

# Tropical data centre proof-of-concept

Le, Duc Van; Liu, Yingbo; Wang, Rongrong; Tan, Rui

2019

Le, D. V., Liu, Y., Wang, R., & Tan, R. (2020). Tropical data centre proof-of-concept.

<https://hdl.handle.net/10356/137780>

---

© 2020 Tropical data centre proof-of-concept. All rights reserved.

*Downloaded on 13 Mar 2024 16:08:36 SGT*

**Technical Report**

# **Tropical Data Centre Proof-of-Concept**

Duc Van Le\*  
Yingbo Liu\*  
Rongrong Wang  
Rui Tan<sup>†</sup>

*School of Computer Science and Engineering  
Nanyang Technological University  
Singapore*

\*The first two authors made equal contributions.

<sup>†</sup>Corresponding author.

Version 1.1: December 14th 2019

## Changelog

**Version 1.0** November 15th 2019. Initial version.

**Version 1.1** December 14th 2019. Add a section on future research. Add page breaks between sections.

## CONTENTS

Contents	2
1 Introduction	5
2 Testbed Design and Benchmark Tests	7
2.1 Design and Construction of Testbed	7
2.1.1 Design of Testbed	7
2.1.2 Construction of Testbed	9
2.1.3 Configuration of Testbed	11
2.2 Design of Experiments on the Testbed	12
3 Sensor Data Analytics	15
3.1 Benchmark Tests and Collected Data	15
3.2 Summary of Test Results	16
3.3 CPU Performance	17
3.3.1 Settings	17
3.3.2 Analysis	17
3.4 HDD Performance	20
3.4.1 Settings	20
3.4.2 Analysis	21
3.5 Main Memory Performance	21
3.5.1 Settings	21
3.5.2 Analysis	23
3.6 Node Performance	23
3.7 Energy profiles	23
4 Thermal and Power Modeling	26
4.1 Psychrometric Modeling	26
4.2 Power Consumption Models	28
4.3 Server Outlet Temperature Model	29
4.3.1 Modeling	29
4.3.2 Parameters identification	32
4.3.3 Validation of $M_A$ on Testbed	34
4.3.4 Validation of $M_B$ on Testbed	38
4.3.5 Summary and Discussion	40
5 Air-side Free-Cooled DC Control and Cost Savings	42
5.1 Ventilation Control and Energy Saving	42
5.1.1 Fixed ventilation	42
5.1.2 Adaptive ventilation	42
5.1.3 Energy savings	43
5.2 Advanced Control by Deep Reinforcement Learning	44
5.2.1 Approach Overview	44
5.2.2 MDP Formulation	45
5.2.3 Offline Training of DRL Agent	46
5.2.4 Sensor Requirement	47
5.2.5 DRL Agent Training and Execution	47
6 IT Equipment Failures and Mitigation Approaches	50
6.1 Summary of Failures	50
6.1.1 Phase 1: From July 2018 to February 2019 (8 Months)	50
6.1.2 Phase 2: From March 2019 to October 2019 (8 Months)	50

6.2	Investigation on Failure Reason	51
6.2.1	Our Own Investigation on Server Room Condition	52
6.2.2	Vendor's Fault Analysis	54
6.2.3	Summary and Discussions	55
6.3	Recommendations for Singapore	55
7	Discussions and Conclusion	59
8	Future Research	61
	References	62

## **ACKNOWLEDGEMENT**

The results in this report were generated in a collaborative project between Info-Communications Media Development Authority (IMDA) and Nanyang Technological University (NTU) of Singapore. The project, which spanned from December 15th 2017 to December 14th 2019, was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Green Data Centre Programme. The authors of this technical report acknowledge Dr. Lek Heng Ngoh (IMDA), Yew-Wah Wong (NTU), and Dr. Yonggang Wen (NTU) for their valuable inputs. The authors acknowledge the contributions from the following partners (names not listed in order): Dell EMC, Epsilon, ERS Industries, Fujitsu, Future Facilities, Hewlett Packard Enterprise, Huawei, IX Technologies, Keppel Data Centres, Singtel, Intel, Micron, National University of Singapore, The Green Grid, and Uptime Institute.

## 1 INTRODUCTION

Singapore is a data center (DC) hub in Southeast Asia [1]. However, Singapore’s year-round high temperatures and humidity levels introduce significant challenges for the local DC operators in improving the energy efficiency of their infrastructures. As Singapore’s DCs spend more energy in cooling, their average power usage effectiveness (PUE), which is 2.07 [15], is higher than the global average of 1.7 [9]. In the United States, the DC sector accounted for 1.8% of the country’s total electricity consumption in 2014 [16]. In Singapore, this percentage is up to 7% [1]. Thus, technologies that can improve DC energy efficiency in the tropics will further enhance Singapore’s attractiveness as a regional data center hub. They are also important to Singapore’s energy sustainability and commitment to Paris Agreement.

Air-side free cooling that utilizes outside cold air to cool the information technology (IT) equipment has been increasingly used to improve the energy efficiency of DCs [7]. However, air-side free cooling in the tropics has been long thought infeasible from the intuition that the high temperature and relative humidity (RH) of the air supplied to the servers will undermine their performance and reliability. On the other hand, the American Society of Heating, Refrigeration and Air-Conditioning Engineers (ASHRAE) has been working for years on expanding its suggested allowable temperature and RH ranges for IT equipment. For instance, the servers compliant with ASHRAE’s Class A3 can operate continuously and reliably when the temperature and RH of the supply air are up to 40°C and 90%. This sheds light on the possibility of air-side free-cooled DCs in Singapore, since the record temperature in Singapore is 37°C only and the ambient RH is in general lower than 90%.

To investigate the feasibility of air-side free cooling in Singapore, together with multiple partners in DC industry and research, we designed, constructed, and experimented with an air-side free-cooled DC testbed consisting of three server rooms located in two local DC operators’ premises. The testbed hosts 12 server racks with 60 kW total power rating. This technical report holistically introduces this project, presents the measurement results, discusses the experiences and learned lessons obtained from the project. Specifically, the main results of this project are summarized as follows.

- We perform 18-month microbenchmark tests to investigate the performance of the servers under different partially controlled environmental conditions. The measurement results show that the servers can operate without computing performance degradation when the supply air temperature is up to 37°C.
- The total energy consumed by cooling and IT equipment is decreased by about 45% when the supply air temperature setpoint is increased from 25°C to 33°C. Moreover, the adaptive ventilation using fans only to cool the IT equipment can reduce the PUE down to 1.05.
- We build and evaluate various analytical and neural network models based on meta information and real sensor data collected from the testbed to characterize the dynamic state of the temperature, RH, and power consumption of the IT equipment and the testbed’s supporting infrastructure for cooling, ventilation, and heating.
- We develop a deep reinforcement learning (DRL)-based advanced control algorithm that can maintain the supply air temperature and RH below respective specified thresholds for the sake of IT hardware reliability. Extensive simulations are performed to evaluate the performance of the control algorithm.
- We detail our investigation on the reasons of the server failures occurred during the 18 months’ tests. We recommend several mitigation approaches to the designs and operations of similar air-side free-cooled DCs in Singapore’s tropical environments.

The rest of this technical report is organized as follows. Section 2 presents the design of the testbed and the benchmark tests on the IT equipment. Section 3 presents the sensor data analytics. Section 4 presents the mathematical modeling on the power consumption and thermal aspects of the testbed. Section 5 presents various operating procedures for air-side free-cooled DCs and the related cost saving quantification based on the successful IT equipment tests on our testbed. Section 6 presents the IT equipment failures we encountered on the testbed during the project and proposes mitigation approaches. Section 7 discusses several issues and provides the conclusion remarks. Section 8 discusses future research.

## 2 TESTBED DESIGN AND BENCHMARK TESTS

### 2.1 Design and Construction of Testbed

2.1.1 *Design of Testbed.* We design the testbed with three objectives.

- (1) On the testbed, we should be able to maintain the condition of the air supplied to the IT equipment at a certain setpoint for a period of time (e.g., several days). The condition includes three aspects that are often considered important for IT equipment performance and reliability, i.e., temperature, RH, and air volume flow rate. The setpoint can be adjusted within a wide range, such that we can evaluate the performance of the IT equipment under various conditions. In other words, we can run the testbed in a *controlled* mode. However, we later found that RH control in a wide range is difficult, which will be discussed shortly.
- (2) We can run the testbed in an *uncontrolled* mode, in that we just use the outside air without adjusting its condition to take away the heat generated by the IT equipment. We aim to run the testbed in this uncontrolled mode for an extended period of time to understand the direct impact of the outside air on the IT equipment and the achievable energy saving.
- (3) The testbed should include a standard server room with well controlled conditions to generate the baseline results.

To meet the above three objectives, we design a testbed consisting of three server rooms, denoted by Room-A, Room-B, and Room-C. Room-A and Room-B are two side-by-side purposely built server rooms to support the aforementioned controlled and uncontrolled experiments. The side-by-side arrangement makes sure that they will inhale outside air with the same condition, enabling comparative experiments. We planed to build these two server rooms in the premise of a commercial colocation DC operator that is referred to as Operator-A. As such, we may leverage the domain expertise of Operator-A in facility management, 24/7 monitoring, security assurance, emergency response, and etc. Room-C is a standard server room operated by another commercial colocation DC operator that is referred to as Operator-B.

The original testbed design objectives include RH control capability. However, from the discussions with facility suppliers and our study, we found that for Room-A and Room-B, implementing RH setpoints in a wide range in Singapore’s tropical condition is costly and technically challenging. First, as Room-A and Room-B will continuously inhale outside air, from our industrial partner with DC facility expertise, the commercially available dehumidifier and humidifier cannot sustain the RH and air volume flow rate setpoints specified in the experiment plan (cf. Section 2.2). Note that

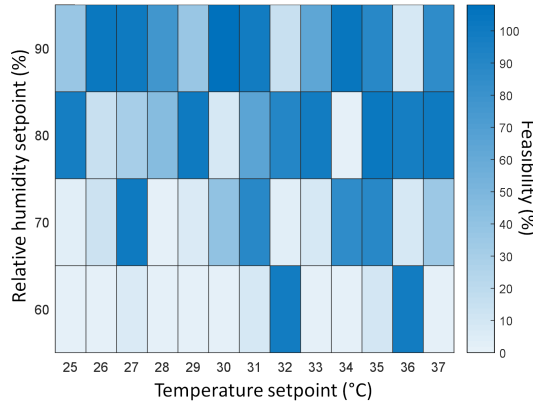


Fig. 1. Feasibility of temperature/RH setpoints during Jul and Aug 2018 in Singapore.

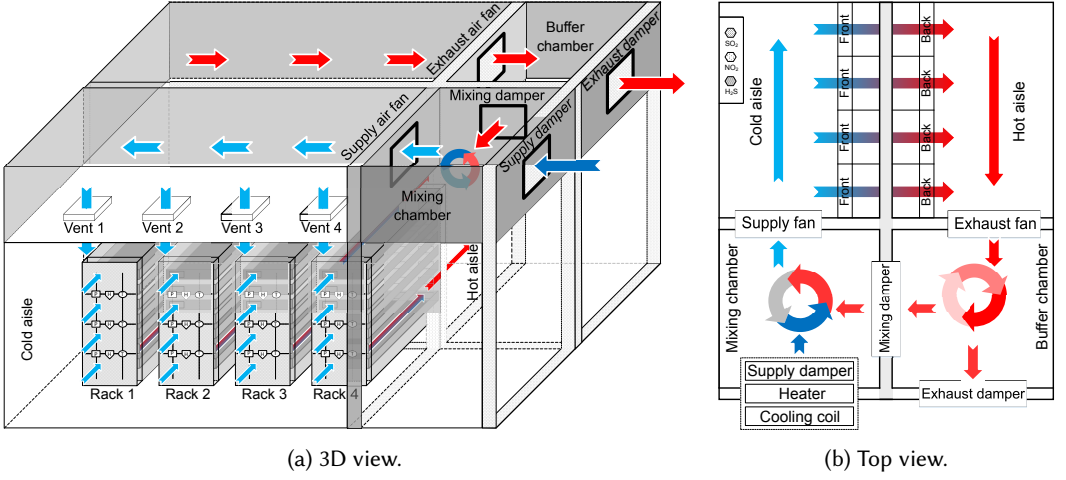


Fig. 2. Design of Room-A/B. Room-B does not have heater. Arrows represent the air flows.

typical data centers often have enclosed environment, in that the air is circulated within the data center building. As they inhale a limited amount of air from the outside, they have low dehumidification demands. Second, we have also studied a possible energy-efficient *cooling-then-mixing* dehumidification approach. Specifically, it uses a cooling coil to condense and remove the water vapor contained in the air entering the server room and then mixes the dried cold air with a controlled portion of the hot air generated by the IT equipment to maintain the temperature of the air supplied to the IT equipment at the setpoint. However, for a total IT load of 20 kW in a server room, our simulation studies show that the ability of this dehumidification approach in maintaining the temperature and RH setpoints highly depends on the temperature and RH of the outside air. The gray scale in Fig. 1 shows the percentage of time in July and August 2018 in Singapore, during which the corresponding temperature and RH setpoints on the  $x$ - and  $y$ -axis, respectively, can be maintained by the cooling-then-mixing approach. We can see that it is difficult to maintain low temperature and RH setpoints simultaneously for long periods of time.

Given the challenges in controlling RH in a wide range, we focus on maintaining the temperature and air volume flow rate setpoints in the design of Room-A and Room-B. Our design is as follows. Each of Room-A and Room-B is equipped with a cooling coil and multiple fans to move the air through the room. Fig. 2 shows the design of Room-A. Figs. 2a and 2b show the 3D and top views of Room-A. The room has two layers, where each layer is divided into four chambers. The outside air is continuously inhaled into the mixing chamber on the top layer. A cooling coil and an air heater are installed in the mixing chamber to process the outside air before supplying it to the servers. Two fans (i.e., supply air fan and exhaust fan) are installed on the top layer to move air. Moreover, there are three dampers (i.e., supply damper, exhaust damper, and mixing damper) as shown in Fig. 2. By setting their openness, we can control the air flow paths. After the supply air fan, the air enters a chamber and then goes down to the cold aisle chamber on the bottom layer through four vents. This design improves the evenness of the cold air volumes passing through the vents. Four 42U server racks are installed on the bottom layer, sitting between the cold aisle and hot aisle chambers. Our design well separates the cold air supplied to the servers and the hot air generated by them. This enables the precise control of the condition of the air supplied to the servers. The hot air from the servers goes to a chamber across the bottom and top layers

and is then moved by the exhaust air fan into a buffer chamber. Depending on the settings of the three dampers, the hot air is exhausted and/or recirculated to the mixing chamber. By setting the openness of the three dampers, we can control the percentage of the hot air generated by the IT equipment that will be mixed with the cold, relatively humid outside air to form warm, relatively dry air for the IT equipment. This design gives a certain level of RH control capability that can be used to reduce the negative impact of airborne contaminants on the reliability of the IT equipment. The details of this RH control will be presented in Section 5.2.

The design of Room-B is almost same as that of Room-A. The only difference between Room-B and Room-A is that, Room-B does not have a heater. This reduces the equipment cost and does not impede our experiments, because we can assign the controlled experiments with high temperature setpoints to Room-A. After the designs of Room-A/B are generated, we contract a third-party company to build a computational fluid dynamics (CFD) model based on our designs and perform extensive simulations to check whether the thermal properties of the two server rooms meet our requirements. Note that after the testbed is commissioned, the CFD model is improved by another third-party company based on the data traces generated by the testbed to achieve a root mean square error (RMSE) of about 1.2°C in predicting temperatures in the server rooms.

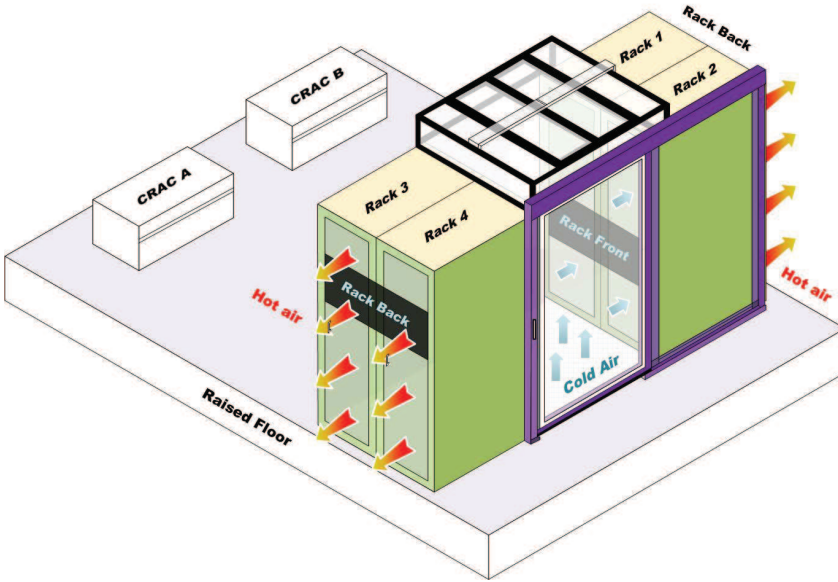


Fig. 3. Design of Room-C and cold air containment.

Room-C is a standard private vault in the premise of Operator-B. It follows the typical raised floor design and has two computer room air conditioning (CRAC) units that duty-cycle to provide cooling. We purposely improve its energy efficiency to make it an optimistic baseline by adding a cold air containment design as illustrated in Fig. 3. The figure also illustrates the layout of the four IT racks and the air flows.

**2.1.2 Construction of Testbed.** The construction of Room-A/B undertaken by a contractor took about four months. Fig. 4(a) shows the two side-by-side storage rooms located within the premise of Operator-A that were later retrofitted into Room-A and Room-B. Figs. 4(b) and (c) show the exterior of Room-A and Room-B during and after the construction, respectively. As seen in Fig. 4(c),

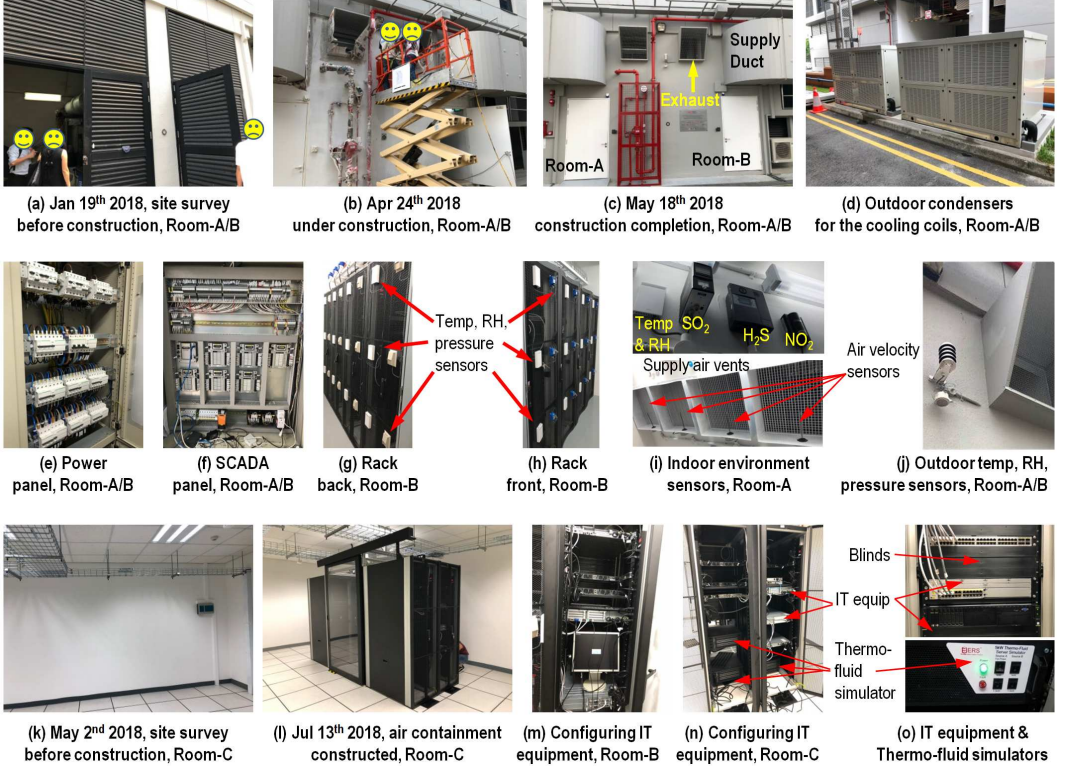


Fig. 4. Construction and configuration of Room-A, Room-B, and Room-C.

two supply air ducts were constructed such that there is sufficient space separation between the air inhaled and exhausted by Room-A/B. Air filters of Class MERV-6 were installed in the air ducts to prevent PM10 and larger particles from entering the server rooms. The red pipelines shown in Fig. 4(c) belong to a fire protection system. Note that, as Room-A and Room-B would experience high temperatures at their hot aisles, the testbed must have a fire protection system with 24/7 monitoring. Fig. 4(d) shows the outdoor condensers for the cooling coils installed in Room-A/B. The distance from these condensers to Room-A/B is about 30 meters to reduce the heat recirculation from the condensers to the two rooms. Fig. 4(e) and (f) show the power and Supervisory Control and Data Acquisition (SCADA) panels for Room-A/B. Each branch in the power panel has a smart meter for branch-level monitoring. All sensors and actuators deployed in Room-A/B are SCADA slaves communicating with a SCADA master using Modbus TCP protocol. The SCADA master runs on a workstation computer that is located within a conditioned room. Fig. 4(g)-(j) show various sensors deployed in Room-A and Room-B. Note that understanding the air flow field is important for DC monitoring. However, air flow field can only be measured using indirect methods. We deployed air velocity sensors at the vents that supply air to the cold aisle. The air velocity measurements in m/s can be converted to air volume flow rate in  $\text{m}^3/\text{h}$  based on the cross section area of the vents. On the IT racks, we deployed differential pressure sensors to measure the pressure drop across the racks. The pressure drop measurements help understand the spatial distribution of the air volume flows over the cross section of the racks. We also deployed sensors

to monitor the concentration of  $S_2O$ ,  $H_2S$ , and  $NO_2$ , that are often considered the major corrosive gases threatening server hardware. A total of 85 sensors in various modalities were deployed on the testbed. Fig. 4(k) shows an empty private vault provided by Operator-B to be retrofitted as Room-C. Fig. 4(l) shows the four racks we deployed in Room-C with the constructed cold air containment.

In each server room, we deployed four 42U IT racks. Thus, our testbed of three server rooms hosts a total of 12 racks. The planned power rating for each rack is 5 kW. If all the racks are fully populated with servers, the capital expenditure (Capex) for IT equipment will be twice of the Capex for constructing all the supporting facilities shown in Fig. 4. We received a total of 33 on-loan IT devices from four major IT equipment manufacturers as their contributions to this research project. We deployed the same set of 11 IT devices in each server room, as shown in Figs. 4(n) and 4(o). As the racks are not fully populated, to increase cooling efficiency, we applied blinds as shown in Fig. 4(o) on the empty rack slots. Moreover, to increase the power consumption of the IT racks for realism of the experiments, we deployed four in-rack thermo-fluid simulators in each of Room-A and Room-B, and eight in Room-C, as shown in Fig. 4(n). The thermo-fluid simulator can be configured manually to consume a certain power among multiple discrete levels up to 5 kW. With the thermo-fluid simulators, we can reduce the Capex of the testbed, while maintain its realism in terms of power consumption and heat generation. Thanks to Operator-A's and Operator-B's provision of the spaces as their contributions to the project, the operating expenditure (Opex) of the testbed is mainly the energy charge. The Opex of the testbed over about 1.5 years is about 10% of the Capex for constructing the testbed.

**2.1.3 Configuration of Testbed.** We configured all servers and network switches/routers so that we can easily control their operations for experiments. Moreover, as all the three server rooms are located in the premises of Operator-A and Operator-B, it is desirable that we can access all IT equipment and the supporting facilities remotely from our university campus. The remote access should be configured prudently with cybersecurity always in the mind. Although the IT equipment on the testbed will not be used for production, we have a major concern regarding cyber-attacks that take over the SCADA system to damage the costly supporting facilities and/or use the facilities to create safety incidents (e.g., fires by the heater).

We installed the unmodified CentOS v6.9 GNU/Linux on all the servers and configured the switches to form an Intranet in a fat tree topology. We configured three routers on our testbed to use three public IPv4 addresses. Once we made our routers publicly accessible, we observed multiple rounds of port scanning from the Internet, which is often the first step of cyber-attacks. We applied a whitelist of accessible ports and remote host IP addresses to restrict the access. The SCADA master provides a password-protected web interface to access real-time or historical sensor data and adjust the setpoints of actuators (heater, supply/exhaust fans, air dampers, and cooling coils). The SCADA master was configured to use HTTPS for the web interface to ensure the integrity and confidentiality of the communications between the testbed and our campus.

We developed a set of BASH scripts to control and monitor servers' running status. The details are as follows.

- (1) For CPU status control, we use `cpulimit` v0.2 to maintain the utilization of each physical core of a CPU at a specified level. Then, we use a customized LINPACK benchmark provided by the CPU vendor to measure the CPU performance.
- (2) For hard disk drive (HDD) status control, we use the `cgroups` to maintain the read/write throughput of the HDDs configured to operate in the RAID0 mode. Then, we use `fio` to generate HDD read/write requests.
- (3) For memory status control, we use `memtester` to generate test traffic and find memory faults.

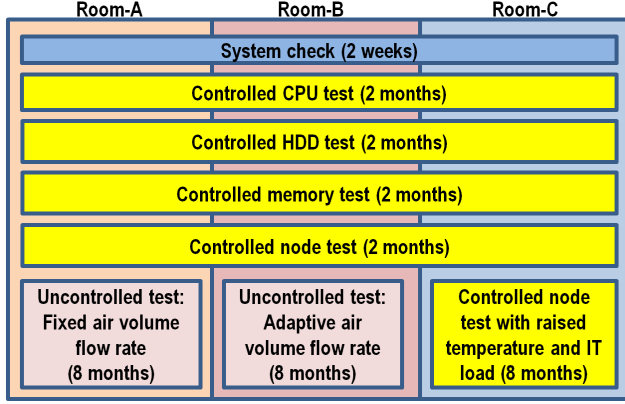


Fig. 5. The planned experiments.

- (4) For server status monitoring, we use nine tools: `cpupower`, `edas-utils`, `impitool`, `sar`, `rsyslog`, `smartmontools`, `lm_sensor`, `bmc`, and `foo`.

Note that many of server-level parameters (e.g., inlet temperature, server errors and server power) are based on the Intelligent Platform Management Interface (IPMI). The collected data traces are uploaded periodically to Google Cloud Storage. During the combined tests of all the scripts we developed, we found that when we tried to maintain the CPU utilization at 100%, IMPI’s sampling experienced significant jitters, degrading the quality of the server status monitoring. Thus, in our planned experiments (cf. Section 2.2), the highest CPU utilization that will be maintained for extended period of time is 90%. We only conducted short-period experiments for 100% CPU utilization.

On the SCADA master, our contractor used a script language to implement the following algorithms. First, they implemented the proportional-integral-derivative (PID) control for the supply and exhaust fans to maintain the air volume flow rate setpoint based on the measurements of the air velocity sensors shown in Fig. 4(i). The control error is within 5%. Second, they implemented the bang-bang control for the coiling coil and heater to maintain the temperature of the air supplied to the IT racks at a setpoint. The control error is about 1°C.

In the planned experiments (cf. Section 2.2), the operations of the servers and the supporting facility need to be coordinated. Thus, we configured the NTP clients of the servers and the SCADA master on our testbed to synchronize their clocks with the NTP servers in Singapore’s NTP pool. The second-accurate clock synchronization of NTP over Internet suffices for the needed coordination.

## 2.2 Design of Experiments on the Testbed

We planned to conduct two groups of experiments: controlled tests and uncontrolled tests. Fig. 5 shows the planned experiments. The time periods shown in Fig. 5 are net test times. From our experience, there were also various overheads that consumed the project time, such as preparation of the test scripts, repair of faulty devices, additional tests to verify results, facility maintenance, and etc. We planned to complete all tests in Fig. 5 in a duration of 20 months.

A controlled test focuses on a key component of the server, i.e., CPU, hard disk drive (HDD), and memory. Specifically, during a *unit test* of a controlled test, the ambient condition (temperature and air volume flow rate) and the operating status of the tested component are maintained at a certain level for one hour. A controlled test consists of hundreds of unit tests with all combinations

Table 1. Experiment settings for controlled experiments.

Parameters	Minimum	Maximum	Step Size
Inlet air temperature	25°C	37°C <sup>†</sup>	1°C
Air flow rate*	2500 m <sup>3</sup> /h	12500 m <sup>3</sup> /h	2500 m <sup>3</sup> /h
Servers' CPU utilization(U) <sup>#</sup>	10%	90%	20%
Hard disk read/write speed	10 MB/sec	100 MB/sec	20 MB/sec
Memory block size	8 kB, 16 kB, 32 kB, 64 kB, 128 kB, 256 kB		

\*Applicable for Room-A and Room-B only.

<sup>#</sup> $U = 10\%$  or  $U = 80\%$ , *Step size* = 10%, 10% is reserved for OS and benchmark tools.

<sup>†</sup>The highest temperature record of our city is 36.7°C.

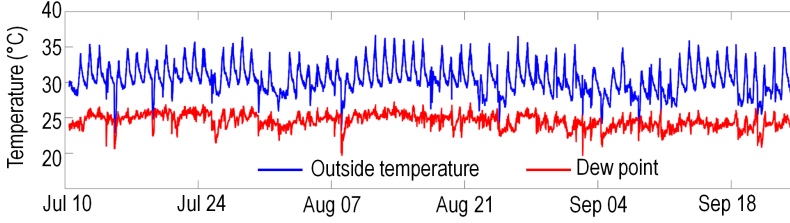


Fig. 6. Outside temperature and dew point in Jul, Aug, and Sep of 2018, in Singapore.

of the server room ambient condition and server component status each swiping the respective range summarized in Table 1. Note that the maximum temperature setpoint of 37°C is the record maximum ambient temperature in Singapore. During the controlled node test, we simultaneously vary the operating status of CPU, HDD, and memory. For the first four controlled tests in Room-C, the temperature setpoint for the return hot air is set to be 20°C as suggested by Operator-B. The CRAC unit controls the volume flow rate of the cold air supplied to the four racks. In the last controlled test in Room-C, we vary the temperature setpoint from 21°C to 35°C with step size of 1°C and the total power of eight thermo-fluid simulators within [10 kW, 20 kW, 30 kW, 35 kW]. The controlled tests allow us to understand the performance and thermal safety of the IT equipment under various conditions.

There are two uncontrolled tests in which the air inhaled by Room-A and Room-B are not conditioned by cooling coils and heater. Thus, the servers experience the ambient temperature and RH. In the uncontrolled test in Room-A, we fix the air volume flow rate to a setting that ensures no overheating on the IT racks. This setting is determined from the test results obtained in the controlled tests with the most extreme condition (i.e., 37°C and full utilization of servers). In the uncontrolled test in Room-B, we adapt the air volume flow rate to the outside temperature. The adaptation logic is designed based on the controlled test results. The uncontrolled tests allow us to assess the energy saving that can be achieved by the air-side free cooling design in Singapore's tropical condition. The results of the fixed ventilation and adaptive ventilation experiments are presented in Section 5.

*Dew point prevention.* During the controlled experiments, the cooling coils are used to maintain the temperature at the cold aisle at the setpoint. When the outside air is hot and humid (e.g., before an afternoon rainfall), the temperature of the cooled air leaving the cooling coil may reach the dew point. As such, the saturated cold air may condense on a colder surface. If such condensation occurs on the printed circuit boards (PCBs) of the IT equipment, the resulted short circuits

may damage the IT equipment. Although this concern can be mitigated by the fact that the heat generated by the IT equipment will increase the temperature and thus decrease the RH of the air passing through the IT equipment, for the safety of the IT equipment, we implemented a dew point prevention mechanism in the control algorithms for the cooling coils. Specifically, if the temperature setpoint is more than 3°C lower than the outside air dew point that can be calculated based on outside air temperature and RH, we stop conditioning the inhaled air. Fig. 6 shows the outside temperature and dew point in about two months. We can see that the dew point fluctuates at around 25°C, which is the minimum temperature setpoint during our tests (cf. Table 1). Thus, this mechanism only disallowed the tests with low temperature setpoints for limited time duration. With this mechanism and the heat generated by the IT equipment, the RH at the cold aisle is capped at 90%.

### 3 SENSOR DATA ANALYTICS

This section presents the analysis of the sensor data obtained from the testbed during the controlled experiments.

#### 3.1 Benchmark Tests and Collected Data

We perform microbenchmarks to investigate the performance of the servers under different environmental conditions. We separately test the CPUs, main memories, and HDDs, which are the main components related to servers' computing performance. For each component, we vary the cold aisle temperature, the room air flow rate, and a certain operating setpoint of the computing component in their respective ranges as shown in Table 1. We use the CPU tests to illustrate. There are a total of 13 temperature levels (from 25°C to 37°C), 5 room air flow speed levels (from 2500 to 12500 m<sup>2</sup>/h), and 7 CPU utilization levels (from 10% to 100%). Thus, there are  $13 \times 5 \times 7 = 455$  unit tests for CPU. Each unit test lasts for one hour. For comparison, we conduct the same number of unit tests in the baseline room, in which the cold aisle temperature is fixed at 21°C.

We have developed various shell scripts that control the facility and the servers to automate the execution of the unit tests. In each unit test, we collect data from the deployed sensors and various monitoring software tools installed on the servers. The sensors are sampled every five seconds. All sensors are wired to an embedded Supervisory Control and Data Acquisition (SCADA) master deployed within the chamber beneath the mixing chamber. Each sensor reading is timestamped using the SCADA master's globally synchronized clock upon its arrival at the SCADA master. Since the SCADA system uses a dedicated wired network, the communication delay from a sensor to the SCADA master is negligible. We install a number of software packages including PMI, SAR, syslog and lm-sensor to access the servers' built-in sensors and performance counters. All servers are synchronized to three time servers in our city's NTP pool.

We collected the following datasets during the project period:

- (1) **CPU test dataset:** This dataset was collected during the CPU test in which we varied the supply air temperature, air volume flow rate, and CPU utilization. Main purpose of this dataset is to investigate how the temperature and air flow rate affect the CPU's computing performance and thermal safety.
- (2) **HDD test dataset:** This dataset was collected during the HDD test in which we varied the supply air temperature, air volume flow rate, and HDD random read/write speed. Main purpose of this dataset is to investigate how the temperature and air flow rate affect the HDD's performance in terms of IOPS and response time.
- (3) **Memory test dataset:** This dataset was collected during the memory test in which we varied the supply air temperature, air volume flow rate, and the block size setting in copying data from one memory area to another memory area of the tested memory. Main purpose of this dataset is to investigate how the temperature and air flow rate affect the memory's performance in terms of data copying bandwidth.
- (4) **Node test dataset:** This dataset was collected during the node test in which we varied the supply air temperature, air volume flow rate, CPU utilization, HDD random read/write speed, and block size setting in copying data in the memory. Main purpose of this dataset is to investigate how the temperature and air flow rate affect the server's computing performance when the CPU, memory, and HDD are in use simultaneously.
- (5) **Fixed ventilation test dataset:** This dataset was collected in the uncontrolled test adopting fixed air volume flow rate, while the cooling coil is not used. Main purpose of this dataset is to investigate the energy saving when a simple air flow rate control (i.e., fixed flow rate) is used.

- (6) **Adaptive ventilation test dataset:** This dataset was collected in the uncontrolled test adapting adaptive air volume flow rate, while the cooling coil is not used. Main purpose of this dataset is to investigate the energy saving when an advanced air flow rate control (i.e., adaptive flow rate) is used.

The collected measurement data in each dataset can be divided into four domains as follows.

- (1) **Performance:** The performance data includes server computing throughput, CPU and memory utilization, core frequency, HDD read/write throughput, and etc.
- (2) **Reliability:** The reliability data includes correctable and un-correctable memory errors, HDDs' latent sector errors and self-monitoring, analysis, and reporting technology (SMART) records.
- (3) **Environment:** The environmental data relate to the conditions of the test room, servers and outdoor weather. Examples include room/rack temperature and RH, air velocity, servers' inlet/outlet temperatures, processor core and disk temperatures, etc.
- (4) **Energy:** The energy data includes power measurements of cooling coils, room fans, heater, server racks and other IT equipment.

### 3.2 Summary of Test Results

Now, we provide a high-level summary of the controlled test results on the servers' performance in Room-A/B. The detailed analysis on the measurements is presented in the following subsections.

- (1) **CPU test results.** We measured giga floating point operations per second (GFLOPS) to characterize the CPU performance. We also monitored the CPU core frequency to pinpoint performance degradation caused by frequency throttling. The tests show that, for all CPUs in Room-A and Room-B, the temperature setpoint has little/no impact on GFLOPS and core frequency when (1) the temperature setpoint is from 25°C and 37°C, (2) the CPU utilization is from 10% to 90%, and (3) the air volume flow rate is 2500 m<sup>3</sup>/h and above. We also investigated the thermal safety of the tested CPUs. The vendor of the tested CPUs specifies  $\bar{T}_{case}$  for each CPU model, which is the upper limit of the CPU *case temperature* for thermal safety. However, each CPU only has a built-in digital thermal sensor to measure  $T_{core}$ , which is the *core temperature* on the die. During the tests, the measured  $T_{core}$  was always below  $\bar{T}_{case}$ . As the case temperature is always lower than the core temperature, the case temperature, although inaccessible, must be lower than  $\bar{T}_{case}$ . Thus, all the tested CPUs were thermally safe during the CPU tests in Room-A and Room-B. This also explains the absence of core frequency throttling in the tests. An expert representative from the CPU vendor agreed the above results.
- (2) **HDD test results.** We measured the input/output operations per second (IOPS) and response time during random read and write accesses to characterize the HDD performance. The tests show that, for all HDDs in Room-A and Room-B, the temperature setpoint has little/no impact on IOPS and response time when (1) the temperature setpoint is from 25°C to 37°C, (2) the HDD random read/write speed is from 10 MB/s to 100 MB/s, and (3) the air volume flow rate is from 2500 m<sup>3</sup>/h to 12500 m<sup>3</sup>/h. The results also show that the HDD random read/write speed has little impact on the server energy consumption.
- (3) **Memory test results.** We measured the speed of copying a large amount of data from a user space memory area to another area using various block sizes to characterize the memory performance. We use cyclic redundancy check (CRC) to verify the integrity of the data copying. The tests show that, for all memories in Room-A and Room-B, the temperature setpoint has little/no impact on memory speed when (1) the temperature setpoint is from 25°C to 37°C, (2) the block size setting is from 8 kB to 256 kB, and (3) the air volume flow rate

Table 2. Specification of tested CPU models (64-bit instruction set).

Model	$f_{\text{base}}$ (GHz)	$f_{\text{turbo}}$ (GHz)	$T_{j\text{max}}$ (°C)	$T_{c\text{max}}$ (°C)	Cores	L3 cache (MB)	TDP (W)
Model1	2.2	2.6	89	78	10	13.75	85
Model2	2.3	2.6	90	78.9	10	25	105
Model3	3.8	4.0	100	n.a.	4	8	72
Model4	2.2	2.5	95	80	12	30	105

$f_{\text{base}}$ : base frequency;  $f_{\text{turbo}}$ : turbo frequency for all cores; TDP: thermal design power  
 $T_{c\text{max}}$  is from CPU datasheet;  $T_{j\text{max}}$  is from the coretemp Linux kernel driver.

is from 2500 m<sup>3</sup>/h to 12500 m<sup>3</sup>/h. No CRC verification errors occurred during the tests. The results also show that the memory speed has little impact on the server energy consumption.

- (4) **Node test results.** We tested the CPU, HDD, and memory simultaneously under a total of six server status levels. At the first level where the server has light workload and the sixth level where the server is stressed, the CPU utilization, HDD read/write speed, and memory block size in data copying are {10%, 10 MB/s, 8 kB} and {90%, 100 MB/s, 256 kB}, respectively. The test results show that the performance metrics of CPU, HDD, and memory are similar to those tested separately, except that the memory speed is affected by CPU utilization setpoint. This is because CPU cycles are needed to copy data for testing the memory. In contrast, the HDD performance is not affected by CPU utilization setpoint, because HDD is a low-speed apparatus compared with CPU and memory. All the CPUs were also thermally safe, although CPU, HDD, and memory generate heat simultaneously.

### 3.3 CPU Performance

**3.3.1 Settings.** The six servers in each test room are equipped with twelve CPUs in four models as shown in Table 2. The models of these CPUs were launched between 2014 and 2017. We use `cpulimit v0.2` to set a usage percentage for a process and maintain the utilization of each physical core of a CPU at the levels shown in Table 1. Then, we run a customized LINPACK benchmark provided by the CPU vendor. Note that LINPACK is a widely adopted floating point computation benchmark.

Moreover, it is threaded to effectively leverage multiple CPU cores. To faithfully measure the performance of the physical cores, we disable the hyper threading feature of all CPUs in the servers' basic input/output system (BIOS) settings. The LINPACK reports the GFLOPS as the CPU computing throughput. We also use the `cpupower v3.10.0` and `coretemp v2.6.32` to monitor the core frequencies and temperatures that are measured by a digital temperature sensor embedded in each core.

**3.3.2 Analysis.** The performance of CPUs is evaluated under various cold aisle temperatures and CPU utilization levels using three metrics, including CPU computing throughput (i.e., GFLOPS), core frequency and temperature.

Fig. 7 shows the GFLOPS of four tested CPU models under various cold aisle temperatures and CPU utilization levels. Under the same temperature and CPU utilization, each result is the average GFLOPS of five unit tests, each of which is conducted using a different air flow rate from 2500 m<sup>3</sup>/h to 12500 m<sup>3</sup>/h. The relative standard deviation (RSD) of each GFLOPS result is also included to show the effect of the air flow rate on the CPU throughput. As shown in Fig. 7, the GFLOPS of all tested CPUs remain stable across different cold aisle temperatures. These results imply that the cold aisle temperature from 21°C to 37°C has little impact on the GFLOPS of all tested CPU

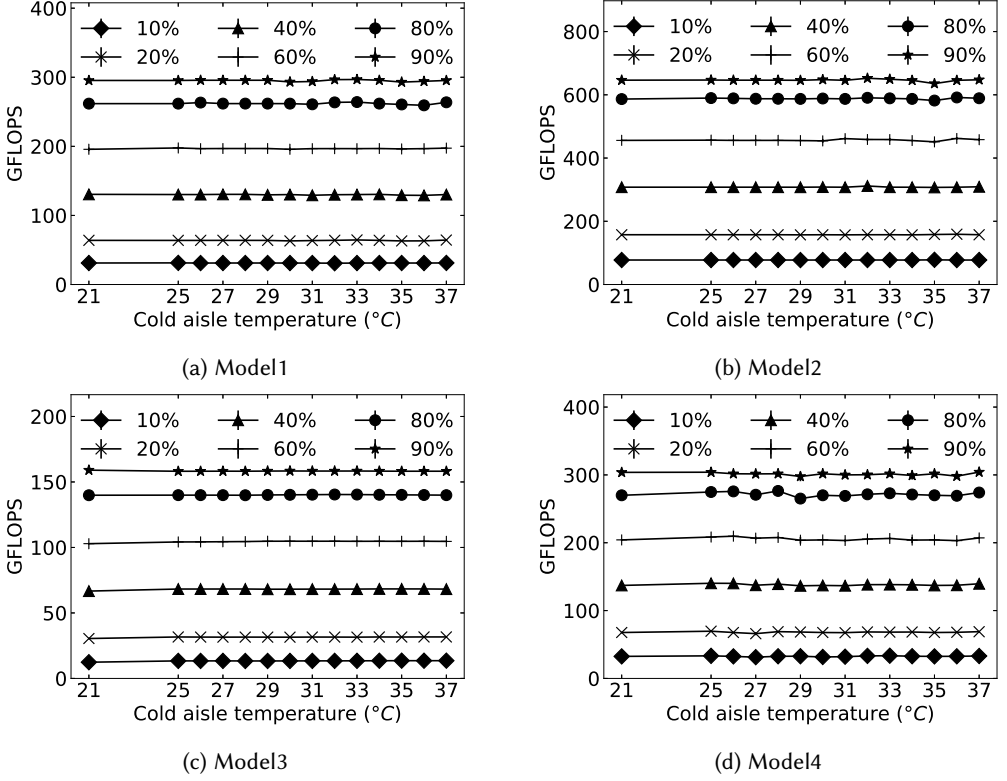


Fig. 7. GFLOPS under various cold aisle temperatures and CPU utilization levels. Each error bars represent the relative standard deviation (RSD).

models. In other words, increasing the cold aisle temperature from 21°C up to 37°C does not cause significant degradation of CPU computing performance.

From Fig. 7, the CPU utilization is the main factor affecting the GFLOPS. The GFLOPS of four CPU models consistently increases with the CPU utilization as shown in Fig. 8. This is because a higher CPU utilization level allows a higher amount of CPU's resource for the computing task.

Fig. 9 shows the core frequencies of the four CPU models under various cold aisle temperatures from 21°C to 37°C. The core frequencies of tested CPU models do not change over different temperatures up to 37°C. Under each temperature, the variation of the core frequency represented by the error bars as shown Fig. 9 are caused by the transient phases the beginning and end of the LINPACK benchmark processes. From Table 2 and Fig. 9, we can see that Model 2 and Model 4 run at their turbo frequencies, whereas Model 3 runs at its base frequency. To perform an intensive computing task, the CPU can normally run its cores at clock rates (i.e., the core frequency) from the base frequency up to the turbo frequency, provided that the power and temperature do not exceed the design limits.

Fig. 10 shows the average temperature of cores in four CPU models under various cold aisle temperature in cases that the CPU utilization is 10% and 90%. The horizontal lines in Fig. 10 represent the  $T_{jmax}$  and  $T_{cmax}$  of the CPU models as shown in Table 2. From Fig. 10, the core temperature increases with the increase of the cold aisle temperature and the CPU utilization. Moreover, the

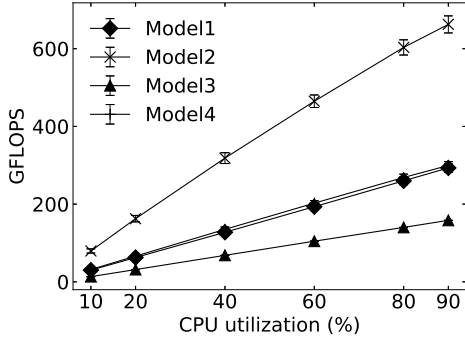


Fig. 8. GFLOPS under various CPU utilization levels.

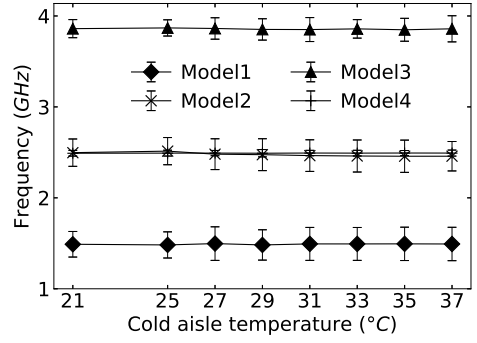
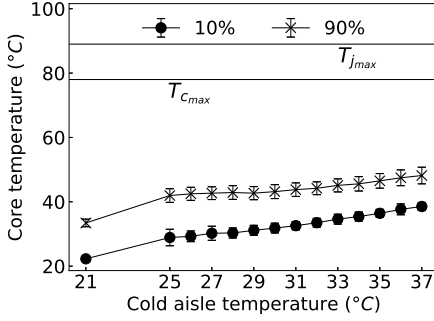
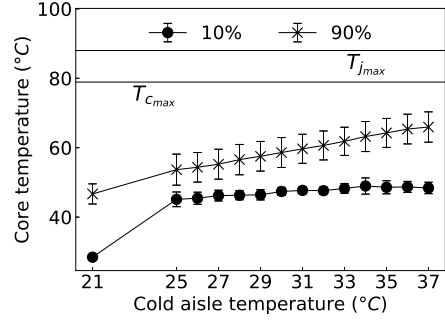


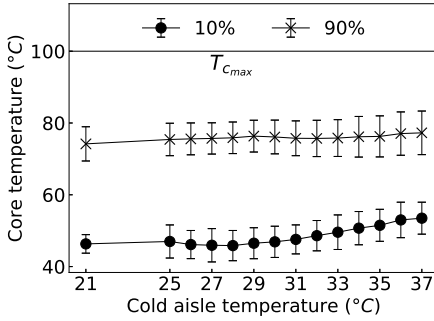
Fig. 9. Core frequency under various cold aisle temperatures.



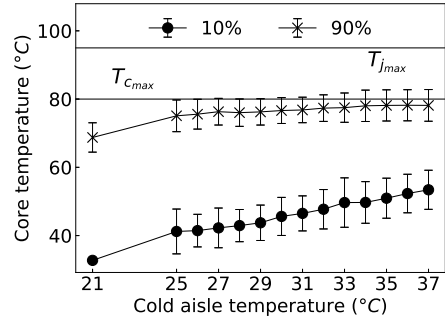
(a) Model1



(b) Model2



(c) Model3



(d) Model4

Fig. 10. Core temperature under various cold aisle temperature and CPU utilization. Each result is the average core temperature of five unit tests. The error bars represent the RSD.

core temperatures of each CPU model is always less than than the  $T_{jmax}$  and  $T_{cmax}$ , which are represented by the horizontal lines when the cold aisle temperature and the CPU utilization level are up to 37°C and 90%, respectively.

Note that the  $T_{\text{cmax}}$  is the upper limit of the CPU case temperature specified by the manufacturers, whereas the  $T_{\text{jmax}}$  is the upper limit of the CPU junction temperature derived by the coretemp Linux kernel driver. If these upper limits are exceeded for extended periods of time, the reliability of the CPU will be compromised. The core temperature measured by a digital temperature sensor embedded in the CPU die is normally higher than the temperatures at the junction between the die and CPU’s PCB and the case. Therefore, the CPU is thermally safe if the core temperature is lower than  $T_{\text{jmax}}$  or  $T_{\text{cmax}}$ . As shown in Fig. 10, except for Model 4, there are considerable gaps between the core temperatures and  $T_{\text{jmax}}$  or  $T_{\text{cmax}}$ . The core temperatures of Model 4 at 90% are close to the  $T_{\text{cmax}}$ . However, they are still about 17°C lower than the  $T_{\text{jmax}}$ .

In summary, the core temperatures of all CPU models are within the safe range when the cold aisle temperature is from 25°C to 37°C and the CPU utilization is from 10% to 90%. This is the main reason why the CPU computing performance in terms of GFLOPS and core frequency does not significantly drop when the temperature and the CPU utilization are up to 37°C and 90%, respectively.

### 3.4 HDD Performance

Table 3. Specification of tested HDD models.

Model	Operating temperature (°C)	Rotation speed (RPM)	Max throughput* (MB/s)	Capacity (GB)
HDD1	[5, 55]	10,520	129 to 224	600
HDD2	[5, 55]	15,000	151 to 202	300
HDD3	[5, 55]	15,030	175 to 250	300

\*The maximum throughput of single disk in the sequential read mode.

**3.4.1 Settings.** The six servers in each test room are equipped with twelve HDDs in three different models as shown in Table 3. To test the HDDs, we first specify the read/write throughput of the HDD (i.e., the maximum read/write speed) using the cgroups, a Linux kernel feature that can limit the resource usage of a collection of processes. Then, under a certain read/write throughput level, the performance of HDDs is evaluated using a random access method which access the locations on the HDDs in a non-contiguous manner. The random access method is adopted because (1) it stresses the HDDs by a lot of movements of the disk’s header to write/read data in different locations, (2) it is a common method widely used by many block-based applications such as database, mail and file servers.

To implement the random access method for controlling the HDD activity, we use the fio tool with the random and mixed read and write mode. The input/output operations per second (IOPS) reported by the fio tool is recorded as one of the HDD performance metrics. The relationship between the IOPS and the write/read throughput is:

$$\text{IOPS} = \frac{\text{throughput}}{\Phi_{IO}},$$

where  $\Phi_{IO}$  is the size of an input/output request. In our tests, we fix the  $\Phi_{IO}$  at 16KB and vary the throughput from 10 MB/s to 100 MB/s as shown in Table 1. In addition, we also use the sar tool to monitor the average response time for completing the read/write requests generated by the fio. Note that the response time is the end-to-end latency from the time when the read/write request is

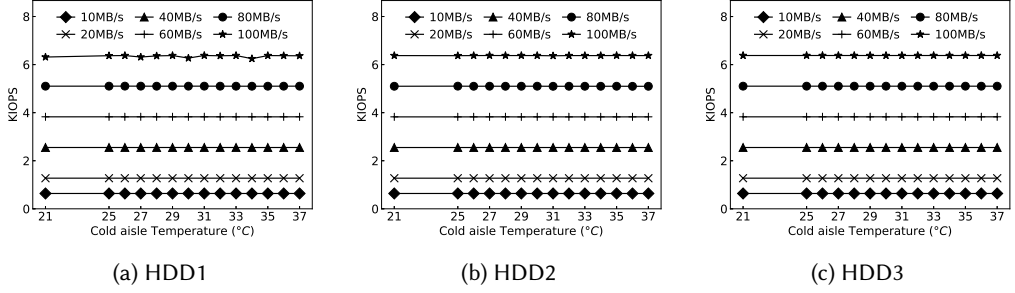


Fig. 11. KIOPS under various cold aisle temperatures.

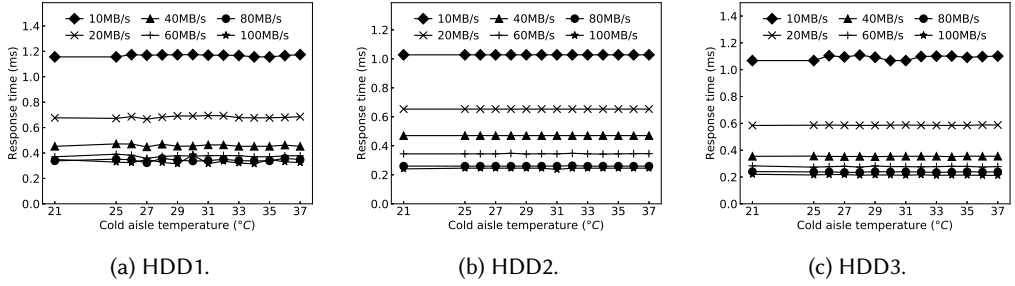


Fig. 12. Response time under various cold aisle temperatures.

generated to the time when the request is completed. It consists of latency of the HDD's hardware and system's software operations.

**3.4.2 Analysis.** We use the IOPS and response time as two metrics to evaluate the performance of the tested HDDs under various cold aisle temperatures and write/read speeds.

Fig. 11 shows the kilo IOPS (KIOPS) of three tested HDD models under various cold aisle temperatures from 21°C to 37°C and write/read speed from 10 MB/s to 100 MB/s. From Fig. 11, the KIOPS of all tested HDDs remains stable when the cold aisle temperature changes from 21°C to 37°C. In other words, the cold aisle temperature up to 37°C does not cause significant effect on the maximum sustained bandwidth that the HDD can handle. In addition, the KIOPS only depends on the write/read speeds of the HDDs. Specifically, over all HDD models, the KIOPS consistently increases with the increase of the HDD's speed.

Fig. 12 shows the response time of three HDD models over the variation of the cold aisle temperature and write/read speed. The response time is the total latency from the time that the I/O request is generated and placed at the I/O queue to the time that it is completed. From Fig. 12, we can see that the cold aisle temperature also does not cause significant effect on the response time of the HDDs. Moreover, the response time greatly decreases with the increase of the write/read speed from 10 MB/s to 20 MB/s. Then, the response time remains mostly the same when the speed is from 60 MB/s to 100 MB/s.

### 3.5 Main Memory Performance

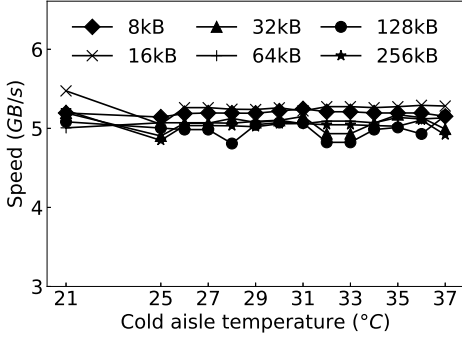
**3.5.1 Settings.** Our servers are installed with two different memory models as shown in Table 4. To test the performance of the memory, we developed a shell script program that combines two memory test tools *memtester* and *ubw* to copy data from a location to another location in the main

Table 4. The tested memory models (64-bit instruction set).

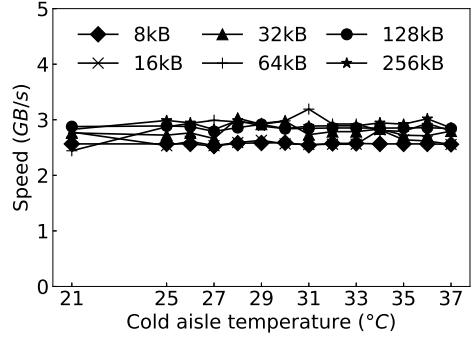
Model	DIMM Type	Size (GB)	Operating Frequency (MHz)	Num. of DIMM	Num. of CPU	Arch.
MEM1	DDR-4	16	2,400	1	1	NUMA
MEM2	DDR-4	64	2,400	2	2	NUMA

<sup>1</sup> DIMM: Dual In-line Memory Module.

<sup>2</sup> NUMA: Non-uniform memory Access.

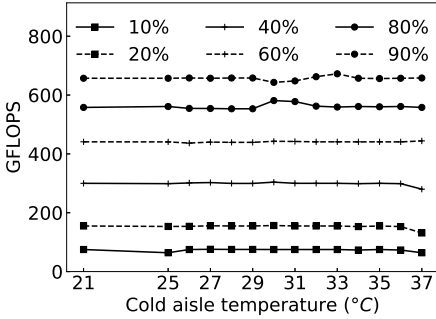


(a) MEM1

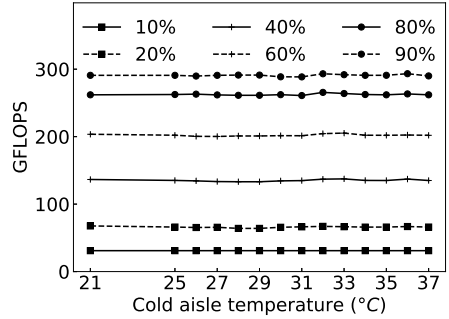


(b) MEM2

Fig. 13. Memory speed under various cold aisle temperatures and memory block sizes.



(a) Model2



(b) Model4

Fig. 14. GFLOPS under various cold aisle temperatures and CPU utilization levels in the node test.

memory. Specifically, a chunk of data with a size of  $\Phi = n \times B$  is first generated at a random location in the memory. Then, the data chunk is copied to another random location for  $n$  times, each of which moves a data block with a size of  $B$  bytes to the destination location. During a memory unit test, multiple data chunks are moved between random locations in the main memory. We use the memory speed, denoted and calculated by  $v = \frac{\Phi}{T}$ , as a metric to evaluate the memory performance

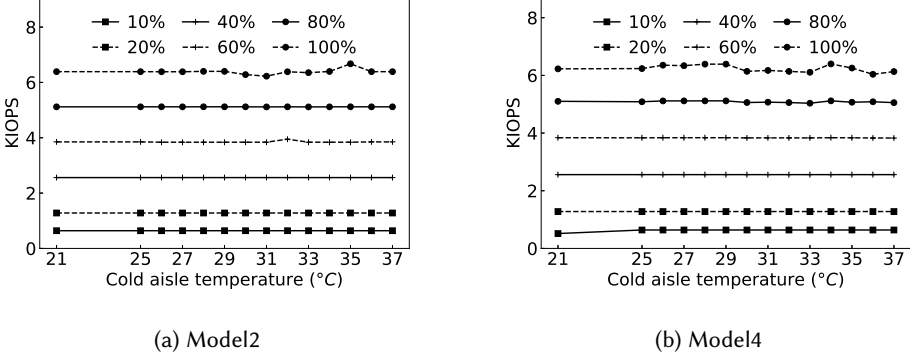


Fig. 15. HDD IOPS under various cold aisle temperatures and throughput in the node test.

under various cold aisle temperatures and values of block size  $B$ , where  $T$  is the latency for copying a data chunk from the source position to the destination position.

**3.5.2 Analysis.** Similarly, we test the performance of the main memory under various cold aisle temperatures from 21°C to 37°C. Under each temperature, we vary the block size  $B$  from 8 kB to 256 kB as shown in Table 1.

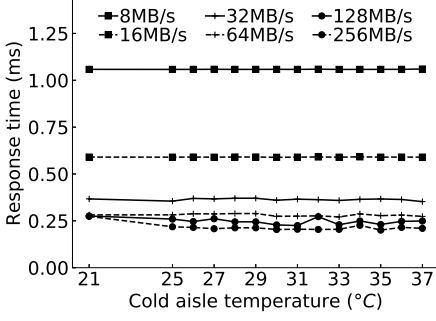
Fig. 13 shows the average memory speed of two memory models, MEM1 and MEM2 under various cold aisle temperatures and memory block sizes. With the same block size, the memory speed of MEM1 and MEM2 is slightly varied over different temperatures. However, the effect of temperature on the variation is inconsistent. This is because during the test, the memory resource is not dedicated only to the memory test. Operating system scheduler and other server components such as CPU, caches, and system bus can randomly occupy the memory resource for their operations. As a result, the speed of data copying may vary over time, depending on the available of the memory resource. In other words, the increased cold aisle temperature does not cause the variation of the memory speed. The temperatures 21°C to 37°C have little impact on the memory speed.

### 3.6 Node Performance

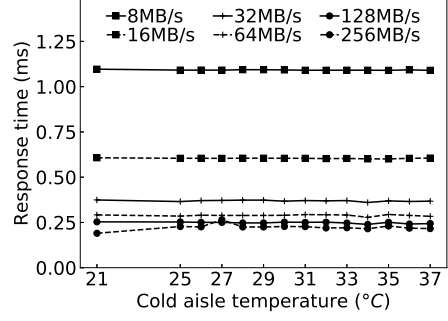
We also perform the node test that stresses CPU, HDD and main memory simultaneously under different environmental conditions. A total of 330 unit tests were conducted, where CPU utilization, HDD speed and memory block size are simultaneously varied in their respective ranges as shown in Table 1. The effects of cold aisle temperature on CPU GFLOPS, HDD IOPS and response time, and memory speed are presented in Figs. 14, 15, 16 and 17, respectively. The node tests show similar CPU, HDD and memory performance results to those of tests which stress the CPU, HDD and memory of the server separately. The cold aisle temperature up to 37°C does not cause significant effect on performance of CPU, HDD and memory.

### 3.7 Energy profiles

We conducted a set of experiments to understand the energy consumption profile of Room-A/B. Fig. 18(a) shows the energy consumption of cooling coil, heater, and server racks in Room-A when the temperature setpoint was varied from 25°C to 37°C during a 13-hour experiment. Each data point in the figure is the energy consumption during one hour. When the temperature setpoint was greater than 33°C, the outside temperature was lower than the setpoint. Thus, the cooling coil stopped working and the heater started operation. The energy consumption of the server racks

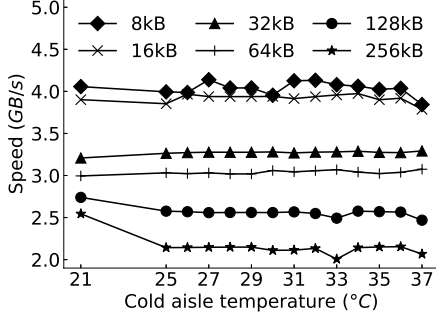


(a) Model2

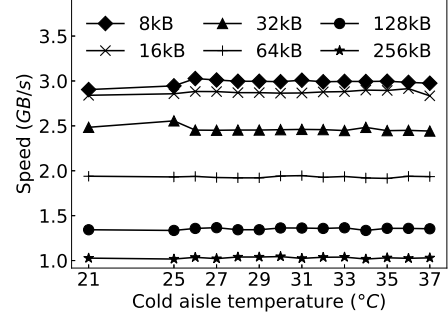


(b) Model4

Fig. 16. HDD response time under various cold aisle temperatures and throughput in the node test.



(a) Model2



(b) Model4

Fig. 17. Memory Speed under various cold aisle temperatures and block sizes in the node test.

increased by 6% when the temperature setpoint was varied from 25°C to 37°C. This is because the server enclosure's built-in fans rotate faster when the inlet temperature increases. Fig. 18(b) shows the total energy drop of Room-A by about 45% when the temperature setpoint was increased from 25°C to 33°C. This suggests that a significant energy saving can be achieved by air-side free cooling. The curve in Fig. 18(b) raises when the temperature setpoint is greater than 29°C. This is because there was an outside temperature increase after we completed the test with the temperature setpoint of 28°C.

Fig. 18(c) shows the total server energy consumption in Room-A when the CPU utilization was varied from 10% to 90% and the temperature setpoint was increased from 25°C to 32°C. Each point is the energy measurement over one hour. We can see that, although the server energy in general increases with the temperature setpoint due to the faster server fan rotation, CPU utilization is a major factor affecting the server energy in a linear manner.

Fig. 18(d) shows the energy consumption of the fans in Room-A when the air volume flow rate setpoint was increased from 2500 m<sup>3</sup>/h to 12500 m<sup>3</sup>/h and the temperature setpoint was fixed at 26°C. The fans consumed 5.4% to 22.6% of the total energy consumption of Room-A. Our controlled experiments over eight months show that air volume flow rate of 2500 m<sup>3</sup>/h suffices for

each of Room-A and Room-B to prevent overheating. This implies that if the air-side free cooling design using fans only is successful, the PUE can be reduced to about 1.05. Our adaptive ventilation uncontrolled test results reported in Section 5.1 echo this result.

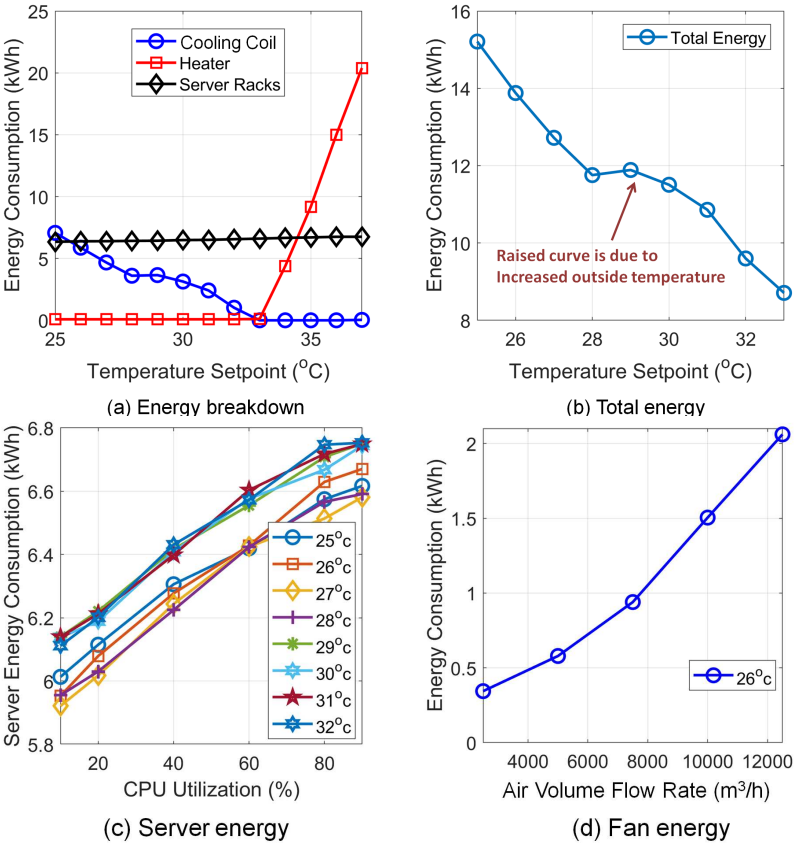


Fig. 18. Energy profile of Room-A. The measurements in (a) and (b) were collected during a 13-hour experiment.

## 4 THERMAL AND POWER MODELING

This section develops various mathematical models to characterize the relationships between various computing and thermal quantities on the testbed.

### 4.1 Psychrometric Modeling

In this section, we perform psychrometric analysis for the four steps of the air processing in the air-side free-cooled DC, i.e., *heating* in the server room, *buffering* in the buffer chamber, *cooling* by the cooling coil, and *mixing* by the damper system. Time is divided into intervals with identical duration of  $\tau$  seconds. The system state, denoted by  $x$ , is a vector  $x = [t_s, \phi_s, p_{IT}, t_o, \phi_o]$ , where  $t$  and  $\phi$  respectively represent temperature and RH, the subscripts  $s$  and  $o$  respectively represent supply air and outside air, and  $p_{IT}$  represents the total power consumption of all IT equipment in the server room. The  $p_{IT}$  determines the amount of heat generated in the server room. The supply and exhaust fans admit air volume flow rate setpoints. To achieve steady state without control errors, the setpoints for the two fans should be identical; otherwise, the server room will be in the dynamic process of pressurization/depressurization or a steady state with control errors. Let  $\dot{v}_s \in [0, \dot{v}_{\max}]$  denote the air volume flow rate setpoint for the two fans, where  $\dot{v}_{\max}$  is the maximum achievable air volume flow rate. The cooling coil admits a setpoint  $\Delta t$  that represents the reduction of temperature, i.e.,  $\Delta t = t_o - t_p$ , where  $t_p$  represents the temperature of the processed air leaving the cooling coil. Let  $\Delta t_{\max}$  represent the maximum temperature reduction that can be achieved by the cooling coil. Thus,  $\Delta t \in [0, \Delta t_{\max}]$ . Let  $\alpha \in [0, 1]$  denote the setpoint for the damper system, which is the fraction of the recirculated hot air in the supply air. Thus,  $1 - \alpha$  is the fraction of the outside air in the supply air. A setpoint  $\alpha$  can be achieved by controlling the openness of the three dampers. For example, to achieve  $\alpha = 0$ , the supply and exhaust dampers should be completely open and the mixing damper should be completely closed; to achieve  $\alpha = 1$ , the supply and exhaust dampers should be completely closed and the mixing damper should be completely open. The control action, denoted by  $a$ , is a vector  $a = [\dot{v}_s, \Delta t, \alpha]$ . We aim to construct a Markovian computational model to characterize the psychrometric dynamics of the testbed given the control action  $a$ :

$$t_s[k+1], \phi_s[k+1] = f(t_s[k], \phi_s[k], t_o[k], \phi_o[k], \dot{v}_s[k], \Delta t, p_{IT}[k], \alpha),$$

where  $k \in \mathbb{Z}$  represents the index of time step.

We define the following notation:  $\dot{m}$  is mass flow rate,  $h$  is enthalpy,  $w$  is moisture content; for the above psychrometric variables, we use the subscripts  $\cdot_s, \cdot_h, \cdot_r, \cdot_p, \cdot_o$  to refer to the supply air in the cold aisle, the hot air generated by the servers, the recirculated hot air from the buffer chamber to the mixing chamber, the processed air leaving cooling coil, and the outside air provided to the cooling coil, respectively. The four steps are as follows:

- (1) **Heating:** Servers generate heat and introduce no extra moisture. Thus, the air enthalpy at the hot aisle is higher than that at the cold aisle, while the moisture contents at the two aisles are identical. Denoting by  $\eta$  the servers' heat rate transfer coefficient, the psychrometrics of the server room is

$$\begin{cases} \dot{m}_s h_s + \eta p_{IT} = \dot{m}_h h_h, \\ \dot{m}_s = \dot{m}_h, \\ w_s = w_h. \end{cases} \quad (1)$$

- (2) **Buffering:** The hot aisle air is transported into the buffer chamber by the exhaust fan. Under the setpoint  $\alpha$  for the damper system, the buffer chamber is characterized by

$$\dot{m}_r = \alpha \dot{m}_h. \quad (2)$$

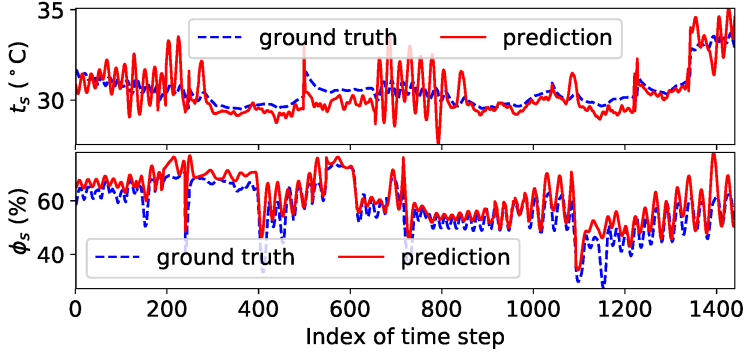


Fig. 19. Prediction results by the psychrometric model (RMSEs for  $t_s$  and  $\phi_s$  are  $0.83^\circ\text{C}$  and  $5.3\%$ , respectively).

- (3) **Cooling:** The total energy of ideal gas is the sum of dry air's energy and water vapor's energy. Without condensation, the cooling coil does not change moisture content of air passing through. Moreover, it does not change mass flow rate. Thus, the condition of the air leaving the cooling coil is given by

$$\begin{cases} h_p = c_p(t_o - \Delta t) + w_p(c_{pw}(t_o - \Delta t) + l), \\ w_p = w_o, \\ \dot{m}_p = \dot{m}_o, \end{cases} \quad (3)$$

where  $c_p$  and  $c_{pw}$  respectively represent the specific heat of dry air and water vapor which are constants; another constant  $l$  represents the evaporation heat. Note that  $c_p(t_o - \Delta t)$  is the enthalpy of the dry air leaving the cooling coil;  $w_p(c_{pw}(t_o - \Delta t) + l)$  is the enthalpy of the water vapor leaving the cooling coil.

- (4) **Mixing:** The air leaving the cooling coil and the recirculated hot air are mixed in the mixing chamber. Governed by the conservation of mass and energy, the psychrometrics of the mixing process can be characterized by

$$\begin{cases} (1 - \alpha)h_p + \alpha h_r = h_s, \\ (1 - \alpha)w_p + \alpha w_r = w_s, \\ \dot{m}_p + \dot{m}_r = \dot{m}_s. \end{cases} \quad (4)$$

Taking the moisture contents of the two influxes as boundaries, Eq. (4) suggests that the outflow's moisture content will be in between, which is the basis of the RH control through adjusting  $\alpha$ .

The above models in Eqs. (1)-(4) are for enthalphy, moisture content, and mass flow rate. These quantities can be converted to temperature, RH, and volume flow rate according to the equations presented in [4]. The aforementioned Markovian computational model is as follows. By initializing the  $h_s$  and  $w_s$  in Eq. (1) with the current state of the supply air condition (i.e.,  $t_s[k]$  and  $\phi_s[k]$ ), we use the remaining equations in Eqs. (2)-(4) to update  $h_s$  and  $w_s$ . The updated values are then used to initialize the  $h_s$  and  $w_s$  in Eq. (1) again and then solve Eqs. (2)-(4). This process is iterated until  $h_s$  and  $w_s$  converge; the converged values are converted to  $t_s[k + 1]$  and  $\phi_s[k + 1]$ . Thus, the Markovian computational model has no closed-form expression.

We use data traces collected during the controlled experiments on the testbed to evaluate the psychrometric models presented. The inputs to the model are  $t_o$ ,  $\phi_o$ ,  $p_{IT}$ ,  $\dot{v}_s$ ,  $\Delta t$ , and  $\alpha$ ; the outputs are the predicted  $t_s$  and  $\phi_s$ . We use root mean squared error (RMSE) between the prediction and the

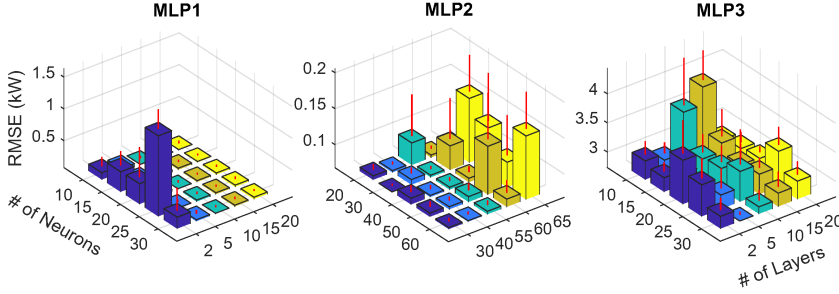


Fig. 20. Impact of hyperparameters on MLPs' performance.

ground truth as the evaluation metric. Fig. 19 shows the prediction results over a time duration of 24 hours. We can see that the prediction by the psychrometric model well tracks the ground truth. The RMSEs for  $t_s$  and  $\phi_s$  are just  $0.83^\circ\text{C}$  and  $5.3\%$ , respectively, over an evaluated period of 24 hours.

#### 4.2 Power Consumption Models

We design three multi-layer perceptrons (MLPs) to model the following powers averaged over the next period of  $\tau$  seconds: (1) IT power  $p_{\text{IT}}[k+1]$ , (2) total power of supply and exhaust fans  $p_f[k+1]$ , and (3) cooling coil power  $p_c[k+1]$ . The average non-IT power consumed in the next time period is  $p[k+1] = p_f[k+1] + p_c[k+1]$ . The MLPs use the respective power measurements in the past  $K$  time periods as a part of the input to address the autocorrelation of power consumption. Moreover, the MLPs use additional inputs that will be discussed below. Note that the hyperparameters of the MLPs (e.g., the number of layers and neurons) are designed based on real traces collected from the testbed.

The first MLP (MLP1) modeling  $p_{\text{IT}}[k+1]$  additionally takes  $t_s[k]$  and  $\dot{v}_s[k]$  as inputs. This is because (1) higher temperatures lead to higher rotation speeds of server fans and CPU fans, (2) air flow generates forces on the fan blades. The second MLP (MLP2) modeling  $p_f[k+1]$  additionally takes  $\dot{v}[k]$  and  $t_s[k]$  as inputs. This is because (1) fan power increases with fan speed, (2) with a higher temperature, materials exhibit higher strength, resulting in the increase of stresses on rotating components. The third MLP (MLP3) modeling  $p_c[k+1]$  additionally takes  $\Delta t[k]$  and  $\dot{v}_s[k]$  as inputs. This is because (1) the setpoint  $\Delta t$  determines the cooling capacity needed, (2) the cooling coil consumes more power when it processes a larger volume of air.

We evaluate the three MLP models for predicting IT power, cooling power, and fan power. Each MLP is trained, validated and tested using 1375, 700 and 1080 data samples, respectively. The settings of  $K$  (i.e., the respective power measurements in the past  $K$  periods used for prediction) for the three MLPs are 5, 1, and 1. For all MLPs, the training batch size is set to 128; the training time is 3,000 epochs. The Adam optimizer with a learning rate of 0.001 is used for training. Moreover, we use the rectified linear units (ReLUs) as the activation function for input and hidden layers; we use linear units for output layer. We conduct extensive evaluation to choose the number of hidden layers and neurons for each MLP to minimize the prediction RMSEs. The evaluation for a certain combination of hyperparameter settings is repeated 5 times to account for the randomness of the training.

Fig. 20 shows the error bars for testing RMSEs with various hyperparameter settings of the number of hidden layers and the number of neurons. MLP1 achieves the smallest RMSE of  $0.10 \pm 0.07$  kW with 5 hidden layers, each of which has 20 neurons. MLP2 achieves the smallest RMSE of

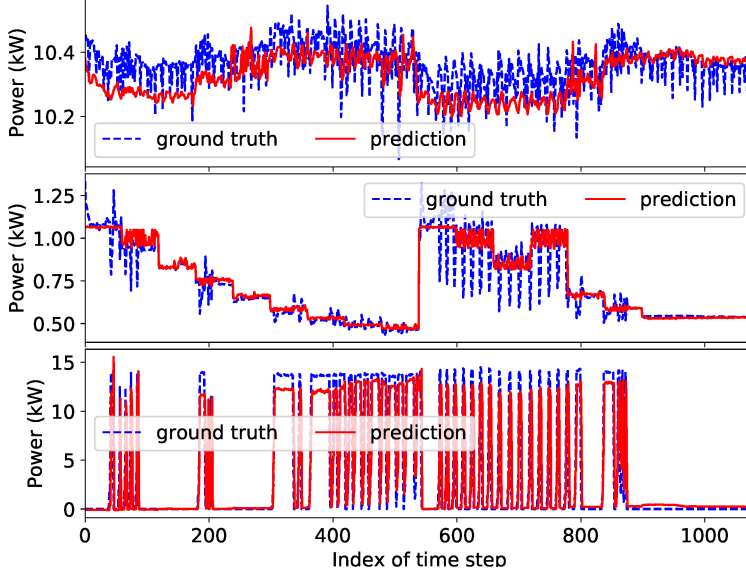


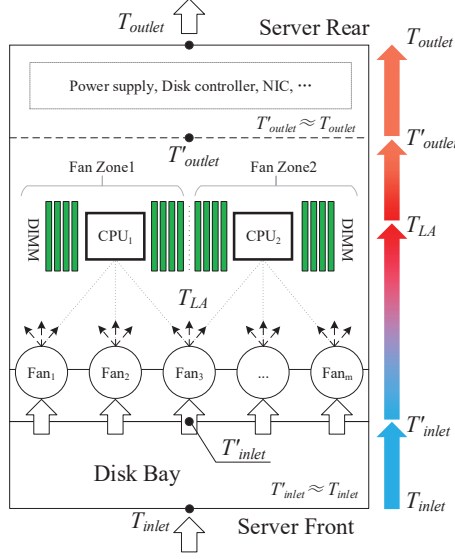
Fig. 21. Prediction results of MLPs. Top: IT power; middle: fan power; bottom: cooling coil power.

$0.07 \pm 0.002$  kW with 60 layers, each of which has 40 neurons. MLP3 achieves the smallest RMSE of  $2.73 \pm 0.03$  kW with 5 layers, each of which has 30 neurons. Fig. 21 shows the ground truth and the prediction by the three MLPs with the chosen hyperparameters over a time duration of 18 hours. Overall, the predictions well track the ground truths.

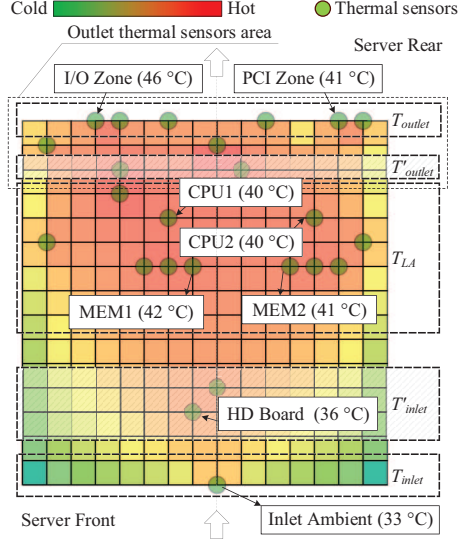
### 4.3 Server Outlet Temperature Model

We investigated all of the servers used in the tests and found they are almost sharing the same abstract architecture as shown in Fig. 22a. According to the path of the airflow passed through a server, we could divide temperatures into five points from server front to rear: 1)  $T_{inlet}$ , the point is closed to the cold aisle; 2)  $T'_{inlet}$ , the point is at the position where airflow enters into the server fan areas; 3)  $T_{LA}$ , the point that is closed to processors; 4)  $T'_{outlet}$ , the point locates at the back in which airflow leaves through the CPU, and  $T_{outlet}$  that is closed to hot aisle. We evaluate the temperature distribution on a real server as shown in Fig. 22b. In the figure, we can see: 1) although there are  $3^{\circ}\text{C}$  to  $4^{\circ}\text{C}$  differences among the temperature points that are classified, the measured temperatures are in accord with the hypothesis of airflow path; 2) there are  $5^{\circ}\text{C}$  variations in different positions. The following discussion will be based on the abstract server model and the observations of temperature distribution.

**4.3.1 Modeling.** Before starting the thermal activity description, we suppose that the airflow through one server is an aggregate of the flows generated by all of the fans. Thus, a steady-state description (Eq. (5)-(12)) of the relationship between a server and its running environment inside the TDC can be represented by many compiled effects factors (e.g., CPU core temperature, server



(a) Rack-based server architect



(b) A snapshot of heat distribution

Fig. 22. An abstract rack-oriented server model that is widely used in a modern data center that has hot-aisle and cold-aisle. (a) Top view of heat transferring through a rack server. The airflow blowing from cold aisle enters into the server front and then is absorbed by server fans. (b) A snapshot of the heat distribution graph of HPE server [10] when the server is idle, and the inlet temperature is 33°C in our test bed.

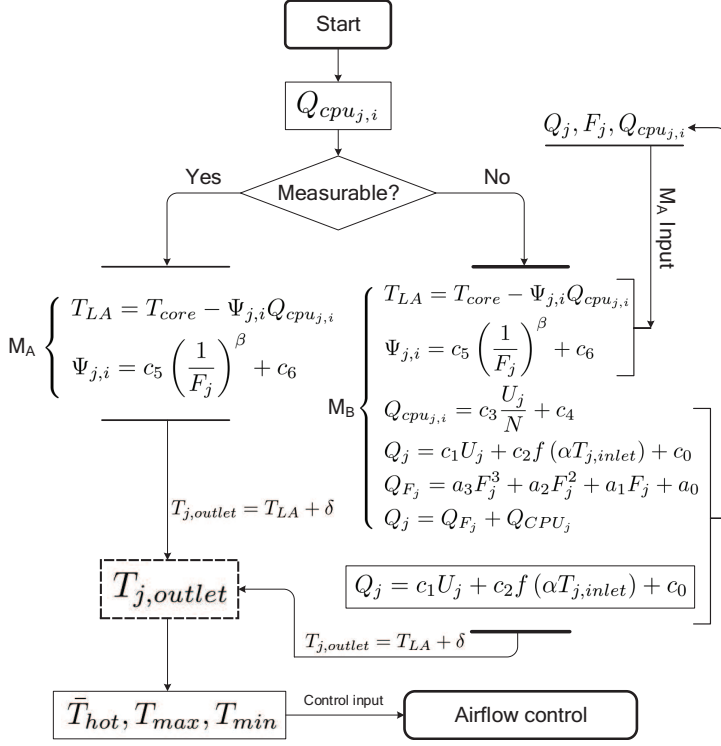


Fig. 23. Application based on the measurement of  $Q_{cpu,j,i}$

fan speed, server power consumption and CPU power consumption). The steady state can be described by the following equations.

$$\dot{V}_{total} = K \sum_{j=1}^N V_{j,req} \quad (5)$$

$$T_{j,outlet} = T_{LA} + \delta \quad (6)$$

$$T_{LA} = T_{core} - \Psi_{j,i} Q_{cpu,j,i} \quad (7)$$

$$Q_{cpu,j,i} = c_3 \frac{U_j}{N} + c_4 \quad (8)$$

$$\Psi_{j,i} = c_5 \left( \frac{1}{F_j} \right)^\beta + c_6 \quad (9)$$

$$Q_j = c_1 U_j + c_2 f(\alpha T_{j,inlet}) + c_0 \quad (10)$$

$$Q_{Fj} = a_3 F_j^3 + a_2 F_j^2 + a_1 F_j + a_0 \quad (11)$$

$$Q_j = Q_{Fj} + Q_{CPUj} \quad (12)$$

The equations listed in Eq. (5) to Eq. (12) are briefly described as following:

- Eq. (5) is used to assess the total airflow volume rate  $\dot{V}_{total}$  supplied to a room in which there are N servers installed. It could be calculated by the sum of the airflow rate required to cool heat load generated by each server  $j$ .

- Eq. (6) is that we used to estimation the outlet temperature of a  $j$  server. In view of the complex situation in a server (Figs. 22a and 23), a penalty  $\delta$  is introduced to compensate for the prediction errors between  $T'_{j,outlet}$  and  $T_{j,outlet}$ . A rack server that is small in size and low in cost has a relatively small number of components than high-end servers. In this case, CPU is the dominant heat source in a server, hence,  $T_{LA}$  can be a proxy of  $T_{j,outlet}$ , i.e.,  $\delta \rightarrow 0$ .
- As shown in previous sections, the server power consumption is not only depended on CPU usages but also on the inlet temperature that has a relationship with server fans. Thus, we modeled the server power consumption with Eq. (10), where  $Q_j$  is the total heat load of a server  $j$  when the CPU utilization is at  $U_j$ ;  $c_0$  could be considered as the power consumption when the server is idle;  $c_1$  and  $c_2$  are constant coefficients; Therefore, considering inlet temperature that contributes to server power (in essence, it is related to server fan power), we introduce inlet temperature affecting function  $f(\alpha T_{j,idle})$ , where  $\alpha$  is defined as a curve shape like previous research [22]. This change causes a little difference from the model, which only correlates the server power to CPU utilization in the earlier studies [6, 8, 17, 22].
- Eq. (8) is the model of CPU power consumption, in which  $Q_{cpu,i}$  is the utilization of a CPU  $i$  in a server  $j$ ; scope  $c_3$  is a constant coefficient, and  $c_4$  is the consumption when a CPU is idle.
- Eq. (7) is employed to estimate the server outlet temperature, where  $\Psi_i$  is the case-to-local ambient thermal characterization parameter<sup>1</sup>(ATCP) of the  $i$ th CPU with a heat sink within a server  $j$ ;  $T_{LA}$  is the average ambient temperature entering the processor heat sink fin section, and  $Q_{cpu,j,i}$  is the heat transferred per unit of time between a single CPU and the local ambient air.
- In Eq. (9), ATCP  $\Psi_i$  is related to server fan speeds.  $c_5$  and  $c_6$  are constants related to the fluid and material properties of the air, the CPU package and the heat sink. The parameter  $\beta$  also defines the shape of the ATCP curve that is a function of the airflow rate [22]. Referring to ASHRAE [3], heat sink convective thermal performance is proportional to the inverse of airflow, thus, the  $\beta$  is set to  $-1$  in this paper.
- Fan power consumption (Eq. (11)) is approximately a cubic function of the rotational speed of the rotor given in revolutions per minute (RPM) [22]. Therefore, parameters  $a_0$ ,  $a_1$  and  $a_2$  can be identified by curve fitting using data samples.

Eq. (7) to Eq. (9) are for servers in which CPU power is measurable while Eq. (10) to Eq. (12) are for those that CPU power cannot be measured. Thus, applications could be divide into two models:  $M_A$  and  $M_B$ , as seen in Fig. 23. In  $M_A$ , intuitively, it is simple to work out  $T_{j,outlet}$  using  $Q_{cpu,j,i}$ ,  $T_{core}$  and server fan. In these three variables,  $Q_{cpu,j,i}$  is not always available to measure fro all CPU models. However only CPUs (e.g., *Intel Sandy Bridge* and later series processors) that implement the RAPL (Running Average Power Limit) interface can directly get power readings with finer time resolution and higher accuracy. Thus,  $M_B$  is considered as a solution to the conditions when CPU power cannot be measured. Therefore, Eq. (10), Eq. (11) and Eq. (12) are introduced and jointly combined to estimation  $Q_{cpu,j,i}$ . In fact, the process of  $M_B$  reveals the physical relationship between the  $T_{j,outlet}$  and CPU power consumption, which reflects the thermal change due to workloads [11]. It also show the server fan activities correlated to the inlet temperature via  $\Psi$ , as show in the equation system Eq. (5) to Eq. (12).

**4.3.2 Parameters identification.** In this part, we describe the details of how the parameters were identified and also demonstrate how to employ the servers sensors to collect necessary data to

<sup>1</sup><https://www.intel.com/content/dam/www/public/us/en/documents/guides/xeon-scalable-thermal-guide.pdf>

determine the parameters. Some methods of parameters identification are similar to that used in [22]. For comparison and validation, two types of models are employed. Model1 ( $M_1$ ) stand for relatively cheaper in the market, and Model2 ( $M_2$ ) represents a high-end server. In our model, all the parameters and coefficients mentioned above are divided into two categories: 1) the ones that can be directly measured, and 2) the other ones that could be only calculated by curve fitting.

The measurable parameters including:

- **CPU temperature  $T_{core}$ :** Due to the widespread availability of on-die temperature sensors, CPU core temperature can be obtained with no additional instrumentation. On-die counters are exposed exclusively through Model-Specific Registers (MSRs) that are only readable from kernel space. Nevertheless, many tools can be used to read the MSR in the kernel and make the values accessible to applications, such as `msr-tools` in Linux, which can read the MSR in a million seconds.
- **CPU utilization  $U_j$ :** As OS provides a dynamic real-time view of a running system, such that it is easier to collect  $U_j$  with no additional drivers, for example, on Linux platform, OS comes with various utilities to report CPU utilization, such as `top` and `sar`.
- **Server inlet temperature  $T_{j,inlet}$ :** The inlet temperature of servers is one of the most critical parameters in the operation of a data center. It should effectively and consistently meet the manufacturers' server inlet temperature requirements. For the sake of safety, in modern rack-based servers, vendors are commonly provided inlet thermal sensors to collect real-time state, and also offer accessibility services by implementing IPMI interface. Previous research has shown the variation of temperature could incur the reliability problems of server components [5]. Therefore, based on their observations, we suppose that to operating a data center in a normal way, the inlet temperature of a server should not be changed frequently, then, the query response time and time resolution of this value via `IPMITools` is accurate and fine enough for our models.
- **Server power consumption  $Q_j$  and fan speed  $F_j$ :** Since these two variables are available in modern rack-oriented servers, we firstly use the tools like `IPMITools` to obtain sample data. When we work out the coefficients of the corresponding models discussed in Section 4.3, then, they can be dynamically calculated from the models.
- **CPU power consumption  $Q_{cpu,j,i}$ :** During the tests, we employ RAPL [18] to cap CPU power, because all of the tested servers support RAPL, thus, we can get the real-time power consumption from corresponding MSRs using `Turbostat`<sup>2</sup>. Some other tools are also available such as `Power Governor`<sup>3</sup>.

Coefficients identification for Eq. (7) to Eq. (12):

- **Server power consumption model:** We run a series of experiments on our testbed to identify the coefficients of Eq. 10. In the beginning, the inlet temperature was set to 25°C, and the CPU utilization was set to specific value started from 10% for 1 hour, and then pushed to next setpoint for another period of 1 hour until the CPU utilization is up to 90% while we always reserved 10% for our sampling tools. The experiments were repeated for temperatures between 25°C and 35°C, with step of 1°C. Metrics parameters, such as the CPU utilization, server power, inlet temperature and CPU power, were collected during the tests and utilized in later experiments to determine the coefficients of other equations. After the experiments were finished, curve fitting method (`curve_fit` in python) was used to identify  $c_1$  and  $c_2$ ;  $c_0$  is obtained by leaving the server idle; and the shape  $\alpha$  of the server power is also obtained by curve fitting as shown in Fig. 24 for Model1 and Fig. 25 for Model2, and it can also be tuned

<sup>2</sup><https://github.com/torvalds/linux/tree/master/tools/power/x86/turbostat>

<sup>3</sup><https://software.intel.com/en-us/articles/intel-power-governor>

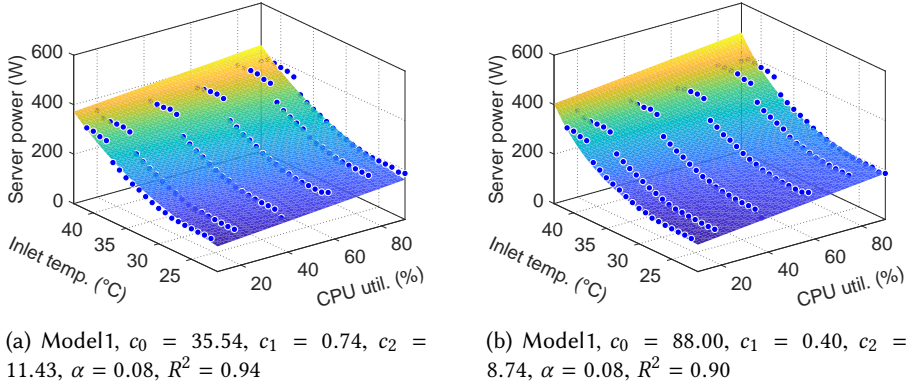
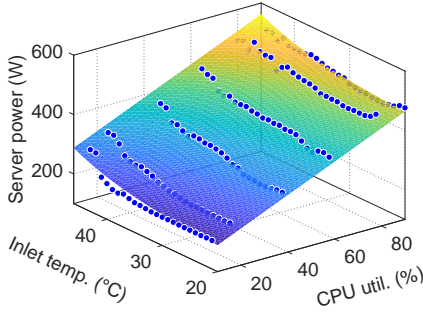


Fig. 24. Model1, server power as a function of inlet temperature and CPU utilization.

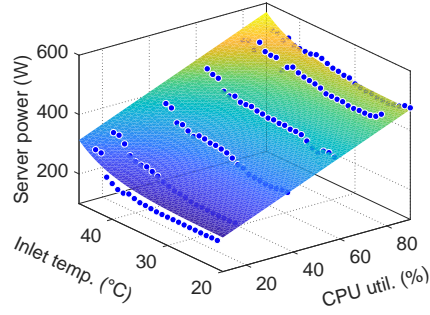
by annually setting the proper value based on the curve, as shown in Fig. 24b and 25b. As listed in Table 5, we provide several characteristic functions to observe the effects of  $T_{j,inlet}$  on server power in different forms. As seen from the table, by properly choosing  $e^{\alpha T_{j,inlet}}$ , the  $R^2$  of both server Model1 (0.90) and Model2 (0.97) can fit experimental data well. It indicates the server power has a positive correlation with inlet temperature and capture the potential relationship on them. Moreover, we also check that most of the servers in testing are complying with the standard of ASHRAE 3. That suggests the maximum inlet temperature of a server is limited up to 45°C. From Fig. 24 and 25, the experiments already cover the inlet temperature range of servers, therefore. Jointly considering the linear relationship with CPU utilization, this model can be used to predict server power when the inlet temperature is from 25°C to 45°C, and CPU utilization is from 0% to 100%.

- **Processor power consumption model.** The CPU power can be represented as a linear function of the CPU utilization, as shown in Eq. 8 in which the slope  $c_3$  and intercept  $c_4$  are derived from experimental data collected previously. Fig. 26a illustrates the results of curve fitting and shows fine goodness of Model1 ( $R^2 = 0.99$ ) and Model2 ( $R^2 = 0.99$ ). We can clearly observe the linear correlation between the CPU power and utilization.
- **CPU thermal capacity model.** Most of the vendors equipped their servers with many kinds of sensors (e.g., the thermal sensors for I/O Zone and PCI Zone of a server, as shown in Fig. 22b) to measure the server outlet temperature  $T_{j,outlet}$ . If such a kind of sensors exists, we approximately let  $T_{j,outlet} = T_{LA} + \delta$ . Otherwise, we can use the sensor installed in the rear of racks as a proxy of  $T_{LA}$ . Thus the thermal resistance can be identified by using the relationship described in Eq. 7. Since the fan speeds  $F_j$  are known, we can approximate these two coefficients through curve fitting. Fig. 27a shows the results of Model1 and Model2 when  $\beta$  is assigned to 1, which is suggested by ASHARE. The variable  $\beta$  can affect the shape of curves and can be tuned manually. Fig. 27a gives different behaviors when  $\beta$  is at several possible values.

**4.3.3 Validation of  $M_A$  on Testbed.** As shown in Fig. 23,  $M_A$  is a scenario where a CPU vendor provides its products with a mechanism of power monitoring (e.g., Intel RAPL). In this scenario, we only need to identify all the coefficients of equations described in Eq. 7-9. Then we can validate the  $M_A$  by performing a comparison of predicting outlet temperature and real data. To gather data related to  $M_A$ , we do several experiments by changing CPU utilization and cold aisle temperature as Table 1 listed. The experiments can be divided into two groups: 1) case 1, in which CPU utilization is

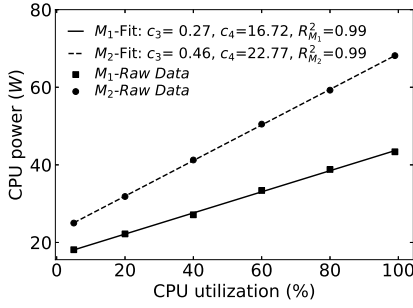


(a) Model2,  $c_0 = 144.21$ ,  $c_1 = 3.43$ ,  $c_2 = 1.81$ ,  $\alpha = 0.09$ ,  $R^2 = 0.98$

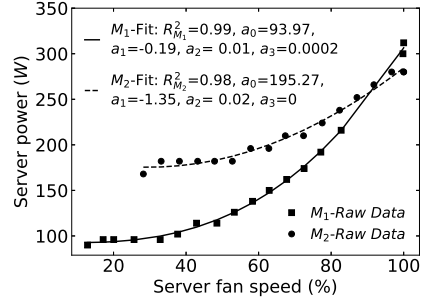


(b) Model2,  $c_0 = 182$ ,  $c_1 = 3.20$ ,  $c_2 = 0.03$ ,  $\alpha = 0.18$ ,  $R^2 = 0.97$

Fig. 25. Model2, server power as a function of inlet temperature and CPU utilization.



(a) Processor power as a function of CPU utilization.



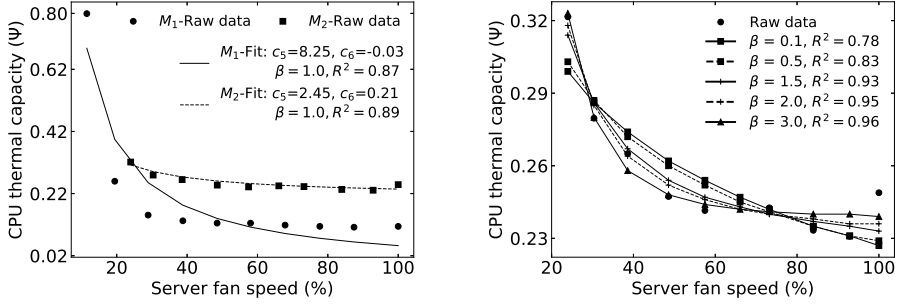
(b) Server fan power as a function of server fan speeds.

Fig. 26. Parameters identification of CPU and fan power model on server Model1 ( $M_1$ ) and Model2 ( $M_2$ )

manually set in order of 10%, 20%, 40%, 60%, 80%, and 90%, and maintain the cold aisle temperature at a specific point that should be in the range of [25°C, 37°C] with each 2°C; 2) case 2, in which CPU utilization is set to a constant level while varying temperature from 25° to 37°C with a step size of 2°C. Fig. 28 and Fig. 29 show the prediction results of two server models on these two cases.

Table 5. Characteristic functions for Eq. 10

$f(\alpha T_{j,inlet})$	Model1					Model2				
	$c_1$	$c_2$	$\alpha$	$R^2$	RMSE	$c_1$	$c_2$	$\alpha$	$R^2$	RMSE
$e^{\alpha T_{j,inlet}}$	3.20	0.18	0.02	0.97	18.62	0.76	22.55	0.06	0.94	20.17
$\alpha T_{j,inlet}$	3.23	0.32	1.62	0.93	27.55	0.44	3.10	2.16	0.72	43.43
$(\alpha T_{j,inlet})^2$	2.75	3.81	0.19	0.81	46.77	0.72	0.56	0.57	0.92	22.58
$(\alpha T_{j,inlet})^3$	1.60	2.19	0.13	0.14	99.79	0.77	0.76	0.19	32.37	0.84
$\alpha \ln(T_{j,inlet})$	3.41	-2.06	-0.68	0.93	28.92	0.63	10.60	12.73	68.18	0.30



(a) CPU thermal capacity as a function of server fan speeds.

(b) CPU thermal capacity under different curve shapes.

Fig. 27. Parameters identification

*Validation in Case 1.* Fig. 28 is the testing results when CPU utilization is separately set to 10% and 90%. The sub-figures (a) (b) (c) and (d) always show seven cycles. Each cycle represents a cold temperature setpoint that was testing for one day. The total seven setpoint, which appear in the same order in tests, are, from left to right: 25°C, 27°C, 29°C, 31°C, 33°C, and 37°C. For the sake of comparison, we separate each sub-figure into two parts: the top and bottom part. The top part of each sub-figures shows the real-time outlet temperatures (Ground Truth, *GRD-OTemp*) of servers and the predicted outlet temperature (Estimated, *EST OTemp*) in  $M_A$ . The bottom part shows two curves: one is for the server inlet temperature, and another is for a server fan activity.

As seen from Fig. 28a and Fig. 28c, when CPU utilization is maintained at 10%, the predicted outlet temperatures of both two models can properly follow the trends of real-time data. The results show that the root mean square error (RMSE) of the outlet temperature obtained by the  $M_A$  is less than 1.3°C. When the CPU utilization is pushed to 90%, the prediction curve can track the real data line with RMSE almost equal to 1.2°C in Model1, but the RMSE of Model2 is about 3.6°C. The possible reason lies in that the higher CPU utilization causes the environment inside a server changed a lot, especially the temperature. This is indicated from Fig. 28c and 28d where the fan activity of Model2 in Fig. 28d is almost close to 100% compared with 76% of Model1. Therefore, Due to the higher temperature of servers, more components are affected by the higher CPU temperature. It is also hard for  $M_A$  to make a more accurate prediction in this kind of server.

*Validation in Case 2.* Fig. 29 is the testing results at the temperatures of 25°C and 37°C separately. The sub-figures (a) (b) (c) and (d) also show six cycles. Each cycle represents a CPU utilization setpoint that was testing for one day. The setpoints, which appear in the same order in tests, are, from left to right: 10%, 20%, 40%, 60%, 80%, and 90%. The top part of each sub-figures also show the real-time outlet temperatures of servers and the predicted outlet temperature by using  $M_A$  model. The bottom part shows two curves: one is for the server inlet temperature, and another is for a server fan activity.

As shown in the bottom of all sub-figures in Fig. 29, the inlet temperature almost tracks the setpoints at 25°C and 37°C. We can also see the predicted trend of Model1 have a growth as the CPU utilization increased, which can be qualified by  $d_{25} = 0.37$  and  $d_{37} = 0.23$  ( $d_t = \frac{OT_{90} - OT_{10}}{OT_{10}}$ ), but the real-time outlet temperatures are not significantly affected by CPU utilization with  $d_{25} = 0.04$  and  $d_{37} = 0.02$ . One of the reasons is that the predicted model ( $M_A$ ) assumes the power of CPU is following the liner relationship with CPU utilization, and another possible reason has fully relied on server fan control. The control algorithm might be varied vendor by vendor. To dig out more

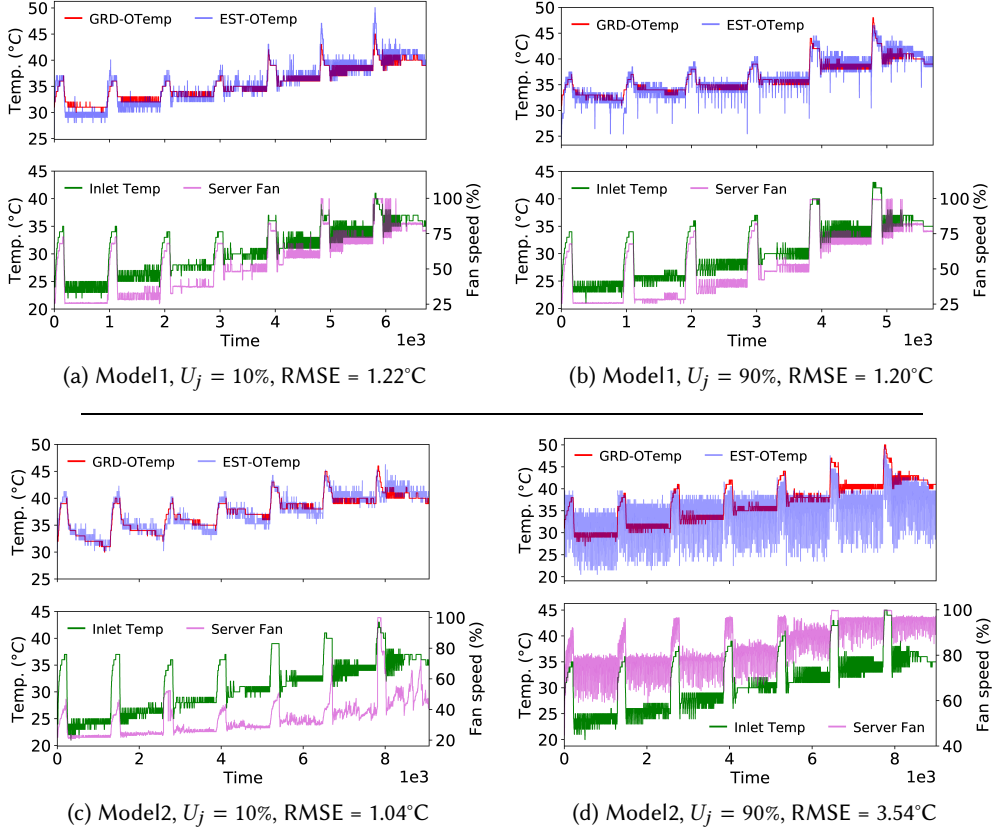


Fig. 28. Case 1, using  $M_A$  described in Fig. 23 to approximately estimate server outlet temperature while keeping CPU utilization at 10% and 90% respectively.

details of a server fan control is out of the scope of this paper. We are more concerned with server fan activity that leads to outlet temperature prediction.

The predicted results also show that when the inlet temperature of Model1 is at 25°C, the RMSE is about 3.41°C, which has 1.41°C higher than that of 37°C. Compared to the activities of Model1, we can see the ground truth temperature of Model2 in Fig. 29c has little drops when CPU utilization is at 10% with  $d_{25} = -0.09$ . We can find the answer to this dropping from the server fan trend at the bottom of the same figure. As seen from this figure, once the CPU utilization is up to 60%, the server fan speeds suddenly increase up to 65%, and the  $d$  indicator of the real-time outlet temperature is about -0.09 compared with 0.02 at 37°C in Fig. 29d, in which the real outlet temperature has a very little increase with  $d_{37} = 0.05$ . Recalling the  $M_A$  model, it assumes the outlet temperature is higher than the inlet temperature. That's why when the inlet temperature at 25°C, the RMSE has a higher RMSE. Even this, we can estimate the outlet temperature with fluctuations within 4°C for this server.

In summary, from our experiments, we can perform outlet temperature prediction using the  $M_A$  model with an average RMSE of 1.75°C of Case 1 and 3.02°C of Case 2. By referring to the difference in vendors' server layout, there exists some variation of prediction errors. Owing partly to the fact of fewer numbers of servers ( $N$ ) and air leakage in our testbed, for  $M_A$  and  $M_B$ , the

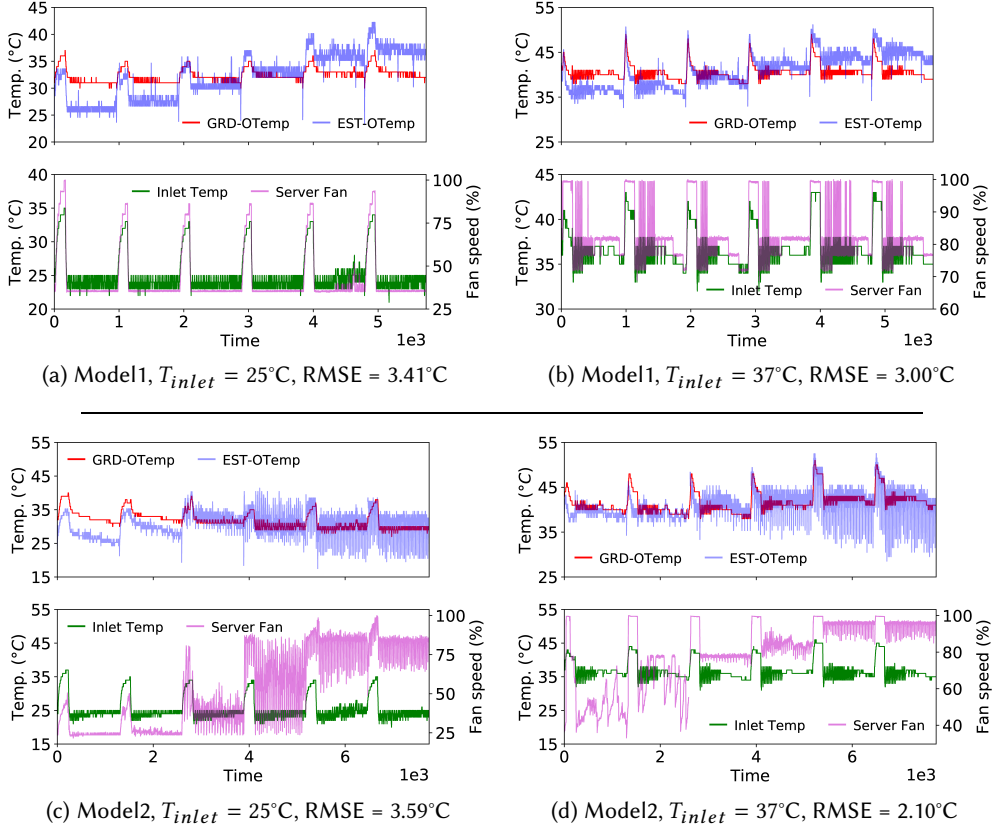


Fig. 29. Case 2, using  $M_A$  in Fig. 23 to approximately estimate the outlet temperature while keeping cold aisle temperature at 25 $^{\circ}\text{C}$  and 37 $^{\circ}\text{C}$  respectively.

outlet temperature also might be affected by air recirculation from the cold aisle, which would be one of the factors that lead to prediction accuracy. However, our approach proposed a proper solution for the out temperature prediction between 25 $^{\circ}\text{C}$  and 37 $^{\circ}\text{C}$  with a mean of 2.34 $^{\circ}\text{C}$ .

**4.3.4 Validation of  $M_B$  on Testbed.** We also valid the prediction approach using  $M_B$  (shown in Eq.10-12) with same settings as  $M_A$ . Firstly, set CPU utilization at the specific level of 10% and 90% respectively, then we observe the outlet temperature by changing inlet temperature from 25 $^{\circ}\text{C}$  to 37 $^{\circ}\text{C}$  with a step size of 2 $^{\circ}\text{C}$ . Each setpoint lasts for one hour. After that, inversely, change the setting order by independently maintaining inlet temperature at 25 $^{\circ}\text{C}$  and 37 $^{\circ}\text{C}$ . At each temperature setpoint, we set CPU utilization to the level of 10%, 20%, 40%, 60%, 80%, and 90%. There is a difference from the figures in the previous section for the validation of  $M_A$ . In each sub-figures of Fig. 30 and Fig. 31, we also show the server fan prediction in the bottom part of the corresponding sub-figures.

**Validation in Case 1.** In Fig. 30, when the CPU utilization is at the level of 10% of both Model1 and Model2, the predicted values can properly follow the real-time outlet temperature (see Fig. 30a and Fig. 30c). The results of these two figures show that the RMSE of the outlet temperature obtained by the  $M_B$  is less than 1.3 $^{\circ}\text{C}$ . However, when the CPU utilization is pushed to 90%,  $M_B$  has a different behavior. In Model1, the RMSE is 1.22 $^{\circ}\text{C}$  compared with 3.81 $^{\circ}\text{C}$  in model2. An obvious observation

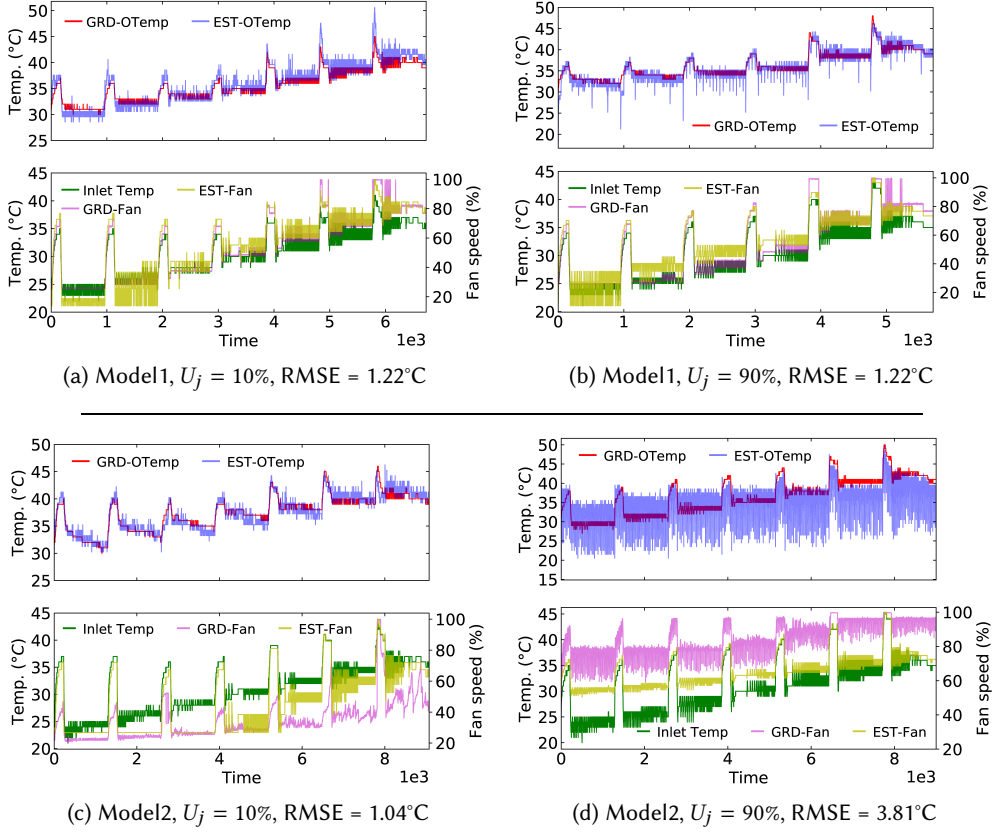


Fig. 30. Case 1, using  $M_B$  in Fig. 23 to approximately estimate the outlet temperature of Model1 and Model2 while keeping CPU utilization at 10% and 90% respectively.

in Fig. 30d is that the  $EST - OTemp$  prediction curve of Model2 has obvious jitters compared to the same level of Model1. The reason can be mainly ascribed to the server fan control due to more heat generated from CPU, as seen from Fig. 30d.

**Validation in Case 2.** The sub-figures of Fig. 31 are the results of testing at the 25°C and 37°C respectively. As discussed in previous sections, Case 2 fixes the inlet temperature. Thus  $M_A$  presents a positive correlation with CPU utilization. That's because there is a relationship between  $T_{LA}$  and CPU utilization. Due to the assumption,  $M_B$  also shows this trend in Fig. 31. However, Fig. 31c shows a difference from Fig. 31a, Fig. 31b and Fig. 31c. Intuitively, a little drop is shown in Fig. 31c, but the  $d$  of it is only about -0.19°C. This little drop could be explained by the server fan activities. As shown in the bottom part of Fig. 31c, when the CPU utilization is up to 60%, the server fan also goes up to 55% while the predicated server fan is about 70%. Then we know the growth is about 25%-30%. It offsets the effects of prediction with  $M_B$ . That's the reason that when the inlet temperature is at 25°C, the outlet temperature remains almost the same. Therefore, it causes the max variation ( $RMSE = 4.27^\circ\text{C}$ ) compared to other sub-figures in Fig. 31.

In summary, we can observe from the results that the same server model from the same vendor has almost the same outlet temperature characteristics. For example, typically in Model1, the difference between the minimum temperature of 25°C and the maximum of 37°C is 0.37°C. Therefore,

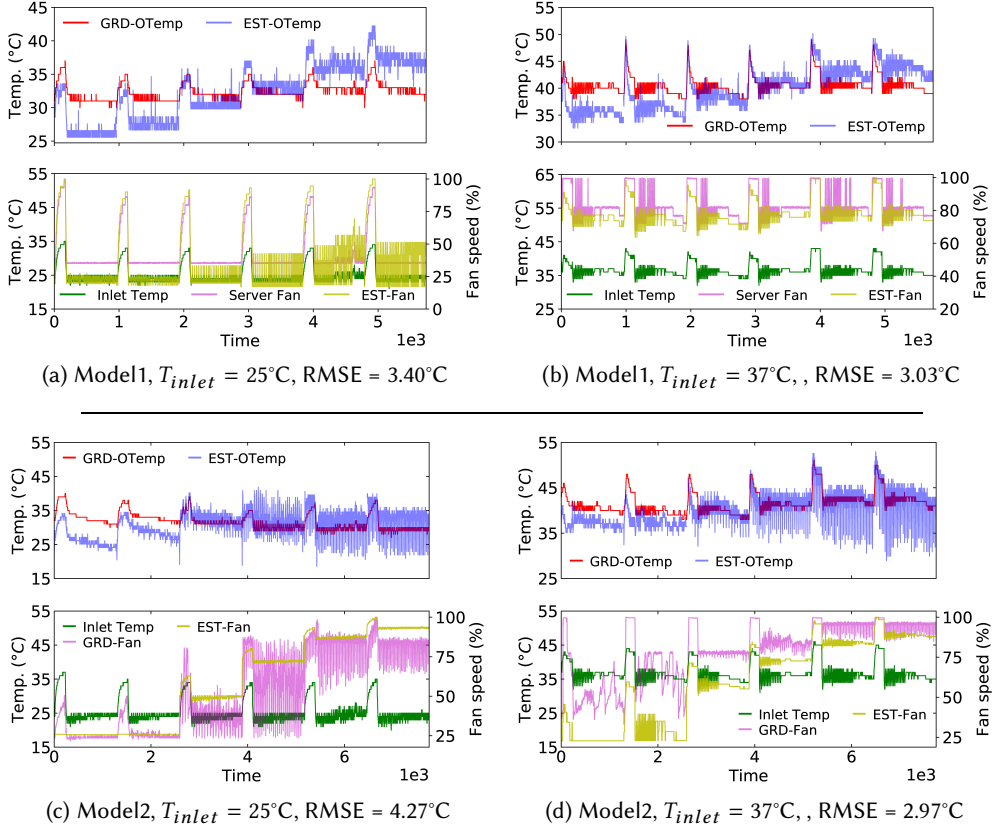


Fig. 31. Case 2, using  $M_B$  in Fig. 23 to approximately estimate the outlet temperature of Model1 and Model2 while keeping cold aisle temperature at  $25^{\circ}\text{C}$  and  $37^{\circ}\text{C}$  respectively.

we can infer that the prediction of other temperatures in this range has the same trend. Although the RMSE is relatively higher, once we obtain characteristics of a specific server, then we can manually add compensation ( $\epsilon = 3^{\circ}\text{C}$  for Model1 and Model2 in Fig. 31) to make up for the variation. Thus we could get a more accurate prediction.

**4.3.5 Summary and Discussion.** The outlet temperature of servers is directly related to heat exchange in the DC room. Inversely, it is also the point that could reflect the dynamic room temperature states. In the tropical air-side free cooling environment, this is also the necessary and essential parameter to track. On the other hand, we also rely on the outlet temperature to perform server states steering and airflow control. In modern state-of-art rack-based servers, most of the vendors offer thermal sensors that are built into their products. Thus users who are interested in the value can interface with the onboard manage system (e.g., iBMC of Huawei, iLO of HPE, iDrac of Dell, etc.) to query the real-time readings. Nonetheless, there are also some servers that they do not provide such kinds of sensors, such that we have to resort to the model approaches to build a model to estimate the temperature appropriately.

However, it might take much longer to query the management system (e.g., BMC) by using a protocol like *IPMI*. Then, the incurred problem could be that the querying response time is not satisfying the requirements of fine airflow control. Besides, as the CFD is widely used in modern

data center design and optimization, it also requires a higher resolution of outlet temperature. It also drives us to build such numeric models to predict outlet temperature based on the given information, which could be treated as fixed such as inlet temperature and the high-resolution parameters including CPU core temperatures and CPU utilization. The models adopted in this paper could be utilized as a general method for servers in the data center. Even if we only validated our model approaches on two types of servers, they could be easily extended to the others. Benefitting from these models, we could implement more fine airflow control and CFD model tuning.

## 5 AIR-SIDE FREE-COOLED DC CONTROL AND COST SAVINGS

In this section, we first present the configuration and results of the ventilation tests which use only fans to cool servers. Then, we present a deep reinforcement learning (DRL)-based advanced control algorithm with a main objective of minimizing the total cooling energy consumption while maintaining the supply air temperature and RH below desired thresholds. The energy saving of each approach is also evaluated.

### 5.1 Ventilation Control and Energy Saving

An air-side free-cooled DC mainly leverages room fans to blow the outside air into the server room to remove the heat generated by IT equipment. The cooling devices such as water-chiller based cooling systems and compressor-based air conditioners are turned off by default and turned on only when required. Therefore, we conduct ventilation tests to investigate the server's behaviors and energy savings when only room fans are used to cool the servers.

**5.1.1 Fixed ventilation.** We run servers with fixed ventilation whose goal is to provide a sufficient and constant air volume flow rate to make sure no overheating, regardless of outside condition and server workload. Specifically, we maintain an air flow rate at  $5000 \text{ m}^3/\text{h}$  in an uncontrolled environment condition where cooling coils and heaters are turned off. All servers are fully utilized with 100% CPU utilization level.

Fig. 32 shows 30-days' trace of maximum cold and hot air temperatures under the fixed ventilation. Note that the cold air temperature is close to the outside air temperature since the cooling coils and heaters are not used. From Fig. 32, the maximum hot air temperature is always less than  $43.6^\circ\text{C}$  over a period of 30 days. With a fixed air flow rate of  $5000 \text{ m}^3/\text{h}$ , no server's overheating is found as the maximum cold air temperature is always within a range of  $[28^\circ\text{C}, 34.7^\circ\text{C}]$ . Note that the server will be automatically shutdown to avoid overheating once the inlet temperature reaches to a threshold of  $45^\circ\text{C}$ .

**5.1.2 Adaptive ventilation.** To minimize the fan energy consumption, we develop and implement a control logic to provide right enough and dynamic air volume flow rate to the servers, in response to outside condition and server workload changes. Specifically, the room air flow rate is controlled such that the maximum hot aisle air temperature does not exceed a threshold.

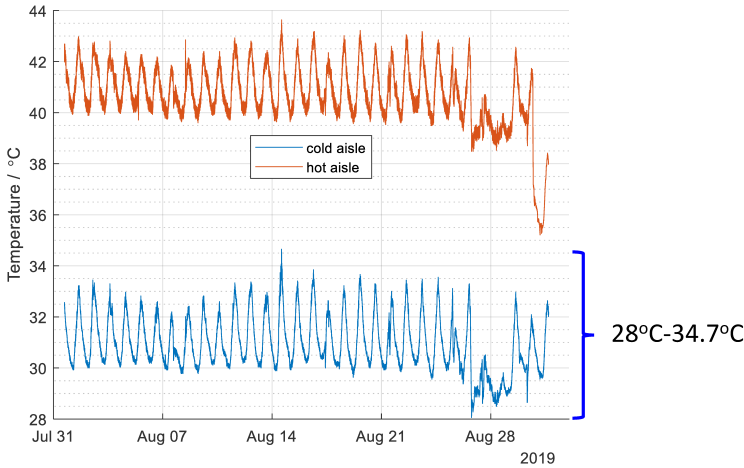


Fig. 32. Maximum cold and hot air temperatures under fixed ventilation.

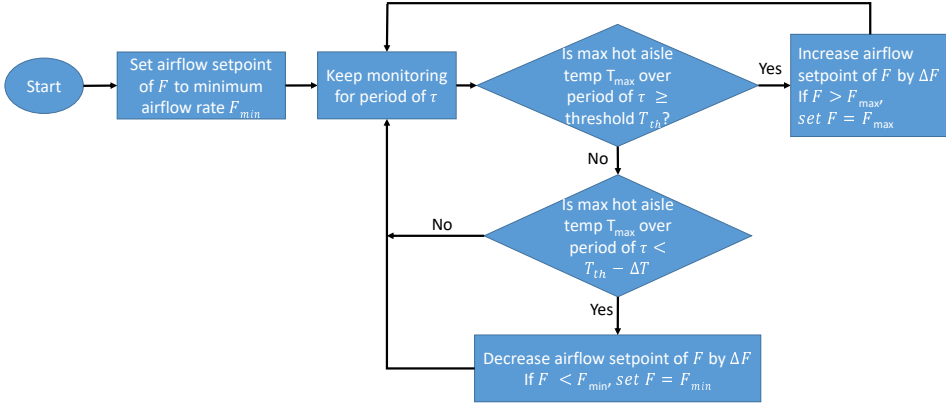


Fig. 33. Adaptive ventilation control logic.

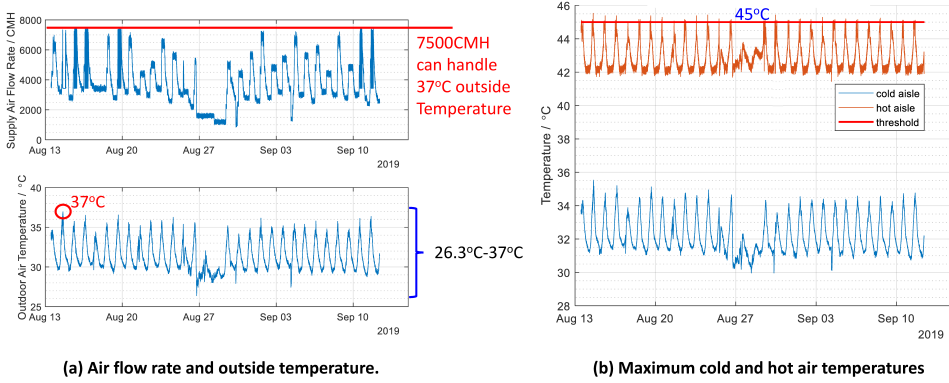


Fig. 34. Execution results of adaptive ventilation.

Fig. 34 shows the flowchart of the control logic. The air flow rate, denoted by  $F$ , is initially set to a minimum value  $F_{min}$ . At the beginning of every period of  $\tau$ , if the current maximum hot aisle temperature, denoted by  $T_{max}$  is higher than a threshold of  $T_{th}$ ,  $F$  is increased by a certain step  $\Delta F$ . In case  $T_{max} < T_{th} - \Delta T$ , where  $\Delta T \geq 0$ ,  $F$  is decreased by  $\Delta F$ . Otherwise,  $F$  remains at its current value. Note that the air flow rate can be increased only up to the maximum air flow,  $F_{max}$  that the room fans of the TDC testbed can provide.

Fig. 34 shows the execution result of the adaptive ventilation test under the developed control logic over a 30-day test period with the following settings:  $\tau = 1\text{min}$ ,  $F_{min} = \Delta F = 200\text{ m}^3/\text{h}$ ,  $T_{th} = 45^\circ\text{C}$ ,  $\Delta T = 3^\circ\text{C}$ , and  $F_{max} = 12500\text{ m}^3/\text{h}$ . From Fig. 34(a), the air flow rate can be adaptively controlled in response to the change of the outside air temperature. The maximum air flow rate is  $7500\text{ m}^3/\text{h}$  when the outside air temperature reaches to  $37^\circ\text{C}$ . From Fig. 34(b), the maximum hot temperature is always maintained below  $45^\circ\text{C}$ . This suggests that the control objective is well achieved.

**5.1.3 Energy savings.** Fig. 35 shows that PUEs of controlled and uncontrolled (i.e., ventilation) tests. As shown in Fig. 35, the controlled tests have high PUEs since major energy is consumed by cooling coils and fans to maintain a certain cold air temperature setpoint. For instance, the maximum PUE is 7.81 when the cold air temperature setpoint is  $25^\circ\text{C}$  and the room air flow rate

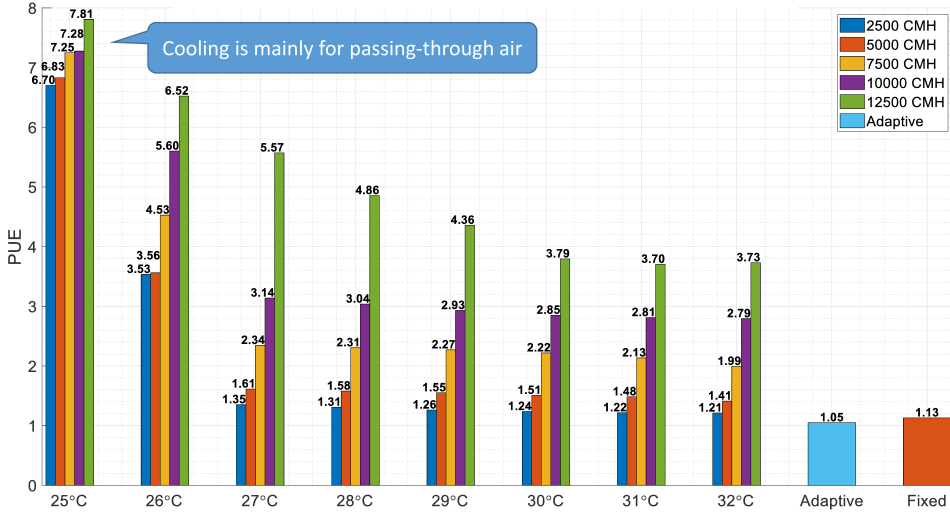


Fig. 35. PUEs under various cooling modes.

is  $12500 \text{ m}^3/\text{h}$ . The PUE greatly decreases with the increase of the cold air temperature setpoint. Compared to the controlled tests, the ventilation tests can achieve lower PUEs. Specifically, the adaptive ventilation has a very low PUE of 1.05 as shown in Fig. 35. Note that typical DCs with traditional water-chiller based cooling systems have average PUEs of 1.7 [9]. This result demonstrates that the high energy efficiency can be obtained by the air-side free cooling which uses the outside air for cooling.

## 5.2 Advanced Control by Deep Reinforcement Learning

**5.2.1 Approach Overview.** To achieve desired server's computing performance and reliability, we need to maintain the supply air temperature below the reliability thresholds of the servers and RH below a certain level (e.g., the lowest deliquescent RH of the particulate and gaseous contaminants present in the outside air). The temperature threshold can be obtained by controlling cooling coils to process the outside air before supplying to the servers. However, the RH control is a challenging task in tropics as presented in Section 2.1.1. Traditionally, dehumidification is achieved by a cooling-then-reheating process. Specifically, a certain amount of moisture is condensed out from the humid air by cooling the air below its dew point. Then, the cold air is reheated to the desired temperature. However, the cooling and reheating processes consume significant energy. In our work, to reduce the supply air RH in an energy-efficient manner, we recirculate a portion of the hot return air and mix it with the fresh outside air to form supply air. The mixing can be implemented by controlling the openness of the three dampers as illustrated in Fig. 2. Note that, without condensation, the hot return air and the fresh outside air have the same absolute humidity. From psychrometrics, the hotter mixed air will have lower RH compared with the fresh outside air. However, when the fresh outside air is hot, the hotter mixed air to achieve the desired low RH may exceed the servers' reliability thresholds. In this case, the cooling coil should be activated to reduce the temperature of the incoming air.

We develop control algorithms for the supply and exhaust fans, the cooling coil, and the dampers such that the energy consumption of the non-IT facility is minimized subject to that the temperature and RH of the air supplied to the servers are below respective specified thresholds for the

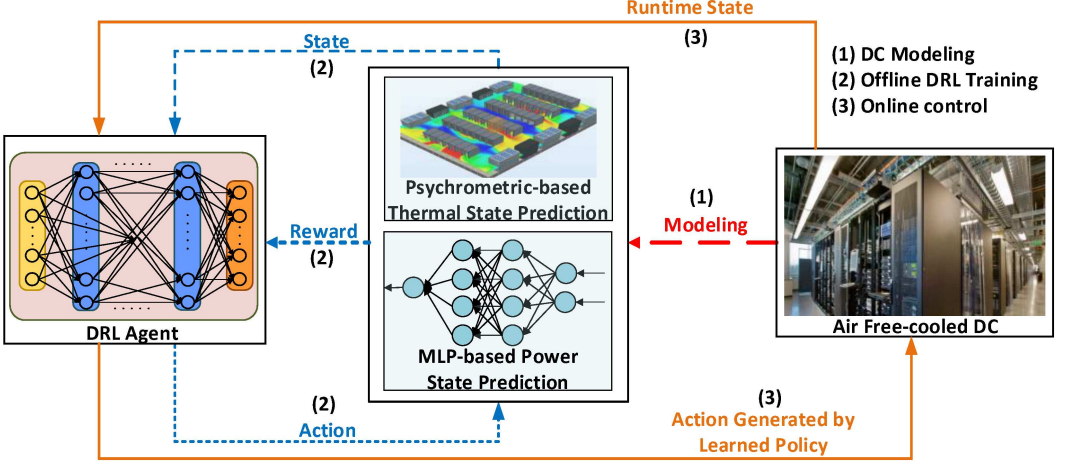


Fig. 36. Design workflow of DRL-based DC control.

sake of IT hardware reliability. The system will operate in the presence of exogenous disturbances, i.e., the time-varying ambient condition and heat from servers.

**5.2.2 MDP Formulation.** Time is divided into intervals with identical duration of  $\tau$  seconds. In this study, we consider the secondary controls (i.e., adjustment of setpoints) for the actuators. The beginning time instant of a time interval is called a *time step*. Control action is performed at every time step. Thus, the  $\tau$  is referred to as *control period*. We do not consider the details of the primary controls of the actuators; we assume that the actuators can implement the setpoints decided by the secondary controls using their closed-loop primary controls and the system has reached the steady state by the end of every control period. In practice, the setting of the control period can be chosen with the consideration of the dynamics of the primary controls to ensure the above assumption. Under the above setting, the temperature and RH of the supply air at next time step depend only on the system's state (conditions of outside and supply air, servers' powers) and the control action at the current time step. Therefore, the control problem can be modeled as a Markov decision process (MDP). We now define the terminologies of the MDP formulation.

**System state:** The system state, denoted by  $x$ , is a vector  $x = [t_s, \phi_s, p_{IT}, t_o, \phi_o]$ , where  $t$  and  $\phi$  respectively represent temperature and RH, the subscript  $s$  and  $o$  respectively represent supply air and outside air, and  $p_{IT}$  represents the total power consumption of all IT equipment in the server room. The  $p_{IT}$  determines the amount of heat generated in the server room.

**Control action:** The supply and exhaust fans admit air volume flow rate setpoints. To achieve steady state without control errors, the setpoints for the two fans should be identical; otherwise, the server room will be in the dynamic process of pressurization/depressurization or a steady state with control errors. Let  $\dot{v}_s \in [0, \dot{v}_{\max}]$  denote the air volume flow rate setpoint for the two fans, where  $\dot{v}_{\max}$  is the maximum achievable air volume flow rate. The cooling coil admits a setpoint  $\Delta t$  that represents the reduction of temperature, i.e.,  $\Delta t = t_o - t_p$ , where  $t_p$  represents the temperature of the processed air leaving the cooling coil. Let  $\Delta t_{\max}$  represent the maximum temperature reduction that can be achieved by the cooling coil. Thus,  $\Delta t \in [0, \Delta t_{\max}]$ . Let  $\alpha \in [0, 1]$  denote the setpoint for the damper system, which is the fraction of the recirculated hot air in the supply air. Thus,  $1 - \alpha$  is the fraction of the outside air in the supply air. A setpoint  $\alpha$  can be achieved by controlling the openness of the three dampers. For example, to achieve  $\alpha = 0$ , the supply and

exhaust dampers should be completely open and the mixing damper should be completely closed; to achieve  $\alpha = 1$ , the supply and exhaust dampers should be completely closed and the mixing damper should be completely open. The control action, denoted by  $a$ , is a vector  $a = [\dot{v}_s, \Delta t, \alpha]$ .

**Reward function:** When a control action  $a$  is performed at the current time step with a system state of  $x$ , let  $p(x, a)$  denote the average power consumed by the supply and exhaust fans to maintain the air volume flow rate  $\dot{v}_s$  and the cooling coil to lower the temperature by  $\Delta t$  Celsius degree over the next control period of  $\tau$  seconds; let  $t_s(x, a)$  and  $\phi_s(x, a)$  denote the supply air temperature and RH, respectively. We define a penalty function as follows:

$$q(x, a) = \lambda_1 \cdot \max(t_s(x, a) - t_{th}, 0) + \lambda_2 \cdot \max(\phi_s(x, a) - \phi_{th}, 0), \quad (13)$$

where  $t_{th}$  and  $\phi_{th}$  are the temperature and RH thresholds for the long-term reliability of the IT hardware equipment;  $\lambda_1$  and  $\lambda_2$  are configurable weights. From the definition of  $q(x, a)$ , if the supply air temperature and RH do not exceed their respective thresholds, no penalty will be applied. The immediate reward, denoted by  $r(x, a)$ , is defined as

$$r(x, a) = -p(x, a) - q(x, a). \quad (14)$$

where  $p(x, a)$  is the average cooling power consumed in the next control period. Note that the cooling power is a sum of powers consumed by cooling coils and fans i.e.,  $p = p_f + p_c$ . Thus, the reward is defined based on the weighted sum of the cooling power consumption and the degrees of supply air temperature and RH requirement violations. The impact of  $\lambda_1$  and  $\lambda_2$  on the control performance will be evaluated and discussed in Section 5.2.5.

**Air-side free-cooled DC control problem:** At every time step, the system controller observes the system state  $x$ . Then, it decides and executes a control action  $a$  to operate the supply and exhaust fans, cooling coil and dampers in the next control period of  $\tau$  seconds. At the end of the next control period, the system controller can receive an immediate reward  $r(x, a)$  as a feedback signal. The control design objective is to find a control policy that determines  $a$  based on  $x$  to maximize the expected reward over a long run, i.e.,  $\mathbb{E}[r]$ .

In general, it is difficult to design a closed-form control policy to maximize  $\mathbb{E}[r]$  because the state evolution of the system (i.e.,  $t_s(x, a)$  and  $\phi_s(x, a)$ ) is complex. Model-predictive control (MPC) is a widely adopted approach to solve MDP problems (e.g., [13] for HVAC control). However, the optimization of MPC is computationally expensive and often for a limited time horizon only. DRL is an emerging approach to deal with the above challenges. In the interactions between the DRL agent and the environment (i.e., the controlled system), the agent will learn the optimal control policy from the historical data including system states, control actions, and the resulted immediate rewards. With sufficient interactions, the DQN learned by the agent can well capture the highly complex system dynamics. Moreover, the learned control policy approaches optimality for a long time horizon comparable to the time duration of the training phase. In what follows, we will present the detailed design of our DRL system to address the air-side free-cooled DC control problem.

**5.2.3 Offline Training of DRL Agent.** We adopt the learning framework in [14] to train offline a DQN for the control agent to capture a good control policy to address the problem formulated in Section 5.2.2. Specifically, the DQN is trained through interacting with the computational model developed in Section 4.1 for  $N$  episodes, each of which consists of  $T$  time steps. An episode starts with a state chosen randomly from the training data. Then, at the  $k$ th time step, an action  $a[k]$  is selected for state  $x[k]$  according to the  $\epsilon$ -greedy algorithm [19] based on action-values given by the DQN. Given the selected action  $a[k]$ , the  $t_s[k+1]$ ,  $\phi_s[k+1]$  and  $p_{IT}[k+1]$  are estimated using the psychrometric model and IT power model ((cf. Section 4)), where the outside air condition (i.e.,  $t_o[k+1]$  and  $\phi_o[k+1]$ ) are taken from real traces. To calculate the immediate reward  $r[k]$ , powers

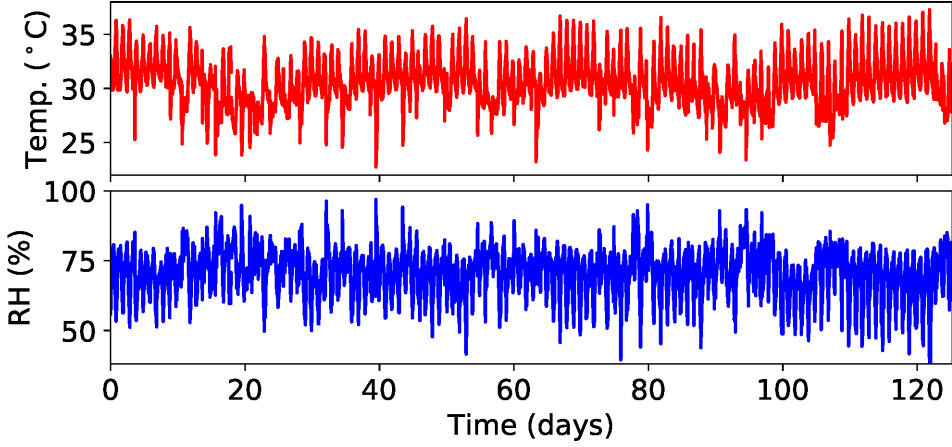


Fig. 37. Outside air condition over 125 days.

of fans and cooling coil with respect to the selected  $\dot{v}_s[k]$  and  $\Delta t[k]$  are determined using MLP2 and MLP3, respectively.

During the learning phase, two mechanisms, i.e., experience replay and target Q-network, are used to update the weights of the DQN (denoted by  $\theta$ ) every time step. For the target Q-network mechanism, we use the soft target update method [12] to update the weights  $\theta'$  of the target Q-network by setting  $\theta' = \beta\theta + (1-\beta)\theta$  with  $\beta \ll 1$ . The soft target update often gives better learning stability than the hard target update of the original DQN training.

**5.2.4 Sensor Requirement.** To run the trained DRL agent, the essential sensors include: (1) temperature and RH sensors to monitor the outside air and supply air conditions; (2) a meter to monitor the total power consumption of the IT equipment. Moreover, to implement the primary controls of the supply/exhaust fans, the cooling coil, and the damper system, we need the following sensors: (1) air volume flow rate sensors to monitor the air entering the cold aisle and the air passing the mixing damper; (2) a temperature sensor measuring the air leaving the cooling coil. To collect training data for the offline learning of DRL, meters to measure the power consumption of supply and exhaust fans, as well as the cooling coil are needed in addition to the sensors mentioned above.

**5.2.5 DRL Agent Training and Execution.** We build fully connected deep neural networks as the DQNs (i.e., the primary and target action-value functions). Each network consists of an input layer, four hidden layers and a linear output layer, where each hidden layer has 20 ReLUs. From our extensive trials, the choice of four-layer perception achieves satisfactory convergence performance for the control of the simulated testbed. The DRL agent admits a system state and chooses an action  $a = [\dot{v}_s, \Delta t, \alpha]$  from a discrete action space:  $\dot{v}_s$  is from 1000 m<sup>3</sup>/h to 5000 m<sup>3</sup>/h with step size of 500 m<sup>3</sup>/h;  $\Delta t$  is from 0°C to 10°C with step size of 1°C; and  $\alpha$  is from 0 to 1 with step size of 0.1. These step sizes are from the physical constraints of the supply/exhaust fans, the cooling coil, and the damper system. The size of the action space is  $9 \times 11 \times 11 = 1089$ . We set the RH threshold  $\phi_{th} = 65\%$ , which is the deliquescent RH of many contaminants [2]. We set  $t_{th} = 45^\circ\text{C}$ , which is the reliability temperature of ASHRAE Class A4 servers. The control period is one minute. For the offline training of the DQN, we adopt the following settings: training batch size is 64; replay memory size is 50000; discount factor  $\gamma = 0.99$ ; sort target update weight  $\beta = 0.01$ ; Adam optimizer's learning rate is 0.001; the  $\varepsilon$  of the  $\varepsilon$ -greedy method reduces linearly from 1 to 0.1.

Fig. 37 shows 125 days' outdoor air conditions of the testbed area. We use the first 95 days' data for training the DRL agent and the remaining data for evaluating the trained agent. The offline

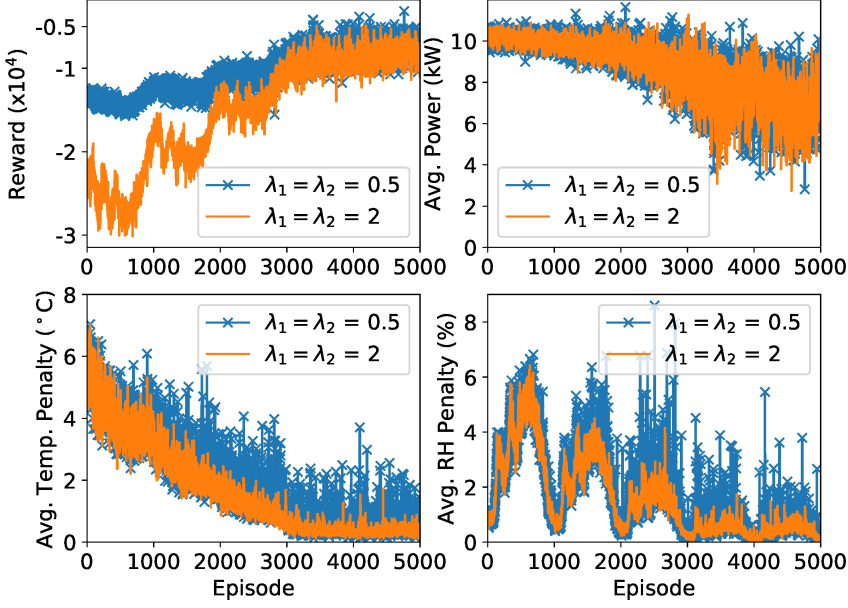


Fig. 38. DRL training convergence and penalty factors.

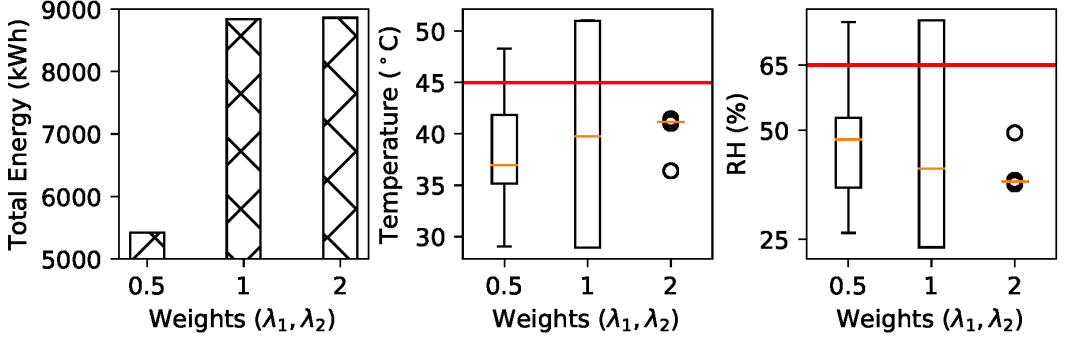


Fig. 39. Execution results under various  $\lambda_1$  and  $\lambda_2$  settings. The red lines in 2nd and 3rd subfigures represent  $t_{th}$  and  $\phi_{th}$ .

training is for  $N = 5000$  episodes, each of which consists of  $T = 1000$  control periods. At the beginning of each episode, we select a batch of 1,000 samples of outside air condition to drive the training. During the training, the system state is determined based on the action taken by the agent, the psychrometric model, and the power models in Section 4.

The weights  $\lambda_1$  and  $\lambda_2$  in Eq. (13) affect the trade-off between power consumption and compliance to the temperature/RH requirements. We evaluate the convergence of the DRL agent training under various settings for  $\lambda_1$  and  $\lambda_2$ . Fig. 38 shows the training traces of reward, average power, and average temperature and RH penalties (i.e.,  $\max(t_s(x, a) - t_{th}, 0)$  and  $\max(\phi_s(x, a) - \phi_{th}, 0)$ ) over an episode of 1000 time steps. Along the training episodes, the reward becomes flat; the power consumption has increasing variance but decreasing overall trend. Both temperature and RH penalties drop during training. With  $\lambda_1 = \lambda_2 = 2$ , the penalties are close to zero after 5,000 episodes. Differently, with  $\lambda_1 = \lambda_2 = 0.5$ , the penalties are higher. We also train the DRL agent

under various settings for temperature and RH thresholds as well as weights. The results show that the training of the DRL agent is convergent after a certain number of training episodes (e.g.,  $N = 5000$ ) with learning curves similar to those shown in Fig. 38.

We evaluate the execution of the trained DRL agent for controlling the system in trace-driven simulations over a period of 30 days. The last 30 days' outdoor air condition trace shown in Fig. 37 is used to drive the simulations. Fig. 39 shows the total energy consumption and boxplots for the distributions of supply air temperature and RH over the execution period of 30 days with the DRL agents trained with various  $\lambda_1$  and  $\lambda_2$  settings. From the 1st subfigure, the energy consumption increases with the weight. This is because, with smaller  $\lambda_1$  and  $\lambda_2$ , the agent is trained towards saving more power. However, from the 2nd and 3rd subfigures, when the two lambdas are no greater than 1, the temperature and RH may exceed their thresholds. When the two lambdas are 2, the temperature and RH do not exceed their thresholds during the 30-day test period. The above results show the trade-off between the energy consumption and the temperature/RH requirement compliance. In practice, grid search can be applied to choose the settings of  $\lambda_1$  and  $\lambda_2$  based on training and validation data, to achieve compliance of the temperature and RH requirements.

## 6 IT EQUIPMENT FAILURES AND MITIGATION APPROACHES

In this section, we present the IT equipment failures that occurred on the testbed during the tests. We also present our investigation on the reasons of the failures. Lastly, we discuss the mitigation approaches, as a recommendation to the operations of such air-side free-cooled DCs in tropical environments.

### 6.1 Summary of Failures

In Table 6 and Table 7, we summarize all the failures encountered during the reporting periods. Note that all failures occurred on servers. The networking equipment (i.e., switches and routers) have no failures.

*6.1.1 Phase 1: From July 2018 to February 2019 (8 Months).* All tested HDDs and memories work successfully during this reporting period. One CPU fault was caused by over voltage from a malfunctioned mainboard. All other CPUs work successfully during this reporting period.

We encountered server failures in Room A and Room B during the tests. Specifically, among a total of 6 servers from Vendor 1 that are deployed in Room A and Room B, four of them failed in October and November of 2018. We requested Vendor 1 to fix the failed servers and also requested fault analysis. Vendor 1 fixed them in January of 2019. Specifically, the mainboards of three servers were replaced; the CPU on one of the three servers was damaged due to mainboard malfunction and was replaced. The backplane of the fourth server with fan alarms was replaced. After the repairs, all servers from Vendor 1 operate normally. Vendor 1's lab analysis confirmed that the previous batch of faults were not caused by high temperature; they were caused by the dust and corrosion.

*6.1.2 Phase 2: From March 2019 to October 2019 (8 Months).* All tested HDDs and memories work successfully during this reporting period. All CPUs work successfully during this reporting period.

We encountered five additional server faults in this test period.

- Four of them occurred on four Vendor 1 servers that had faults before and were fixed in Jan of 2019. As mentioned previously, Vendor 1's lab analysis confirmed that the previous batch of faults were not caused by high temperature; they were caused by the dust and corrosion.

Table 6. Server Status Overview as of Feb 28th 2019.

Server Model	Room A	Room B	Room C
Vendor 1 Model1	Faulty, fixed <sup>1</sup> , operating	Operating	Operating
Vendor 1 Model2	Faulty, fixed <sup>2</sup> , operating	Operating	Operating
Vendor 1 Model3	Faulty, fixed <sup>3</sup> , operating	Faulty, fixed <sup>4</sup> , operating	Operating
Vendor 2	operating	Operating	Operating
Vendor 3	operating	Faulty <sup>5</sup>	Operating
Vendor 4	operating	Operating	Operating

<sup>1</sup>Replaced the mainboard (Jan 4th 2019)

<sup>2</sup>Replaced the mainboard (Jan 4th 2019);

<sup>3</sup>Front fan failed; Rear fan working normally; Lots of noise.

1) Replaced the faulty fan; Engineers suggested to replace the backplane (Jan 4th 2019);

2) Replaced the backplane (Jan 15th 2019);

<sup>4</sup>Replaced the mainboard (Jan 4th 2019);

<sup>5</sup>Automatically halted. (The vendor did not responded to our failure report)

Table 7. Server Status Overview as of Sep 23rd 2019

Server Model	Room A	Room B	Room C
Vendor 1 Model 1	Faulty, mainboard replaced, faulty again after 4 months <sup>1</sup>	Operating	Operating
Vendor 1 Model 2	Faulty, mainboard replaced, faulty again after 6 months <sup>2</sup>	Faulty, mainboard replaced, faulty again after about 2 months <sup>4</sup>	Operating
Vendor 1 Model 3	Faulty, fan backplane replaced, faulty again <sup>3</sup>	Faulty, fixed <sup>4</sup> , operating	Operating
Vendor 2	operating	Operating	Operating
Vendor 3	operating	Faulty <sup>5</sup>	Faulty <sup>5</sup>
Vendor 4	operating	Operating	Operating

No faults with 12 switches and 3 routers.

<sup>1</sup>Replaced the mainboard (Jan 4th 2019); failed again (May 11th 2019)

<sup>2</sup>Replaced the mainboard (Jan 4th 2019); failed again (July 16th 2019)

<sup>3</sup>Front fan failed; Rear fan working normally; Lots of noise.

1) Replaced the faulty fan; Engineers suggested to replace the backplane (Jan 4th 2019);

2) Replaced the backplane (Jan 15th 2019); Front fan failed again and the server cannot be powered on;

<sup>4</sup>Replaced the mainboard (Jan 29th 2019); failed again (March 20th 2019)

<sup>5</sup>Automatically halted. (The vendor did not responded to our failure reports)

Since no additional countermeasures against dust and corrosion were applied for the fixed servers, this new batch of server faults should have the same reason as before, i.e., they were caused by dust and corrosion.

- The remaining one fault occurred on a Vendor 3 server. As a result, all two servers from Vendor 3 on the testbed have failed. Vendor 3 has not responded to our server fault report and fault analysis request. From our investigation, the NO<sub>2</sub> concentration levels on the testbed are 5 times higher than Vendor 3's requirement. This can be a main reason of the failure.

The corruptions caused by the co-presence of corrosive gas, dusts, and considerable relative humidity were the cause of the server faults. Severe corruptions can be observed on the CD drives of the two failed Vendor 3 servers, as shown in Fig. 40. Fig. 41 shows the dusts resting on the PCBs of the motherboards of the faulty servers. Note that the vendor of the network equipment conveyed that the switches/routers have anti-corrosion coating. This can be the main reason that the network equipment on the testbed, though being exposed to the same environment condition as the servers, had no failures.

## 6.2 Investigation on Failure Reason

This section presents the observations and investigation regarding several server faults on the testbed. During the controlled tests, a number of servers in Room-A and Room-B had faults and could not be booted. The remaining servers and all network equipment, forming a large portion of all tested IT equipment, were healthy after 16-month operation, at the time when this report was prepared. The vendors performed on-site examination for the faulty servers, replaced the faulty components to revive the servers, and sent the faulty components for lab-based fault analysis. The fault analysis results show that (1) corrosion caused by airborne contaminants on the motherboards and other supporting PCBs is the main reason of the faults; (2) the server faults are not



Fig. 40. Corrosion observed on the CD drive of a server in Room A in January and September of 2019. The rightmost figure shows the CD drive of the same model of server in Room C.

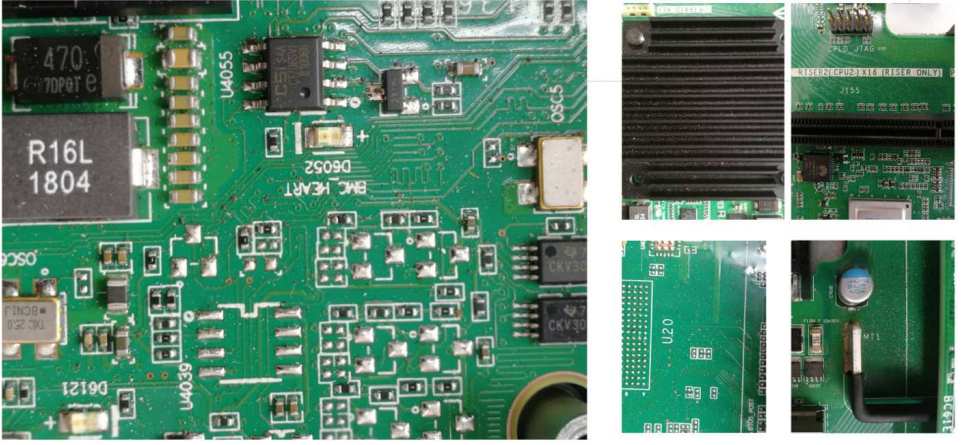


Fig. 41. Dusts rest on the PCBs of the motherboard.

caused by CPUs, HDDs, and memories; and (3) high temperature is not a reason of the server faults. In what follows, we provide some detailed information of the vendors' fault analysis and our investigation on the server room condition.

**6.2.1 Our Own Investigation on Server Room Condition.** We have performed investigation on the potential reasons of the server failures. The results of our own investigation as follows.

- (1) We investigated the temperature and RH conditions over the testing period and compare them with the specifications of the servers. Our investigation shows that the temperature and RH conditions do not violate the temperature and RH requirements of servers. Here are the detailed results:
  - **Temperature:** Two out of three Vendor 1 models are A4-compliant, i.e., they can continuously operate under a condition of 45°C temperature and 90% relative humidity; Servers from Vendor 2, 3, and 4 are A3-compliant, i.e., they can continuously operate under a condition of 40°C temperature and 90% relative humidity. Except for a single model from Vendor 1, no temperature violation occurred during the testing period. A further investigation on the Vendor 1 model shows that, this model can host a 4-node system and therefore requires a lower temperature of 35°C. But the server of this model on our testbed only hosts a 2-node system. Thus, this server will generate less heat and can tolerate higher

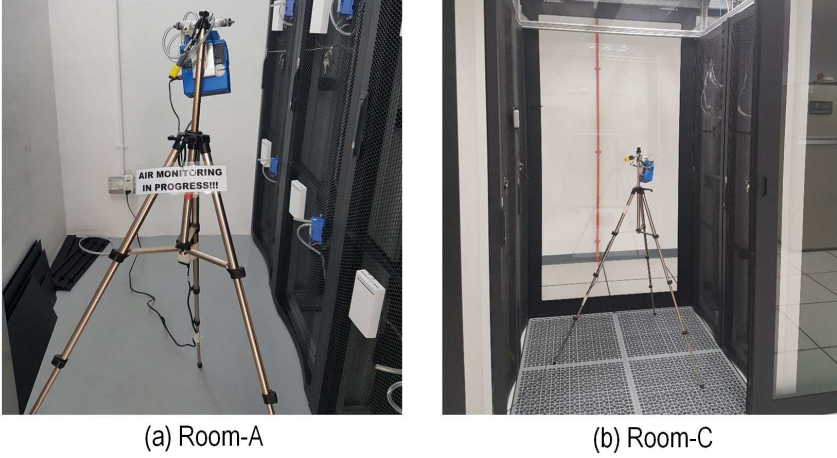


Fig. 42. Simultaneous precise one-day measurement of corrosive gases concentrations in Room-A and Room-C.

temperature. Given that the maximum temperature during the tests is  $37^{\circ}\text{C}$ , temperature is not the reason for the failures of the servers.

- **RH:** From the servers' datasheets, each server requires that the RH is lower than a threshold among 85%, 90%, and 95%. Fig. 43 shows the traces of outside air RH and the cold air RH in Room-A during three months before the server faults occurred. Note that because we varied the cold air temperature in Room-A during the controlled experiments, the cold air RH changed accordingly as shown in Fig. 43. We can see that the most stringent RH requirement of 85% was violated for limited time periods, while the other two RH requirements of 90% and 100% were never violated. As a comparison, we also investigated the cold air RH in Room-C. Following the common practice, Operator-B sets  $20^{\circ}\text{C}$  and 50% as the temperature and RH setpoints for the hot return air that is inhaled by the CRAC unit. The temperature and RH within the cold air containment is about  $17^{\circ}\text{C}$  and 70% that is represented by the solid horizontal line in Fig. 43. From the figure, we can see that, in fact, the RH of the cold air of Room-C is close to and higher than the average RHs of Room-A's outside and cold air, respectively. Since there is no fault in Room-C, we think high RH alone is not the reason of the faults. It is important to note that the RH requirements specified by the vendors assume that the air supplied to the servers is clean.
- (2) We investigated the air quality by analyzing the measurements of the corrosive gases concentration. Table 8 shows a server vendor's requirements and our measurements. We can see that the  $\text{SO}_2$  concentration is slightly higher than the requirement and the  $\text{NO}_2$  concentration is up to 5x higher than the requirement. Since the gas sensors we deployed on the testbed as shown in Fig. 4(i) are designed for real-time long-term monitoring but with less accuracy, we contracted a third-party company with gaseous contaminants monitoring expertise to perform one-day measurements in Room-A and Room-C simultaneously. Fig. 42 shows the company's measurement apparatuses in the two rooms. Table 9 shows the measurement results. We can see that the  $\text{NO}_2$  concentration in Room-A is at least 4.9x higher than that in Room-C. By comparing our  $\text{NO}_2$  concentration measurements with those reported in ASHARE's whitepaper [2], we conclude that the estimated corrosion rate is at

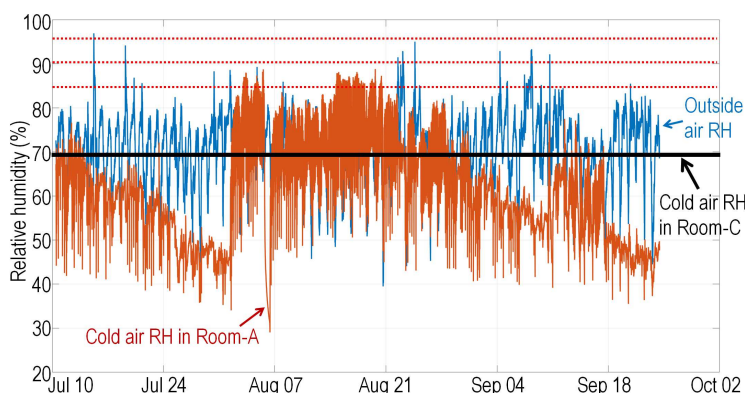


Fig. 43. RH of outside air, cold air in Room-A, and cold air in Room-C before the server faults in Room-A/B. The three horizontal dash lines represent the servers' maximum allowable RH levels that specified in their datasheets.

Table 8. A server vendor's requirement and our measurement (unit: ppb).

Gas	Required*	Reading
H <sub>2</sub> S	< 3	≈ 0
SO <sub>2</sub>	< 10	≈ 15
NO <sub>2</sub>	< 50	100-250

\*The concentration upper bounds are based on RH ≤ 50%.

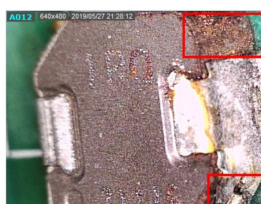
Table 9. Simultaneous precise one-day measurement (unit:  $\mu\text{g}/\text{m}^3$ ).

Gas	Room-A	Room-C
H <sub>2</sub> S	13	< 12
SO <sub>2</sub>	< 10	< 10
NO <sub>2</sub>	49	< 10

"<" means that the actual value is below the measurement resolution.



Disconnection caused by corrosion



Corrosion



Corrosion

Fig. 44. Microscopic images generated during vendor's fault analysis.

least 12 times higher than Vendor 1's requirement. This result is consistent with the observed corrosion on the metal parts of the servers in Room A and B. Note that past research showed that high RH will enhance the corrosion effects of corrosive gases [20].

- (3) As Room-A and Room-B are about 100 meters from a major highway in Singapore, we also suspect that the car exhaust gas is a major source of the NO<sub>2</sub>. Room-C has clean air because DC operators filtrate the air entering the DC building to remove the corrosive gases.

**6.2.2 Vendor's Fault Analysis.** The key findings of Vendor 1's fault analysis include the following:

- The faults of the two mainboards that failed during Phase 1 were caused by dusts and/or corruptions. These two mainboards functions properly at room temperature after being shipped

to the lab of Vendor 1. The test engineer applied liquid nitrogen spray and successfully replicated the faults, suggesting that the fault was caused by dust. This is because when the moisture in the air condenses on the motherboard, the dust on the motherboard absorbs the condensed moisture and causes short circuits. Rust on the PCBs can be observed under microscope. After cleaning the motherboard using alcohol, the motherboard restored, passed 48-hour CPU and memory pressure tests at room temperature, and survived liquid nitrogen spray tests. This behavior points to dust-induced failures.

- A faulty CPU is confirmed to be caused by the over voltage due to a failed power supply chip on the mainboard. Thus, the CPU failure is a cascading failure. It is not caused by overheating. The reason of the chip failure is not specified by the fault analysis.
- Vendor 1 confirmed that high temperature is not the cause of the server failures. Vendor 1 concluded that the short circuits and disconnections due to dusts and corrosions found on the printed circuit board (PCB) surface are the main causes. Fig. 44 shows several pinpointed corrosion problems that caused the mainboard failures. Vendor 1 conveyed that the faulty servers do not have anti-corrosion coating and suggested that anti-corrosion coating on PCBs can prevent the faults we encountered on the testbed.

**6.2.3 Summary and Discussions.** From an existing study [21], corrosion on metal materials is a joint effect of corrosive gases and RH, because the corrosive gases will absorb moisture in the air to form acids. Particulate contaminants can also attack the metal materials in a similar way or cause short circuit if the ambient RH exceeds the deliquescent RH of the contaminants [2]. Note that dust can be seen on the faulty motherboards under microscope during the server vendors' lab-based fault analysis. Therefore, the server faults in Room-A and Room-B can be attributed to (1) the co-presence of NO<sub>2</sub>, dust, and high RH, (2) the lack of anti-corrosion coating for the PCBs in the faulty servers. ASHRAE's whitepaper [2] mentions that sea salt carried by winds can also damage electronic devices in coastal areas. As there are no mature off-the-shelf sensors to monitor salt concentration in the air, our current research falls short of telling whether sea salt contributed to the server faults. But this issue is of great interest for future research.

### 6.3 Recommendations for Singapore

From the investigation of the failure reason, several general recommendations to mitigate the impact of corrosion on the server reliability include:

- Build natural air cooled DCs in clean areas;
- Apply dust and gaseous pollutant filtration;
- Apply anti-corrosion coating on the servers' PCBs exposed to air to increase server reliability in natural air cooled DCs.

We also performed an analysis to make a recommendation on the selection of server equipment and the temperature control. We consider a temperature control approach as follows. When the outside temperature exceeds a specified threshold, the system activates the standby cooling coil to process the outside air and maintain the temperature of the air supplied to the servers at the threshold. During this process, the air temperature decreases while the RH increases. When the

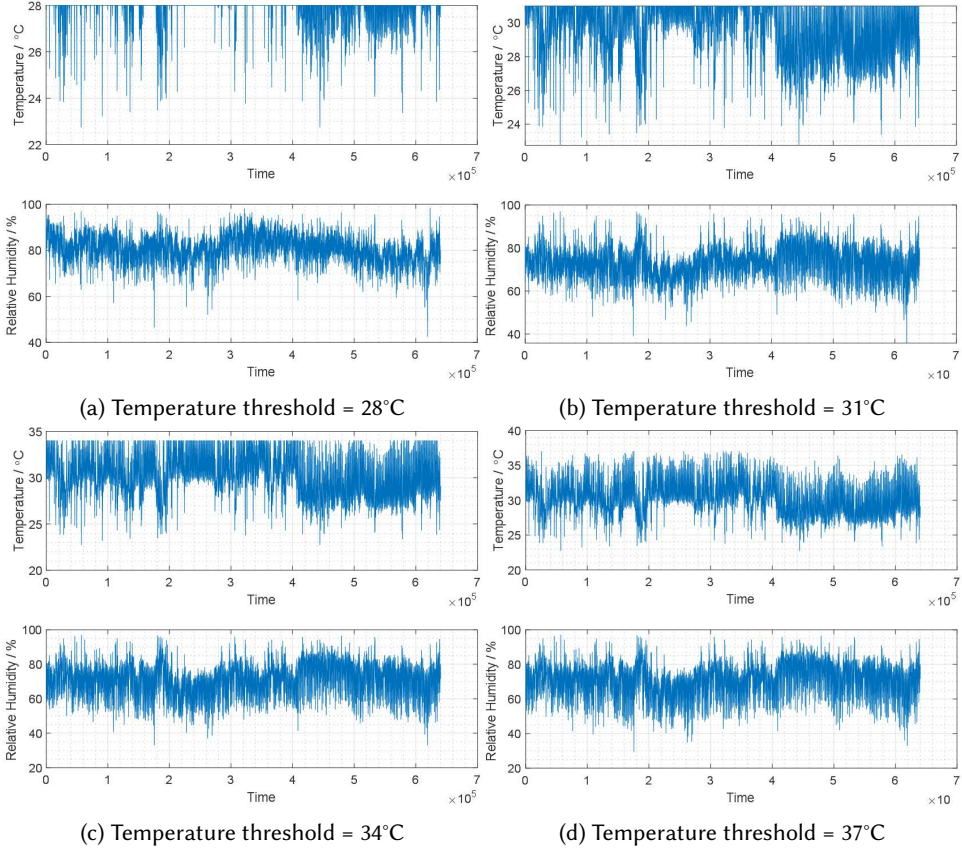


Fig. 45. Simulation results of the temperature control.

cooling coil is activated, the RH of the air supplied to the server can be calculated as follows:

$$P_{W1} = f(T_{outside} + 273.15), \quad (15)$$

$$P_{W2} = f(T_{supply} + 273.15), \quad (16)$$

$$RH_{supply} = RH_{outside} \times \frac{P_{W1}}{P_{W2}}, \quad (17)$$

$$f(T) = e^{-\frac{5.8 \times 10^{-3}}{T} + 1.392 - 4.864 \times 10^{-2} \times T + 4.176 \times 10^{-5} \times T^2 - 1.445 \times 10^{-8} \times T^3 + 6.546 \times \log T}, \quad (18)$$

where  $T_{outside}$  and  $T_{supply}$  are the outside and supply air temperatures in °C,  $RH_{outside}$  and  $RH_{supply}$  are the outside and supply air RH. Based on the above equations, we perform a simulation using Singapore's daily outdoor air temperature and mean RH in the last 5 years (Jan 2014 to Oct 2018) to compute the temperature trace and RH trace in the server room when four different temperature thresholds (28°C, 31°C, 34°C, 37°C) are used. Fig. 45 shows the simulated temperature and RH traces for the four temperature threshold settings. Fig. 46 shows the cumulative distribution function of the simulated RH under the four temperature thresholds.

Then, we cross check the simulated temperature and RH traces against ASHRAE's four classes regarding IT equipment's temperature/RH tolerance. Fig. 47 shows the relationship between different levels of temperature control and ASHRAE's classes.

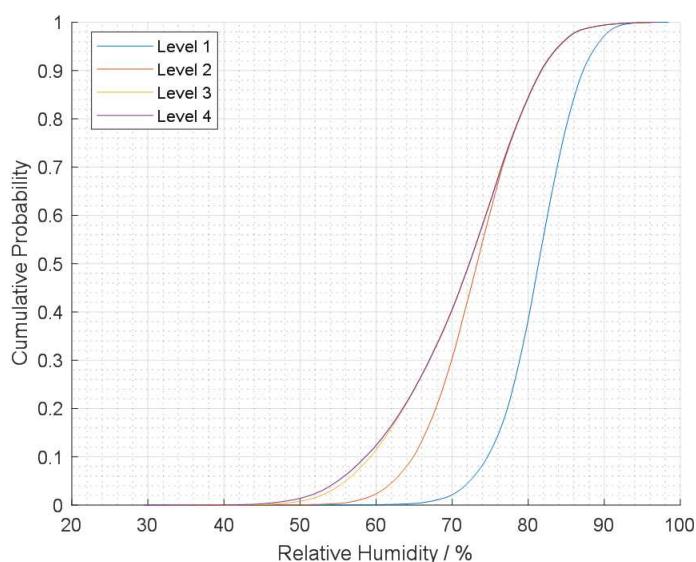


Fig. 46. Cumulative distribution function of the simulated RH. Temperature thresholds corresponding to Level 1, Level 2, Level 3, and Level 4 are 28°C, 31°C, 34°C, and 37°C.

TDC Level	ASHRAE Classes							
	A1		A2		A3		A4	
	[15°C, 32°C]	Compliance with RH of [20%, 80%]	[10°C, 35°C]	Compliance with RH of [20%, 80%]	[5°C, 40°C]	Compliance with RH of [8%, 85%]	[5°C, 45°C]	Compliance with RH of [8%, 90%]
Level 1 (≤ 28°C)	✓	38.3%	✓	38.3%	✓	78.4%	✓	97.2%
Level 2 (≤ 31°C)	✓	84.3%	✓	84.3%	✓	96.5%	✓	99.4%
Level 3 (≤ 34°C)	✗	84.3%	✓	84.3%	✓	96.5%	✓	99.4%
Level 4 (≤ 37°C)	✗	84.3%	✗	84.3%	✓	96.5%	✓	99.4%

Fig. 47. Relationship between different levels of temperature control and ASHRAE's classes.

Table 10. Relative energy savings achievable with the energy consumption at 25°C as the baseline.

Level	Supply air temperature threshold	ASHRAE X-factor	Relative energy saving
Level 1	28°C	1.36	31.33%
Level 2	31°C	1.45	56.77%
Level 3	34°C	1.52	53.36%
Level 4	37°C	1.6	53.48%

We have also investigated the energy saving corresponding to the four temperature settings based on the energy profiles shown in Fig. 18. Table 10 shows the ASHRAE's reliability X-factor and the relative energy saving (with the energy consumption at 25°C as the baseline. From the table, we can see that the energy saving achieves a knee point when the temperature control threshold is

31°C. With higher temperature threshold, the energy saving has a slight reduction. This is because the servers' built-in fans will rotate faster, resulting in higher energy consumption.

By jointly analyzing the results in Fig. 47 and Table 10, we can achieve a rule-of-thumb recommendation of adopting ASHRAE Class A3 IT equipment and running the air-side free-cooled DC with a temperature upper limit of 31°C. With this configuration, we can fully meet ASHRAE's A3 temperature requirement, and meet A3 RH requirement for 96.5% of time, and achieve the highest energy saving of 56.77%. Note that many latest servers (e.g., all Dell's gen14 servers and all HPE's DLx gen9 servers) are compliant with the A3 requirement. This rule-of-thumb recommendation does not override the need of clean air and/or anti-corrosion measures, since the ASHRAE's A3 requirement assumes clean air.

## 7 DISCUSSIONS AND CONCLUSION

As the first systematic trial of air-side free cooling for DCs in Singapore, our research has generated various valuable experiences and information for DC-related entities. Some of them are in the form of learned lessons that the future research and industrial practice should consider. We summarize them as follows.

- **Singapore's high temperatures do not impede air-side free cooling:** During the project period, no computing performance degradation was observed on our testbed with cold aisle temperature up to 37°C. Investigation shows that the server faults on our testbed were not caused by temperature. As ASHRAE has been working on extending the recommended allowable temperature and RH ranges for IT equipment, many latest servers (e.g., all Dell's gen14 servers and all HPE's DLx gen9 servers) are compliant with the A3 requirement to be able to tolerate 40°C. Thus, the temperatures of Singapore with a record maximum of 37°C will not impede air-side free cooling. From the results of this project, we make a rule-of-thumb recommendation of adopting A3-compliant IT equipment and running an air-side free-cooled DC when the ambient temperature is no greater than 31°C. Only when the ambient temperature is higher than 31°C, standby cooling is activated to maintain the temperature at 31°C. This operating guideline ensures that the A3 requirements on temperature and RH are met for most of time in Singapore. Moreover, it gives a satisfactory saving in overall energy consumption.
- **Server hardening vs. airborne contaminants removal:** We believe that by only deploying hardened IT equipment with anti-corrosion coating on the PCBs exposed to air, hardware faults caused by corrosion and conductive dust will be resolved. Alternatively, better airborne contaminants filtration can be employed. The following two categories of filtration approaches can be considered:
  - **Passive filtration:** This project uses Class MERV-6 to remove PM10 and larger particles. Filters in higher classes can be used instead to remove finer particles. For corrosive gases, the hot air generated by the servers can be recirculated and mixed with the outside cold air to form warm air with lower RH to be supplied to the servers. The lower RH will reduce the corrosive gases' attack capabilities. This approach requires no extra energy and exploits the higher temperature tolerance of the latest servers. The details of this approach are described in Section 5.2. The speed control logic of server built-in fans may need adjustment to avoid fast wear and tear due to unnecessarily high rotation speeds in high temperatures. Note that the server fan speed control logic update can be implemented using a shell script and deployed easily.
  - **Active filtration:** Electrostatic air cleaner can be employed to strengthen the particle removal. Traditional chemical approaches can be applied to remove corrosive gases. However, these approaches will consume energy.Therefore, this project has narrowed the feasibility problem of air-side free cooling in Singapore down to the effectiveness of airborne contaminants removal and its associated Capex and Opex. The choice of server hardening and better filtration is a design problem that will depend on specific configurations and constraints of the DC. For example, server hardening may not be feasible for colocation DCs. We note that carefully choosing the location for cleaner ambient air may significantly ease the design of an air-side free-cooled DC.
- **Implication for existing DCs:** Our results also suggest that the existing DCs operated in enclosed buildings can consider increasing their temperature setpoints for better energy efficiency. These DCs will not have the airborne contamination problem, owing to the enclosed

design and the deployed air filtration systems. By simply raising the server room temperature setpoint, the difference between the temperatures of the chilled water entering and leaving the CRAC units of the server room will increase, resulting in decreased volume flow rate of the chilled water for the room. Existing studies have shown that the increase of chilled water temperature difference will lead to less energy usage of the chiller plant and the CRAC units. For a colocation DC that hosts many private vaults for different customers who desire different temperature setpoints, a pricing scheme that fairly attributes the energy usage of the shared chiller plant infrastructure to the customers is still an open issue. If the heat exchange between the private vaults through the physical separations cannot be ignored, the problem becomes more challenging. Such issues are interesting topics for future research.

## 8 FUTURE RESEARCH

This project has generated two operational recommendations for air-side free cooling with server performance, reliability, and energy efficiency into consideration:

First, Section 5.2 proposes to use mixing to maintain the supply air temperature sufficiently high (but still within the IT equipment's thermal safety threshold, e.g., 40°C of A3 equipment) such that the RH is below the deliquescent RH of the air-borne contaminants. We apply the deep reinforcement learning to learn how to operate the fans and mixing dampers based data-driven models. This approach aims to mitigate the impact of air-borne contaminants. Whether such a control strategy can avoid the corrosion issue is still not experimentally studied.

Second, Section 6.3 provides a simple and practicable recommendation for Singapore's air-side free-cooled DCs without RH and contamination controls. It recommends to use IT equipment compliant with the ASHRAE A3 requirement and run the DCs with supply air temperature not exceeding 31°C. When the ambient temperature is higher than 31°C, standby cooling is activated to maintain the supply air temperature at 31°C. Therefore, the temperature and RH requirement of ASHRAE A3 class are satisfied with high probability. However, the air-borne contaminants may still cause IT hardware corrosion with the help of RH. Therefore, in this recommendation, the IT equipment should have anti-corrosion measures.

Based on extensive knowledge obtained in this project, the future research can use the above two operational recommendations to design, analyze, and verify a control approach to achieve reliable, high-temperature air-side free-cooled DCs in Singapore's tropical environment. The new control approach can use the air mixing to maintain the supply air temperature sufficiently high as in the approach presented in Section 5.2, use electrostatic air cleaner to reduce dusts entering the room, and adjust the server fan control logic to avoid fast wear and tear. The objective is to operate air-side free-cooled DCs reliably with normal servers that are not specifically hardened with anti-corrosion measures. This project achieved a PUE of about 1.05 under the adaptive ventilation approach. With additional air processing units (e.g., electrostatic air cleaner), a higher PUE of 1.1 can be a reasonable target of the new control approach. Moreover, artificial intelligence (AI)-powered joint control of the environment and computation can also push the Pareto frontier of the performance, reliability, and energy efficiency. The study presented in Section 5.2 has shown the great potential of the deep reinforcement learning in the control of the air-side free-cooled data centres with complex cyber-physical dynamics. Future research can implement the new control approach and perform verification experiments for an extended period of time.

## REFERENCES

- [1] 2018. Singapore is top data center hub in SE Asia: report. <https://bit.ly/2LEeV2B>
- [2] ASHRAE. 2011. Gaseous and Particulate Contamination Guidelines for Data Centers. *White Paper* (2011).
- [3] ASHRAE. 2012. *IT Equipment Thermal Management and Controls*. Technical Report. ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment. <https://bit.ly/2quzc1D>
- [4] ASHRAE. 2013. Psychrometrics. In *2013 ASHRAE Handbook: Fundamentals* (SI).
- [5] Nosayba El-Sayed, Ioan A Stefanovici, George Amvrosiadis, Andy A Hwang, and Bianca Schroeder. 2012. Temperature management in data centers: why some (might) like it hot. *ACM SIGMETRICS Performance Evaluation Review* 40, 1 (2012), 163–174.
- [6] Xiaobo Fan, Wolf-Dietrich Weber, and Luiz Andre Barroso. 2007. Power provisioning for a warehouse-sized computer. In *ACM SIGARCH computer architecture news*, Vol. 35. ACM, 13–23.
- [7] Hainan Zhang, Shuangquan Shao, Hongbo Xu, Huiming Zou, Changqing Tian. 2014. Free cooling of data centers: A review. *Renewable and Sustainable Energy Reviews* 35 (2014), 171–182.
- [8] Taliver Heath, Ana Paula Centeno, Pradeep George, Luiz Ramos, Yogesh Jaluria, and Ricardo Bianchini. 2006. Mercury and freon: temperature emulation and management for server systems. In *ACM SIGARCH Computer Architecture News*, Vol. 34. ACM, 106–116.
- [9] Kevin Heslin. 2015. Data Center Industry Survey Results. <https://bit.ly/2K1UDz4>.
- [10] HPE. 2018. HPE iLO 4 User Guide. [https://h20628.www2.hp.com/km-ext/kmcsdirect/emr\\_na-c03334051-19.pdf](https://h20628.www2.hp.com/km-ext/kmcsdirect/emr_na-c03334051-19.pdf).
- [11] Adam Wade Lewis, Soumik Ghosh, and Nian-Feng Tzeng. 2008. Run-time Energy Consumption Estimation Based on Workload in Server Systems. *HotPower* 8 (2008), 17–21.
- [12] Timothy P. Lillicrap et al. 2016. Continuous control with deep reinforcement learning. In *Proc. 4th ICLR*. <http://arxiv.org/abs/1509.02971>
- [13] Y. Ma, F. Borrelli, B. Hencsey, B. Coffey, S. Bengae, and P. Haves. 2012. Model Predictive Control for the Operation of Building Cooling Systems. *IEEE Trans. Control Syst. Technol.* 20, 3 (May 2012), 796–803.
- [14] V Mnih et al. 2015. Human Level Control Through Deep Reinforcement Learning. *Nature* 518, 7540 (2015), 529–533.
- [15] Jiacheng Ni and Xuelian Bai. 2017. A review of air conditioning energy performance in data centers. *Renewable and sustainable energy reviews* 67 (2017), 625–640.
- [16] Eric O’Shaughnessy, Chang Liu, and Jenny Heeter. 2016. *Status and Trends in the U.S. Voluntary Green Power Market (2015 Data)*. Technical Report NREL/TP-6A20-67147. National Renewable Energy Laboratory. <https://doi.org/10.2172/1329462>
- [17] Ramya Raghavendra, Parthasarathy Ranganathan, Vanish Talwar, Zhikui Wang, and Xiaoyun Zhu. 2008. No power struggles: Coordinated multi-level power management for the data center. *ACM SIGOPS Operating Systems Review* 42, 2 (2008), 48–59.
- [18] E. Rotem, A. Naveh, A. Ananthakrishnan, E. Weissmann, and D. Rajwan. 2012. Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge. *IEEE Micro* 32, 2 (March 2012), 20–27. <https://doi.org/10.1109/MM.2012.12>
- [19] Richard S. Sutton and Andrew G. Barto. 1998. *Introduction to Reinforcement Learning* (1st ed.). MIT Press, Cambridge, MA, USA.
- [20] Jan-Erik Svensson and Lars-Gunnar Johansson. 1993. A Laboratory Study of the Effect of Ozone, Nitrogen Dioxide, and Sulfur Dioxide on the Atmospheric Corrosion of Zinc. *Journal of The Electrochemical Society* 140 (1993), 2210–2216.
- [21] J-E Svensson and L-G Johansson. 1993. A laboratory study of the effect of ozone, nitrogen dioxide, and sulfur dioxide on the atmospheric corrosion of zinc. *Journal of the Electrochemical Society* 140, 8 (1993), 2210–2216.
- [22] Zhikui Wang, Cullen Bash, Niraj Tolia, Manish Marwah, Xiaoyun Zhu, and Parthasarathy Ranganathan. 2009. Optimal fan speed control for thermal management of servers. In *ASME 2009 InterPACK Conference collocated with the ASME 2009 Summer Heat Transfer Conference and the ASME 2009 3rd International Conference on Energy Sustainability*. American Society of Mechanical Engineers, 709–719.