IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received September 17, 2019, accepted October 3, 2019, date of publication October 9, 2019, date of current version November 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946342

A Novel Methodology to Improve Cooling Efficiency at Data Centers

KANG HAO CHEONG[1,2], (Member, IEEE), KENNETH JIAN WEI TANG[1], JIN MING KOH[1], SIMON CHING MAN YU[3], U. RAJENDRA ACHARYA[4,5,6], AND NENG-GANG XIE[7]

[1]Science and Math Cluster, Singapore University of Technology and Design SUTD, Singapore 487372
[2]SUTD-MIT International Design Centre, Singapore 487372
[3]Interdisciplinary Division of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong
[4]Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore 599489
[5]Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore 599494
[6]School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, Subang Jaya 47500, Malaysia
[7]School of Management Science and Engineering, Anhui University of Technology, Ma'anshan 243002, China

Corresponding authors: Kang Hao Cheong (kanghao_cheong@sutd.edu.sg) and Neng-Gang Xie (xienenggang@aliyun.com)

This work was supported in part by the Singapore University of Technology and Design Start-up Research Grant under Grant SRG SCI 2019 142, and in part by the Science and Technology Major Project of Anhui Province under Grant 17030901037.

ABSTRACT Data centers are mission-critical infrastructures. There are strict service level requirements and standards imposed on operators and maintainers to ensure reliable run-the-clock operation. In the context of thermal management and data hall environmental control, the formation of hot and cold spots around server cabinets are especially undesirable, and can result in overheating, lifespan reductions, and performance throttling in the former and condensation damage in the latter. In this paper, we present a comprehensive multi-pronged methodology in data center environmental control, comprising computational fluid dynamics (CFD) simulation-aided predictive design first-stage approach, and a complementing Internet of Things (IoT) reactive management system that autonomously monitors and regulates fluctuations in thermal parameters. The novel hybrid methodology is demonstrated on various test scenarios derived from real-world context, and prototypes of the IoT system have been experimentally validated. The approach is shown to be efficient in eliminating unfavourable environmental variations and provides an enhanced understanding of common design problems and respective mitigation measures.

INDEX TERMS Optimization, CFD modelling, simulation, cooling.

I. INTRODUCTION

Cooling processes serve a critical role in the effective operation of data centers [1]–[3]. There are strict temperature and humidity requirements that co-location (Colo) service providers must adhere to, often stipulated within the terms of a service level agreement (SLA) between the provider and the client leasing the data hall premises for the placement and operation of server infrastructures [4], [5]. To ensure that these factors are adhered to, rack-mounted temperature sensors and ceiling-dropped space sensors are deployed at strategic locations to monitor environmental parameters in the hall space and alert the operator of any non-compliance, which may degrade the safe operation and resourcing of these oftentimes mission-critical computing equipment.

The implications of ineffective cooling include decreased server reliability caused by hardware exceeding designed thermal limits [6]–[10], diminished server performance due to thermal throttling [11]–[13], increased energy consumption consequent of operation at non-optimal temperature ranges [14] and aggravated internal fan load to dissipate heat, and in the extreme case, server room fires caused by equipment failures and overheated hardware [15]. These consequences easily extend beyond the data center into the global digital ecosystem [16], [17]. Studies have also indicated rapid deterioration of hardware longevity at increased operating temperatures [18]–[20]—the lifespan of general electronics is expected to diminish by 50% for every 10°C increase in temperature over a typical ambient of 21°C, and hard-disk failure rates are expected to double for every 15°C increase [21]. Hardware failure results not only in inflated maintenance costs but also uncertain reliability in the long-term.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Li.

While stringent measures are typically taken in the design and operation of data centers to ensure that key data hall environmental parameters are within acceptable levels, undesired temperature variations in the form of hot and cold spots may nevertheless arise. These can be transient in nature [22] or caused by unforeseen on-site infrastructural constraints or loads, and are therefore difficult to suppress through traditional *a priori* planning. As such, innovative approaches to address these needs are extremely important and it is the intent of this paper to explore methods to optimize environmental control, such that thermal regulatory failure and related effects can be efficiently avoided. While existing literature and industry best practices already encompass computational simulation-aided design methodologies [23]–[26] and dynamic load-balancing techniques at various software and hardware abstraction levels [27], [28], the fusion of different approaches have remained practically limited, and the exploitation of smart networks also presents much untapped potential [29], [30]. In this paper, we propose a multi-pronged approach comprising a *predictive* infrastructure design methodology utilizing computational fluid dynamics (CFD) and heat transport simulations, coupled with an Internet of things (IoT) *reactive* management system to sense transient fluctuations in environmental parameters and effect adjustments in near real-time. An accurate three-dimensional representation of a data hall along with its infrastructural sub-components is constructed, on which the proposed approach is demonstrated. The approach is shown to be effective for various scenarios, encompassing differing equipment placement and configurations, and differing airflow and thermal conditions. A practical engineering prototype of the IoT management system and its reactive operational modes are also demonstrated.

We first present a theoretical review of data center environmental control principles and guidelines and relevant thermal-airflow phenomena in Section II. We then present the implementation and results of the predictive first stage of our approach in Section III and the reactive second stage in Section IV, followed by a discussion in Section V.

II. THEORETICAL REVIEW

A. TECHNICAL GUIDELINES

We make critical reference to the ASHRAE TC9.9 Thermal Guidelines for Data Processing Environments white paper [31] in this study, in order to set appropriate environmental control objectives for the design and evaluation of the proposed CFD/IoT predictive-reactive approach. The ASHRAE TC9.9 white paper was first written by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) technical committee (TC) in 2004 to create a common set of environmental parameter guidelines across hardware manufacturers supplying mission-critical computing hardware, and has been periodically revised to address changes in technology and to enhance the comprehensiveness of the guidelines.

Prior to the release of these guidelines, environmental parameters required for the operation of data center computing hardware were unregulated and manufacturer-dependent, often resulting in issues of ambiguity and consistency over correct operating standards [32]. In response to the industrial void on consistent and unambiguous operating conditions, these guidelines provide a common environmental interface for the equipment and its surroundings, guidance on the evaluation and testing of the operational health of data centers, and a standardized methodology for reporting the environmental characteristics of computer systems.

With reference to these guidelines [31], a data center operating environment consisting of mission-critical server infrastructure is classified under class A1 with regard to environmental control measures. It is recommended that operating temperatures of these mission-critical server infrastructures be within a temperature range of 18–27°C at 60% typical relative humidity, with an allowable limit of 15–32°C at 20–80% relative humidity. Parameters out of the upper boundaries of these specifications are classified as hot spots while parameters out of the lower boundaries of these specifications are classified as cold spots [31].

B. HOT AND COLD SPOTS

Unfavourable thermal conditions in a data center may be binarily classified as hot and cold spots. Hot spots are defined as localized regions of hot air at the intakes of network and server equipment, in this study taken to be greater than temperature values recommended by the ASHRAE TC9.9 guidelines. Prolonged exposure to hot spots is detrimental to the reliability, performance and longevity of electronic hardware, and may void warranties and maintenance agreements of hardware manufacturers in more severe cases [33].

Likewise, cold spots refer to localized regions of cold air resulting from excessive cooling. Prolonged exposure to cold spots results in increased risks of corrosion or liquid damage to electronic equipment resulting from the condensation of moisture [34], [35]. For the purposes of this study, we take the minimum allowable temperature of cool air to be consistent with the ASHRAE TC9.9 guidelines. In light of the potential damage to equipment and degradation in performance, it is imperative that both hot and cold spots be avoided, or otherwise quickly addressed once discovered.

C. PRINCIPLES OF COOLING IN DATA CENTERS

Server cabinets in a data center are typically arranged in rows with consistent orientations, such that all air intakes face a cold aisle where cool air is supplied, and all air exhausts face a hot aisle where heated air is removed. A sectional view of a data hall arranged in such a hot-cold aisle configuration is shown in Figure 1. In data centers with multiple rows of cabinets, the orientations of the racks are arranged to alternate, such that the cold-hot aisle configuration is preserved. Cold air supplied from computer room air conditioning (CRAC) units is channeled through the raised floor plenum to servers via perforated floor tiles located along the cold aisle;

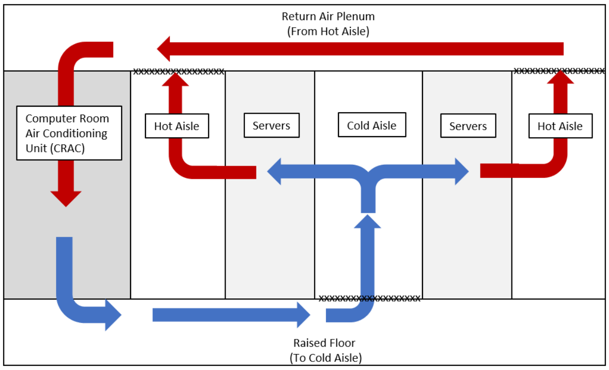


FIGURE 1. Simplified airflow schematic in a hot-cold aisle configuration.

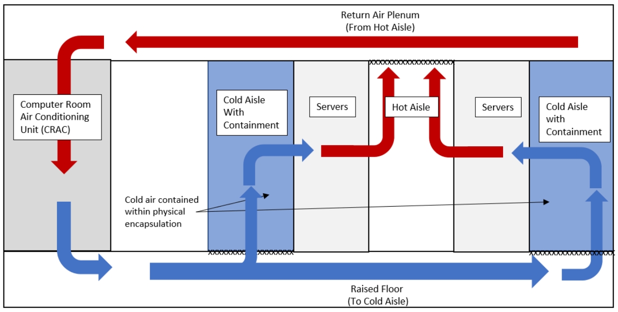


FIGURE 2. Cold Aisle Containment.

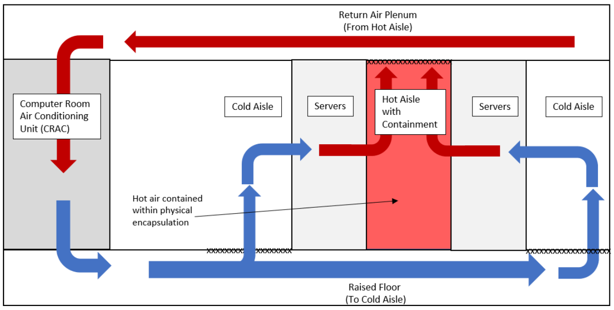


FIGURE 3. Hot Aisle Containment.

the conditioned air is drawn into the server racks to cool the internal electronic components, and then egested into the hot aisle, where a slight pressure gradient aided by natural convection returns the heated air into the return plenum and back to the CRAC units for cooling and subsequent recirculation [36].

D. COLD AND HOT AISLE CONTAINMENT

The effectiveness of cooling can be further enhanced with cold and hot aisle containment measures as shown in Figures 2 and 3. These methods operate based on the physical encapsulation of cold or hot aisles, thus preventing the contamination or mixing of conditioned and exhausted air. This results in improved cooling efficiency and performance, at the expense of more complex physical infrastructure. In the cold aisle containment, the physical encapsulation of the cold aisles is erected through the installation of barriers, thereby completely isolating the cold aisle from the rest of the data hall. This ensures that only the conditioned air is ingested by the cabinets and renders the surrounding environment effectively a hot aisle, therefore subjecting equipment not mounted in the cabinets and people working in the data halls to warmer air. In hot aisle containment, a complete encapsulation of hot aisles is implemented through similar means, ensuring that exhaust air can only flow towards the return ceiling plenum. This, likewise, renders the surrounding space effectively a cold aisle, and subject the non-cabinet equipment to colder air.

E. EXISTING INDUSTRIAL PRACTICES AND RELEVANT LITERATURE

Potential inconveniences may arise when data centres configured with non-adjustable perforated floorboards or manually operated volume control damper grates (VCDG) undergo underlying airflow configuration changes. These changes are a result of preventive maintenance activities when air conditioner units in computer room are operated in a different order or due to cooling equipment downtime resulting in the change of underlying hall temperatures, airflow pressures and directions. This results in the need for data centre operators without prior installed aisle containment measures to request for access into client sensitive halls to adjust computer room air conditioning parameters or manually reconfigure damper configurations and monitor hall temperature parameters until it has satisfactorily stabilised with reference to the client's service level agreement. Owing to the multi-variable nature of cooling, this often results in a repetitive approach which has operational and security implications. This requires innovative solutions to improve work flow and operational efficiency which our paper seeks to close this critical gap.

Existing literature on improving cooling efficiency primarily comprises hot-cold aisle containment, blanking panels and works involving the hardening of servers to tolerate higher temperatures to the use of liquid cooling in lieu of air [37]–[39]. Some of these measures may be impractical for retrofitting based on existing infrastructure limitations, costs, downtime, strict security protocols, approval processes and operational constraints of a live data centre environment. This is where our hybrid approach proposed in the following sections can play a vital role in working alongside existing measures to augment and improve cooling.

III. APPROACH 1: COMPUTATIONAL FLUID DYNAMICS

In this section, we detail the predictive first stage of the proposed environmental control methodology, utilizing virtual facilities and computational fluid dynamics simulation tools to model an accurate representation of the operational modes of a data hall. This approach tackles potential design problems on cooling capacity, such as inadequate or excessive cooled airflow to server racks, allowing for an enhanced *a priori* planning of equipment and cooling arrangements within the data hall premises. Offline virtual analyses can also be performed (and indeed demonstrated) to simulate changes in configurations to server racks, aisle floorboards, and CRAC unit settings to understand their effects on the data hall environment, before implementing the changes in a live hall setting, and appropriate response plans can also be drafted for cases of reconfiguration or equipment failure based on these simulations.

A. SIMULATION SOFTWARE

We utilize *Future Facilities* 6 *Sigma Room*, a specialized virtual facility (VF) and computational fluid dynamics (CFD) software for the simulation of the data center environment. The software iteratively solves the many simultaneous equations representing the conservation equations and Navier-Stokes equations for fluid dynamics and heat transport. The numerical scheme defines mass, momentum and energy on a three-dimensional automatically-generated mesh [40]. In comparison to other conventional CFD software choices such as *Ansys Fluent* and *Autodesk CFD*, *6 Sigma DC Room* is extensively used by data center operators globally [25], and its ease of convenience of use with built-in libraries of data center-specific hardware that are individually modelled, tested and validated to match real-world operational behaviours renders it preferable in our context. The accuracy of the software has been independently validated by end-users and leading academic institutions prominent in performing audits of CFD tools [41] to yield error differences within 5% of real-world results, thereby lending confidence to the predictions made through simulations.

B. SIMULATION ENVIRONMENT

Detailed simulations are conducted on a 91 m^2, 28-cabinet capacity data hall with 4 CRAC units arranged in a hot-cold aisle configuration (Figure 4). This configuration was chosen as it represents the most commonly used arrangement in the data center industry, and conforms to the industrial best practices for efficient cooling of servers in a data hall setting—specifically, CRAC units at each end of server rows, hot aisle width of approximately 1000 mm and cold aisle width of approximately 1200 mm [31], [42]. Specific conditions imposed on the infrastructural components of the data hall, such as floorboards, servers, and CRAC units, are presented in Table 1. Likewise, these represent typical values expected in modern data centers.

C. SCENARIOS & RESULTS

Simulation variables and scenarios are aplenty and readily explorable in a virtual setting. We shall focus on common hot-cold spot scenarios encountered in the industry, as well the feasibility and effectiveness of plausible mitigation measures, in this study. Three distinct scenarios are demonstrated—cold aisle contamination (Section III-C1), inadequate cooling in a semi-occupied data hall (Section III-C2), and cold-hot aisle containment (Section III-C3). For all demonstrated scenarios, baseline simulations are first run to yield a comparison basis, and suggested mitigation measures are evaluated against these.

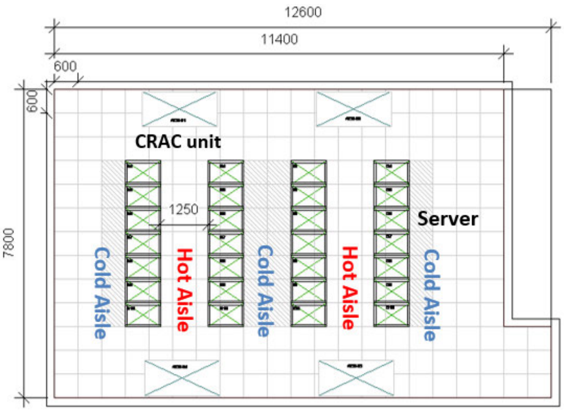


FIGURE 4. Floor plan of the modelled data hall.

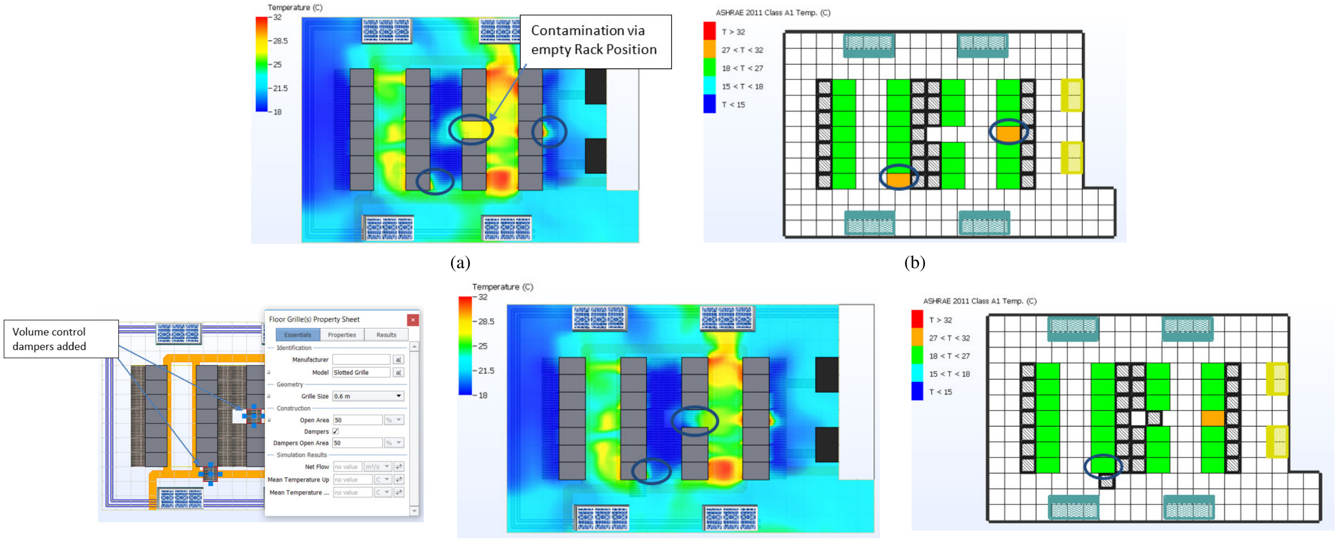
TABLE 1. Simulation constants for the computational fluid dynamics (CFD) study. Symbols H_{rm}, H_{rf}, Q_{nom}, U_{air}, T_{air}, P_{med} and P_{high} denote room height, raised floor height, nominal cooling load, air flow rate, air supply temperature set point, low-density thermal power, and medium-density thermal power respectively. The 42U cabinet form factor corresponds to 2050 mm in height by 600 mm in width by 900 mm in depth.

Infrastructure	Attributes	Conditions
Data Center Hall	$H_{\mathrm{rm}} = 4.5\,\mathrm{m}$ $H_{\mathrm{rf}} = 0.64\,\mathrm{m}$	No heat loss via walls and floors.
Floorboards (Perforated)	Up to 50% opening	No phantom leakages from non-perforated floorboards.
Computer Room Air Condition Unit (CRAC)	$Q_{\mathrm{nom}} = 60\,\mathrm{kW}$ $U_{\mathrm{air}} = 3.6\,\mathrm{m}^3/\mathrm{s}$ $T_{\mathrm{air}} = 18\,°\mathrm{C}$	Fixed air flow.
Servers	42U Universal Rack $P_{\mathrm{low}} = 3\,\mathrm{kW}$ $P_{\mathrm{med}} = 5.5\,\mathrm{kW}$	Constant homogeneous heat generation from cabinets; cabinets completely filled.

1) COLD AISLE CONTAMINATION

This scenario investigates undesired temperature variations caused by the contamination of the cold aisle. Such contamination can be caused by the improper placement of servers, or back-flow resulting from excessive heated air unable to be promptly returned to the ceiling plenum and recirculated. The contamination of the cold aisle increases the temperature of ingested air into neighboring server racks, and diminishes cooling efficiency. A baseline simulation run on the data hall with an empty rack position reveals two sites of contamination in the central cold aisle and the formation of a hot spot (Figure 5a), with two server cabinets found to be in non-compliance with ASHRAE 2011 Class A1 standards (Figure 5b). This necessitates corrective action to prevent thermal overload on the affected cabinets.

A plausible solution to the observed undesired thermal scenario is to introduce perforated floorboards integrated with



(a) (b)

(c) (d) (e)

FIGURE 5. **(a) Resultant heat map of the baseline simulation with one empty cabinet position, showing two sites of contamination in the central cold aisle and one hot spot (circled), and (b) the corresponding ASHRAE temperature compliance map. (c) The addition of two VCDG floor grills to mitigate cold aisle contamination. (d) Resultant heat map of the post-mitigation simulation with the two cold aisle contamination sites eliminated (circled), and (e) the corresponding ASHRAE temperature compliance map with a non-compliant cabinet eliminated (circled). All heat maps and temperature compliance maps were captured at a height of ≈1 m from the raised hall floor.**

volume-control damper grates (VCDGs) to suppress hot air infiltration into the cold aisle. In this implementation, two perforated floorboards integrated with VCDG were placed at the flow paths of identified contamination sites (Figure 5c), with the floor grills and constituent dampers set to 50% opening areas. The corresponding simulation results (Figure 5d) shows a noticeable decrease in temperatures of approximately 13%, with contamination of the cold aisle largely eradicated and the elimination of one non-compliant server cabinet (Figure 5e).

2) INADEQUATE COOLING OF SEMI-OCCUPIED DATA HALL

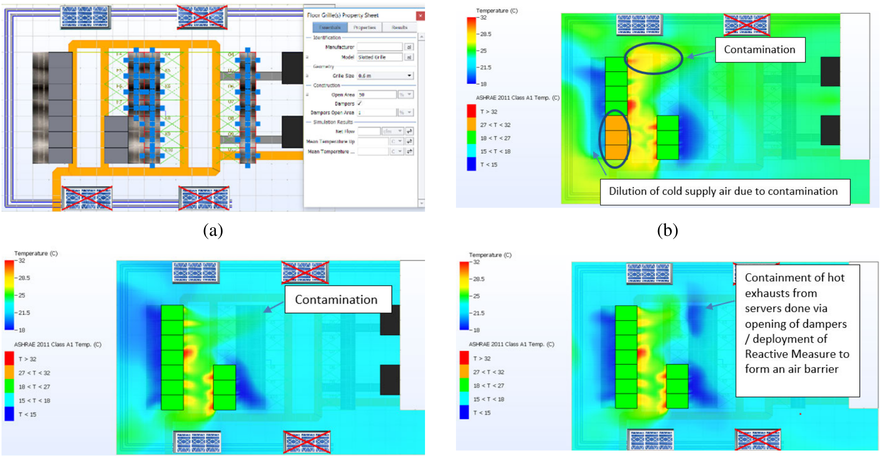
This second scenario concerns the inadequate cooling of a semi-occupied data center hall. Such occurrences are commonly the result of commissioning and decommissioning processes in data center environments, where the staggered deployment or removal of servers may lead to an underestimation of cooling requirements, as the contamination of cold and hot aisles are not easily taken into account.

As illustrated in the baseline simulation schematic (Figure 6a), only a single 60 kW CRAC unit was powered on to provide cooling to the servers, in emulation of typical real-world operating conditions where heat load in the room is estimated to be low and naïvely calculated to be sufficiently supportable by one CRAC unit. Perforated floorboards with integrated VCDGs are manually opened only at rack positions with live servers (low-density rated at 3 kW each) to ensure adequate volumetric air flow and prevent cool air wastage. The baseline simulation reveals significant contamination of the entire data hall, with the indoor environment averaging 25°C in conjunction to extensive dilution of the cold aisles (Figure 6b). Three servers were also found to be in non-compliance with ASHRAE 2011 Class A1 recommended standards, requiring corrective action to prevent thermal overload.

A plausible solution to the unfavourable thermal scenario seen in the baseline simulation is to activate an additional CRAC unit located across the hall to provide additional cooling. Corresponding simulation results (Figure 6c) shows the solution to be effective in eliminating the three instances of non-compliance, and in reducing dilution of the cold aisles. Indoor environment temperatures are also observed to be significantly cooler, averaging approximately 22°C, corresponding to a 14% improvement to the baseline. While the results of this solution is adequate, mild contamination can still be observed. This can be further improved through the opening of two VCDGs (Figure 6d) to introduce air flow and serve as an air barrier to contain hot server exhausts within the hot aisle.

3) COLD/HOT AISLE CONTAINMENT

This third scenario explores the utilization of hot and cold aisle containment methods [1], [43], [44] in reducing undesired temperature variations in a data hall. A baseline simulation comprising 5.5 kW medium-density servers, four CRAC units, an empty cabinet position, and an absence of containment measures was run. Note that in the previous two simulations, 3 kW low-density servers were modelled—this change is to demonstrate the effects that higher-density server racks, a trend that the data center industry in general is moving towards, have on airflow conditions. The results of the baseline simulation are illustrated in Figure 7a, revealing



Contamination

Dilution of cold supply air due to contamination

Contamination

Containment of hot exhausts from servers done via opening of dampers / deployment of Reactive Measure to form an air barrier

(a) (b) (c) (d)

FIGURE 6. **(a) Simulation schematic of the semi-occupied data hall. (b) Resultant heat map of the baseline simulation, showing inadequate cooling of three server cabinets and dilution of the cold aisles (circled). (c) Lower temperatures and reduced cold aisle contamination were observed with the activation of an additional CRAC unit. (d) Further improvement is observed with the utilization of VCDGs to serve as an air barrier for containment of hot exhaust at hot aisles. All heat maps and temperature compliance maps were captured at a height of ≈1 m from the raised hall floor.**

significant dilution of the cold aisles caused by excessive exhausted heat. Two servers were found to be in non-compliance with ASHRAE 2011 Class A1 standards, with one unit exceeding Class A1 allowable limits (Figure 7b), requiring immediate action.

Simulation results with containment measures implemented are presented in Figures 7c–f. The utilization of hot and cold aisle containment is evidently effective in reducing hot spots. The accelerating demands on computational power and their superior space efficiency has seen rapidly increasing adoption of higher-density server racks in modern times, and this demonstration suggests that cold or hot aisle containment measures, while not strictly essential with lower-density equipment, becomes important at high densities. For data centers already in operation, this presents a conundrum to operators—to invest in containment measures to boost cooling efficiency and shoulder the potential disruption in service and capital costs during transition, or to remain with the non-containment status quo.

IV. APPROACH 2: AN INTERNET OF THINGS (IOT) MANAGEMENT SYSTEM

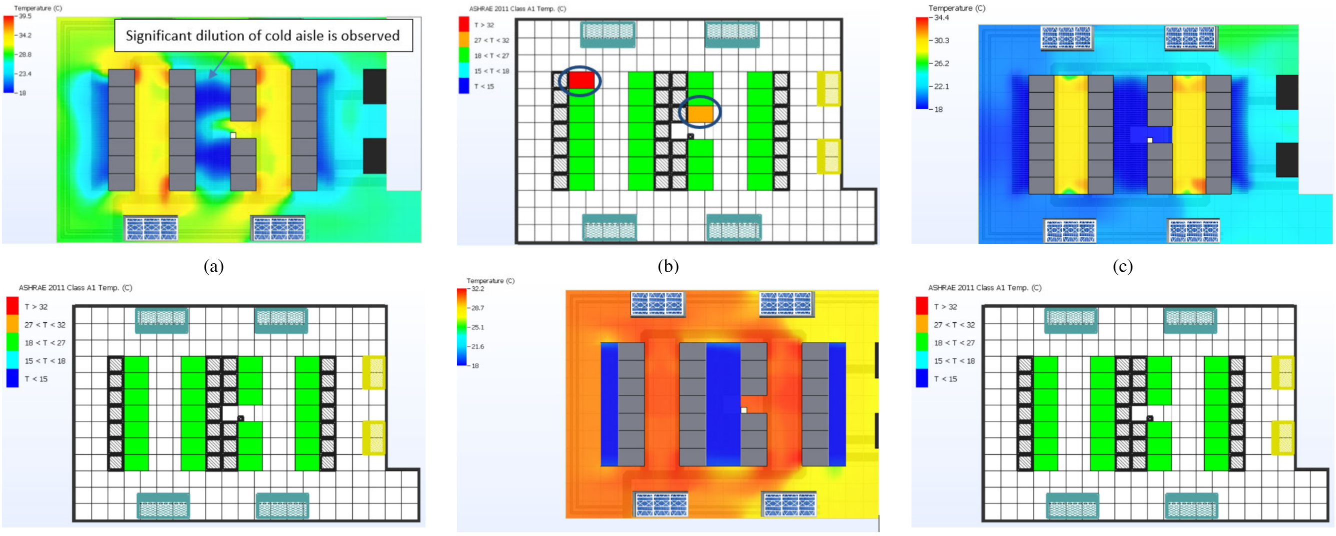
The simulation-based predictive approach presented enables better-informed decision-making in the planning and design of data centers, and also allows changes in configuration to be virtually evaluated before actual implementation in a live setting. While this reduces the likelihood of static design flaws in environmental control, transient effects due to fluctuations in server load and airflow, and unforeseen constraints or changes in infrastructure cannot be readily dealt with using such an *a priori* method. A closed feedback control is best suited for these.

We present an autonomous Internet of things (IoT) management system to serve this role, comprising a chain of IoT-enabled sensing and actuating devices to monitor and regulate environmental parameters such as temperature and humidity within a data center environment with manual overrides available as a safe guard measure. Specifically, we develop the IoT management system based on the simulation results of Section III-C1 and III-C2, in which the modulation of airflow via volume-control damper grates (VCDGs) at strategic locations were shown to be effective in suppressing cold and hot spots. This concept was inspired by automated fresh air dampers typically found in mechanical ventilation, heating ventilation and air conditioning systems. In its primary operational mode, the IoT management system autonomously controls the percentage area opening of VCDGs throughout the data center, thereby eliminating the need for manual tuning of floor grill dampers and enabling fast response times to changes in thermal conditions. This reduces operational man-hour requirements, and can lead to improved cooling efficiency and data center performance.

The implementation demonstrated in this study utilizes a network of IoT-enabled VCDGs connected to a central communication device, which processes and publishes collated data to an online IoT portal specially developed on Node-red and linked to Google Sheets for remote monitoring, control and logging purposes. In addition, as a redundant fail-safe measure, the system is also configurable to send regular data updates to operators via the Short Messaging System (SMS), or otherwise send alerts when environmental parameters are beyond acceptable ranges. Within this network, device-to-device communication is achieved through the Message Queuing Telemetry Transport (MQTT) protocol.

A. SYSTEM OVERVIEW

The construction of the proposed IoT management system requires the convergence of numerous technical processes and components that must work in unison to per-



(a) (b) (c)

(d) (e) (f)

FIGURE 7. **(a) Baseline simulation heat map on a data hall with 5.5 kW medium-density server cabinets and one empty cabinet position, and (b) the corresponding ASHRAE temperature compliance map showing two non-compliant cabinets and one beyond the allowable temperature range. (c) Post-mitigation simulation heat map utilizing hot aisle containment, and (d) corresponding temperature compliance map. (e) Post-mitigation simulation heat map utilizing cold aisle containment, and (d) corresponding temperature compliance map. All heat maps and temperature compliance maps were captured at a height of $\approx$1 m from the raised hall floor.**

TABLE 2. **Functional overview of the IoT-enabled VCDGs. *T* denotes air temperature sensed by sensing nodes in proximity to the VCDG device. Data publishing interval can be shortened or lengthened from 3600 sec as necessary.**

Damper Position	Temperature Limits	Within Limit	Damper Activation	Publishing Interval
Stationary	$20^\circ\text{C} < T < 25^\circ\text{C}$	Yes	No	3600 sec
Adjusted	$T \geq 20^\circ\text{C}$ (Open) $T \leq 20^\circ\text{C}$ (Close)	No	Yes	3600 sec

form its intended function reliably. A high-level conceptual overview of the intended operation of the IoT-enabled VCDGs is shown in Table 2, implemented programmatically in *Python*, and Figure 8 presents schematics for the VCDG feedback control system, and the network topology between the various IoT components. The following subsections (Sections IV-A1–IV-A4) details technical specifications of the key IoT devices.

1) CONTROLLER

A Raspberry Pi 3 Model B serves as the controller for each IoT node. The Raspberry Pi 3 platform was selected due to its relative affordability and the expansive range of connectivity options, inclusive of Bluetooth, tethered USB, and 802.11 Wi-Fi, available on-board without additional hardware. The Raspberry Pi can also be remotely connected to via secure shell (SSH) to access its operating system and general-purpose input and output (GPIO) ports, thereby easing maintenance in an IoT context. The computational power available on the Raspberry Pi also enables a larger degree of scalability before additional controllers need to be added.

Debian was installed as the operating system, with the necessary Ada-Fruit libraries, Paho-MQTT client and GAMMU SMS Daemon for purposes of temperature-humidity monitoring, device-to-device communication, and remote SMS capabilities.

2) ACTUATORS

VCDG damper actuation was achieved through a mounted 5V 28BYJ-48 stepper motor coupled to a ULN2003 motor controller. The motor was selected for its small size, and its history of use in the heating, ventilation and air conditioning (HVAC) industry, illustrating similarities in the intended mode of utilization. In our implementation, a half-step motor excitation sequence is utilized.

3) SENSORS

VCDG actuator modules are coupled to sensing nodes comprising a DHT11 temperature-humidity sensor, capable of a temperature detection range of 0–50°C, accurate to ±2°C, and a relative humidity detection range of 20–80%, accurate to ±5%, both of which completely encompasses the ranges stipulated by the ASHRAE guidelines and are well within the plausible parameters of a typical data center hall space. During each polling interval, sensor parameters and damper position status are transmitted via MQTT to a central IoT communications device, which collates and publishes the data to the IoT platform via Wi-Fi.

4) SMS MODEM

Notification of actuator operation statuses and environmental parameter data can be sent via SMS to data center operators.

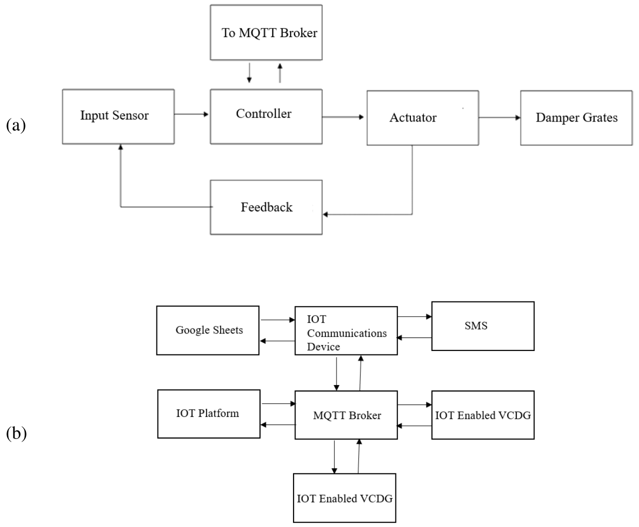


FIGURE 8. **(a) Flowchart illustrating the VCDG feedback control loop, and (b) a schematic illustrating the network connections between different IoT devices.**

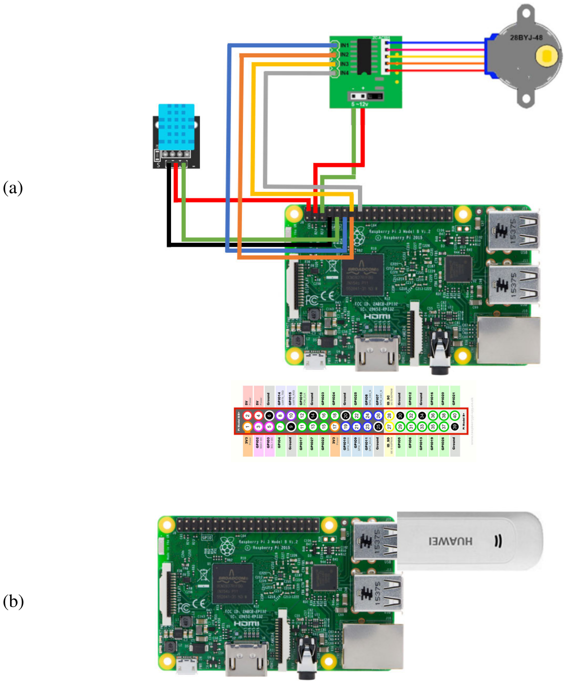


FIGURE 9. **Schematic and interconnections for (a) the IoT-enabled VCDG devices and (b) the central IoT communications device.**

This is especially useful if Internet coverage (via Wi-Fi) is lost and the IoT portal cannot be assessed, or when the operator is not at his/her workstation. A Huawei E1550 USB 3G dongle, necessary interfacing drivers and the GAMMU SMS Daemon was installed on the central communications device to enable this functionality.

B. PROTOTYPE

A practical engineering prototype of the proposed IoT management system was constructed for demonstration and evaluation purposes. Schematics of the IoT-enabled VCDGs and central communication node are shown in Figure 9, illustrating connections to the GPIO ports and peripherals on the controller. A proof-of-concept prototype of a VCDG device is shown in Figure 10, in both open and closed damper positions, and the IoT portal graphical interface is shown in Figure 11, illustrating data and status reporting features available to the operator.

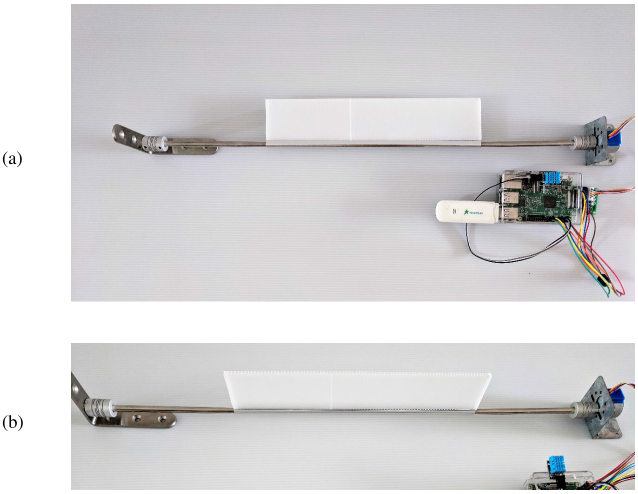
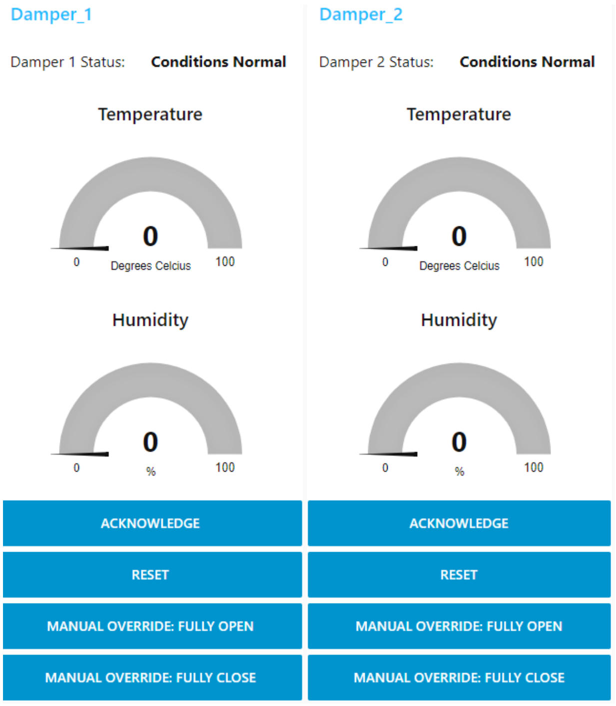


FIGURE 10. **Constructed VCDG damper prototype with controller board in (a) the closed position, and (b) the opened position. Several units actuate in synchrony to control airflow over the grille area.**



Damper_1

Damper 1 Status: **Conditions Normal**

Temperature

0

0 Degrees Celcius 100

Humidity

0

0 % 100

ACKNOWLEDGE

RESET

MANUAL OVERRIDE: FULLY OPEN

MANUAL OVERRIDE: FULLY CLOSE

Damper_2

Damper 2 Status: **Conditions Normal**

Temperature

0

0 Degrees Celcius 100

Humidity

0

0 % 100

ACKNOWLEDGE

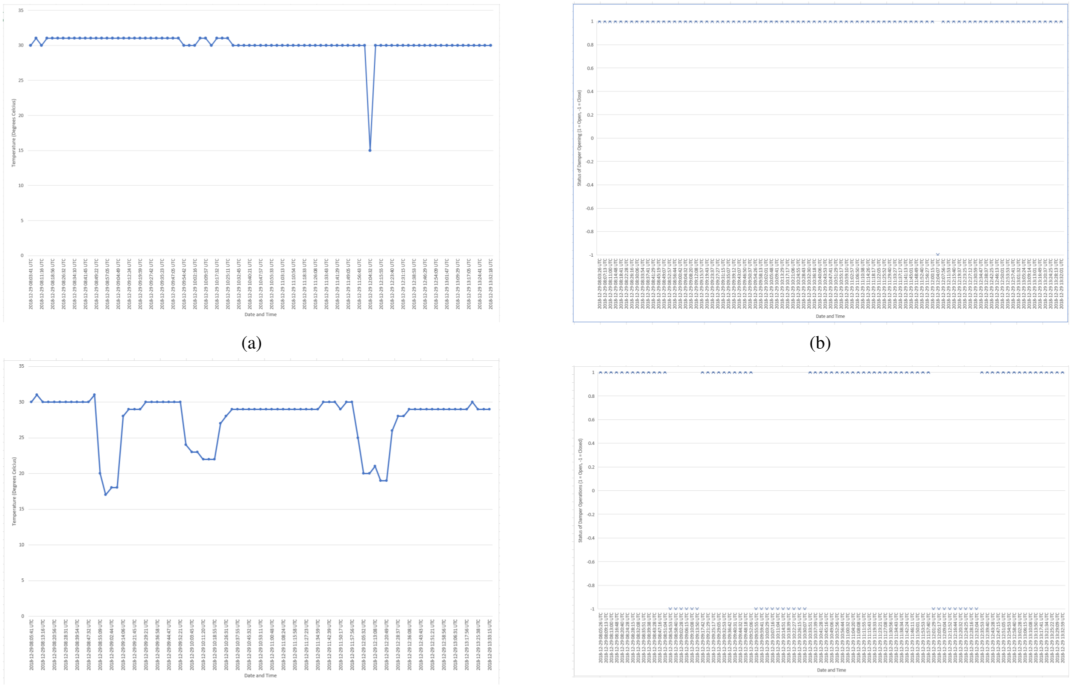
RESET

MANUAL OVERRIDE: FULLY OPEN

MANUAL OVERRIDE: FULLY CLOSE

FIGURE 11. **A screenshot of the developed IoT platform interface, showing collated sensed environmental data and damper status graphs.**

An experimental validation of the constructed prototype had been carried out. The operation of the VCDG dampers had been programmed simplistically, to fully open when sensed air temperature exceeds a set point of 29°C, and to fully close when sensed air temperature falls below a set point of 27°C. When air temperature is between these specified



(a) (b) (c) (d)

FIGURE 12. **(a) Temperature data collected over the validation period for IoT-enabled VCDG device #1, and (b) corresponding damper operation status over the same time period. (c) Temperature data for IoT-enabled VCDG device #2, and (d) corresponding damper operation status. A +1 damper state represents an opened position, and A-1 damper state represents a closed position.**

limits, the VCDG remains steady at its prior position. The system can be trivially modified to adjust the damper to an interpolated position between 0-100% effective airflow, say, via a linear mapping or a simulation-determined nonlinear map; this simplistic scheme was chosen for demonstrative clarity. Results of the validation are presented in Figure 12, illustrating the autonomous response of two IoT-enabled VCDGs towards changing data center thermal conditions. The results indicate the correct operation of the VCDGs as specified, and successful data transmission and logging functions of the IoT management system.

Through the integration of the various IoT components detailed in Section IV-A, real-time sensing and autonomous algorithm-based control of environmental parameters is achieved, remote monitoring capabilities are realized via the developed IoT platform and redundant SMS notification, and the enabled data logging options can be exploited for analysis and subsequent data-based improvements on data center cooling and computing infrastructure. The VCDG system can potentially interface with the CRAC units for optimal control of set point temperatures and air flow rates; and this motivates future work. In such integrated systems, additional redundancy measures ought to be taken, in consideration of operational needs and the possibility of human errors caused by increased system complexity.

V. CONCLUSION

In this paper, we have demonstrated a multi-pronged approach to enhance and optimize data center cooling, firstly through a simulation-based predictive method leveraging on computational fluid dynamics (CFD) and heat transport computational analyses, and secondly through a novel reactive Internet of things (IoT) feedback-controlled autonomous management system. The use of computational fluid dynamics simulations enables accurate predictions of data center operational conditions in a virtual setting, and can provide a better understanding of design flaws as well as the effects of configuration changes, before implementation in a live data hall. In the various archetypal scenarios considered (Section III-C), simulational analysis was demonstrated as an efficient tool both to identify potential thermal non-compliance and to evaluate plausible mitigation measures, across diverse circumstances of cold aisle contamination, inadequate cooling during commissioning and decommissioning processes, and hot-cold aisle containment for higher-density server equipment. Evidently, across the wide spectrum of possible problems in data hall environmental control, there exists no one-size-fits-all solution. Simulation tools such as that utilized here enable the rapid exploration of possible resolution strategies and alternative infrastructure configurations, thereby aiding in efficiently eliminating unfavourable operating conditions.

This *a priori* simulational branch of the approach is supplemented by a reactive IoT management system that autonomously monitors and regulates environmental parameters, enabling good robustness to transient fluctuations and unforeseen conditions within the data center. A feasible prototype of the IoT system has been demonstrated on affordable hardware, and its operational behaviour validated on rudimentary real-world tests. Used in conjunction, these predictive and reactive branches can enable effective cooling and environmental control in data centers under a wide range of scenarios. The development of these tools is especially important amidst the current movement towards higher-density data centers that must simultaneously satisfy stringent computational performance and energy sustainability requirements [45]–[48]. In our work, we have studied the practical operational situations in data centre processing environment and the impacts it has on cooling. This is reflected in the first approach (predictive) where various conditions are simulated and assessed for its effectiveness. The second approach (reactive) further contributes to the research via the development of an IoT-based solution which enhances operator monitoring, work flow and cooling deficiencies through automation of volume control damper grates that adjusts its opening according to hall conditions. This is an advancement over the current industrial data centre norms that are primarily non-adjustable or manually adjustable. In optimizing thermal control, not only is hardware performance, reliability and lifespan enhanced, but energy expenditure can also be reduced as well.

ACKNOWLEDGMENT

The authors wish to thank the handling editor and two anonymous reviewers for their constructive comments.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] B. Fakhim, M. Behnia, S. Armfield, and N. Srinarayana, "Cooling solutions in an operational data centre: A case study," *Appl. Therm. Eng.*, vol. 31, pp. 2279–2291, Oct. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1359431111001578

[2] J. Shuja, K. Bilal, S. A. Madani, M. Othman, R. Ranjan, P. Balaji, and S. U. Khan, "Survey of techniques and architectures for designing energy-efficient data centers," *IEEE Syst. J.*, vol. 10, no. 2, pp. 507–519, Jun. 2016.

[3] Y. Joshi and P. Kumar, "Introduction to data center energy flow and thermal management," in *Energy Efficient Thermal Management of Data Centers*, Y. Joshi and P. Kumar, Eds. Boston, MA, USA: Springer, 2012, pp. 1–38. doi: 10.1007/978-1-4419-7124-1_1.

[4] H. Geng, *Data Center Handbook*. Hoboken, NJ, USA: Wiley, 2014.

[5] V. Josyula, M. Orr, and G. Page, *Cloud Computing: Automating the Virtualized Data Center*. Cisco, San Jose, CA, USA,, 2011.

[6] C. J. M. Lasance, "Thermally driven reliability issues in microelectronic systems: Status-quo and challenges," *Microelectron. Rel.*, vol. 43, no. 12, pp. 1969–1974, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026271403001835

[7] J. D. Parry, J. Rantala, and C. J. M. Lasance, "Enhanced electronic system reliability—Challenges for temperature prediction," *IEEE Trans. Compon. Packag. Technol.*, vol. 25, no. 4, pp. 533–538, Dec. 2002.

[8] P. Lall, M. Pecht, and E. B. Hakim, *Influence of Temperature on Microelectronics and System Reliability: A Physics of Failure Approach*. Boca Raton, FL, USA: CRC Press, 1997.

[9] C.-C. Teng, Y.-K. Cheng, E. Rosenbaum, and S.-M. Kang, "ITEM: A temperature-dependent electromigration reliability diagnosis tool," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 16, no. 8, pp. 882–893, Aug. 1997.

[10] M. G. Pecht and F. R. Nash, "Predicting the reliability of electronic equipment," *Proc. IEEE*, vol. 82, no. 7, pp. 992–1004, Jul. 1994.

[11] R. Rao, S. Vrudhula, and C. Chakrabarti, "Throughput of multi-core processors under thermal constraints," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, Aug. 2007, pp. 201–206.

[12] R. Rao and S. Vrudhula, "Performance optimal processor throttling under thermal constraints," in *Proc. Int. Conf. Compil. Archit. Synth. Embedded Syst.*, New York, NY, USA, 2007, pp. 257–266. doi: 10.1145/1289881.1289925.

[13] D. Brooks and M. Martonosi, "Dynamic thermal management for high-performance microprocessors," in *Proc. Int. Symp. High-Perform. Comput. Archit.*, Jan. 2001, pp. 171–182.

[14] M. K. Patterson, "The effect of data center temperature on energy efficiency," in *Proc. 11th Intersoc. Conf. Therm. Thermomech. Phenomena Electron. Syst.*, May 2008, pp. 1167–1174.

[15] N. El-Sayed, I. A. Stefanovici, G. Amvrosiadis, A. A. Hwang, and B. Schroeder, "Temperature management in data centers: Why some (might) like it hot," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 1, pp. 163–174, 2012. doi: 10.1145/2318857.2254778.

[16] H. S. Gunawi, M. Hao, R. O. Suminto, A. Laksono, A. D. Satria, J. Adityatama, and K. J. Eliazar, "Why does the cloud stop computing?: Lessons from hundreds of service outages," in *Proc. 7th Symp. Cloud Comput.*, New York, NY, USA, 2016, pp. 1–16. doi: 10.1145/2987550.2987583.

[17] S. Pertet and P. Narasimhan, "Causes of failure in Web applications," Pittsburgh, PA, USA, Carnegie Mellon Univ., Tech. Rep., CMU-PDL-05-109, 2005.

[18] J. Lienig and M. Thiele, *Fundamentals of Electromigration-Aware Integrated Circuit Design*. Cham, Switzerland: Springer, 2018.

[19] J. R. Black, "Electromigration—A brief survey and some recent results," *IEEE Trans. Electron Devices*, vol. ED-16, no. 4, pp. 338–347, Apr. 1969.

[20] E. Pinheiro, W.-D. Weber, and L. A. Barroso, "Failure trends in a large disk drive population," in *Proc. FAST*, 2007, vol. 7, no. 1, pp. 17–23.

[21] J. D. Moore, J. S. Chase, P. Ranganathan, and R. K. Sharma, "Making scheduling 'Cool' temperature-aware workload placement in data centers," in *Proc. USENIX Annu. Tech. Conf., Gen. Track*, 2005, pp. 61–75.

[22] M. Kummert, W. Dempster, and K. McLean, "Transient thermal analysis of a data centre cooling system under fault conditions," in *Proc. 11th Int. Building Perform. Simulation Assoc. Conf. Exhib. Building Simulation*, 2009, pp. 1302–1305.

[23] R. Romadhon, M. Ali, A. M. Mahdzir, and Y. A. Abakr, "Optimization of cooling systems in data centre by computational fluid dynamics model and simulation," in *Proc. Innov. Technol. Intell. Syst. Ind. Appl.*, Jul. 2009, pp. 322–327.

[24] K. C. Karki, A. Radmehr, and S. V. Patankar, "Use of computational fluid dynamics for calculating flow rates through perforated tiles in raised-floor data centers," *HVAC&R Res.*, vol. 9, no. 2, pp. 153–166, 2003.

[25] J. Cho, J. Yang, and W. Park, "Evaluation of air distribution system's airflow performance for cooling energy savings in high-density data centers," *Energy Buildings*, vol. 68, pp. 270–279, Jan. 2014.

[26] A. Almoli, A. Thompson, N. Kapur, J. Summers, H. Thompson, and G. Hannah, "Computational fluid dynamic investigation of liquid rack cooling in data centres," *Appl. Energy*, vol. 89, no. 1, pp. 150–155, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261911001012

[27] C. B. Bash, C. D. Patel, and R. K. Sharma, "Dynamic thermal management of air cooled data centers," in *Proc. Therm. Thermomech. 10th Intersoc. Conf. Phenomena Electron. Syst.*, May/Jun. 2006, p. 8.

[28] R. K. Sharma, C. E. Bash, C. D. Patel, R. J. Friedrich, and J. S. Chase, "Balance of power: Dynamic thermal management for Internet data centers," *IEEE Internet Comput.*, vol. 9, no. 1, pp. 42–49, Jan. 2005.

[29] Q. Liu, Y. Ma, M. Alhussein, Y. Zhang, and L. Peng, "Green data center with IoT sensing and cloud-assisted smart temperature control system," *Comput. Netw.*, vol. 101, pp. 104–112, Jun. 2016.

[30] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Bus. Horizons*, vol. 58, no. 4, pp. 431–440, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0007681315000373

[31] "Thermal guidelines for data processing environments—Expanded Data center classes and usage guidance," ASHRAE Tech. Committee, Atlanta, GA, USA, 2011.

[32] (2017). *About ASHRAE 9.9 Mission Critical Facilities, Data Centers, Technology Spaces Electronic Equipment*. Accessed: Jan. 8, 2019. [Online]. Available: http://tc0909.ashraetcs.org/about.php

[33] P. Lin, "How to fix hot spots in the data center," Schneider Electr. White Paper Library, Schneider Electr. Data Center Sci., Center, West Kingston, RI, USA, Tech. Rep. SPD_VAVR-9GNNGR_EN, 2014.

[34] T. Evans, "Humidification strategies for data centers and network rooms," APC by Schneider Electr., West Kingston, RI, USA, Tech. Rep. SPD_NRAN-5TV85S_EN, 2004.

[35] R. A. Steinbrecher and R. Schmidt, "Data center environments: ASHRAE's evolving thermal guidelines," *ASHRAE J.*, vol. 53, no. 12, pp. 42–49, 2011.

[36] W. Lintner, B. Tschudi, and O. VanGeet, "Best practices guide for energy-efficient data center design," U.S. Dept. Energy, Washington, DC, USA, Tech. Rep. NREL/BR-7A40-47201, 2011.

[37] (2018). *Fujitsu Launches Liquid Immersion Cooling System That Lowers Total Cost of Ownership*. Accessed: Aug. 16, 2019. [Online]. Available: https://www.fujitsu.com/global/about/resources/news/press-releases/2018/0906-01.html

[38] InfoComm Development Authority of Singapore, "Green data centre technology roadmap," Strategy Group Prime Minister's Office, Singapore, 2014.

[39] (2016). *Singapore Trials World'S First Tropical Data Centre For a Smart Nation*. Accessed: Aug. 16, 2019. [Online]. Available: https://www2.imda.gov.sg/news-and-events/Media-Room/archived/ida/Media-Releases/2016/singapore-trials-worlds-first-tropical-data-centre-for-a-smart-nation

[40] (2018). *SigmaRoom*. [Online]. Available: https://www.futurefacilities.com/products/6sigmaroom/

[41] (2016). *How accurate is 6SigmaDCX*. [Online]. Available: https://www.futurefacilities.com/resources/whitepapers/how-accurate-is-6sigmadcx/

[42] K. Heslin, "Data center cooling: Crac/crah redundancy, capacity, and selection metrics," *Uptime Inst. J.*, vol. 3, 2014, Accessed: Jan. 08, 2019. [Online]. Available: https://journal.uptimeinstitute.com/data-center-cooling-redundancy-capacity-selection-metrics/

[43] J. Niemann, K. Brown, and V. Avelar, "Impact of hot and cold aisle containment on data center temperature and efficiency," Schneider Electr. Data Center Sci. Center, West Kingston, RI, USA, White Paper SPD_DBOY-7EDLE8_EN, 2011.

[44] R. Schmidt, A. Vallury, and M. Iyengar, "Energy savings through hot and cold aisle containment configurations for air cooled servers in data centers," in *Proc. ASME Pacific Rim Tech. Conf. Exhib. Packag. Integr. Electron. Photonic Syst.*, 2011, pp. 611–616.

[45] C. D. Patel, R. Sharma, C. E. Bash, and A. Beitelmal, "Thermal considerations in cooling large scale high compute density data centers," in *Proc. 8th Intersoc. Conf. Therm. Thermomech. Phenomena Electron. Syst.*, May/Jun. 2002, pp. 767–776.

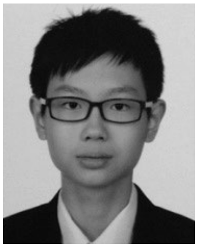
[46] R. R. Schmidt, E. E. Cruz, and M. Iyengar, "Challenges of data center thermal management," *IBM J. Res. Develop.*, vol. 49, no. 4.5, pp. 709–723, Jul. 2005.

[47] M. Marwah, P. Maciel, A. Shah, R. Sharma, T. Christian, V. Almeida, and C. Araújo, E. Souza, G. Callou, B. Silva, S. Galdino, and J. Pires, "Quantifying the sustainability impact of data center availability," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 4, pp. 64–68, 2010.

[48] P. Mahadevan, S. Banerjee, P. Sharma, A. Shah, and P. Ranganathan, "On energy efficiency for enterprise and data center networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 94–100, Aug. 2011.

KANG HAO CHEONG (M'18) received the B.Sc. degree (Hons.) from the Department of Mathematics and University Scholars Programme, National University of Singapore (NUS), in 2007, the Ph.D. degree from the Department of Electrical and Computer Engineering, NUS, in 2015, and the postgraduate Diploma degree in education from the National Institute of Education, Singapore. From 2016 to 2018, he was an Assistant Professor with the Engineering Cluster, Singapore Institute of Technology. He is currently an Assistant Professor with the Science and Math Cluster, Singapore University of Technology and Design (SUTD).

KENNETH JIAN WEI TANG received the B.Eng. degree (Hons.) in sustainable infrastructure engineering (building services) and the M.EngTech. degree from the Singapore Institute of Technology (SIT), in 2018 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Science and Math Cluster, Singapore University of Technology and Design.



JIN MING KOH received the NUS High School Diploma (High Distinction), in 2016. Since 2017, he has been undertaking research projects offered by K. H. Cheong. He is currently with the California Institute of Technology.

SIMON CHING MAN YU received the B.Eng. degree (Hons.) ACGI and Ph.D. DIC degree from the Department of Mechanical Engineering, Imperial College London, London, in 1987 and 1991, respectively. He joined Nanyang Technological University, Singapore, immediately after his graduation as a Lecturer. He became a Senior Lecturer, in 1996 and an Associate Professor, in 2000. He was also very involved in the university administration, since 2000: Principal Staff Officer (President's Office), from 2000 to 2003, University Council Member, from 2001 to 2003, the Vice Dean of the Admission Office, from 2003 to 2006 and the Head of the Division of Aerospace Engineering, School of Mechanical and Aerospace Engineering, from 2008 to 2013. He moved over to the Singapore Institute of Technology as Professor and the Programme Director, in 2013, to establish one of the first engineering degree programmes offered solely by the university. He is currently the Head of the Interdisciplinary Division of Aeronautical and Aviation Engineering (AAE), The Hong Kong Polytechnic University. He has published more than 200 research articles in archive journals and conferences. He managed to secure more than SGD 50 million external grant during his tenure in Nanyang Technological University, especially during the period as the Head of the Division of Aerospace Engineering.

U. RAJENDRA ACHARYA received the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India, and the D.Eng. degree from Chiba University, Japan. He is currently a Senior Faculty Member with Ngee Ann Polytechnic, Singapore. He is also an Adjunct Professor with Taylor's University, Malaysia, also an Adjunct Faculty with the Singapore Institute of Technology, University of Glasgow, Singapore, and also an Associate faculty with the Singapore University of Social Sciences, Singapore. He has published more than 400 articles, in refereed international SCI-IF journals (345), international conference proceedings (42), and books (17) with more than 24,000 citations in Google Scholar (with h-index of 82). He ranked in the top 1% of the Highly Cited Researchers for the last four consecutive years (2016, 2017, 2018, and 2019) in Computer Science according to the Essential Science Indicators of Thomson. He has worked on various funded projects, with grants worth more than SGD 2 million. He has three patents and in the editorial board of many journals. He has served as guest editor for many journals. His current research interests include biomedical signal processing, biomedical imaging, data mining, visualization and biophysics for better healthcare design, delivery, and therapy.

NENG-GANG XIE received the Ph.D. degree in engineering from Hehai University, in 1999. He is currently a Professor and a Ph.D. Supervisor with the School of Management Science and Engineering, Anhui University of Technology.

• • •