Renewable and Sustainable Energy Reviews 171 (2023) 112991

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Research and Technologies for next-generation high-temperature data centers – State-of-the-arts and future perspectives

Yingbo Zhang [a], Kui Shan [a], Xiuming Li [a], Hangxin Li [a,**], Shengwei Wang [a,b,*]

[a] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong
[b] Research Institute for Smart Energy, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Keywords:
Data center
High-temperature operation
Energy saving
Reliability
Cooling technology
IT equipment

ABSTRACT

Data centers have attracted increasing attention worldwide over the last decades due to their high energy consumption. Cooling accounts for about 30–40% of the total energy consumption of data centers. High-temperature data centers could save large amounts of cooling energy by changing their cooling mechanism. More effective use of “free cooling” is the basic and effective means for high-temperature data centers to reduce cooling energy consumption. It is possible to build chiller-less or even chiller-free data centers. They require less capital investment for cooling and allow more hours of “free cooling”. However, a few essential concerns need to be addressed before the wide application of high-temperature data centers, particularly the technical bottlenecks and the reliability and performance of servers and IT equipment. Though many reviews on data centers exist in the existing research, a systematic review of high-temperature data centers, particularly on the above essential concerns, is still unavailable. This paper is intended to fill in these gaps and provide a comprehensive review of these critical aspects. The main benefits and the major bottlenecks for implementing high-temperature data centers as well as the existing efforts and latest technologies to tackle the bottlenecks are categorized and analyzed systematically. In addition, a through review of the main temperature-sensitive IT components (e.g., hard disk drives and CPU) is done, and their current states and potential solutions are analyzed. Finally, the paper elaborates on future perspectives for the development and applications of the high-temperature data center.

1. Introduction

The global demand for data centers has recently dramatically increased over the last few decades due to the increasing demand for information services. Meanwhile, the energy consumed by data centers has increasingly become a serious concern worldwide. The global energy use of data centers accounts for 1%–1.5% of global electricity consumption [1,2]. According to a report from the US Environmental Protection Agency [3], the information and communication industry accounts for about 2% of the CO_2 emission, equivalent to those from the airline industry [4,5]. Data centers are characterized as energy-intensive buildings compared with conventional buildings. The energy flux dissipated in the newly developed data centers is expected to be 6458–10764 W/m^2 [6], while the cooling load of conventional HVAC (Heating, Ventilation and Air Conditioning) systems in conventional commercial buildings is 40–86 W/m^2 [7]. It means that the cooling load density of the new-generation data centers is almost 100 times that in conventional commercial buildings.

The main power consumers in date centers are servers, network equipment, power distribution equipment and cooling equipment [8]. Cooling typically constitutes 30%–40% of the total power consumption in data centers [9,10]. Scientists and engineers are attempting various means to enhance the efficiency of data center cooling systems. Currently, approaches associated with efficiency enhancement have been widely investigated, including airflow distribution optimization [11,12], containment of aisles [13,14], supply and demand match of cooling energy [15,16], and performance improvement in servers [17]. Although efficiency enhancement is an effective means to reduce energy consumption, it can only save energy to a certain amount. As a fundamental solution, increasing the space temperature in data centers (the ambient temperature for servers) can dramatically reduce cooling energy demand by adopting free cooling.

At the earlier stages, data centers were typically operated in a

* Corresponding author. Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong.
** Corresponding author.
E-mail addresses: hangxin.li@polyu.edu.hk (H. Li), beswwang@polyu.edu.hk (S. Wang).

https://doi.org/10.1016/j.rser.2022.112991
Received 29 June 2022; Received in revised form 25 September 2022; Accepted 9 October 2022
Available online 27 October 2022
1364-0321/

Abbreviations:

HVAC	Heating, Ventilation and Air Conditioning
CRAH	Computer room air handler
DIMM	Dual Inline Memory Module
PSU	Power Supply Unit
HDD	Hard Disk Drive
SSD	Solid State Drive
SAS	Serial Attached Small Computer System Interface
BMC	Burst Mode Controller
EEPROM	Electrically Erasable Programmable read only memory
CPU	Central Processing Unit
PCH	Paging Indicator Channel
MEM	Memory Device
PCB	Printed Circuit Board
NIC	Network Interface Card

temperature range of 20–21 °C with a common notion of "cold is better" [18], and even some were as cold as 13 °C [19]. Due to the conservative suggestions by manufacturers and conventional wisdom, engineers and operators tend to follow conventional practices. Usually, mechanical cooling is needed to maintain a low space temperature in data centers, but it consumes huge amount of energy. Thus, raising the space temperature in data centers is expected to reduce the mechanical cooling energy by shortening chiller operating hours. It has been reported that increasing the space temperature in data centers by 1K results in the saving of the total power consumption by 4–5% [20,21].

The high-temperature data center is regarded increasingly as a trend in the data center profession and industry. In 2011, ASHRAE (American Society of Heating, Refrigeration, and Air conditioning Engineers) further expanded the range of allowable environmental conditions of IT equipment and adopted two new classes in their updated guidelines (class A3 and class A4) based on the updated critical information (IT design and failure data) provided by IT manufacturers [22]. The new Class A3 environment has the upper temperature bound of 40 °C and the Class A4 environment even has the upper temperature bound of 45 °C, as shown in Fig. 1.

In fact, if IT equipment could withstand the long-term operation of up to 45 °C, the worldwide deployment of highly economized and even chiller-free data centers can be realized [23]. Some data center operators are considering operating their data centers at higher temperatures [24]. Beaty et al. [25,26] reported comprehensive analysis and discussions on the workflow for raising the space temperature from the cooling perspective. Baidu reported that it is possible to design a kind of server that can support an ambient temperature of 50 °C according to the temperature specification of key components in servers [27]. Intel and Microsoft reported that most servers worked well under higher ambient temperatures and outside air [28]. A chiller-less economized data center test conducted by ASHRAE shows that short-term high-temperature had little impact on data centers [18]. The data center utilized the outdoor air to cool their computer rooms throughout the year and worked well. In addition, some studies reported that temperature variations of outdoor air could be well regulated via proper control [29] and have little impact on server operations [30].

Although the concept of high-temperature data center has been put forward, a comprehensive review and systematic analysis on high-temperature data centers is still lacking. This study therefore aims to fill in this gap and provide a comprehensive review and systematic analysis from the point of view of the benefits, bottlenecks, existing efforts and latest technologies as well as the future perspectives for implementing high-temperature data centers. The main contributions of this work include: *i.* Analyze and categorize the main benefits and concerns of implementing high-temperature data centers. *ii.* Investigate and elaborate the main limitations and bottlenecks of implementing high-temperature data centers, considering system reliability, power consumption and performance under high-temperature operation; *iii.* Investigate and elaborate the main temperature-sensitive IT components, such as hard disk drives and CPU, and analyze their current states and potential solutions; *iv.* Systematically review the existing research and development efforts associated with current technologies and feasible solutions to tackle existing bottlenecks from four levels (i.e., chip level, server level, rack level, and room level); *v.* Elaborate the future perspectives for the development and applications of high-temperature data centers. This work could help decision makers and various stakeholders to understand the existing efforts and the latest development in high-temperature data centers.

2. Main benefits and bottlenecks for high-temperature data centers

Fig. 2 summarizes main benefits and concerns of high-temperature data centers today. There are four main advantages and four main concerns associated with raising the space temperature in data centers. These main advantages and concerns are discussed and elaborated as follows.

i. More free cooling hours and great energy-saving potential: Free cooling is widely employed in large data centers in mild and cold regions by utilizing outdoor cold air (when the outside temperature is lower than of indoor space in data centers) [31]. "Free cooling is the single biggest opportunity for efficiency, greater than all other regions combined",

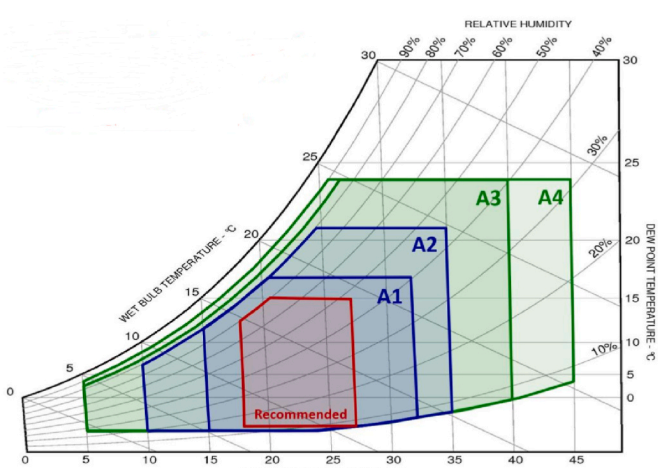

Fig. 1. Ambient temperature classification of IT equipment by ASHRAE [22].

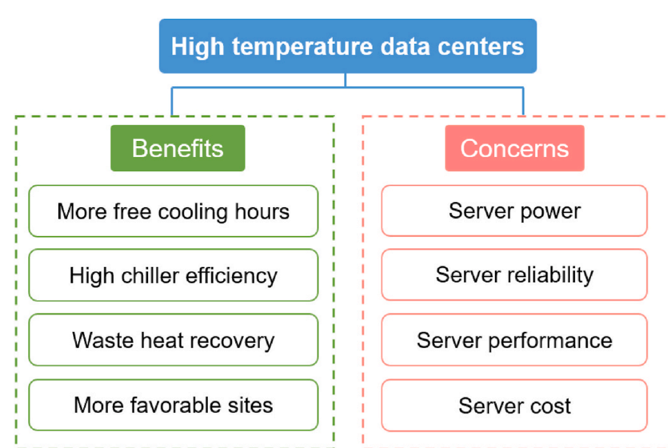

Fig. 2. Main benefits and concerns of high-temperature data centers.

according to the data center director of Google [23]. A study suggested that, when the ambient temperature of servers in data centers rose from 27 °C to 35 °C, the airside free cooling hours in a year increased from 80% to 99 % in Europe, as shown in Fig. 3 [32]. Large Internet companies, like Google, Microsoft and Intel, have been aggressive in operating their data centers at high temperatures (i.e., greater than 27 °C). Microsoft reported that raising the space temperature in data centers by 2–4 °C in one of their Silicon Valley data centers could save $250,000 in annual energy cost [19]. It was also estimated that the space temperature of 35 °C combined with containments could lead to 22% energy savings [33].

ii. Higher chiller efficiency: Higher space temperature in data centers allows higher chilled water temperature. Every degree (K) increase in the chilled water temperature can bring about 2% improvement for the chiller efficiency [34,35].

iii. More favorable site options: Because of more free cooling hours, the high-latitude area in the Pan-Arctic region has become a hotspot for data centers site selection [36,37]. Google has deployed a number of its data centers in high latitudes near the north pole to minimize the cooling cost [38]. It was reported that 91–99 % data centers in North America, Europe and Japan could use free cooling throughout the year if Class A3 equipment was used (dry-bulb temperature range of Class A3: 5–40 °C) [39]. Thus, raising the ambient temperature of servers could enable the deployment of highly economical and even chiller-less data centers worldwide.

iv. Improved waste heat recovery: The main barrier to recover waste heat from data centers is that the heat collected, although plentiful, is of too low quality [2]. Increasing the space temperature in data centers makes it possible to recover more heat of higher quality or temperature. The heat can be used by the neighboring buildings or district heating systems.

Despite these advantages, it is observed that very few data centers have attempted the high-temperature operation as reported in one survey [40]. This is due to some concerns, mainly system reliability, computing performance, power consumption and cost issues related to high-temperature operation [41]. For example, it is usually considered that high temperature will affect the reliability of servers, subsequently impacting the overall system reliability. If these concerns are well addressed, the high-temperature environment will become a revolutionary turning point in the date center industry. The arguments against high-temperature data centers are discussed in detail in the following Section 3.

3. Temperature-sensitive IT components, implementation limitations and existing efforts

Servers are the core elements of a data center. There are four main limitations in implementing high-temperature data centers concerning servers, as depicted in Fig. 4, including: *i.* the impact on server reliability; *ii.* the impact on server performance; *iii.* the impact on system power consumption; *iv*. the tradeoff between server cost premium and operating cost saving.

3.1. Server reliability vs high-temperature

The reliability of servers is the main bottleneck for high-temperature date centers. A common understanding is that a high-temperature environment would reduce the reliability and availability of servers. In this regard, the latest Thermal Guidelines for Data Processing Environments [22] published by ASHRAE reported that the relative failure rate of volume servers is function of server inlet temperature (the ambient temperature of servers) for 7 × 24 × 365 continuous operating conditions, as shown in Table 1. Where, X-factor denotes the relative

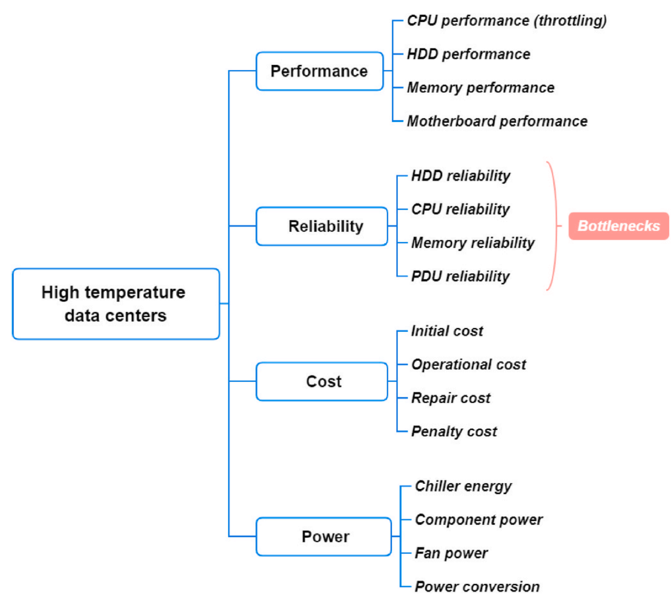

Fig. 4. Key challenges to the implementation of high-temperature data centers.

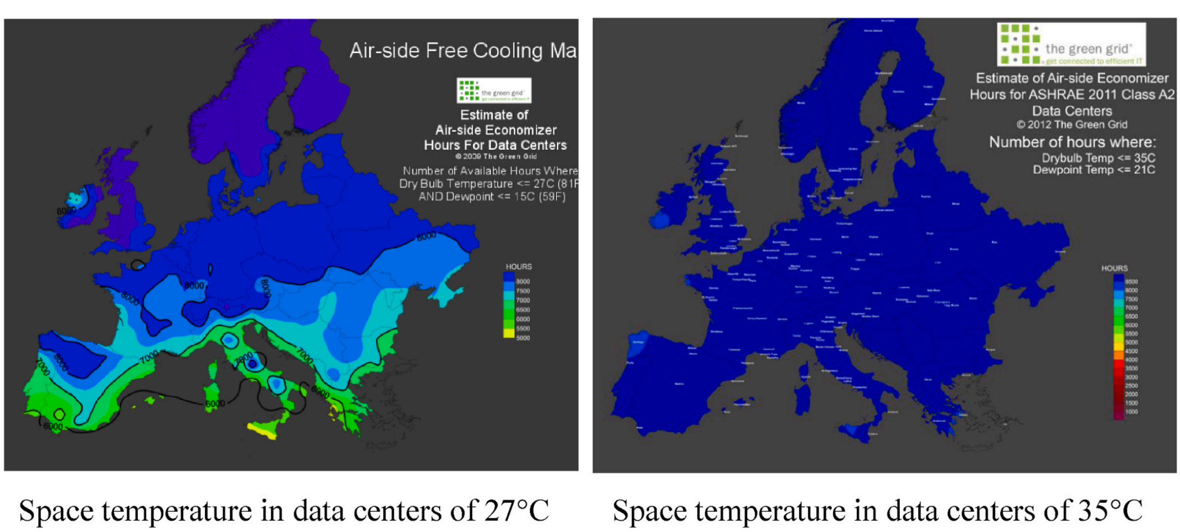

Fig. 3. Air-side free cooling map of Europe [32].

Table 1
The relative failure rate of volume servers as a function of server inlet temperature [22].

Dry Bulb Temperature (°C)	X-Factor of Average Failure Rate	Lower Bound of X-Factor	Upper Bound of X-Factor
15	0.72	0.72	0.72
17.5	0.80	0.87	0.95
20	0.88	1.00	1.14
22.5	0.96	1.13	1.31
25	1.04	1.24	1.43
27.5	1.12	1.34	1.54
30	1.19	1.42	1.63
32.5	1.27	1.48	1.69
35	1.35	1.55	1.74
37.5	1.43	1.61	1.78
40	1.51	1.66	1.81
42.5	1.59	1.71	1.83
45	1.67	1.76	1.84

failure rate compared to the baseline of a server inlet temperature of 20 °C (i.e. X-Factor is 1.00 at server inlet temperature of 20 °C). It is noticeable that the average failure rate increases as the server inlet temperature rises.

In reality, there is specific high-temperature period (excursion time) for a particular location, mainly in hot summer. Thus, the impact on the server reliability is much more benign than the 7 × 24 × 365 continuous high-temperature operating conditions, even for the harshest scenario (the chiller-free operation for achieving free cooling throughout the year).

In addition, the guideline pointed out that the reliability data was meant to capture most of the volume server market (belonging to Class A2 rated to 35 °C [23], and very few products currently exist for the Class A3 environment [18]). However, some specific server information was not disclosed. For example, there is no definite data on the failure rate of servers that could function under high-temperature with improved heat sinks and advanced materials.

The impacts of high temperature on the system reliability are further analyzed in section 3.1.1 and 3.1.2 from two aspects: the impacts of high-temperature excursion on data centers, and the temperature-sensitive components in servers.

3.1.1. *High-temperature excursion*

High-temperature "excursion time" is the period requiring special attention for the operation of high-temperature data centers. This has been studied by some IT companies. Google tested the temperature limits of its hardware by running the servers at higher temperature. Results show that the servers ran just fine without more failures [42]. A Google data center in Dublin was optimized to use fresh air to cool tens of thousands of servers throughout the year without any air-conditioners or chillers. Despite of a period of high temperature in a year and annual temperature fluctuation, the servers worked well. It means that servers are much sturdier than previously imaged and could bear high-temperature excursion for a certain time [43]. Intel conducted a comparative study over a period of 10 months, using about 900 production blade servers [44]. The test room was divided equally into two side-by-side compartments. One used conventional air-conditioners, while the other used air economizers to cool down servers with 100% outside air at temperatures of up to 33 °C. Servers in the economizer compartment were subjected to considerable variation in temperature and humidity as well as poor air quality. The failure rate was 3.83% and 4.46% for the conventional air-conditioned compartment and the air economizer compartment, respectively, with insignificant difference [44].

3.1.2. *Temperature-sensitive components*

To further understand the reliability of servers, the reliability of individual components of servers should be investigated, especially the components that are prone to failures. Table 2 shows the detailed temperature limitations of components in servers, referring to Intel's Romley Sandy Bridge-EN platform design guide and component specification [45]. It can be seen that the component with the lowest temperature limit is hard disk drives (HDDs).

A few investigations also show that hard disk drives (HDDs) and memories are the most frequently replaced components in modern data centers, and typically 1%–5% of drives in data centers need to be replaced in a year [46,47]. Microsoft Corporation observed and recorded actual data on different kinds of failure types over two years from their typical large-scale data centers housing more than 100,000 servers, as shown in Fig. 5 [48]. HDDs accounted for 71% of the known failures, being the most dominant failing part.

Similarly, Baidu corporation recorded the statistical data of storage servers running in their data centers between the periods of April–June 2012, as illustrated in Fig. 6 [45]. The failure rate of HDDs was the highest, representing approximately 98% of all system failures, followed by memories and motherboards. The failure rates of other system components are close to zero.

3.1.2.1. *Hard disk drives (HDDs)*.

Given that HDDs are the most critical components, studies have been conducted to investigate the relation between temperature and the failure rate of HDDs. A specific test conducted by Microsoft Corporation shows that the increase of server inlet temperature had certain impact on the failure rate of HDDs, but redesigning the layout of server components could keep the HDDs failure rate at a very low level even though the inlet temperature was up to 40 °C [30]. A recent study of Google presented a very interesting result showing that very low temperatures were actually more detrimental to disk reliability compared with higher temperatures [21]. Pinheiro et al. [49] has similar observation and found a rapid drop in disk failure rates when increasing the temperature, followed by a slight increase when temperatures were above 45 °C, as shown in Fig. 7.

Nosayba et al. [21] collected the data on HDD replacements from January 2007 to May 2009 at 19 different data centers of Google, covering 5 different HDD models. The failure rates of two models increased significantly when the temperature rose from 40 °C to 55 °C, while the failure rates of other three models increased slightly, as shown in Fig. 8. This implies that different models or types of HDDs have different responses to the high-temperature operation. A study of Alibaba group reported that the HDD inlet temperature of 50–55 °C was a reasonable range to get an acceptable reliability level for high-temperature data centers [41].

The failures of HDDs are probably caused by the vibration due to the rotation of server fans. It was reported that HDD performance degradation was proportional to the speed of server fans, and an unoptimized fan control algorithm could reduce HDD performance by up to 60%

Table 2
Temperature limits of components in typical servers [45].

Key Components	Temperature in SPEC (°C)
HDD (Hard Disk Drive)	60
SAS (Serial Attached Small Computer System Interface)	70
Crystal Oscillator	70
Clock Generator	70
BMC (Burst Mode Controller)	70
EEPROM (Electrically Erasable Programmable read only memory)	85
CPU@95W TDTS	89
PCH (Paging Indicator Channel)	92.7
MEM@DDR3 (Memory Device)	95
PCB (Printed Circuit Board)	105
NIC (Network Interface Card)	123
Regulator	125
Ferrite Bead	125
Diode/Triode/Transistor	125

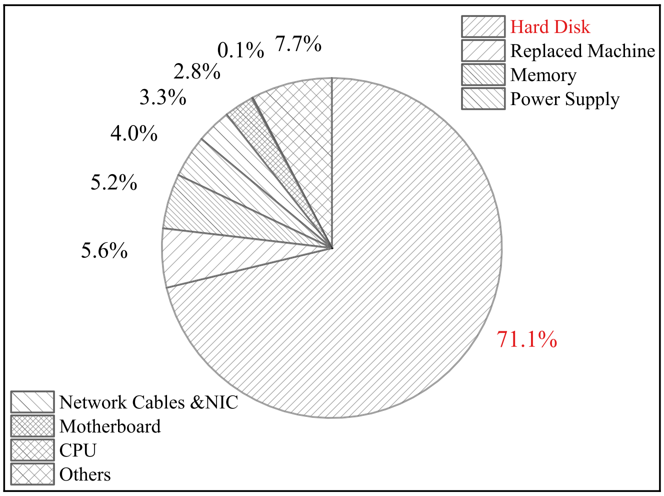

Fig. 5. Breakdown of hardware component errors in a large datacenter - Microsoft.

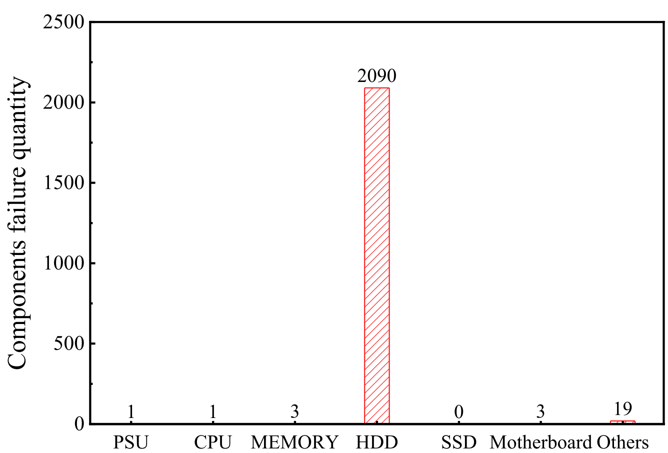

Fig. 6. Breakdown of hardware component errors in a datacenter - Baidu.

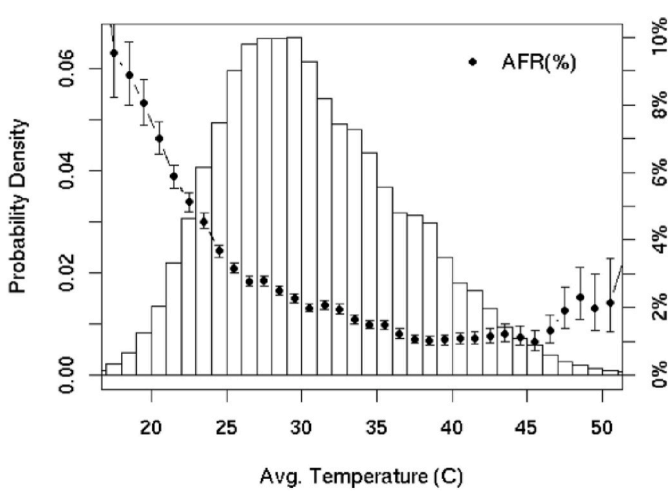

Fig. 7. Distribution of average temperatures and failures rates of HDDs [49].

[50]. A study on the solid-state drives (SSDs) shows SSDs had better reliability compared with HDDs [51], but the price of SSDs is 3–5 times that of HDDs. In this regard, Chan et al. proposed a hybrid SSD-HDD system and a novel fan control method to minimize the impact of the

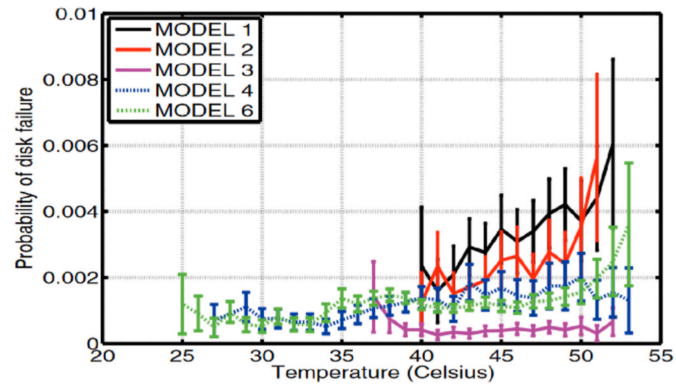

Fig. 8. The monthly probability of HDD failure as a function of temperature [21].

vibration [52]. It is also suggested by Zhang et al. [41] that a solution to reduce the failure rate of servers is to replace HDDs by SSDs for high-temperature data centers.

3.1.2.2. Memories. High temperature is considered to have negative impacts on the reliability of hardware components in memories due to the large physical change on materials. Some studies show that the high-temperature operation was expected to increase leakage current in memory chips and lead to a higher likelihood of flipped bits in the memory array [53,54]. However, Schroeder et al. [55] reported that temperature had a surprisingly low effect on memory errors.

Table 3 shows the results of a comprehensive study conducted by Baidu corporation concerning the effect of high temperature on all key components of servers [27]. It can be observed, when raising the ambient temperature to 50 °C, that the failure rates of both motherboards and memories were doubled. However, the ambient temperature of 35 °C–40 °C is acceptable for high-temperature data centers. Overall, high temperature indeed affects the reliability of some servers, but some latest IT products show the possibility to implement high-temperature data centers around 35 °C or even 40 °C [23].

3.2. Server performance vs high-temperature

Server performance mainly refers to the performance of central processing unit (CPU), memory and hard disk drive. It is generally considered that the component performance degradation, such as frequency reductions for CPUs and memories, is associated with higher ambient temperatures. However, a few studies show that it was not the case.

IBM conducted a particularly well-defined and controlled study, focusing on the server performance within the Class A3 environment. They selected nearly 70 different CPUs as test samples, with server inlet temperature baseline of 25 °C for tests of each piece of equipment, and then repeated the tests at 35 °C (Class A2 upper limit) and at 40 °C (class A3 upper limit). As summarized in Table 4, the performance when operating at the 35 °C and 40 °C were almost the same as that at 25 °C, with a maximum reduction of 2% in some case. Obviously, the results indicate that there was no performance degradation at higher

Table 3
Evaluated components AFR (average failure rate) [27].

Ambient temperature	35 °C–40 °C	50 °C
Components	AFR (%)	AFR (%)
Motherboard	2.62	5.06%
Memory	0.32	0.70%
HDD	0.73	0.92%
FAN	1.79	2.50%
Power Supply	2.92	2.92%

Table 4
Server performance at different inlet temperatures [24].

Test conditions			Server inlet temperature		
System	Model	Exerciser	25 °C	35 °C	40 °C
X3550M4	130W	Linpack Turbo-on	1.00	1.00	1.00
	115W	Linpack Turbo-on	1.00	1.00	1.00
X3650M4	115W Best	Linpack Turbo-on	1.00	1.00	1.00
	115W Best	SPECjbb2005	1.00	1.00	0.99
	115W Best	SPECint_rate2006	1.00	1.00	1.00
	115W Best	SPECfp_rate2006	1.00	1.00	1.00
X3650M4	130W	Linpack Turbo-On	1.00	1.00	1.00
X240	130W Best	Linpack Turbo-On	1.00	0.99	0.99
	130W Best	SPECjbb2005	1.00	0.99	0.99
	130W Best	SPECint_rate2006	1.00	1.00	1.00
	130W Best	SPECfp_rate2006	1.00	1.00	1.00
	130W Worst	Linpack Turbo-On	1.00	0.99	0.98
	130W Worst	Linpack Turbo-Off	1.00	1.00	1.00
	130W Worst	SPECjbb2005	1.00	0.99	0.99
	130W Worst	SPECint_rate2006	1.00	1.00	0.99
	130W Worst	SPECfp_rate2006	1.00	1.00	0.99
dx360M4	115W Worst	Linpack Turbo-On	1.00	1.00	1.00
	115W Worst	SPECjbb2005	1.00	1.00	1.01
	115W Worst	SPECint_rate2006	1.00	1.00	1.00
	115W Worst	SPECfp_rate2006	1.00	1.00	1.00
AVERAGE			**1.00**	**1.00**	**1.00**

temperatures up to certain extent [24]. Similarly, Wang also studied the impact of CPU temperature on server performance, and concluded that CPU temperature has no obvious impact on server performance [56].

A study conducted by University of Toronto tested the performance of HDDs, CPUs, and memories with a wider range of workloads and temperatures [21]. The test temperatures were much higher than the IBM tests, so performance degradation became easier to be identified if outside the range of normal statistical error. Tests were conducted inside a thermal chamber in which temperature could be controlled in 0.1K increments from −10 °C to 60 °C (much wider than typical temperature range in data centers today). The results show that, for CPU and memory performance, there was not any throttling down on any of benchmarks up to 55 °C. As for the hard drives, they found a decrease in "throughput" at the ambient temperature of 40 °C for the Seagate 73 GB and Hitachi SAS drives, at 45 °C for the Fujitsu and Seagate 500 GB SAS drives, and at 55 °C for the Hitachi Deskstar. It means different types of products have different responses to the high-temperature operation.

Alibaba Group also studied the performance of key components of servers operating at different ambient temperatures, as shown in Fig. 9. Where, the thermal margin means additional operating margin after meeting the safety limit requirements. As seen, all key components,

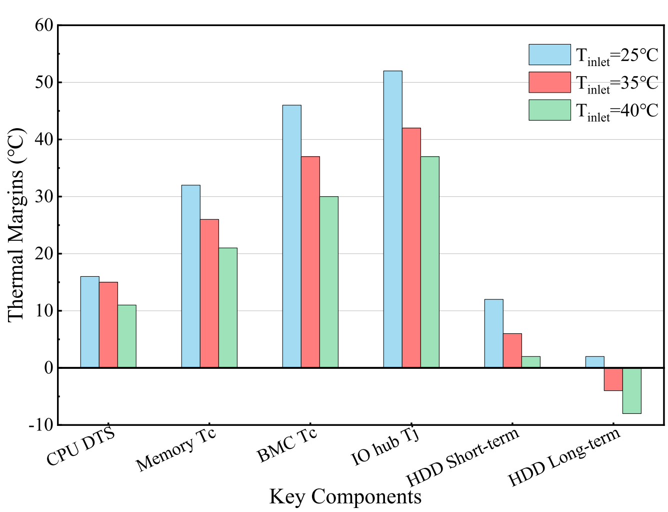

Fig. 9. Thermal margin of key components in servers [41].

except HDD at long-term operation, kept sufficient thermal margin even at a server inlet temperature of 40 °C [41]. These results agree with the results reported by the University of Toronto.

Baidu corporation carried out one-month experiments about the relation between the performance of key components in servers and ambient temperatures. They found that, when the ambient temperature rose from 25 °C to 50 °C, the CPU, memory and HDD performance remained stable. However, when HDD ran at 54–57 °C, the performance declined by approximately 30%. When the HDD ran at 57 °C, the performance immediately dropped by an average of about 40%–80% [27].

Based on these test results, it can be concluded that the HDD is a key limitation for high-temperature operation. It seems that different types of HDDs have different responses to high-temperature operation. Hence, HDDs with a wider/higher temperature range are necessary for high-temperature data centers.

3.3. *Power consumption vs high-temperature*

3.3.1. *Server power*

When the ambient temperature of servers increases, the energy consumption of server fans will increase to keep all components below certain critical thermal thresholds. It was reported that server fans contribute up to 14% of total power consumption in data centers [57]. In addition, the power consumption of components in servers also increases slightly as the inlet temperature rises due to the increase in leakage current for some silicon devices [56].

ASHRAE studied the relation between server power and server inlet temperature, as shown in Fig. 10 [18]. For Class A2 servers, the server power would increase by 4%–8% when server inlet temperature increased from 15 °C to 30 °C. The server power could increase by 7%–20% when the inlet temperature increased from 15 °C to 35 °C. The server power rise of Class A3 servers could be similar as Class A2 servers, while Class A3 servers would very likely need improved heat sinks and/or fans to properly cool the components.

3.3.1.1. Fan power rise. In the latest thermal guidelines for data processing environments from ASHRAE [22], the relation between server air flow rate and ambient temperature was tested, as shown in Fig. 11. The server air flow rate increases as the ambient temperature rises. The blue area is mainly caused by different types of servers configured with different fan control algorithms. The guidelines also listed two typical fan control algorithms, including fixed fan speed control and variable fan speed control, as illustrated in Fig. 12. In most cases, servers with variable fan speed control show better energy-saving potential.

Schneider [58] divided IT server fan behavior into 3 categories according to server configurations, including: i). Servers configured with a high-stress load, ii). Servers configured with a moderate load, and iii). Servers configured with a light load. They concluded that servers configured with a light load usually were equipped with the constant fan speed control algorithm, the other two types were often equipped with the variable fan speed control algorithm. According to Fan Affinity Laws, shown in Equation (1), the fan power is in cubic growth of the rotational speed [59]. For example, the power consumption of the fan will increase by 8 times if the speed of the fan is increased by 2 times. Thus, fan power will increase significantly when the server detects an increase in ambient temperature.

$$\text{fan power} \propto (\text{fan speed})^3 \tag{1}$$

3.3.1.2. Leakage power. Patterson et al. [60] reported that the temperature has little effect on the dynamic power of CPUs but would drive the amount of leakage. Leakage power is mainly induced by the leakage current for some silicon devices. In the past, it did not draw much attention since the leakage current was an order of magnitude less than the chip's normal operating current [61]. In later generations of CPU (e.g. 0.35–0.18 μm technologies), the leakage was also a small fraction of

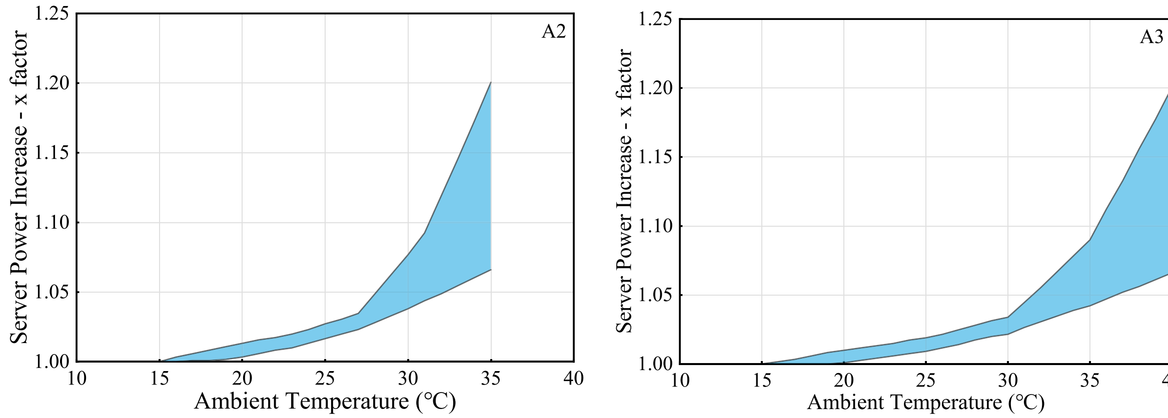

Fig. 10. Server power rise *vs* Ambient temperature range for classes A2 and A3 servers [18].

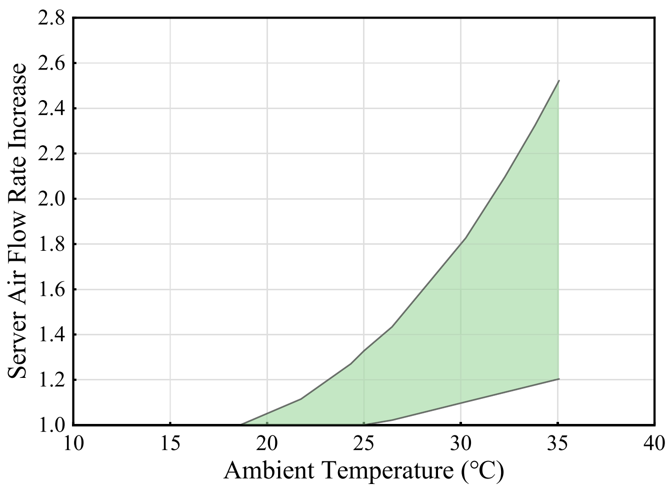

Fig. 11. Server flow rate increase versus ambient temperature increase.

the total power. However, as the scale of semiconductor technology enters the nanometer era recently, the continued shrinking geometries in the silicon have caused the leakage to increase as high as 50%, due to the leakage paths being shorter [62–64]. Fallah et al. [65] reported that every 1K increase across multiple process generations would cause a consistent 2% leakage increase. As technology moves to 45 nm and beyond, these temperature-dependent leakage rates are expected to increase further. Thus, the leakage power needs to be considered when raising the ambient temperature.

Fig. 13 shows the effect of the ambient temperature on the power of key components in servers. As the ambient temperature increases, fan power increases substantially, while the CPU and memory powers increase slightly and the HDD power decreases a bit [45]. It can be concluded that fan power is the main contributor to server power rise when the ambient temperature increases.

3.3.2. *Total power consumption*

Generally, the power consumption of chiller plants will decrease when higher temperatures are set in computer rooms because of the higher chilled water temperature supplied. Reduced chiller energy consumption results from the increased chiller efficiency due to a higher supply chilled water temperature, less cooling load due to higher space temperature, and less chiller operation time due to increased free cooling time.

However, server fan power and computer room air handler (CRAH) fan power may increase. CRAH must supply more cooling air to servers to match the increased airflows of server fans. Thus, server power and CRAH fan power may offset the power saving in chillers plants.

Moss et al. [58] found that, when the server inlet temperature was 25–27 °C, the total energy consumption was the lowest for a chilled water system. However, their study did not consider the free cooling and advanced servers which can withstand higher temperatures with low fan power. Seaton [66] studied server power and cooling energy consumption at different server inlet temperatures without considering free cooling. The results also show that server inlet temperature within 25–27 °C had the lowest total energy consumption, in line with David's results. Similarly, Muroya et al. [67] found that total power

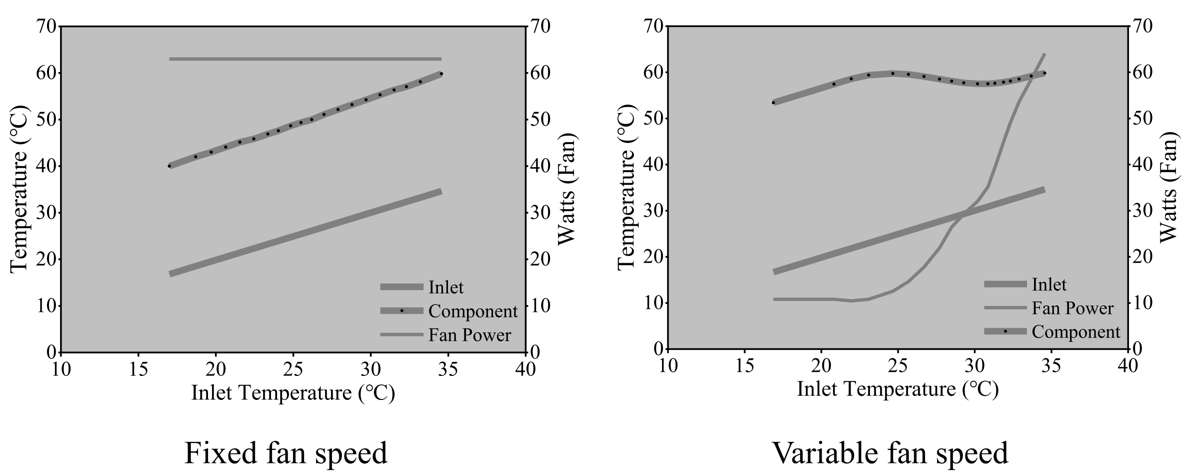

Fig. 12. Inlet and component temperatures with two kinds of server fan behavior [22].

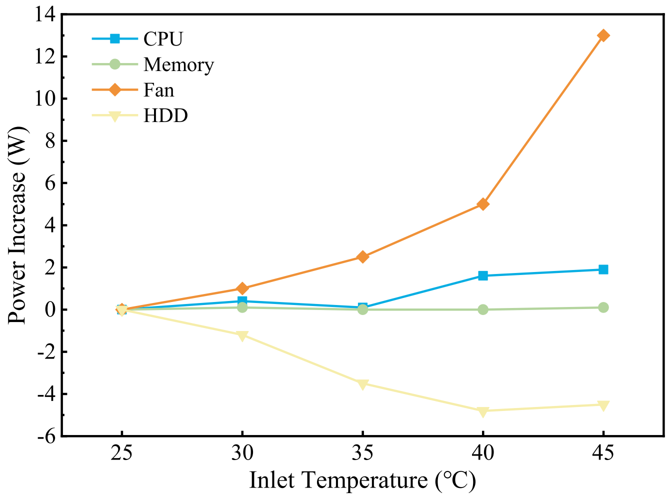

Fig. 13. Key Component Power versus temperature.

consumption of data centers reached a minimum when server inlet temperature was around 25 °C, as shown in Fig. 14.

These studies considered the facility energy consumption when adopting mechanical cooling only. In practical applications, free cooling will be adopted to save more cooling energy. Seaton [66] further estimated the cooling costs of increased inlet temperatures considering air-side free cooling, and found that total cooling energy consumption would further decrease when the server inlet temperature was up to 40 °C.

3.4. System cost vs server inlet temperature

3.4.1. Server cost

There are four classes of servers including class A1, A2, A3, and A4. The new server classes (A3 and A4) that support wider environmental envelopes are usually associated with higher prices and are less applied. Currently, class A1 servers can be found in low-price markets and on old data centers constructed long time ago, with the upper temperature limit of 30 °C [68]. Class A2 servers are, currently, the standard base products, with the upper temperature limit of 35 °C. Class A3 servers cost a premium over class A2 servers and are becoming the standard, with the upper temperature limit of 40 °C. Class A4 servers cost more premium compared with class A3 servers, with the upper temperature limit of 45 °C. High-temperature data centers require class A3 and class A4 servers due to their wider temperature limits.

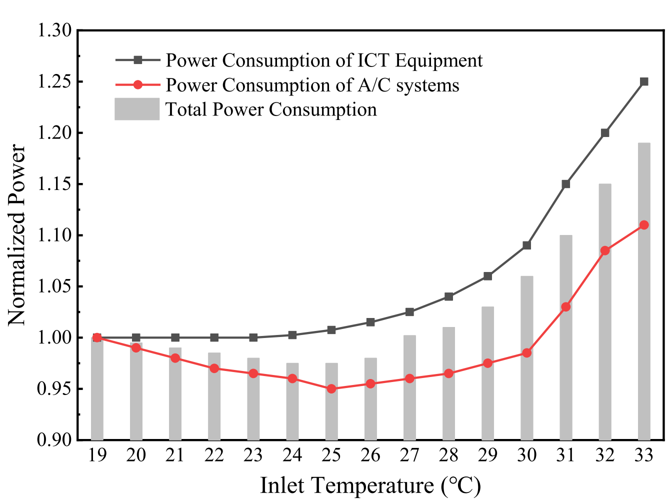

Fig. 14. Total power estimation in data centers.

For high-temperature data centers, we could not ignore the premiums of higher-class servers which are usually equipped with optimized heat removal mechanisms and more robust components. For example, the improved heat sink design and advanced materials are available but very expensive [18]. It is estimated that, if adopting cooling improvements to maintain the server performance, there will be a 1%–2% premium for class A3 servers compared with class A2 servers. There will be a 5%–10% premium for class A4 servers compared with class A2 servers. In addition, given that the cooling system might not be improved due to the volume constraints of servers, many server designs might need improved non-cooling components, such as by adopting advanced materials, to achieve class A3 or class A4 operation. In those cases, there would be a cost premium of 10%–15% over a class A2 server [18].

3.4.2. Total life-cycle cost

Although the capital investment of high-temperature data centers will increase due to the premiums of higher-class servers, the operating cost of high-temperature data centers will decrease due to more free cooling hours. The increase of capital cost mainly consists of server premiums. The reduction of capital cost mainly involves the investment of chiller plants, especially for chiller-less plants. The reduction of operating cost mainly associated to the increased chiller efficiency and reduced chiller operating hours.

Baidu corporation conducted a total cost of ownership analysis, considering cooling operating cost and maintenance cost [45]. The results show that the cooling system investment cost would be greatly reduced and the overall cost benefit would increase by 7.48% if the chillers were not used and air-side free cooling mode was used. Rubenstein et al. [69] investigated the operating cost of data centers when increasing the space temperature in computer rooms, and concluded that increasing temperature indeed reduced the chiller power consumption, but the overall cost savings may not be realized due to the increase in server power and decreased reliability.

These studies analyzed the correlation between the operating cost and the space temperature in data centers without considering the premiums of higher-level servers, the capital cost of chiller-free cooling plants and the operating cost savings due to more free cooling time. It can be observed that current research lacks a detailed economic analysis on high-temperature data centers.

4. Current technologies and feasible solutions

This section summarizes current technologies and feasible solutions to implement high-temperature data centers from room level, rack level, server level and chip level in section 4.1, 4.2, 4.3 and 4.4, respectively. Current available optimization, improvement and innovation methods at each level are introduced and elaborated, as shown in Fig. 15. In addition, the current state and the future perspectives of data centers are presented and elaborated in section 4.5.

4.1. Room level

Optimizing airflow management is an effective way to enhance cooling efficiency in data centers, which could reduce hot air recirculation and cold air bypass. Most data centers have hot spots problems that the temperature of local areas is significantly hotter than the average temperature in the computer rooms. The temperature of the hot spots could be 8–10K higher than the average temperature in the computer rooms [21]. Therefore, more cooling energy and lower supply air temperature are needed for the reliable operation of all servers, due to “short board effect”.

Good airflow management allows servers to operate at a higher ambient temperature. With good airflow management, a 24 °C cold air from computer room air handlers can result in a maximum server inlet temperature of 25–26 °C somewhere in the data center (without hot

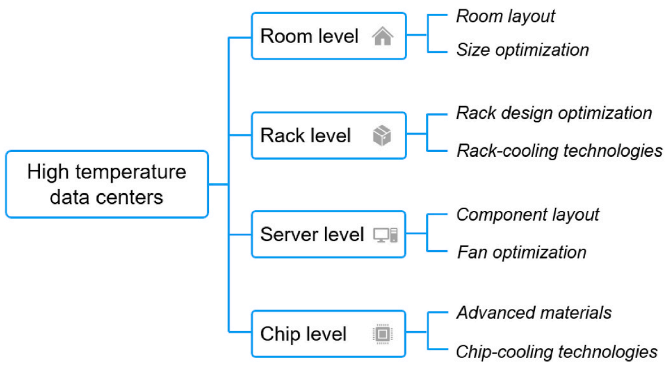

Fig. 15. Main feasible solutions at different levels for implementing high-temperature data centers.

spots), whereas with very poor airflow management, a 13 °C cold air could easily result in server inlet temperatures ranging anywhere from 25 °C up to over 32 °C due to hot spots. Thus, the improvement methods at room level are mainly to optimize airflow management and to eliminate hot spots.

4.1.1. Containment strategy

Aisle containment strategy, including cold aisle containment and hot aisle containment, is an effective means to improve airflow distribution in data centers, which has been widely used [70]. Containment could reduce the mixing effects of hot and cold airstreams that would cause unwanted temperature rises at rack inlets. Fig. 16 shows a typical hot-aisle/cold-aisle configuration. The cold air goes from raised floors to the room through perforated tiles, is aspired by racks on the front side, gains the heat created by racks, and is expelled at the rear of each rack. The hot air is then aspired by the computer room air conditioning units and is rejected to the exterior environment.

An experiment in a data center with a floor area of 52 m^2 shows that cold aisle containment could dramatically remove hot spots at rack inlet compared with conventional open aisle configuration [72]. With cold aisle containment, the rack inlet temperature could be reduced by up to 40% without changing the room layout [73]. Other studies found that the supplied air temperature could be raised by 3K [74], or raised from 18 °C to 22 °C by using aisle containment [75]. It is also found that hot-aisle containment could save 43% in annual cooling system energy cost and 15% reduction in annualized PUE, comparing with cold-aisle containment [71].

4.1.2. Room layout

Room layout mainly involves room shape optimization, rack layout, and the location of computer room air handlers (CRAHs) [76]. A study found that locating CRAH units perpendicular to rack rows had better cooling efficiency than locating it in line with rack rows [77]. Schmidt et al. [78] studied, overhead/raised-floor air supply systems, and found that raised-floor design could achieved a higher temperature of supply air. Another study also found that the raised-floor air supply with overhead return system had the best cooling performance among four different systems, since this system could use the lowest supply air flow rate to achieve the same average temperature at the same heat density and supply air temperature [79].

4.1.3. Structural size optimization

Structural size optimization mainly includes the optimization of plenum size, ceiling height and the shape of the perforated tile. An optimized structure allows computer room air handlers to supply higher temperature cold air to computer rooms.

A study shows that a 12K decline in server inlet temperature could be achieved by optimizing the plenum depth, ceiling height, and cold aisle location [80]. Lu et al. [81] studied the effect of geometry configurations on the thermal performance of data centers, and found that the plenum with gradient cross-sections created by inclined partitions could reduce the inlet and outlet air temperature by 1.7–2.9K. Rack inlet temperatures could also be affected by the raised floor depth and perforated tile-free areas. The results show that for a very slight variation of open tile areas (less than 8%), some rack inlet air temperatures saw a reduction of as much as 10K [82]. Open area of perforated tiles affects flow uniformity alongside the tiles. 25% open area of perforated tiles was recommended to maximize perforated tile airflow uniformity and temperature distribution uniformity [83]. Tile orientation also affects air flow distribution. The sensitivity in bulk flow rate and variations in flow distribution of different tile orientations has been studied [84]. The perforated tile geometry also has considerable impacts on the airflow characteristics in an aisle [85].

4.2. Rack level

At rack level, the thermal performance is affected by the rack design [86] and the server placements within racks [87]. Rack level modifications and optimization could improve heat dissipation by up to 50% and decrease temperature variability by 60% under the same cooling infrastructure [86,87]. In addition, the innovative technologies applied in rack-level cooling is also an effective way to enhance cooling efficiency.

4.2.1. Rack design and optimization

The server configuration is found to have a strong correlation to failure rates when raising the inlet temperature. The correlation between failures and the location of servers within a rack should be considered in when setting the inlet temperature of servers [30]. A study on server rack design proposed to decouple the fans and power supply units from the node and to place all fans on the rear side of the rack. A new rack design lowered the power consumption by 10% while allowing key components to keep enough margin to maintain the reasonable

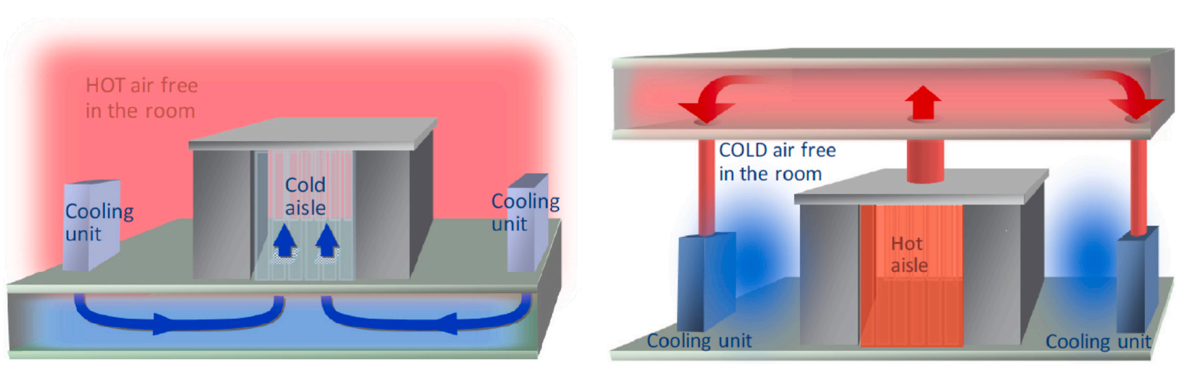

Fig. 16. Hot-aisle/cold-aisle configuration [71].

reliability levels, even under an ambient temperature of 40 °C [41]. Another similar design achieved a power reduction of 66.7% compared with general racks under the same test conditions [88].

It was also found that the heat generated by the IT equipment could be efficiently discharged with higher porosity. By increasing the porosity ratio of rack doors from 25% to 50%, the thermal environment inside the rack could be improved significantly [89].

Placement of servers affect rack thermal performance significantly. Wang et al. [90] proposed a drawer-type rack design with much hot aisle space and less cold aisle space. The design could reduce hot air recirculation and cold air bypass, and decrease the rack maximum inlet temperature by up to 13.3K. Chu et al. studied the effect of server layout and heat exchanger layout on the airflow uniformity in a vertical direction [11]. Results show that the placement of servers and heat exchangers had a great impact on the airflow distribution at the server inlet. To obtain a reasonable server arrangement, a similar test was conducted to discover the relationship between servers' placements and airflow distribution. It was recommended to uniformly mount the servers from the middle of the rack to the upper for low load rate cases [91].

4.2.2. Innovative rack-cooling technologies

Yu et al. [92] applied the solid sorption heat pipe coupled with direct air cooling technology for rack-level cooling in data centers, and found that this cooling technology could reduce the peak temperature of servers from 75.8 °C to 68.8 °C. Such reduction of server peak temperature means high space temperature could be allowed in data centers. Li et al. [93] proposed a multi-split backplane cooling system at the rack level. A maximum of 12% energy consumption can be saved by a reliable operation control strategy. Tian et al. [94] designed an internally cooled rack with a two-stage heat pipe. Their test results show that the new cooling solution basically eliminates undesired air mixing and hot spots, and the annual cooling energy consumption is reduced by about 13%.

4.3. Server level

Today's data center cooling systems are developed from traditional building HVAC systems. The main difference is that the serviced objects of the data center cooling systems are mainly servers, while the serviced objects of traditional offices and residential buildings are mainly people. The comfortable temperature range of the human body is narrow, but the server is man-made and can be improved from the aspects of structure and materials. Redesigning servers is a feasible and effective means implementing for high-temperature data centers.

A server chassis consists of sheet metal casings, CPUs, motherboards, power supplies, fans, memories, HDDs and the cables connecting these components [95]. The layout of servers and the position of each component on the chassis are optimized considering its overall efficiency and cost.

Redesigning the server layout is an effective way to increase heat dissipation efficiency. Numerical study on a hybrid air/liquid-cooled server shows that a basic server layout optimization such as changing the memory module angles and spacing could enhance both the cooling effectiveness and the potential for waste heat recovery from the air stream. Server layout optimization could also decrease entropy generation by 15% [96].

Sarma et al. [97] conducted a CFD simulation about the thermal design of a 2-rack unit high computing server, and summarized existing problems and reasonable thermal design. Hybrid cooling system are proposed for servers. It combines air cooling for low power components (e.g., power supplies, storage disk drives, and printed circuit boards) and water cooling for higher power components (e.g., microprocessors and dual inline memory module (DIMM)). This kind of server could accept water of up to 45 °C and air of up to 50 °C into the node, and saved 25% energy consumption in total [98].

To investigate the impact of the internal design of servers on their performance and cooling efficiency, servers had been tested in a chimney exhaust rack. Internal recirculation due to unreasonable design could cause a hot air leakage from the server into the cold aisle, and subsequently affect the performance of adjacent servers. Because of the leakage, the power consumption of adjacent servers increased by around 20% [99].

4.4. Chip level

Chip-level cooling objects usually include processors and hard disks and memories [100]. Generally, the schemes at the chip level could be divided into two categories: advanced chip materials and chip cooling.

4.4.1. Chip materials

The third-generation semiconductors with a wide bandgap, such as GaN/SiC equipment are considered to be of high performance, which further improves the energy efficiency [36]. The wider bandgap allows the material to operate at a higher temperature, stronger voltage, and faster switching frequency [101]. When applied in data centers, such advanced silicon materiel will significantly improve energy efficiency in the data center industry.

4.4.2. Innovative chip-cooling technologies

Chip cooling enhancement technologies mainly include high conductivity thermal interface materials [102,103] and innovative chip cooling technologies [104]. Chip cooling enhancement technologies could increase effectively heat dissipation of heat sources (mainly processors (60–75W each), DIMMs (6W each) [2]).

Nanofluids [105,106], such as TiO_2 nanofluid and SiC nanofluid, are found to have a high-efficiency cooling potential [107,108]. They could reduce CPU surface temperature by about 3.3K (8.2%) compared with deionized water with cylindrical grooves heat sink [109]. Choi et al. [110] designed a new CPU cooler based on an active cooling heat sink combined with heat pipes. The total fan power consumption could be effectively reduced by 66.2% when using a heat pipe embedded heatsink to replace the conventional heatsink [111].

Chip-scale refrigeration technologies are viable for thermal limited applications, such as thermo-electric coolers [112] and small-scale refrigeration cycle [113]. Deng et al. [114] studied two-stage multichannel liquid-metal cooling system for thermal management of high-heat-flux-density chip array. The experimental results show that the proposed liquid-metal cooling system can accommodate a heat flux of 50–200 W/cm^2 with a convective heat transfer coefficient exceeding 20,000 W/(m^2·K). Liang et al. [115] compared the heat pipe, thermoelectric system and vapor compression refrigeration for electronics cooling. They concluded: *i.* heat pipe is more attractive for cooling large devices at higher temperatures, *ii.* two-stage TE system could be used for cooling devices at lower temperature, and *iii.* VCR system was capable of dissipating much higher heat flux (200 W/cm^2) at lower temperature than all other technologies [115]. Chowdhury [116] integrated the thermo-electric coolers fabricated from nanostructured Bi_2Te_3-based thin-film superlattices into state-of-the-art electronic packages. Naphon et al. [117] studied HDD cooling with vapor chamber, and found that the average HDD temperature with the vapor chamber cooling system was 15.21% lower than those without the vapor chamber cooling system.

5. Status quo, future perspectives and opportunities

The space temperature in data centers is still very conservative, particularly for those for rental purpose. It is observed that very few data centers tried the high-temperature operation due to the concern on reliability though the allowing temperature limit is increased by some professional associations, such as ASHRAE.

Several issues need to be well addressed before adopting the high-temperature operation in practice. The most important issue is to enhance the high-temperature tolerance of server components, especially the HDDs and CPUs. Perhaps an increase of the space temperature in data centers by 5–10K could make a huge difference in cooling energy use. Some advanced materials, such as GiN, the third-generation semiconductor materials, may be the new breakthrough for electronic components. In addition, optimizing data center cooling systems is another approach to make high-temperature operation of data centers possible. The optimal design, control, and choice of IT equipment are the main contributors to cooling efficiency enhancement. In high-temperature data centers, free cooling can be adopted to realize chiller-less or even chiller-free cooling plants or operation.

This paper focuses on summary work on the applicability and development of high-temperature data centers from the perspective of benefits, bottlenecks, current technologies, existing efforts and future perspectives. The key point of the implementation of higher-temperature data centers is the promotion of servers that could bear higher server inlet temperatures without consuming more energy. Therefore, the development and perspective of the cooling technologies on IT equipment in data centers need to be further reviewed systematically. Also, energy-efficient and innovative chip cooling technologies need to be studied and addressed.

This work summaries the existing efforts and the latest development of high-temperature data centers. If operators want to raise server inlet temperature in practice, it is necessary to consult with the manufacturer of the equipment that they want to purchase or they have purchased to understand the performance capability at extreme upper limits of the allowable thermal envelopes.

6. Summary and conclusions

High-temperature data center is a promising means and a major development direction for major saving of cooling energy particularly by changing cooling mechanism fundamentally, i.e., adopting free cooling effectively. However, the choice of space temperature in most data centers is still conservative due to the concern for server reliability and performance. This paper provides a comprehensive review of the existing efforts and latest developments associated to high-temperature data centers. The benefits, concerns and bottlenecks, current technologies and solutions as well as future perspectives for implementing high-temperature data centers are summarized and analyzed. The high-temperature operation would be one of the development directions of the next-generation data centers, and it may result in a revolution in the data centers industry. This review will help decision makers and various stakeholders to understand the current status, benefits, bottlenecks, challenges and future perspectives for high-temperature data centers.

In respect to the most important aspects associated to the implementation of high-temperature data centers, the major conclusions and observations can be drawn as follows. Concerning the main benefits, high-temperature data centers have great energy-saving potential due to more free cooling hours and even chiller-less/chiller-free cooling plants. Meanwhile, the high-temperature operation can provide more opportunity for waste heat recovery. For example, the waste heat could be used in neighboring buildings or in a district heating system. Concerning the bottlenecks and limitations, the most replaced component in servers is the hard disk drive. It needs to be further improved to enhance the reliability of servers placed in high-temperature data centers. As rising the space temperature in data centers involves the increase of initial investment of IT equipment today and the saving of operating cost, there is a need of trade-off between them concerning the life cycle operation cost for developing high-temperature data centers. Concerning the status quo and existing efforts, the reliability and performance of severs are more robust than most people might imagine, particularly with the advent of the class A3 and class A4 servers. The new generation servers equipped with improved heat sink and advanced chip materials have better performance to withstand high-temperature operation. Current technologies and feasible solutions for high-temperature data centers fall into 4 categories, including room level, rack level, server level and chip level. Concerning future perspectives, the keys to implementing high-temperature data centers are the development and enhancement of servers and IT equipment for high temperature operation, the awareness and deployment of new generation servers, and the optimization of cooling systems in data centers at all levels. Thus, energy-efficient and innovative chip cooling technologies need to be studied and addressed in the further research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by a collaborative research fund (C5018-20G) of the Research Grant Council (RGC) of the Hong Kong SAR and a project of strategic importance of The Hong Kong Polytechnic University.

References

[1] Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. Science 2020;367:984–6.
[2] Ebrahimi K, Jones GF, Fleischer AS. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. Renew Sustain Energy Rev 2014;31:622–38.
[3] Deymi-Dashtebayaz M, Valipour-Namanlo S. Thermoeconomic and environmental feasibility of waste heat recovery of a data center using air source heat pump. J Clean Prod 2019;219:117–26.
[4] Belkhir L, Elmeligi A. Assessing ICT global emissions footprint: trends to 2040 & recommendations. J Clean Prod 2018;177:448–63.
[5] Andrews D, Newton E, Adibi N, Chenadec J, Bienge K. A circular economy for the data centre industry: using design methods to address the challenge of whole system sustainability in a unique industrial sector. Sustainability 2021;13:6319.
[6] Nadjahi C, Louahlia H, Lemasson S. A review of thermal management and innovative cooling strategies for data center. Sustain Comput: Inf Syst 2018;19:14–28.
[7] Rambo J, Joshi Y. Modeling of data center airflow and heat transfer: state of the art and future trends. Distributed Parallel Databases 2007;21:193–225.
[8] Cheung H, Wang S, Zhuang C, Gu J. A simplified power consumption model of information technology (IT) equipment in data centers for energy system real-time dynamic simulation. Appl Energy 2018;222:329–42.
[9] Mitchell-Jackson J, Koomey JG, Nordman B, Blazek M. Data center power requirements: measurements from Silicon Valley. Energy 2003;28:837–50.
[10] Luo Y, Andresen J, Clarke H, Rajendra M, Maroto-Valer M. A decision support system for waste heat recovery and energy efficiency improvement in data centres. Appl Energy 2019;250:1217–24.
[11] Chu W-X, Wang C-C. A review on airflow management in data centers. Appl Energy 2019;240:84–119.
[12] Zhang Y, Zhang K, Liu J, Kosonen R, Yuan X. Airflow uniformity optimization for modular data center based on the constructal T-shaped underfloor air ducts. Appl Therm Eng 2019;155:489–500.
[13] Chu W-X, Wang R, Hsu P-H, Wang C-C. Assessment on rack intake flowrate uniformity of data center with cold aisle containment configuration. J Build Eng 2020;30:101331.
[14] Tatchell-Evans M, Kapur N, Summers J, Thompson H, Oldham D. An experimental and theoretical investigation of the extent of bypass air within data centres employing aisle containment, and its impact on power consumption. Appl Energy 2017;186:457–69.
[15] MirhoseiniNejad S, Moazamigoodarzi H, Badawy G, Down DG. Joint data center cooling and workload management: a thermal-aware approach. Future Generat Comput Syst 2020;104:174–86.
[16] Gupta R, Moazamigoodarzi H, MirhoseiniNejad S, Down DG, Puri IK. Workload management for air-cooled data centers: an energy and exergy based approach. Energy 2020;209:118485.
[17] Gandhi A, Harchol-Balter M, Das R, Lefurgy C. Optimal power allocation in server farms. Perform Eval Rev 2009;37:157–68.
[18] ASHRAE Technical Committee (TC). 9.9, thermal guidelines for data processing environments – expanded data center classes and usage guidance. Am Soc Heat Refrigerat Air Conditin Eng 2011.
[19] Miller R. Raise your data center temperature 2008. https://www.datacenterknowledge.com/archives/2008/10/14/google-raise-your-data-center-temperature.

[20] Breen TJ, Walsh EJ, Punch J, Shah AJ, Bash CE. From chip to cooling tower data center modeling: Part I influence of server inlet temperature and temperature rise across cabinet. In: 12th IEEE intersociety conference on thermal and thermomechanical phenomena in electronic systems. IEEE; 2010. p. 1–10. 2010.
[21] El-Sayed N, Stefanovici IA, Amvrosiadis G, Hwang AA, Schroeder B. Temperature management in data centers: why some (might) like it hot. In: Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on measurement and modeling of computer systems; 2012. p. 163–74.
[22] ASHRAE Technical Committee (TC). 9.9 thermal guidelines for data processing environments. fifth ed. American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2021.
[23] Fitch J. Dell's next generation servers: pushing the limits of data center cooling cost savings. DELL Enterprise Reliability Engineering; 2012.
[24] Vallury A, Matteson J. Data center trends toward higher ambient inlet temperatures and the impact on server performance. In: International electronic packaging technical conference and exhibition. American Society of Mechanical Engineers; 2013. V002T09A10.
[25] Beaty DL, Quirk D. De-risking data center temperature increases, Part 1. ASHRAE J 2016;58:74–82.
[26] Beaty DL, Quirk D. De-risking data center temperature increases, Part 2. ASHRAE J 2016;58:70–5.
[27] He Y, Chen G, Wei W, Liu Q, Zhang J, Zhou T, et al. HTA corrosion resistant technology for free cooling. In: 31st thermal measurement, modeling & management symposium (SEMI-THERM)2015; 2015. p. 120–6.
[28] Miller R. Too Hot for Humans, But Google Servers Keep Humming https://www.datacenterknowledge.com/archives/2012/03/23/too-hot-for-humans-but-google-servers-keep-humming. 2012.
[29] Goiri I, Nguyen T, Bianchini R. CoolAir: temperature- and variation-aware management for free-cooled datacenters. ACM SIGPLAN Not 2015;50:253–65.
[30] Sankar S, Shaw M, Vaid K, Gurumurthi S. Datacenter scale evaluation of the impact of temperature on hard disk drive failures. ACM Trans Storage 2013;9:1–24.
[31] Li J, Li Z. Model-based optimization of free cooling switchover temperature and cooling tower approach temperature for data center cooling system with water-side economizer. Energy Build 2020;227:110407.
[32] Kelley C, Singh CH, Smith V. Data center efficiency and IT equipment reliability at wider operating temperature and humidity ranges (White Paper). The Green Grid; 2012.
[33] Ahuja N. Datacenter power savings through high ambient datacenter operation: CFD modeling study. In: 28th annual IEEE semiconductor thermal measurement and management symposium 2012; 2012. p. 104–7.
[34] Dubin FS, Mindell HL, Bloome S. How to save energy and cut costs in existing industrial and commercial buildings: an energy conservation manual. United States: Department of Energy; 1976.
[35] Squillo T. 3 Ways To Increase Chiller Efficiency https://www.facilitiesnet.com/hvac/article/3-Ways-To-Increase-Chiller-Efficiency–17669. 2018.
[36] Liu Y, Wei X, Xiao J, Liu Z, Xu Y, Tian Y. Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers. Global Energy Interconnect 2020;3:272–82.
[37] Daraghmeh HM, Wang C-C. A review of current status of free cooling in datacenters. Appl Therm Eng 2017;114:1224–39.
[38] Fernández-Montes A, Fernández-Cerero D, González-Abril L, Álvarez-García JA, Ortega JA. Energy wasting at internet data centers due to fear. Pattern Recogn Lett 2015;67:59–65.
[39] Harvey T, Patterson M, Bean J. Updated air-side free cooling maps: the impact of ASHRAE 2011 allowable ranges (White Paper). The Green Grid; 2012.
[40] Miller R. Server Failures Don't Rise Along With the Heat 2012. https://www.datacenterknowledge.com/archives/2012/05/29/study-server-failures-dont-rise-along-with-the-heat.
[41] Zhang S, Ahuja N, Han Y, Ren H, Chen Y, Guo G. Key considerations to implement high ambient data center. 31st thermal measurement, modeling & management symposium (SEMI-THERM)2015. 2015. p. 147–54.
[42] Mone G. Redesigning the data center. Commun ACM 2012;55:14–6.
[43] Miller R. In Dublin, Cool Climate Fuels Cloud Computing Cluster 2013. https://www.datacenterknowledge.com/archives/2013/04/08/dublin-free-cooling.
[44] Don Atwood JGM. Reducing Data Center Cost with an Air Economizer 2008. https://www.intel.com/content/dam/doc/technology-brief/data-center-efficiency-xeon-reducing-data-center-cost-with-air-economizer-brief.pdf.
[45] He Y, Chen G, Zhang J, Zhou T, Liu T, Zhu P, et al. Consideration for running data center at high temperatures and using free air cooling. In: International electronic packaging technical conference and exhibition. American Society of Mechanical Engineers; 2015. V001T09A14.
[46] Schroeder B, Gibson GA. Understanding disk failure rates: what does an MTTF of 1,000,000 hours mean to you? ACM Trans Storage 2007;3:8. –es.
[47] Schroeder B, Gibson GA. A large-scale study of failures in high-performance computing systems. IEEE Trans Dependable Secure Comput 2010;7:337–50.
[48] Sankar S, Shaw M, Vaid K. Impact of temperature on hard disk drive reliability in large datacenters. In: IEEE/IFIP 41st international conference on dependable systems & networks (DSN)2011; 2011. p. 530–7.
[49] Pinheiro E, Weber W-D, Barroso LA. Failure trends in a large disk drive population. In: 5th USENIX conference on file and storage technologies. USENIX Association; 2007.
[50] Chan CS, Jin Y, Wu Y-K, Gross K, Vaidyanathan K, Ti Rosing. Fan-speed-aware scheduling of data intensive jobs. In: Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design. Redondo Beach, California, USA: Association for Computing Machinery; 2012. p. 409–14.
[51] Schroeder B, Merchant A, Lagisetty R. Reliability of nand-based SSDs: what field studies tell us. Proc IEEE 2017;105:1751–69.
[52] Chan CS, Pan B, Gross K, Vaidyanathan K, Rosing TŠ. Correcting vibration-induced performance degradation in enterprise servers. Perform Eval Rev 2014;41:83–8.
[53] Al-Ars Z, van de Goor AJ, Braun J, Richter D. Simulation based analysis of temperature effect on the faulty behavior of embedded drams. In: Proceedings international test conference 2001 (Cat No 01CH37260). IEEE; 2001. p. 783–92.
[54] Hamamoto T, Sugiura S, Sawada S. On the retention time distribution of dynamic random access memory (DRAM). IEEE Trans Electron Dev 1998;45:1300–9.
[55] Schroeder B, Pinheiro E, Weber W-D. DRAM errors in the wild: a large-scale field study. Perform Eval Rev 2009;37:193–204.
[56] Wang Y. Evaluating and modeling the energy impacts of data centers. In: Terms of hardware/software architecture and associated environment: ecole nationale supérieure Mines-Télécom Atlantique; 2020.
[57] Zapater M, Tuncer O, Ayala JL, Moya JM, Vaidyanathan K, Gross K, et al. Leakage-Aware cooling management for improving server energy efficiency. IEEE Trans Parallel Distr Syst 2015;26:2764–77.
[58] Moss D, Bean JH. Energy impact of increased server inlet temperature (APC white paper). American Power Conversion; 2009.
[59] Yeo S, Hossain MM, Huang J-C, Lee H-HS. ATAC: ambient temperature-aware capping for power efficient datacenters. In: Proceedings of the ACM symposium on cloud computing. Seattle, WA, USA: Association for Computing Machinery; 2014. p. 1–14.
[60] Patterson MK. The effect of data center temperature on energy efficiency. In: 11th intersociety conference on thermal and thermomechanical phenomena in electronic Systems2008; 2008. p. 1167–74.
[61] Wan J, Gui X, Zhang R, Fu L. Joint cooling and server control in data centers: a cross-layer framework for holistic energy minimization. IEEE Syst J 2018;12:2461–72.
[62] Kim NS, Austin T, Baauw D, Mudge T, Flautner K, Hu JS, et al. Leakage current: Moore's law meets static power. Computer 2003;36:68–75.
[63] Shin D, Kim J, Chang N, Choi J, Chung SW, Chung E. Energy-optimal dynamic thermal management for green computing. IEEE/ACM International Conference on Computer-Aided Design - Digest of Technical Papers2009; 2009. p. 652–7.
[64] Agarwal A, Mukhopadhyay S, Kim C, Raychowdhury A, Roy K. Leakage power analysis and reduction: models, estimation and tools. IEE Proc Comput Digit Tech 2005;152:353–68.
[65] Fallah F, Pedram M. Standby and active leakage current control and minimization in CMOS VLSI circuits. IEICE Trans Electron 2005;88:509–19.
[66] Seaton I. Airflow Management Considerations for a New Data Center – Part 1: Server Power versus Inlet Temperature 2017. https://www.upsite.com/blog/server-power-versus-inlet-temperature/.
[67] Muroya K, Kinoshita T, Tanaka H, Youro M. Power reduction effect of higher room temperature operation in data centers. IEEE Network Operations and Management Symposium; 2010. p. 661–73.
[68] Seaton I. Airflow Management Considerations for a New Data Center – Part 3: Server Cost vs Inlet Temperature 2017. https://www.upsite.com/blog/server-cost-vs-inlet-temperature/.
[69] Rubenstein B, Faist M. Data center cold aisle set point optimization through total operating cost modeling. In: Fourteenth intersociety conference on thermal and thermomechanical phenomena in electronic systems. ITherm); 2014. p. 1111–20.
[70] Nada SA, Elfeky KE. Experimental investigations of thermal managements solutions in data centers buildings for different arrangements of cold aisles containments. J Build Eng 2016;5:41–9.
[71] Niemann J, Brown K, Avelar V. Impact of hot and cold aisle containment on data center temperature and efficiency (APC White Paper). Schneider Electric Data Center Science Center; 2011.
[72] Sundaralingam V, Arghode VK, Joshi Y, Phelps W. Experimental characterization of various cold aisle containment configurations for data centers. J Electron Packag 2015;137:011007.
[73] Gondipalli S, Bhopte S, Sammakia B, Iyengar MK, Schmidt R. Effect of isolating cold aisles on rack inlet temperature. In: 11th intersociety conference on thermal and thermomechanical phenomena in electronic Systems2008; 2008. p. 1247–54.
[74] Gao C, Yu Z, Wu J. Investigation of airflow pattern of a typical data center by CFD simulation. Energy Proc 2015;78:2687–93.
[75] Cho J, Yang J, Park W. Evaluation of air distribution system's airflow performance for cooling energy savings in high-density data centers. Energy Build 2014;68:270–9.
[76] Dunlap K, Rasmussen N. Choosing between room, row, and rack-based cooling for data centers (APC White Paper). Schneider Electric Data Center Science Center; 2012.
[77] Nada SA, Said MA. Effect of CRAC units layout on thermal management of data center. Appl Therm Eng 2017;118:339–44.
[78] Schmidt RR, Iyengar M, Vogel M, Pienta B. Comparison between underfloor supply and overhead supply ventilation designs for data center high-density clusters/discussion. Build Eng 2007;113:115.
[79] Nakao M, Hayama H, Nishioka M. Which cooling air supply system is better for a high heat density room: underfloor or overhead?. In: Thirteenth international telecommunications energy conference-INTELEC. vol. 91. IEEE; 1991. p. 393–400.
[80] Bhopte S, Agonafer D, Schmidt R, Sammakia B. Optimization of data center room layout to minimize rack inlet air temperature. J Electron Packag 2006;128:380–7.

[81] Lu H, Zhang Z. Numerical and experimental investigations on the thermal performance of a data center. Appl Therm Eng 2020;180:115759.
[82] Schmidt R, Cruz E. Cluster of high powered racks within a raised floor computer data center: effect of perforated tile flow distribution on rack inlet air temperatures. In: ASME 2003 international mechanical engineering congress and exposition; 2003. p. 245–62.
[83] VanGilder JW, Schmidt RR. Airflow uniformity through perforated tiles in a raised-floor data center. In: International electronic packaging technical conference and exhibition; 2005. p. 493–501.
[84] Seymour M. Modeling perforated tiles in data centers – What is required? 2015 31st Thermal Measurement, Modeling & Management Symposium (SEMI-THERM) 2015:8–12.
[85] Khalili S, Tradat MI, Nemati K, Seymour M, Sammakia B. Impact of tile design on the thermal performance of open and enclosed aisles. J Electron Packag 2018; 140:010907.
[86] Mulay V, Agonafer D, Irwin G, Patell D. Effective thermal management of data centers using efficient cabinet designs. In: International electronic packaging technical conference and exhibition; 2009. p. 993–9.
[87] Rolander N, Rambo J, Joshi Y, Allen JK, Mistree F. An approach to robust design of turbulent convective systems. J Mech Des 2006;128:844–55.
[88] Pang W, Wang C, Ahuja N, Zhang J, Zhou A, Si P, et al. An advanced energy efficient rack server design. In: 16th IEEE intersociety conference on thermal and thermomechanical phenomena in electronic systems (ITherm). IEEE; 2017. p. 806–14. 2017.
[89] Ling Y, Zhang X, Li S, Han G, Sun X. The numerical simulation and experimental study on rack cooling effect in date center with UFAD system. Int J Low Carbon Technol 2015;10:446–51.
[90] Wang IN, Tsui Y-Y, Wang C-C. Improvements of airflow distribution in a container data center. Energy Proc 2015;75:1819–24.
[91] Jin C, Bai X. The study of servers' arrangement and air distribution strategy under partial load in data centers. Sustain Cities Soc 2019;49:101617.
[92] Yu Y, Wang L. Solid sorption heat pipe coupled with direct air cooling technology for thermal control of rack level in internet data centers: design and numerical simulation. Int J Heat Mass Tran 2019;145:118714.
[93] Li X, Zhang C, Sun X, Han Z, Wang S. Experimental study on reliable operation strategy for multi-split backplane cooling system in data centers. Appl Therm Eng 2022;211:118494.
[94] Tian H, He Z, Li Z. A combined cooling solution for high heat density data centers using multi-stage heat pipe loops. Energy Build 2015;94:177–88.
[95] Siddarth AV. Experimental study on effects of segregated cooling provisioning on thermal performance of information technology servers in air cooled data centers. The University of Texas; 2015.
[96] Sakanova A, Alimohammadi S, McEvoy J, Battaglioli S, Persoons T. Multi-objective layout optimization of a generic hybrid-cooled data centre blade server. Appl Therm Eng 2019;156:514–23.
[97] Sarma ANSVS, Ambali VD. Cooling solution for computing and storage server. In: 16th IEEE intersociety conference on thermal and thermomechanical phenomena in electronic systems. ITherm); 2017. p. 840–9. 2017.
[98] Iyengar M, David M, Parida P, Kamath V, Kochuparambil B, Graybill D, et al. Server liquid cooling with chiller-less data center design to enable significant energy savings. In: 28th annual IEEE semiconductor thermal measurement and management symposium (SEMI-THERM)2012; 2012. p. 212–23.
[99] Khalili S, Alissa H, Nemati K, Seymour M, Curtis R, Moss D, et al. Impact of server thermal design on the cooling efficiency: chassis design. J Electron Packag 2019; 141:031004.
[100] Lin J, Zheng H, Zhu Z, Gorbatov E, David H, Zhang Z. Software thermal management of DRAM memory for multicore systems. Perform Eval Rev 2008;36: 337–48.
[101] Wang L, Wang ZL. Advances in piezotronic transistors and piezotronics. Nano Today 2021;37:101108.
[102] Habibi Khalaj A, Halgamuge SK. A Review on efficient thermal management of air- and liquid-cooled data centers: from chip to the cooling system. Appl Energy 2017;205:1165–88.
[103] Feng S, Yan Y, Li H, He Z, Zhang L. Reprint of: temperature uniformity enhancement and flow characteristics of embedded gradient distribution micro pin fin arrays using dielectric coolant for direct intra-chip cooling. Int J Heat Mass Tran 2020;161:120235.
[104] Kheirabadi AC, Groulx D. Cooling of server electronics: a design review of existing technology. Appl Therm Eng 2016;105:622–38.
[105] Nallusamy D NM. Heat transfer enhancement analysis of Al2O3-water nanofluid through parallel and counter flow in shell and tube heat exchangers. Int J Nanosci 2017;16:1750020.
[106] Nallusamy S. Thermal conductivity analysis and characterization of copper oxide nanofluids through different techniques. J Nano Res 2016;40:105–12.
[107] Sarafraz MM, Arya A, Hormozi F, Nikkhah V. On the convective thermal performance of a CPU cooler working with liquid gallium and CuO/water nanofluid: a comparative study. Appl Therm Eng 2017;112:1373–81.
[108] Zhao N, Guo L, Qi C, Chen T, Cui X. Experimental study on thermo-hydraulic performance of nanofluids in CPU heat sink with rectangular grooves and cylindrical bugles based on exergy efficiency. Energy Convers Manag 2019;181: 235–46.
[109] Zhao N, Qi C, Chen T, Tang J, Cui X. Experimental study on influences of cylindrical grooves on thermal efficiency, exergy efficiency and entropy generation of CPU cooled by nanofluids. Int J Heat Mass Tran 2019;135:16–32.
[110] Choi J, Jeong M, Yoo J, Seo M. A new CPU cooler design based on an active cooling heatsink combined with heat pipes. Appl Therm Eng 2012;44:50–6.
[111] Wang Y, Wang B, Zhu K, Li H, He W, Liu S. Energy saving potential of using heat pipes for CPU cooling. Appl Therm Eng 2018;143:630–8.
[112] Hu HM, Ge TS, Dai YJ, Wang RZ. Experimental study on water-cooled thermoelectric cooler for CPU under severe environment. Int J Refrig 2016;62: 30–8.
[113] Poachaiyapoom A, Leardkun R, Mounkong J, Wongwises S. Miniature vapor compression refrigeration system for electronics cooling. Case Stud Therm Eng 2019;13:100365.
[114] Deng Y, Zhang M, Jiang Y, Liu J. Two-stage multichannel liquid–metal cooling system for thermal management of high-heat-flux-density chip array. Energy Convers Manag 2022;259:115591.
[115] Liang K, Li Z, Chen M, Jiang H. Comparisons between heat pipe, thermoelectric system, and vapour compression refrigeration system for electronics cooling. Appl Therm Eng 2019;146:260–7.
[116] Chowdhury I, Prasher R, Lofgreen K, Chrysler G, Narasimhan S, Mahajan R, et al. On-chip cooling by superlattice-based thin-film thermoelectrics. Nat Nanotechnol 2009;4:235–8.
[117] Naphon P, Wongwises S, Wiriyasart S. Application of two-phase vapor chamber technique for hard disk drive cooling of PCs. Int Commun Heat Mass Tran 2013; 40:32–5.