# Robust data-driven power simulator for fast cooling control optimization of a large-scale computing system

Takashi Shiraishi
Fujitsu Laboratories Ltd., Kawasaki, Japan
shiraishi-ten@fujitsu.com

Hiroshi Endo
Fujitsu Laboratories Ltd., Kawasaki, Japan
endo-hiroshi@fujitsu.com

Takaaki Hineno
Fujitsu Limited, Kamata, Japan
hineno.takaaki@fujitsu.com

Hiroyuki Fukuda
Fujitsu Laboratories Ltd., Kawasaki, Japan
fukuda.hiro@jp.fujitsu.com

## ABSTRACT

Today, the power consumption of large-scale computing systems such as an HPC or a datacenter is a significant social issue. Cooling units consume 30% of the total power. General control policies for cooling units are local (no automatic overall optimization) and static (manual overall optimization nearly once a week). However, free cooling leveraging outside air and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a novel deep neural network (DNN) power simulator that can learn from actual operating logs in the system and can quickly identify the optimum control variables. We demonstrated the power simulator of overall cooling units by using operating logs from an actual large-scale system with 4.7-MW-power IT load and 1.4-MW-power cooling units. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional control policy in the actual system.

## CCS CONCEPTS

•**Computer methodology** → **Machine learning;** Modeling and Simulation;•**Hardware** → *Platform power issues.*

## KEYWORDS

Power Modeling, Cooling Control Optimization, Deep Learning

**ACM Reference format:**

Takashi Shiraishi, Hiroshi Endo, Takaaki Hineno and Hiroyuki Fukuda. 2019. Robust data-driven power simulator for fast cooling control optimization of a large-scale computing system. In *SC'19: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/nnnnnnn.nnnnnnn


*SC'19, November 17-22, 2019, Denver, CO, USA*


## 1 INTRODUCTION

The world's ICT sector, including supercomputers and datacenters, consumes 7% of the global electricity [1]. Recently systems with a total power of over 10 MW are becoming common [2]. Cooling units account for 30% of total energy consumption in the system, and there are still significant challenges in terms of control optimization for cooling systems. One of the challenges is global control optimization. The cooling units consists of hundreds of units such as air handling units (AHU), chillers, and so on, which are usually provided by different vendors. Some units have automatic local control for low power operation. However, control variables that affect other units are not globally controlled. The global control tuning of an overall system is still executed by a building operator nearly once a week or a month. Another challenge is dynamic frequent control. Free cooling that leverages outside air in cold seasons makes the condition of cooling-units sensitive to outside conditions, change hourly [3]. IT load, which also affect the optimum setting of cooling-units, experiences hourly MW-range changes in large supercomputers [4]. Similar trends are forecasted in datacenters due to GPU introduction.

Many researches has tackled the global optimization but there were few studies for the dynamic-frequent, global-control optimization. Reinforcement-learning-based control can learn through actual trial and error [5], but a disadvantage of this approach is its long optimization time. A single trial requires at least few minutes because the cooling units require a stabilization time (from minutes to an hour) after the set points change [6]. Moreover, the trial and error in a running system mostly cannot be acceptable in terms of safety regulation. Computational fluid dynamics (CFD) can calculate cooling-units' dynamics. Simulation-based optimization does not require actual trial in the system, but time-consuming CFD calculation (several hours for just one calculation) has the same problems in optimization time [7, 8].

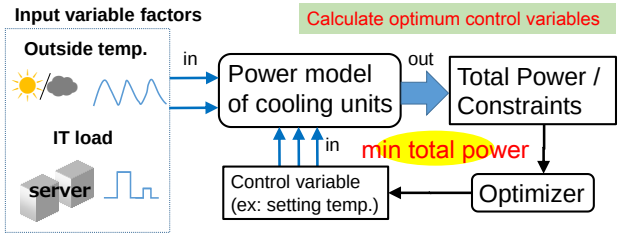


**Fig. 1: Our proposed cooling-units control system**

In this work, we studied a dynamic-frequent, global-control optimization of the cooling units and quantified the effect. We propose overall cooling-unit control system with a power simulator and a control optimizer (Fig. 1). Optimum control variables that minimize total power are dynamically calculated dependent on the input variable factors (outside temperature and IT load fluctuation). The optimization should be fast –within 30 minutes in order to follow the change of the input variable factors.

## 2 MODELING & CONTROL OPTIMIZATION

We demonstrated the overall modeling of an actual operating datacenter that has a modern green cooling system. The building has 9 server rooms with total 4.7-MW IT load. The rooms are cooled by 132 AHUs with air-side free cooling. Water lines are cooled by 4 chillers that are assisted by water-side free cooling of cooling towers. 6 cooling towers were operated in the system. 16 water pumps control the flow amount of the water line. The typical power of the overall water cooling system was 1.4 MW in summer except for the fan power of the AHUs (we did not optimize fan rotation of AHU in this work). 312-time-series logs of the cooling units were monitored every 30 minutes from Jan. 2017 to Mar. 2019. The logs include amount of heat, water temperature, water flow, power, and so on. The half of the period from the beginning was used for the training. The actual power change of IT load was small at less than 5%.

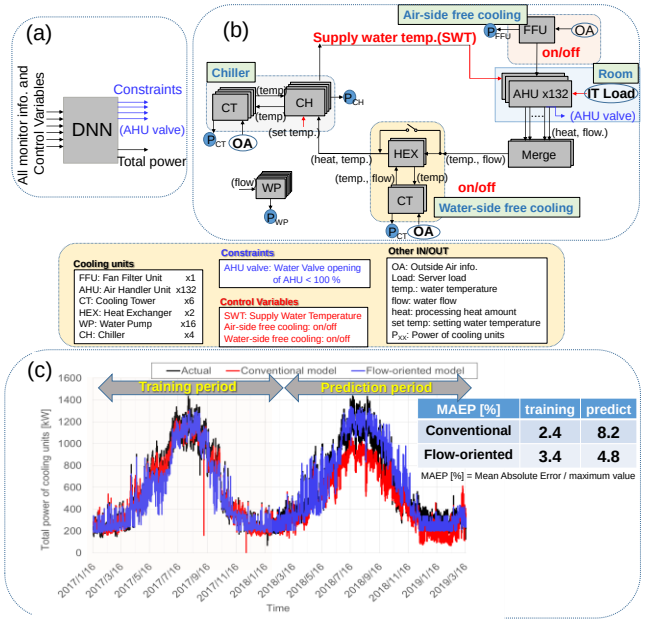


**Fig. 2: (a) Conventional model [9]. (b) Flow-oriented model. (c) Accuracy comparison of total power by each model.**

We adopted DNN as the core regress engine for the modeling because the DNN showed best-class accuracy for the power modeling, compared with other repressors (polynomial regression, support vector regression, and recurrent neural network). Recent progress in the DNN framework also enables easy introduction of a DNN simulator. We compared two DNN modeling approaches by using the same logs. Fig. 2 (a) shows the conventional approach in that all related logs are trained by a single DNN [9]. Fig 2 (b) shows our proposed flow-oriented modeling in that each gray box indicated DNN models. The models were connected based on the actual water flow. Total power was calculated as the sum of each cooling unit power.

Fig 2 (c) shows the calculated total power of the power simulators. A big difference was observed in the prediction-period accuracy. The flow-oriented model achieved a low prediction error of 4.8% as compared to that of 8.2% by the conventional model. The actual control variables of the prediction period were different from those of training period. Conventional model failed to learn the physical dynamics of the cooling units. There are many factors that affect to the total power. Diversity of the actual logs in training was insufficient to learn the correct relationships between those all factors in the conventional model. In the case of the flow-oriented model, input/output of each model was defined based on human knowledge and the DNN can focus on the relatively simple regression. Thus, the flow-oriented model was very robust and it did not require retraining for almost one year. The retraining-less function is very effective in the dynamic control optimization because this DNN training spent three days.

We introduced weight sharing in the AHU models for the AHU to acquire the IT load dependency. Total IT load fluctuation in the system was actually very small but each AHU experienced a different load. To learn IT load dependency, each DNN model of the AHU shared the weight. Thus, one AHU was able to acquire a dependency between IT load, a heat amount and a flow that affected total power, from the logs of other AHUs.

Then, we simulated the control optimization with varying control frequency and compared this with the actual operating policies. The optimization was conducted in three control variables (supply water temperature of the chiller, the water-side free cooling on/off and the air-side free cooling on/off). AHU-valve openings that control water flow in each AHU were predicted as constraints. We also investigated the optimization when a rapid IT load fluctuation exists. The random IT load variation (max: 1.0, min: 0.6, standard deviation=0.16 and frequency = 1 hour, where 1.0 shows actual load on the system) was virtually given to the AHU. This IT load fluctuation is similar to that of a supercomputer site [4].

We calculated a full matrix of the possible control variables (240 conditions) for the optimization from Jan. 2017 to Mar. 2019. The optimization time for a single time point was 80 seconds, which was fast enough for frequent dynamic control. This speed is almost 10,000 times faster than that of the conventional CFD. Fig 3 shows the simulation results. The more the control was frequent, the more the power was reduced. Hourly dynamic control showed a 15% power reduction compared to actual control policies without the load fluctuation (actual load case). We revealed the load fluctuation of 40% increased effectiveness of the frequent control.

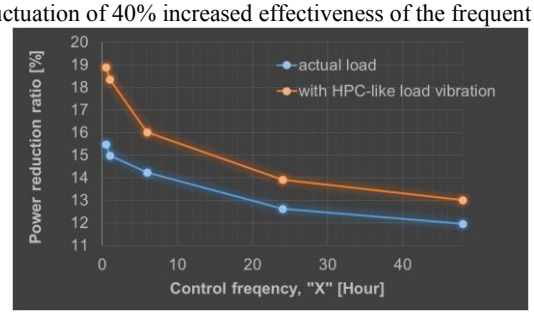


**Fig. 3: Simulated power reduction of the DNN optimization compared to actual policies with/without load variation.**

# REFERENCES


[1] Avgelinou Maria, Bertoldi Paolo, and Castellazzi Luca. Trends in data centre energy consumption under the European code of conduct for datacentere energy efficiency. *Energies*, 10(10):1470, September 2017.

[2] Whitepaper. Energy and power aware job scheduling and power management. *Energy Efficient HPC Working Group*, September 2017.

[3] Jungsoo Kim, Martino Ruggiero, and David Atienza. Free cooling-aware dynamic power management for green datacenters. In *2012 International Conference on High Performance Computing & Simulation (HPCS)*, pages140–146, July 2012.

[4] Shigeto Suzuki, Michiko Hiraoka, Takashi Shiraishi, Hiroyuki Fukuda, Takuji Yamamoto, Shuji Matsui and Atsuya Uno. Power prediction with probabilistic topic modeling for hpc. In *ISC2019 HPC RESEARCH POSTER*, Frankfurt, Germany, June 2019.

[5] Nevena Lazic, Craig Boutilier, Tyler Lu, Eehern Wong, Binz Roy, MK Ryu, and Greg Imwalle. Power prediction with probabilistic topic modeling for hpc. *In Advances in Neural Information Processing Systems 31* (NIPS 2018)

[6] Gianluca Serale, Massimo Fiorentini, Alfonso Capozzoli, Daniele Bernardini, and Alberto Bemporad. Model predictive control(mpc) for enhancing building and hvac system energy efficiency: Problem formulation, applications and opportunities. *Energies*, 11:631, March 2018.

[7] Georgios Varsamopoulos, Michael Jonas, Joshua Ferguson, Joydeep Banerjee, and Sandeep Gupta. Using transient thermal models to predict cyber physical phenomena in data centers. *Sustainable Computing: Informatics and Systems*, 3(3):132–147, September 2013.

[8] Kazumasa Kitada, Yutaka Nakamura, Kazuhiro Matsuda, and Morito Matsuoka. Dynamic power simulator utilizing computational fluid dynamics and machine learning for proposing task allocation in a datacenter. In *CLOUD COMPUTING 2016: The Seventh International Conference on Cloud Computing, GRIDs, and Virtualization*, pages 87–94, 2016.

[9] Jim Gao. Machine learning applications for datacenter optimization. Google White paper, 2014.