energies

MDPI

Article

Evaluation and Optimization of a Two-Phase Liquid-Immersion Cooling System for Data Centers

Cheng Liu [1,2] **and Hang Yu** [1,*]

[1] School of Mechanical and Energy Engineering, Tongji University, Shanghai 201804, China; cliuglobal@163.com
[2] China Mobile Group Shanghai Co., Ltd., Shanghai 200060, China
[*] Correspondence: yuhang@tongji.edu.cn

Abstract: An efficient cooling system for data centers can boost the working efficiency of servers and promote energy savings. In this study, a laboratory experiment and computational fluid dynamics (CFD) simulation were performed to explore the performance of a two-phase cooling system. The coefficient of performance (COP) and partial power usage effectiveness (pPUE) of the proposed system was evaluated under various IT (Information Technology) loads. The relationship between the interval of the two submerged servers and their surface temperatures was evaluated by CFD analysis, and the minimum intervals that could maintain the temperature of the server surfaces below 85 °C were obtained. Experimental results show that as server power increases, COP increases pPUE decreases. In one experiment, the COP increased from 19.0 to 26.7, whereas pPUE decreased from 1.053 to 1.037. The exergy efficiency of this system ranges from 12.65% to 18.96%, and the tank side accounts for most of the exergy destruction. The minimum intervals between servers are 15 mm under 1000 W of power, 20 mm under 1500 W, and more than 30 mm under 2000 W and above. The observations and conclusions in this study can be valuable references for the study of cooling systems in data centers.

Keywords: two-phase cooling; data center; CFD; immersion; optimization

Citation: Liu, C.; Yu, H. Evaluation and Optimization of a Two-Phase Liquid-Immersion Cooling System for Data Centers. *Energies* **2021**, *14*, 1395. https://doi.org/10.3390/en14051395

Academic Editor: Guglielmo Lomonaco

Received: 11 February 2021
Accepted: 24 February 2021
Published: 3 March 2021

1. Introduction

The boosting of digital technology (e.g., Internet of Things, artificial intelligence, big data, 5G, cloud computing) and its extensive applications in many industries, such as transportation [1], communication [2], manufacturing [3], medicine [4], and education [5], demonstrate an increasing need for data processing, storage, and transmission. A data center can be a building or part of a building where data are gathered, processed, and stored [6]. According to a report by the Synergy Research Group, by the end of the third quarter of 2019, there were 504 hyperscale data centers worldwide [7], and another report predicted that this number would increase by 12–14% annually over the next five years [8]. Of all the data centers in 2019, approximately 40% reside in the US, and China, Japan, the UK, Germany, and Australia collectively contain 32%. Data centers typically involve high energy consumption. In the UK, data centers accounted for 1.5% of electricity usage in 2016, with power consumption projected to increase by 20% in 2020 [9]. In China, data centers used 160.8 billion kWh of electricity in 2018, which exceeded the total electricity consumption of all of Shanghai [10].

In addition to power equipment and accessory components, data centers consist of two major energy-consuming parts, IT and heating, ventilation, and air conditioning (HVAC) equipment, which account for approximately 90% of the total energy usage [11]. The HVAC equipment itself has been reported to be responsible for 34% of this total energy [12], because of the need to cool the all-day operation of high-power-density IT equipment and maintain the indoor thermal environment within the appropriate temperature zones [13]. The HVAC system needs to be in operation for approximately 24 h per day. To improve the energy efficiency of a data center, optimizing the HVAC system is one of the key steps. Power usage effectiveness (PUE) is an industry-preferred index for evaluating the infrastructure energy efficiency of data centers [14]. A small PUE approaching 1.0 indicates that the data center approaches optimal energy efficiency.

Data center cooling systems can be classified into two types—air-cooled systems [15,16] and liquid-cooled systems [17]. The first type sends cold air to racks on which servers are placed and then removes the heat generated by the servers. The advantages of this type of system include the following: (1) this traditional air-cooling system is relatively mature and easy to construct and (2) this system has promising potential for the use of free cooling resources, such as natural cold air and water [18], which could substantially benefit the reduction of power usage, decreasing the PUE from a value of 2.01 for a typical computer room air conditioner system to a value as low as 1.1 with free cooling [19,20]. The drawbacks of the air-cooled system are significant, such as a low heat transfer coefficient, difficulty in the design and control of cold and hot air flow, asymmetrical cooling in different servers owing to the geometrical layouts of data centers, large power consumption through the operation of chillers, fans, and pumps, and unstable cooling capacity caused by the outdoor thermal environments [12,17,21,22].

The liquid-cooled system requires the removal of high-density heat, which is too high to be removed by the air-cooled system; This heat is dissipated by high-performance information and communication technology (ICT) devices. Based on the direct contact of the liquid with the heat sources, the liquid-cooled system can be further categorized as direct cooling and indirect cooling. For direct liquid cooling, the dielectric liquid absorbs heat from the electronic components directly; this type of heat transfer can be highly effective [17,23]. Passive two-phase cooling is a fluid-cooled approach in which electronic components are submerged in a phase-changeable liquid bath in a closed box. When the surface temperature of the electronic components exceeds the evaporation temperature of the liquid, the process of boiling heat transfer is triggered, removing the extra heat in the form of vapor bubbles. The produced vapor rises up and further condenses through a water-based condenser installed above the bath. Finally, the heat is absorbed and removed by the coolant; the vapor is then liquefied and drips back to the bath, driven by gravity [24,25]. Compared with the air-cooled system, the two-phase cooling system could reduce the use of accessory equipment, such as chiller pumps and fans, potentially improving the energy efficiency of data centers. In addition, this system could avoid the surface temperature symmetry of electronic devices and guarantee their working environment.

Dashtebayaz and Namanlo [26] studied an air-based cooling system that performs waste heat recovery and reported a coefficient of performance (COP) that varied in the 3–5 range with a PUE of approximately 2.5. Chen et al. [27] applied spray-cooling technology to the cooling of computer centers and suggested that the system COP is highly dependent on the inlet water temperature, with COP possibly changing from 3 to 15 and PUE ranging from 1.45–1.52. Cho et al. [28] proposed many green technologies that can be applied to data center cooling systems under different climatic contexts and determined by applying suitable strategies. The target PUE was in the range of 1.2–1.6. Dong et al. [29] explored the effectiveness of using free-cooling resources to cool data centers and suggested that natural cooling could increase the system COP by 23.7%, rising from 5.9 to 7.3. Lu et al. [30] reported the PUE of data centers in Finland, highlighting its range of variation of 1.2–1.5.

In a two-phase cooling system, the ICT equipment is completely submerged in the dielectric liquid bath; nucleate boiling occurs on the server surface when the server temperature reaches the boiling temperature of the dielectric liquid. Wu et al. [31] evaluated a full-scale two-phase liquid-immersion DC cooling system in a tropical environment. They found that the exergetic efficiency experienced a small rate of decrease, while the supplied power rate increased. The highest efficiency of 69.9% was obtained at zero load, whereas the lowest efficiency of 65.9% was observed at full load. Kanbur et al. [32] studied a two-phase liquid-immersion data center cooling system through experimental and thermo-economic analyses. They found that the optimal COP and PUE values occurred at maximum operation loads of 6.67 and 1.15, respectively, whereas the minimum COP and highest PUE were observed at the minimum operation loads of 2.5 and 1.4, respectively. Choi et al. [33] used two-phase cooling (HFE-7100) for a polymer electrolyte membrane fuel cell. They found that the two-phase boiling heat transfer coefficients of the HFE-7100 in mini-channels were strongly dependent on heat flux and vapor quality but less sensitive to mass flux. However, at present, there are fewer research reports on the mining machine motherboard of the two-phase liquid-immersion cooling system.

Computational fluid dynamics (CFD) simulations have been widely applied to validate the performance of specific cooling systems for data centers and to present optimal designs based on the detailed results of the simulations. Ahmadi et al. [34] verified the energy-saving potential, optimal designs, and operation conditions for the computer room air handling bypass method in cooling data centers by using CFD simulations. Hassan et al. [35] performed CFD simulations in ANSYS Fluent and obtained the temperature, airflow, and pressure distributions for a data center. The prediction provided a three-dimensional (3D) thermal map of the data center and helped to optimize the cooling system. Fulpagare et al. [36] explored transient CFD models to simulate the dynamic requirements of cooling systems to achieve smart control in data centers, and the model performance was validated based on experiments. Using CFD simulations, Nada et al. [37] studied the performance of data centers under different configurations and summarized the effects of the computer room air conditioning unit layout on the thermal performance of the racks. Most previous CFD studies have focused on air-cooled systems; however, studies that focused on two-phase cooling systems are scant. Cheng et al. [38] studied a single-phase immersion cooling system (using 3 M Novec 7100) for single CPU cooling via 3D numerical analysis using ANSYS Fluent. The results of the simulations showed that there was an unbalanced heat distribution around the CPU, and higher flowing speeds of the liquid coolant led to the removal of more heat, resulting in lower CPU temperatures and more balanced heat distribution around the CPU. Ali et al. [39] numerically investigated the thermal performance and stress analysis of enhanced copper spreaders for nucleate boiling immersion cooling of high-power electronic chips. An et al. [40] developed a 3D numerical model using ANSYS Fluent to study two-phase immersion cooling for electronic components. They found that Novec 7000 can support cooling a 5 cm $\times$ 5 cm heat source in a vertical orientation with power as high as 225 W (heat flux of 9 W/cm^2). However, less research was reported using CFD simulation to study the arrangement of the server motherboards.

This study consisted of experiments and a series of CFD simulations. First, an innovative cooling structure and procedure for a two-phase immersion cooling system were developed. Six different cases were investigated to analyze the thermal management performance in the operational load range of 1127–1577 W. The energy efficiency of the system was evaluated based on the partial power usage effectiveness (pPUE) and COP indices, and exergy analysis was performed. Finally, an arrangement of submerged servers was proposed by CFD simulation. This study aimed to formulate an innovative two-phase immersion cooling system for data centers, which includes an optimal arrangement of the servers. Data from the two-phase cooling system were collected; the results and conclusions presented herein may potentially serve as valuable references for researchers and engineers.

2. Methodology

2.1. Experimental Study

A novel, innovative cooling structure and procedure of a two-phase immersion cooling system were established for the cooling of ICT devices. The scheme of the proposed two-phase cooling system is shown in Figure 1, which includes a primary heat exchanger and a tank with coolant and server boards as ICT devices. The ambient air temperature was maintained at 21–23 °C.

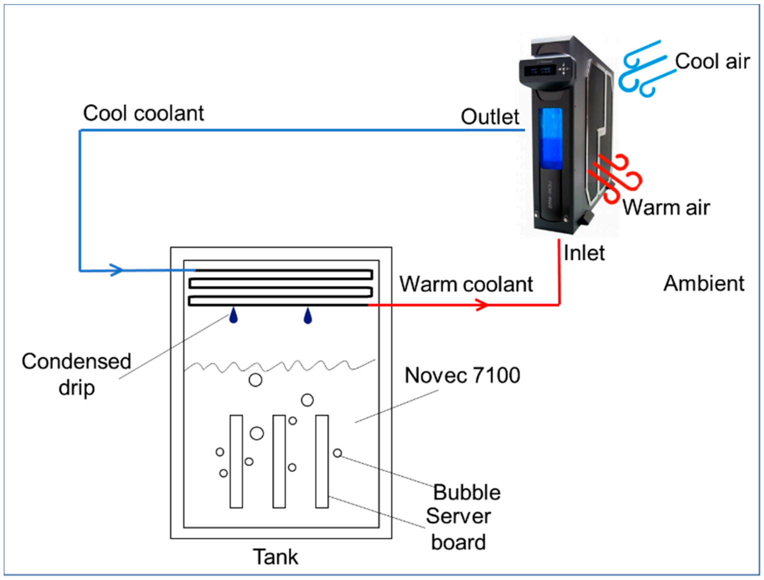

Figure 1. Scheme of the proposed cooling system.

Figure 2a shows the tank where real servers were submerged in the dielectric liquid Novec 7100 supplied by 3M (Minnesota Mining and Manufacturing, Saint Paul, MN, USA) [41]. The dielectric liquid, being expensive, was filled in the tank only to submerge all the servers (Figure 2c) during the experiments. For each experiment, the server power output was maintained at a constant level. As the surface temperature of the servers rose and reached the boiling point of Novec 7100, the surrounding liquid formed bubbles and absorbed heat through evaporation. The produced vapor gathered on the top of the tank, where condenser coils were placed around the wall, as shown in Figure 2c. The heat was transferred from the vapor to the cold coolant circulating inside the condenser coils and was removed from the tank. Thereafter, the vapor condensed into the fluid phase and dripped back onto the liquid bath because of gravity. The heated coolant inside the coil released heat to the environment by a primary heat exchanger, as shown in Figure 2d.

Table 1 lists the thermal characteristics of Novec 7100 at 25 °C, and the specific data are listed in Table A1 in Appendix A. The dimensions of the tank were 650 × 450 × 1050 mm^3 (length × width × height), and the tank was insulated to prevent the influence of the surrounding thermal environment. The servers used in this study were application-specific integrated circuit (ASIC) miners designed to "mine" a specific cryptocurrency. The type of server board was T2T-25T, and the dimensions of each board were 235 × 182 × 8 mm^3. There were 140 T2T CPUs on each board, and the dimensions of the CPU were 8 × 8 × 1 mm^3, with an average CPU TDP of 8.37 W/cm^2. The server boards were placed at 100 mm intervals. This type of server has several modes with rated power outputs ranging from 1000 to 2000 W. The mode of each server can be controlled remotely through a computer program.

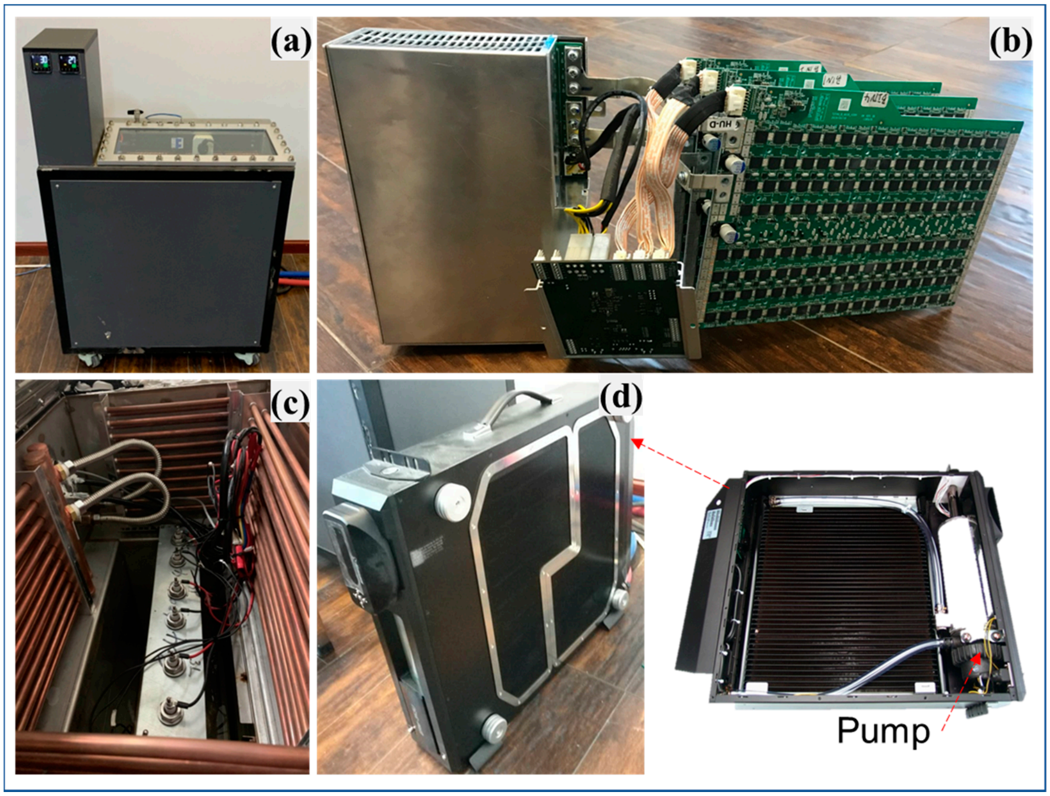

Figure 2. Two-phase cooling tank: (**a**) the tank, (**b**) servers, (**c**) condenser coil, and (**d**) primary heat exchanger.

Table 1. Properties of Novec 7100 at 25 °C.

Boiling Point (°C)	Vapor Pressure (kPa)	Molecular Weight (g/mol)	Density Liquid (kg/m^3)	Dynamic Viscosity (cSt)	Specific Heat (J/kgK)
61	27	250	1510	0.38	1183

During the experiment, all the data were measured when the system was under a stable status, characterized by a lack of fluctuations in the CPU surface temperature. Under each condition, the system was kept running for 60 min before the data were recorded. The temperatures of the server surface, liquid bath, and inlet and outlet coolant inside the coils were recorded using a digital temperature controller (E5CC, OMRON, Kyoto, Japan). The temperature was measured using K-type thermocouples (TT-K-24, OMEGA, Norwalk, CT, USA), which were calibrated using a mercury thermometer with an accuracy of ±0.1 °C. The core temperature of the servers was measured and recorded automatically by the aforementioned computer program. The pressure in the tank was monitored by an automatic pressure relief valve with a pressure sensor (ZSE40AF-01-T, SMC, Kyoto, Japan), with the pressure set to 1.2 kg/cm^2. The heat from the coolant was removed by a heat exchanger (ERM-3K3UC Liquid Cooling System, Koolance, Auburn, AL, USA), and the cooling capacity of the cooling system was 2600 W (8872 BTU/h). The power of the cooling unit (including the pump and the fan) was monitored using a power meter that can monitor and record real-time power consumption. The experiment consisted of six conditions, and the stable data are listed in Table 2.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.