free cooling. The decision on which direction to pursue is determined by the balance between investment cost and energy cost. Even with a continued exponential improvement in cost efficiency (computational performance per unit energy), the total
energy consumption is expected to rise, based on the experience over the past decade (Sect. 1.2.3). Server energy use has doubled between 2003 and 2008, with only a temporary slowdown due to the global economic recession [6]. As this growth continues and energy costs exceed investment cost, a shift in paradigm will arise. Currently, major companies such as IBM are literally creating hot air for an energy bill worth hundreds of millions of dollars, which is an important driver in examining alternative methods of reducing this expenditure or recovering some of the lost exergy.

2.1.1 Total Cost of Ownership (TCO)

The ultimate bottom line to operating a successful business is the compounded cost over the life time of operation, comprising capital investment and operational cost, i.e., the total cost of ownership (TCO). Any technological breakthrough must answer to this criterion. The challenge in getting a comprehensive estimate of cost for operation of a data center is that cost can be moved between IT equipment and infrastructure, and as various investment decisions are made at different times, this can lead to confusion about cost allocation.

As the most commonly used efficiency metric for data centers, the power usage effectiveness (PUE) is related to the TCO. However, the relationship between PUE and TCO is not straightforward, and the definition of PUE itself may need to be updated (Sect. 1.2.2). For a complex operation like a data center, there is a need to develop better TCO cost models. A composite model would be more useful than a single value metric such as PUE.

So although TCO is arguably the main driver behind corporate strategic plans, it is not clear how the TCO cost model is affected by quantifiable parameters (e.g., energy prices, equipment and infrastructure cost) and other less quantifiable aspects (e.g., public perception and legislation).

2.1.2 What drives corporate strategies?

In a rapidly evolving IT market, it is crucial to outline corporate strategies to try to match future customer demands. Companies like Intel employ futurists to help predict what end users will be expecting in the next decade and beyond [30], such as the evolution towards the wide range of form factors seen today, or the rapid expansion of the population carrying mobile devices that has surpassed 75% of the world's inhabitants [31].

As an example of the complex nature of TCO and corporate strategies, Google has been striving towards a zero carbon impact philosophy, offsetting its carbon footprint by buying green energy. Choosing this more expensive green electricity supply seems to be based on more than short term monetary benefit, and reflects the complexity of IT market strategies and evolving public demands. The subjective value of public perception in achieving the lowest PUE value in data centers is an open question.

In reality, achieving a lower PUE value is still a powerful force in driving down overhead cost. With the transition towards cloud computing come new opportunities and new challenges. The introduction of the new ASHRAE standards [24] has led to increased dialog between data center operators and server designers. Instead of using off-the-shelf servers, the servers are reduced to bare-bones cloud-optimized
systems, striving for self-healing IT equipment instead of relying on concurrent maintenance. The infrastructure should be resilient and capable of handling increased variability of inputs. The cooling system and servers are co-designed to reduce the energy consumption cost. The aesthetics are not important; what matters is trading redundancy for resilience. The economics of data center operations is moving away from mainly capital expenditure (CapEx) investment-driven costs to more operational expenditure (OpEx) or energy-driven costs. As a result, there is a shift in focus towards optimizing the overall, integrated system encompassing the IT equipment and utilities, rather than focusing only on component-optimized designs.

This strategy is exemplified in some of Microsoft's largest data centers to date:

- Dublin, Ireland (30,000 m², 22 MW) using air economizers without the need for mechanical chillers, which is scheduled for expansion [32] in 2012 with an additional 10,000 m² and 13 MW.
- Chicago, IL (65,000 m², 60 MW) [33] comprising only stacked containerized units. Unlike the facility in Dublin, chillers are still needed to cover the hottest days, but free cooling is used, outside temperature-permitting.
- Quincy, WA (46,000 m², 27 MW, PUE < 1.2) features a modular containerized design with a roof to protect from the elements, using only outside air and evaporative cooling [34]. The location is chosen partly because of cheap electricity prices due to the nearby Columbia River hydroelectric power plants. The Boydton, VA facility also uses adiabatic (evaporative) cooling instead of chillers.

In this same mindset, future data centers might become more physically integrated with utilities such as electric power plants, opening up alternatives for combined cycle generation, or using alternative forms of energy such as gas-powered fuel cells instead of electricity, that take advantage of a reliable gas supply. The scale of cloud computing might be a determining factor in pushing this evolution.

Although liquid cooling makes sense technologically as a superior heat transfer/transport medium in comparison to air [35], without reusing waste heat it does not make sense from a business perspective for companies like Microsoft and Google. They have chosen to reduce the infrastructure to the bare essentials, following the above strategy. However for alternatively aimed, emerging high-performance, high-density computing technologies [21], liquid cooling may prove to be a more viable option based on the TCO decision criterion.

Historically, the volume server business was controlled by a few companies with near-complete control) on technology development. Cloud computing has made the landscape more flexible. This evolution towards a self-organizing organic market affects even the larger players in their decision-making. While this seems like a natural tendency, care should be taken not to rely solely on random evolutionary mutations of technology. A reluctance to invest in radically new technology, preferring instead well-known air cooling technology in barebones systems, might not prove to be the most profitable long-term solution. Although cloud computing may work to reduce the risk of technology monopolies, in doing so it also risks a narrowing field of view and an industry-wide reluctance to invest in new ideas.

Encouragingly, there is evidence for investment in developing revolutionary technology, both in industry (e.g., IBM) and in academic and government research agencies (e.g., DARPA). If these developments
lead to significant cost reductions and performance gains, especially in the face of a growing cost of energy, cloud computing – by its very nature – will assimilate these changes.

2.2 Technological level

2.2.1 Balance between computation and communication costs

Arguably the main fundamental technological driver for technological progress is the balance between energy costs for computation and for communication. This balance applies both to the macro level (e.g., cloud computing versus local computing) and the micro level (i.e., the evolution towards 3D chip packaging).

The relative contribution of energy costs for communication is strongly affected by the system size. An analysis by IBM [36] shows that 99% of the overall energy consumption in an IT system goes toward communication, with only 1% being used for computation. Three decades ago, this ratio was close to a more sustainable 50/50 ratio [37]. In a typical air-cooled system, about 1 part per million of the volume is taken up by transistors and 96% of the volume is used for thermal transport. As communication costs become more important, so does this 96% waste in system volume. A liquid-cooled system has smaller system volume demands, making it an enabling technology for 3D chip architectures and higher system transistor density.

2.2.1.1 Macro level: Cloud versus local computing

The growing market for mobile handheld devices (e.g., smart phones and tablets) has led to increased data traffic, with only a limited amount of computing performed locally and the majority accomplished remotely in the so-called 'cloud' environment. This evolution should be seen in the context of the balance between computational and communication cost. Currently almost all the processing for a mobile phone is done in a data center, because the costs for transporting ‘electrons’ are not dominating yet, although transport may use up to 99% of the overall energy [37]. The processing itself could be done very efficiently in a small-scale integrated system, but would need to be driven by increased communication costs. In this context, IBM has proposed sugar cube-sized liquid-cooled systems to maximize computational efficiency. The balance between communication and processing cost will invoke a new balance between local and remote computing. In the future, the balance will most likely be much more energy cost-dominated (which may, for example, even lead to data centers being collocated with power plants to minimize energy costs). As either the cost of energy increases, and/or the data transfer volume increases, the balance might shift away from cloud computing towards more local computing.

If ultra-high density systems like IBM’s sugar cube concept take hold and can be implemented in handheld systems, the dependence on remote computing may be reduced. However, if these performance levels become readily available in a cost-competitive manner, data centers of the era would again simply use a cluster of such devices for maximum computational performance. This highlights the positive feedback paradox that may eventually evolve to the same balance as exists today.
The inevitable latency of a cloud system could be avoided by increasing local computing, but some mobile applications will likely always have higher demands than what can be done locally.

Regarding the demand for high performance computing (HPC) in scientific research, the transition to cloud computing is reminiscent of the evolution towards minicomputers some twenty years ago, due to the stranglehold on large-scale HPC facilities by a few large players. As a result, research efforts were scaled down to what could be done on minicomputers, and investments in large-scale systems stalled. However, it is important that the engineering community continue to support new investments in large-scale HPC, as these will enable the most demanding computations addressing grand-challenge problems that may not be achieved with even significant advances in cloud computing.

Figures 5 and 6 show two Sankey diagrams representing two extreme cases: (1) a ‘cloud-optimized’ data center employing air cooling at elevated room temperature to make optimum use of free cooling (Fig. 5), and (2) a fully liquid-cooled data center designed to maximize waste heat energy recuperation (Fig. 6). As in Fig. 2, sections of the paper are indicated in the diagrams where relevant. For the cloud computing case (Fig. 5), the aim of companies like Microsoft and Google is to eliminate the need for mechanical cooling by operating at elevated room air temperatures, allowing these systems to quote very low PUE values. However as discussed in Sect. 1.2.2, these low PUE values do not account for an increase in power consumption within the racks due to more intensive air cooling.

2.2.1.2 Micro level: Towards 3D chip packaging

At the chip level, there is a similar disparity with the energy for communication between cores and memory outweighing that needed for computing operations. The main advantage of 3D packaging, or the vertical integration of chips with different functions (processors, memory, sensors, etc.) is to reduce the communication overhead. Since power dissipation depends on wire length, the energy consumption attributed to communication scales with the square of the system size. By combining the main memory and processor into a single stack, there is no longer a need for cache memory. This (i) reduces the communication latency, and (ii) reduces the energy consumption per floating point operation by more than an order of magnitude.

The evolution of information technology can be seen in two ways: (1) device-centric, where the device performance dominates and evolution is driven by the introduction of better devices, and (2) density-centric, where communication efficiency dominates and evolution is driven towards denser, more efficient systems. This latter viewpoint is relevant for large systems at the petascale and beyond.

Looking at the evolution of IT equipment, the computational density (i.e., number of operations per second per unit volume) is increasing, as is the computational efficiency (i.e., number of operations per unit energy). As an example, the 2008 IBM Roadrunner system in Los Alamos National Laboratory [38] has reached a density of 1 gigaflops per liter at an efficiency of 0.4 gigaflop/J. By comparison, a human brain achieves about $10^4$-$10^7$ times more synapse operations per second per liter, at a $10^3$-$10^6$ times higher efficiency [39]. Although synapse operations cannot be directly compared to floating point operations, the volume- and energy-specific performance characteristics of IT equipment may be thought to be heading towards those of biological brains.
The development of future high performance computing systems at IBM is very much density-centric, with a recent report [37] describing how to scale a 1 petaflops system in 10 liters. The comparison between biological brains and IT equipment can be taken a step further, in that a brain is hot-water cooled, and waste heat is used to raise the body core temperature to achieve better efficiency. The pumping power for the circulatory system, provided by the heart, is about 5% of the total energy budget; a similar level of overhead could be envisaged for such a bio-inspired IT system.

The most recent predictions for high performance computers [23] expect a peak computational performance per system of 1 exaflop per second to be reached by about 2019. Even with continued exponential growth of performance per unit energy, further developments are needed to avoid excessive power consumption levels. Even if the energy consumption per core does not increase, stacking will make for a more difficult thermal management challenge.

### 2.2.2 Multidisciplinary opportunities

In the traditional design process, each component is optimized individually. Only in an integrated design process that optimizes the system as a whole do transport costs become apparent and can affect the overall design optimal.

In terms of driving technological progress, there are typically more opportunities at the interfaces than at the cores of each technical field (e.g., thermal and electrical design; data center and utility operators). Some problems are poorly understood across boundaries which is where synergy is needed. The IT community is diverse and scattered, consisting of end users, data center operators, chip manufacturers, equipment integrators and others, and there is a limited understanding of the perspectives of each. Similarly, conventionally separate electrical design for power consumption reduction (via supply voltage reduction) and improved thermal management technologies have provided only incremental performance extensions to current approaches, whereas electrothermal co-design driven 3D architectures (Sect. 2.2.1.2) may provide more revolutionary gains.

An example of an inherited misconception relates to the operating temperature range for servers. The previous ASHRAE standards specified a very narrow temperature range, which dates back to the days of water-cooled systems when the room was air-conditioned to make it comfortable for people. When air-cooled servers became prevalent, data center operators continued in the previous mode, using ever more powerful air conditioning units to maintain a comfortable working environment. The recent ASHRAE standards allow for a higher room temperature, which enables more energy efficient air-based cooling. This evolution has come about because of a closer dialog between data center operators and server designers.

### 3 Immediate and future challenges

A report by McKinsey & Co. [40] highlights the importance of addressing the energy consumption of IT equipment in the framework of improving the energy efficiency of the U.S. economy as a whole. The report identifies the most important barriers to capturing this efficiency potential: (i) low awareness, in
spite of IT equipment accounting for up to 25% of the electricity usage in the commercial sector, (ii) lack of understanding of energy efficiency compared to other attributes (e.g., price and technical performance), and (iii) poor procurement selection, partly due to misplaced focus on acquisition cost rather than lifecycle cost or scattered budgetary responsibility.

Although the energy consumption issue has been raising concern, it is not yet a dominant factor in the balance between computation and transport costs (see Sect. 2.2). However, it is expected that this balance will shift in the years ahead. This emphasizes the need to address challenges related to implementing more efficient cooling solutions for much denser packaging.

In terms of specific technological challenges, three-dimensional packaging and increasing density means that more developments are needed on the component level (e.g., high conductivity materials and thermal interfaces), as well as the rack and system level (e.g., fluidic and thermal couplings). Currently, air-cooled systems still represent over 95% of data center facilities, where an often overlooked challenge is acoustic noise emission.

From the point of view of designing a cooling system, a holistic approach is more appropriate than focusing on individual device optimization. Certainly for 3D packaging, electro-thermal co-design becomes very critical. One given approach (e.g., air or liquid) does not fit all needs, and continued developments are needed for both high-density (liquid cooled) and low-density (air cooled) approaches.

3.1 Thermal and electrical challenges

3.1.1 A brief history of electronics cooling challenges

Looking back in time, a high level of heat dissipation in itself does not necessarily pose a thermal challenge; rather, it is the heat flux which needs attention. In 1946, the Electronic Numerical Integrator And Computer (ENIAC) [41] built with vacuum tube technology used 174 kW of electrical power. However this only amounted to about 1 kW per square meter of floor space, so the thermal management focus was only at the room level.

As the power density increased, the focus shifted to rack cooling. By the mid 1970s to 1980, the first steps were taken towards thermal management inside the equipment, and the 1980s saw the introduction of liquid cooling of bipolar transistor processors [42], with a maximum bipolar chip power density of about 14 W/cm². After the transition from bipolar to CMOS technology, the period 1985-2000 was governed by enhanced air cooling, with a focus on smaller scales and hot spots. Now that the CMOS chip power density has reached a comparable level as in the bipolar transistor era of 15 W/cm², present and future chip-level thermal challenges will most likely involve (i) hot spots, and (ii) 3D packaging.

As an illustration of the challenge with hot spots, a 40 × 40 mm² quad core processor chip [43] with an average heat flux of 50 W/cm² can experience local heat loads up to 500 W/cm² causing local 30-40°C temperature spikes with thermal gradients up to 20°C/mm. The heat flux in these hot spots is about 1/10 of that on the surface of the sun (6,000 W/cm² at a temperature of about 5,500°C). However, the local heat transfer rate should be high enough to maintain a much lower temperature of about 100°C, a
level which would naturally occur at a heat flux of about 0.1 W/cm² (i.e., solar irradiation on Earth’s surface).

Three-dimensional packaging has inherently more complicated thermal constraints, because for chips within the stack there is no direct conduction path to the outside of the package. The heat density problem has become volumetric, and there is no viable cooling solution that can dissipate the estimated volumetric heat generation levels. For example, 100-1000 W/cm³ (at a temperature of 100°C) [44] is comparable in terms of power density to a liquid metal-cooled nuclear reactor operating at 200-300 W/cm³, yet at a much higher temperature level of 1400°C. For historical comparison, a Cray-3 module had a power density of approximately 39 W/cm³ at a temperature of 30°C using fluorocarbon immersion cooling [45].

Due to the limited available thermal solutions in this landscape, the focus of the next upcoming DARPA thermal management project [46] is on integrated microfluidics in manifolded liquid flow heat sinks, thermal interconnects, and thermal/electrical co-design. The main challenges identified by DARPA include (i) completing the inward migration of thermal packaging, (ii) extracting heat directly from the device, chip and package, and (iii) placing thermal management on an equal footing with functional design and power delivery. Further research is also needed into the physics of failure due to thermal cycling at elevated operating temperatures, which could significantly increase the driving force for cooling, as well as the available exergy for waste heat recovery.

3.1.2 Component level: Challenges for 3D packaging

As the computational performance of HPC systems continues to increase towards exascale computing, the communication bottleneck due to the physical distance between cores and memory becomes more critical, which is driving the development of 3D chip packaging.

In a traditional configuration with separate memory chips and a processor chip (with multiple cores and on-board cache) connected via a motherboard, the majority of the energy (99%) is used for communication compared to about 1% for computation [36]. It takes over 1000 clock cycles to access the main memory. The main advantage of 3D chip integration is to increase the core-to-memory bandwidth by moving the main memory into the processor chip, thereby reducing the length of the connecting wires by several orders of magnitude. Several technological challenges related to novel IC packaging methods (e.g., distributed 3D stacks, 3DIC, and tiled dies) will need to be addressed [47].

In terms of electrical interfacing, different connection methods can be used (e.g., wire bonding, flip chip, 3DIC with through-silicon vias (TSV), or package-on-package) which could all achieve a total package height of less than 1 mm for use in mobile devices. The thermo-mechanical stress induced in these integrated packages will increase compared to traditional single-die packages. Interconnect reliability remains a challenge considering the increased electrical current density to power these high density integrated devices.

The optimal placement of memory and processor cores within the stack is important, and requires intimate co-design of thermal and electrical aspects and possibly splitting up high power components
such as graphical processors into multiple units spread out across the stack. The thermal design will also need to compensate for height differences between neighboring components.

IBM’s proposed solution consists of single-phase interlayer cooling with microchannel pin fin or other micro-structured heat sink layers, interwoven with silicon layers and electrically connected with TSVs that can be electrically insulated from water flow. IBM has been conducting studies on the potential failure modes and optimization for heat transfer performance for these configurations [48,49].

### 3.1.3 Rack and system level

A power density of 20 kW per rack seems to be the threshold for most of today’s air-cooled IT systems. Beyond this threshold, hybrid cooling can be used with selective liquid cooling for high power components and air cooling for the remainder of the equipment. Some examples include IBM’s and Fujitsu’s approach with microchannel liquid cooled heat sinks [50,51], the use of rear-door heat exchangers and multiple internal air-to-liquid heat exchangers to reduce the inlet air temperature for upper rack levels [52]. For the highest power density, full liquid cooling is being used by IBM in the Swiss Federal Institute of Technology (ETH) Zurich [53] and the Leibniz Supercomputing Centre (LRZ) [54]. Alternative methods such as using dielectric sprays in a sealed cabinet have been developed by Bell Labs and Purdue University [55].

### 3.1.4 Robustness and reliability

While rack operation in harsh environments may be motivated in some instances by reduction of overhead energy costs (see Sect. 2.1.1), there is a steep increase in growth of the telecommunications industry in developing countries where temperature/humidity control infrastructure is less prevalent. For these systems, performance becomes increasingly dependent on design for robustness and reliability. From a thermal perspective, heat transfer surfaces are particularly susceptible to fouling either by particulate matter loading or corrosion. Empirical quantification of performance reduction due to fouling is lacking, and when available, it is often tied to a specific geometry. While ASHRAE has defined standard particulate matter compositions for testing (Standards 52.1, 52.2), these are primarily motivated by the air filter industry, and may not reveal the actual thermal performance ramifications for a wide range of environments. Research in this area is required to (i) design surface geometries and coatings for reduced fouling and/or resilience to particulate loading, (ii) realize procedures for effective active and/or passive self-cleaning, and (iii) develop tools for predictive modeling of fouling processes.

### 3.2 Acoustic noise emission

#### 3.2.1 Applicable measurement and operating standards

More so than other fields of engineering, acoustics is well covered by a number of standards. General acoustic noise measurement standards that apply to all products include ISO 3744 (referring to sound power levels in hemi-anechoic chambers) [56], ISO 3741 (referring to sound power levels in reverberation rooms) [57], and ISO 11201 (referring to emission sound pressure level at the location of
the operator and bystander) [58]. Specific standards for IT products include ISO 7779, Ecma-74 and ISO 10302 for measurement of noise levels [59-61], and ISO 9296 for the declaration of noise levels [62].

Acoustic noise standards that prescribe emission limits include the Swedish Statskontoret 26:6 [63] and the European Telecommunications Standards Institute (ETSI) 300 753 [64]. Although these are regional standards, they are adhered to worldwide by many industries.

Other noise exposure limits are enforced by legislation, such as OHSA regulation 1910.95 [65] which limits the maximum exposure to 90 dB(A) for 8 hours and requires preventive actions for exposure to 85 dB(A) for 8 hours. European Directive 89/391/EEC [66] limits exposure to 87 dB(A) for 8 hours, with lower and higher action levels at 80 dB(A) and 85 dB(A) for 8 hours, respectively.

The main source of acoustic noise in IT equipment is fans. The applicable measurement standard for fans is ISO 10302 [61], which gives the noise level as a function of fan load. However fan manufacturers typically only specify noise levels at a distance of 1 m away under zero load conditions, which is often too optimistic an estimate compared to realistic operating conditions.

### 3.2.2 Acoustic challenges for air-cooled IT equipment

Air cooling still remains an attractive thermal management solution for the majority of the IT industry, but the noise level must be reduced by developing high-performance low-noise fans. Noise production varies between fan types, and can be characterized by the intrinsic noise coefficient which is much lower for a centrifugal fan than an axial fan. However, acoustics experts believe that further gains in the development of quiet air-moving devices are very difficult to achieve; alternative solutions are discussed below.

Although there is some opportunity for noise reduction by careful fan selection, simultaneous demands of high flow performance and low noise emission are not easily reconciled.

The noise emission can be reduced passively, but this requires extra space for damping, which contradicts the objective of reducing floor space and communication losses. A possible scenario that would not affect floor space is to increase the rack height to free up space for acoustic damping. Passive acoustic damping at low frequencies is difficult, but effective attenuation of higher frequencies can be achieved with available methods such as acoustic doors.

Acoustic experts are skeptical about active noise cancellation in IT equipment. There are some cases where it might be appropriate but active cancellation remains very challenging for random noise. It also adds a significant cost, as well as the risk of amplifying noise if the controller phase control algorithm drifts.

There is some question about whether acoustic noise remains a primary concern for data centers. Since the new ASHRAE standards have elevated the maximum operating temperatures to 40-45°C at the inlet to the servers, this is no longer a worker-friendly environment, and noise reduction may not be relevant. Moreover, heat transfer is enhanced by strong turbulence, which indicates that the concomitant
increase in noise generation in air-cooled systems is an unavoidable side effect of efficient cooling [67]. There might be room for improvement by alternative air cooling methods, as discussed in Sect. 4.1.

3.3 Allocation of computation and server utilization

Computational loads vary in time, both on a small scale at the chip level and on a large scale at a facility level. At the chip level, hot spots can change 1,000-10,000 times per second. Using information on the current computational load and the readings from embedded temperature sensors, core hopping can be applied (i.e., periodically swapping loads from hot to colder cores) in the millisecond cycle range [68]. For longer cycle times, liquid cooling provides a suitable means to switch between high to low cooling depending on computational load.

At the facility level, companies are striving to increase server utilization as the network capacity improves. However low utilization (less than 10%) is not uncommon. A McKinsey report [69] concludes that data centers tend to focus on acquisition cost rather than lifetime cost, and may be overly concerned about reliability. In this risk-averse mindset, operators tend to overinvest in servers, resulting in low utilization. In typical circumstances, 30% of servers might have utilization levels of less than 3% and yet continue to contribute to the overall electricity consumption [69]. A combination of low utilization and a large number of servers means significant fluctuations in power demand (e.g., 500 kW fluctuations for a 200 MW domain). Operating a data center at 50% utilization costs 2x more than operating at the theoretical maximum of 100%, but on the other hand, a certain margin is necessary for reliable operation. Virtualization is used for multi-server systems to optimally distribute loads and avoid underutilized servers.

3.4 Electrical power delivery

For data centers with traditional direct current (DC) power conversion and delivery systems, throttling the power demand results in poor efficiency. Instead, Google relies on their uninterrupted power supply (UPS) system to run out power spikes, using the batteries as an energy storage buffer. The same strategy can be used to deal with renewable energy sources on site, such as the wind farm near Google’s data center in Hamina, Finland [70].

An alternative to using traditional lead acid batteries in UPS systems could be redox (reduction-oxidation) flow batteries. These act as reversible fuel cells – storing electrical energy in an electrolyte solution – and their efficiency makes them suitable for storing large quantities energy, for instance for renewable power generation applications [71]. By pumping the electrolyte solution to within the chip itself, it could provide DC power at the low voltage required by the chip, while the spent electrolyte can carry heat away. This solution being considered by IBM can potentially lead to much higher efficiencies in the power delivery and thermal management system. Initial studies have shown that the pumping loss due to friction and fuel cell conversion losses would be significantly lower than the Joulean losses in an equivalent electrical power distribution system.

Today’s power distribution systems contain many electrical down- and up-voltage transformations and DC/DC conversions. For a typical data center facility with three-phase 400 V AC incoming power, the
efficiency is about 88% for the UPS system, 93% for the power distribution unit, 79% for the rack or server level power supply units, and 75% for the DC/DC voltage regulators that convert the voltage down to the chip level voltage [36]. These combine to a total efficiency of merely 50%. Using a redox flow battery system, the total conversion efficiency from the incoming facility power to the chip voltage could be up to 70%. This assumes an efficiency of 85% for the incoming AC voltage to DC conversion, 90% for the subsequent DC to electrochemical conversion, a small contribution due to pumping losses, and 90% for the electrochemical to DC conversion on the chip itself [36]. Such a hierarchical electrochemical system with a single macroscopic charging unit and multiple chip-level discharge units might satisfy the congruent demand for electrical power delivery and heat removal.

3.5 Modeling

Thermal simulation and thermal characterization are tools essential to evaluating and optimizing the thermal performance of electronic packages. Thermal characterization can provide a direct way to measure device or package temperature; however a comprehensive characterization of components can be too time-consuming and costly, especially at the data center scale. Furthermore, experimental uncertainty must be considered.

Advancements in software and computing speed make modeling and simulation an effective method in analyzing and predicting thermal performance. However, the uncertainty in material properties, and the simplifications and assumptions made in the modeling and simulation may impact the accuracy of the results. An experienced approach to combined use of both characterization and simulation can make up for the deficiency in each tool and provide realistic and accurate results.

Some aspects, such as acoustics, are particularly time-consuming to simulate. These require four-dimensional simulation of a turbulent flow field with time steps of less than 1 millisecond using a time-marching, weakly compressible, large-eddy simulation method [72]. In each time step, the temperature and flow field are calculated and Fourier-transformed to determine the spectral components. Using commercial computational fluid dynamics (CFD) packages in an industrial development environment, the acoustic field is typically not solved directly with the thermal and turbulent flow field, as this would be computationally prohibitive.

Most of the noise in IT equipment (over 95%) is generated by the fan itself. Aerodynamic noise generated by turbulent flow elsewhere in the system is generally negligible except in regions of high velocity. To account for noise emission in the overall system design, acoustic models should be included with heat transfer and flow modeling, which is computationally very demanding.

4 Implementing advanced cooling solutions

Facing this multiscale thermal management challenge, with a wide range of heat fluxes, characteristic length scales, form factors and device platforms, no single cooling solution can cover the entire spectrum. Successful design typically results from appropriate selection from a palette of solutions.
This section selectively reviews some cooling solutions that can help to address the thermal and energy management challenges outlined in the previous sections. It should be noted that many of these so-called emerging or enabling cooling technologies have been around for more than a decade, yet most have not broken through to off-the-shelf technology. On one hand, this indicates that further research is needed in most areas for these solutions to reach maturity. On the other, a further shift in the balance between computational and communication costs (see Sect. 2.2) might lead to an accelerated introduction of some of these technologies.

As reflected in DARPA's strategy (see Sect. 3.1.1), some key enabling technologies for the coming decade include convective and evaporative microfluidics, enhanced thermal interconnects, and on-chip thermoelectric devices. More than ever, sectors of the market seem ready for active cooling approaches.

### 4.1 Air cooling

#### 4.1.1 Further optimization of proven technology

For a chip manufacturer like Intel, air cooling continues to be the preferred cooling method for mainstream CPU products. This is partly due to the recent diversification of mobile platforms, which has led to a wide range of products from the low cost, power-optimized system-on-a-chip (i.e., memory, processor and graphics processor in a single package) to more desktop-oriented CPUs with one to eight cores. This market diversification brings unique challenges. Although air cooling remains the method of choice, Intel has expressed an interest in enabling innovative energy-efficient cooling solutions, as soon as cost-effectiveness tips the balance.

There is a growing trend towards board-level air flow design and control, as exemplified by Sony's graphics server featuring ducted airflow design to direct flow to the hot spots, temperature-based system fan speed, and individual fan control for processors and memory bank to minimize the total fan power and noise emission [73]. This approach could be extended to rack-level hierarchical active cooling using any of the cooling methods described in the following sections.

Alcatel-Lucent has been researching novel manufacturing techniques to produce monolithic enhanced surface area heat sinks with embedded structures to promote vortex shedding and turbulence generation [74]. This involves many manufacturing issues and can be more costly than standard parallel fin or pin fin heat sink designs, but may be enabled by emerging developments in accurate and low-cost 3D metal printing technologies. The more advanced designs have a lower thermal resistance albeit usually at a penalty of increased pumping power [74]. This coupling between heat transfer and momentum transfer (i.e., pumping power) is related to the Reynolds and Chilton-Colburn j-factor analogies, and has long been a physical bottleneck in thermal management. Some alternative air mover technologies that actively cause temporal regeneration of boundary layers (e.g., piezo fans and synthetic jets - see following sections) may change this coupling between heat transfer and momentum transfer and open up more opportunities than passive methods (spatial regeneration of boundary layers).
As discussed in Sect. 3.2.2, acoustic noise emission is an often-overlooked challenge in air cooled systems. Based on Huawei’s projection of the IT equipment power density evolution, the acoustic noise emission level will increase by about 0.5 dB(A) per year. Since the current emission levels are already close to the limits prescribed by NEBS in North America and ETSI and ISO in Europe [59,64], there is a need for novel noise-reducing technology, although some major proponents of lean air-cooled data centers (e.g., Google, Microsoft) might also be inclined towards a relaxation of the noise emission limits.

To extend the life of air cooling in telecommunication and data centers, the following targets are proposed:

(i) **System level**: more efficient air movers to increase the coefficient of performance (COP) above 20, as well as reduce the noise of single fans to below 60 dB(A). As indicated in Sect. 3.2.2, acoustics experts are not optimistic about meeting these requirements, although there is some room for improvement if the unit cost can increase.

(ii) **Board level**: heat sinks with low pressure drop and thermal resistance, e.g., 0.08 K/W at 140 cubic feet per minute (66 liter per second) for a 1U (rack unit) form factor device. As mentioned above, the coupling between heat and momentum transfer is the main physical restriction for this requirement, yet alternative air moving techniques (see following sections) may provide a way to circumvent this restriction.

(iii) **Component level**: no further increase in component power and a reduction in component thermal resistance.

If air cooling would indeed become untenable, a choice will have to be made between restricting the power density or resorting to liquid cooling. However, considering the potential advances of alternative air cooling techniques, and considering the history of past predictions about the demise of air cooling, this horizon may prove to be quite flexible.

### 4.1.2 Alternative air movers: Piezo fans

For Intel, the main parameters driving cooling technology are thermal performance, cost, system size, acoustics, and energy consumption. In the palette of available solutions, piezoelectrically actuated vibrating cantilever fans (or piezo fans) offer a viable option for small form factor and low power (below 40 W) platforms. Passive and active heat sinks have an overlap zone in the 15-40 W range. A cooling approach using piezo fans that provides cooling in this range and has less system dependency can promote design flexibility.

The main advantages of piezo fans are low cost, low power consumption, low noise, good reliability, and good thermal performance in the low to moderate heat flux range (0.1-1 W/cm²) [75-77]. Piezo fans can be used to replace or augment traditional air flow movers like axial fans, and equivalent fan curves can be established for them [76]. A comparative study by Intel for a chipset heat sink cooled by axial fans and by a number of piezo fans showed that the piezo fans require 50% less power to achieve the same thermal performance and the same noise level as the conventional fan-based cooler [78]. These devices
can be combined with existing speed-controlled axial fans to achieve significant gains in heat transfer performance when the fans are operated at lower speeds.

4.1.3 Alternative air movers: Synthetic jets

As with piezo fans, synthetic jets can significantly increase the convective heat transfer rate by effecting a temporal breakup and thinning of the thermal boundary layer, and enhancing the entrainment of surrounding air. 'Synthetic' refers to the jet flow being synthesized from the surrounding air by periodic suction and blowing of air across an orifice. Like piezo fans, synthetic jet actuators use robust and energy efficient piezoelectric actuators [79], without any rotating parts. Unlike steady jets (such as those used for turbine blade cooling, manufacturing processes, and micro-scale electronics cooling), synthetic jets do not require an external pressurized fluid source, yet a single impinging synthetic jet can achieve the same convective heat transfer rate as a steady impinging jet at the same Reynolds number [80].

Synthetic jets can help to reduce fan speeds by de-coupling heat transfer and thermal transport, where a low-level fan-driven flow carries heated air out of the equipment, while a number of synthetic jet actuators create local cooling on selected components. The magnitude of the jet flow can be easily controlled, making them suitable for integration in actively controlled cooling schemes. Consisting only of a piezoelectric actuator, a cavity and an orifice, a synthetic jet actuator can be used to augment cooling in confined spaces such as low-height arrangements in telecom applications, or other applications with inherently low mass flow.

These devices are scalable, with an optimal operating frequency that varies proportional to the Helmholtz resonance frequency of the cavity and orifice [79]. In air, the typical cooled area varies from below 1 cm² to 25 cm², for a corresponding device volume from 0.1 cm³ to 10 cm³. The desired operating frequency is typically below 200 Hz or ultrasonic (above 20 kHz), as exemplified by a range of small ultrasonic synthetic jet actuators (about 0.8 cm³) manufactured by Murata [81].

In terms of acoustic noise reduction, a single synthetic jet can reduce the noise emission compared to an axial fan while having the same or better thermal performance [82]. Furthermore, a pair of phase-controlled adjacent jets has been used successfully to control the angle of the jet flow [83], which provides even more opportunities for active cooling control. Preliminary experiments on this dual jet configuration have shown significant reductions in noise emission by partial noise cancellation.

4.1.4 Alternative air movers: Ionic wind

Ionic propulsion of air is established by a corona discharge (using direct current) or a barrier discharge (using alternating current), which make the air electrically conductive and cause subsequent air entrainment as charged ions are attracted towards a ground electrode [84,85]. In Ventiva's ionic wind actuator [86], a DC voltage of 3 kV is used with a 1 mm distance between a metal electrode wire and a ground plane mesh.

The main disadvantages are the low pressure head, the ozone production which may pose a health issue, and reliability issues related to dust fouling and wire breakage. Potential applications include LED
lighting with an acoustic emission requirement below 30 dB, mini projectors (no sound generated), and mobile electronic devices. In terms of coefficient of performance, Ventiva's fan takes about 0.25-2 W to cool up to 50 W.

4.2 Liquid cooling

There remains a divide between proponents of air cooling and liquid cooling. One group advocates air-cooled systems operated at an elevated temperature with minimal overhead, taking maximum advantage of low-temperature climates and alternative ultimate heat sinks. Another group advocates liquid-cooled (or hybrid air/liquid-cooled) systems which enable increased power density [87,88], employing waste heat recovery to maximize energy and exergy efficiency and compensate for the higher capital investment costs of liquid-cooled systems.

Given the low cost of data communication, air cooling remains the preferred method for the majority of the industry, including the recent rise in cloud computing facilities that can be located far from populated areas (i.e., at a low floor space cost), in cooler climates, or close to a cheap electricity supply. For barebones air-cooled data centers running at a PUE value of 1.1-1.2, the cooling fans consume less than 10% of the server power consumption, which is small incentive to transition to more expensive liquid-cooled systems.

On the other hand, as the cost of energy and the data traffic volume increase, there is a consensus that future data centers will become dominated by energy cost. IBM's solution is to use water cooling at elevated temperatures. At the chip level, the heat transfer occurs at a temperature difference of 20°C (instead of around 75°C for air cooling) which limits the loss of exergy. At the facility level, the elevated fluid temperature saves on chiller energy and recovers the maximal potential for energy reuse (e.g., domestic heating, and combined heating and cooling). Between IBM's first and second generation hot-water cooled data center systems, the coolant temperature has increased from 60°C for Aquasar in ETH Zurich [53] to 70°C for SuperMUC in LRZ Munich [54].

The long term future for large-scale telecom and data centers from a hardware point of view is with liquid cooling. Liquid may even be brought in direct contact with the electronics, using dielectric water. Complexity in liquid cooling solutions is introduced by using cold plates, manifolds, flexible connections, etc. Immersion cooling remains the simplest approach, and new ways to use passive immersion cooling could be pursued. Water still stands out as an excellent heat carrier and coolant, so it seems natural to take advantage of these compounding benefits. The preconceptions and concerns with water cooling in the telecom industry are subsiding, and in the coming years we may see a gradual introduction of water cooling not only at the high end, but also in low-end systems. The introduction of reliable water cooling technology has changed this mindset. The basic driver behind business decisions remains total cost of ownership.

One of the key practical issues impeding the introduction of liquid cooling at present is a lack of standardization, as indicated by a recent ASHRAE white paper on liquid cooling [89]. There are some
guidelines available on liquid cooling, but only relate to the interface between IT equipment and the facility. AT&T has been pushing for a more comprehensive standard.

### 4.3 Other alternative cooling techniques

Some recent studies have investigated the potential of using a superlattice thermo-electric cooler (TEC) on top of a hot spot, to achieve site-specific cooling. The TEC actively extracts heat from the hotspot area, which allows a significant reduction in pumping power [90, 91], making it a relevant technology for both air cooling and liquid cooling approaches, as well as enabling active local cooling control. High performance TECs are expensive and an overall cost analysis should confirm the economic benefits of this technology for a given application, but it might help extend the life of air cooling as the system density increases. Being more expensive, this technology might be more suitable for increasing performance rather than only reducing direct costs.

Hydrophobic surfaces have a number of potential uses in thermal management: enhanced sustained dropwise condensation for two-phase systems, control of evaporative cooling via electrowetting [92-94], and augmented liquid return in vapor chambers [95]. In particular, vapor-chamber heat spreaders are widely used at the component level for efficient heat spreading and can work against gravity (up to about 30 cm at moderate heat inputs). Significant recent research investment and developments have realized very thin (down to 1 mm) vapor chambers [96-98] that could be useful in telecom applications, where the board-to-board spacing is very limited. Device thickness is mainly limited because of mechanical strength and the maximum achievable vapor chamber power density (i.e., about 800-1200 W/cm² at evaporator heat transfer coefficients of around 300,000 W/(m²K)).

### 5 Conclusions and outlook

In terms of energy efficiency and sustainability, the power utilization effectiveness (PUE) value of data centers has gained a lot of importance in recent years. Although it is not a perfect metric, it has focused the attention of corporate leadership in the IT industry on the importance of energy efficiency. However in spite of positive messages from some major companies, most data centers are still at PUE values of 2 and above. Low-cost yet robust technology is needed to reduce the overall impact of IT on the environment.

The main underlying driver for progress is the total cost of ownership (TCO), which indirectly depends on the balance between computational and communication costs. As the cost of energy and the data traffic volume increase, the IT community might face a paradigm shift, where the energy costs far exceed the capital investment of the equipment and infrastructure.

Moore’s law will evolve and go three-dimensional: Single layer scaling is slowing down, yet stacking of multiple layers and integrating different functions (processing, memory, sensing) allows Moore’s law to be extended into the next decades.
Based on years of incremental evolution and continual improvements to air cooling technology, current IT equipment uses a million-fold larger volume for cooling compared to the size of the actual transistors performing the computations. By thinking outside the box, and reaching across disciplines, future computers might look very different: Liquid cooled systems reusing the majority of the incoming electricity (e.g., IBM Aquasar and SuperMUC) and interlayer-cooled 3D chip stacks, evolving towards bio-inspired designs for maximal energy efficiency which approach the functional density and connectivity of biological brains.

**Glossary of abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DIC</td>
<td>Three-dimensional integrated circuit</td>
</tr>
<tr>
<td>ASHRAE</td>
<td>American Society of Heating, Refrigerating and Air-Conditioning Engineers</td>
</tr>
<tr>
<td>CapEx</td>
<td>Capital expenditure</td>
</tr>
<tr>
<td>CFD</td>
<td>Computational fluid dynamics</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary metal-oxide semiconductor</td>
</tr>
<tr>
<td>COP</td>
<td>Coefficient of performance</td>
</tr>
<tr>
<td>CPU</td>
<td>Central processing unit</td>
</tr>
<tr>
<td>CRAC</td>
<td>Computer room air conditioning</td>
</tr>
<tr>
<td>DARPA</td>
<td>Defense Advanced Research Projects Agency</td>
</tr>
<tr>
<td>DCiE</td>
<td>Data center infrastructure efficiency</td>
</tr>
<tr>
<td>ENIAC</td>
<td>Electronic Numerical Integrator And Computer</td>
</tr>
<tr>
<td>EPA</td>
<td>Environmental Protection Agency</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>HPC</td>
<td>High performance computing</td>
</tr>
<tr>
<td>flop</td>
<td>(Number of) floating-point operations</td>
</tr>
<tr>
<td>flops</td>
<td>(Number of) floating-point operations per second</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated circuit</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>NEBS</td>
<td>Network Equipment-Building System</td>
</tr>
<tr>
<td>OHS A</td>
<td>Occupational Safety and Health Administration</td>
</tr>
<tr>
<td>OpEx</td>
<td>Operational expenditure</td>
</tr>
<tr>
<td>PUE</td>
<td>Power utilization effectiveness</td>
</tr>
<tr>
<td>TCO</td>
<td>Total cost of ownership</td>
</tr>
<tr>
<td>TEC</td>
<td>Thermo-electric cooler</td>
</tr>
<tr>
<td>TSV</td>
<td>Through-silicon via</td>
</tr>
<tr>
<td>U</td>
<td>Rack unit</td>
</tr>
<tr>
<td>UPS</td>
<td>Uninterruptible power supply</td>
</tr>
</tbody>
</table>

**Acknowledgements**

The authors gratefully acknowledge the contributions of the participants and speakers at the *2nd Workshop on Thermal Management in Telecommunication Systems and Data Centers* in Santa Clara, CA, on April 25-26, 2012. Workshop participants included representatives from leading telecommunications
systems and data center operators, equipment manufacturers, and system integrators (Alcatel-Lucent, Celsia Technologies, Coolit Systems, Fujitsu Advanced Technologies, Japan, Huawei North America, Huawei Sweden, Huawei China, Google, IBM, IBM Hudson Valley Acoustics Lab, IBM Research Lab, Switzerland, Intel, Microsoft, Airedale/Modine Manufacturing (US, UK), Nelson Acoustics, Oracle, Ventiva, Hewlett-Packard Labs, and ASHRAE), the Defense Advanced Research Projects Agency (DARPA), as well as academic research groups (Purdue University, University of California-Santa Cruz, and Trinity College Dublin). Dr. Tim Persoons acknowledges funding from the Irish Research Council and FP7 Marie Curie Actions.

References


[22] Huber H. The petascale power challenge. 3rd European Workshop on HPC Centre Infrastructures, Garching, Germany, September 2011.

[23] TOP500 (www.top500.org), June 2012.


[33] Inside one of the world's largest data centers. CNET, November 2, 2009 (Retrieved July 7, 2012 from news.cnet.com/8301-13860_3-10371840-56.html).


[58] ISO 11201:2010 - Acoustics: Noise emitted by machinery and equipment - Determination of emission sound pressure levels at a work station and at other specified positions in an essentially free field over a reflecting plane with negligible environmental corrections. International Organization for Standardization, May 2010.


List of figure captions

Figure 1. Evolution of the worldwide electricity usage from 2000 until 2010: contribution of data centers (△) compared to the sum of all sectors (□). Data obtained from [1-5].

Figure 2. Schematic Sankey diagram for a typical present-day data center facility, representing the distribution of power through three main system components (system and rack level IT equipment, cooling infrastructure, electrical power storage and delivery), partly based on data from [1,2].

Figure 3. Evolution of the average power usage effectiveness (PUE) of data centers from 2000 until 2010: Worldwide average (△) and company-average for Google (○). Data obtained from [1-5] with estimated values for Google from [2].

Figure 4. Word cloud visualization of the round table discussions at the 2nd Workshop on Thermal Management in Telecommunication Systems and Data Centers in Santa Clara, CA on April 25-26, 2012 (Image generated by wordle.net).

Figure 5. Schematic Sankey diagram for a ‘cloud-optimized’ data center using air cooling at an elevated room temperature to maximize the use of free cooling. Compared to a typical data center (Fig. 2), more demanding rack-internal air cooling leads to acoustic noise challenges, which can be partly addressed with advanced air cooling solutions.

Figure 6. Schematic Sankey diagram for a fully liquid-cooled data center, with more efficient high-density rack-level cooling due to the superior heat carrier properties of water. Using water at an elevated temperature, this approach is suitable for waste heat recuperation.
Figure 1

Worldwide electricity usage, TWh/yr

+3%/year

+11%/year

- Sum of all sectors
- Data centers
Figure 2
Figure 3

[Graph showing Power Utilization Effectiveness from 2000 to 2010 for Global average and Google]
Figure 4
Figure 5
Figure 6