Power Prediction for High-performance Computing

Shigeto Suzuki
Fujitsu Laboratories LTD.
Kawasaki Japan
shigeto.suzuki@fujitsu.com

Michiko Hiraoka
Fujitsu LTD.
Kobe Japan
hiraoka.michiko@fujitsu.com

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

Enxhi Kreshpa
Fujitsu Laboratories LTD.
Kawasaki Japan
enxhi.kreshpa@fujitsu.com

Takuji Yamamoto
Fujitsu Laboratories LTD.
Kawasaki Japan
takuji@fujitsu.com

Hiroyuki Fukuda
Fujitsu Laboratories LTD.
Kawasaki Japan
fukuda.hiro@fujitsu.com

Shuji Matsui
Fujitsu LTD.
Kobe Japan
z@fujitsu.com

Masahide Fujisaki
Fujitsu LTD.
Shinagawa Japan
m.fujisaki@fujitsu.com

Atsuya Uno
RIKEN Center for Computational Science
Kobe Japan
uno@riken.jp

ABSTRACT
Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in a K computer as compared to the18.0% obtained using the conventional method.

CCS CONCEPTS
• **Performance Measurement, Modeling, and Tools** → Analysis, modeling, prediction, or simulation methods;• **System Software** → Approaches for enabling adaptive and introspective system software;

Keywords
Power Prediction, HPC, Deep Learning, Job Scheduling

1. Introduction
As their computing performance continues to increase, supercomputers require considerably higher levels of power. The K computer, which was ranked as the most powerful supercomputer according to the TOP500 list in 2011, consumed 12.6 MW of power with a performance of 10-PFlops. Today's most powerful supercomputer is Tianhe-2A that consumes 18.6 MW of power with a performance of 61-PFlops [1]. Power consumption of exascale computers expected in 2021will become a critical limit [2] [3].

Cooling units consume almost 37% power in the whole system [4]. Therefore, reducing cooling-unit power is one approach to alleviate the power limitations of the whole system.

Generally, cooling units are operated at a cooling capacity sufficient for the maximum power of a computing system. This can lead to overcooling in cases of low system utilization. To avoid unnecessary cooling, predictive control has been studied [5, 6, 7].

Because the time constant of cooling-unit stabilization is considerably longer than that after a change in computing system power, key challenge is a highly precise time-series prediction of the power consumed by each job.

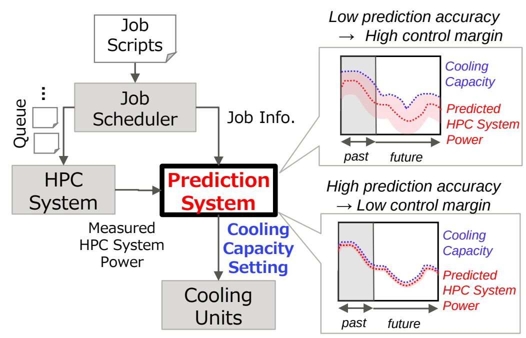

Fig. 1: Proposed HPC system with predictive power control

Fig. 1 shows the concept behind our proposed HPC system based on predictive power control. After jobs are submitted and scheduled in a queue, the queued jobs are executed in order and the HPC system power varies. Cooling units are frequently controlled in advance with adjustments to the variation in predicted HPC system power. As the control margin is determined by the prediction accuracy, we set the target for it less than 10%. Prediction period is set to 30 minutes from the cooling units' time constants, mostly within 30 minutes [6, 7, 8].

2. Proposed method
Most previous studies on HPC power prediction focused on queued-job prediction because their main aims were power-aware

scheduling or power capping. Queued jobs are pre-scheduled before their execution based on the predicted power for this purpose [5,10-12]. On the other hand, predictive control of cooling units allows to update the prediction after a job has shifted to the running state. Fig. 2 illustrates an example of job states at a certain time. In Fig.2, 100% of the total system power consists of running jobs (Job2, Job3, Job6, Job7 and Job8) at current time (T_C). Job9 and Job10 are still in the queue at T_C; however, they will move to the running states during the prediction period. The power at prediction target time (T_{PT}) consists of these queued jobs and current running jobs still executing at T_{PT}. Therefore, it is important to account for both the running jobs and queued jobs for power prediction. In this paper, we propose an effective running-job prediction model. As for queued-job prediction, we adopted the method we proposed in [12].

The proposed running-job prediction model (Fig.3) is combined two prediction methods with different features; one is a power correlated topic model (PCT) and the other is a variable recurrent neural network (VRNN).

PCT is a two-step predicting model. First, while job is in the queue, it uses a topic model to search for the multiple candidates that have similar job information from the past jobs. Second, after the job starts running, it compares the power of each candidate to the actual power history from start to present, and updates the prediction to the candidate with the smallest average error. Then, second step is repeated every time step predefined. In most cases, we can expect that there is at least one candidate in the first selected multiple candidates and the update converges. In case there is no similar candidate, i.e. the case PCT fails, we developed VRNN in addition.

VRNN; recurrent neural network model considering variable network sizes, predicts the future variation of power from the power history of itself. Because it is a time-series regression model, it is hard to train the model by many power histories with various features, especially for the variety of execution time. To improve the convergence, we design VRNN to contain multiple prediction models based on the execution time. As each prediction model is trained by the power histories with almost same execution time, it can improve the regression convergence and prediction accuracy.

From the prediction results of the two models, i.e. PCT and VRNN, one prediction result is selected based on the average error between the prediction and actual histories as same algorithm used in PCT.

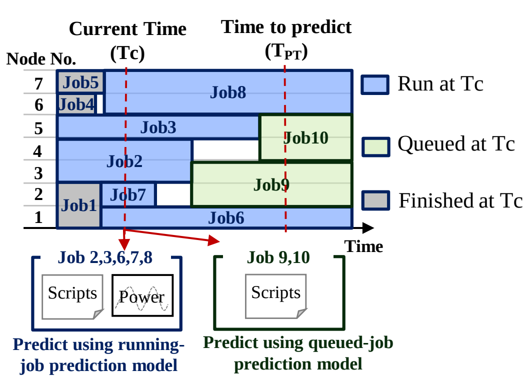

Fig. 2: An example of job states for power prediction.

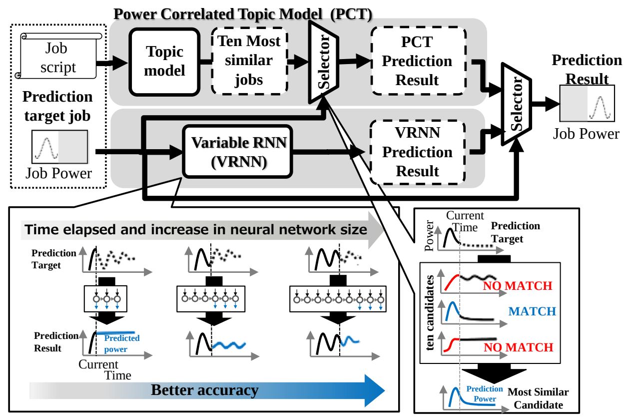

Fig. 3: Block diagram of proposed power prediction system for HPC.

3. Results

Fig. 4 shows the evaluation results of the predicted total job power in the K computer. The evaluation period is July to September 2017. A comparison of the queued-job prediction model (our previous work) [12] with the queued-job and running-job prediction models shows that the prediction accuracy can be improved by 12.3%. A comparison of time-series predictions is shown in Fig.5.

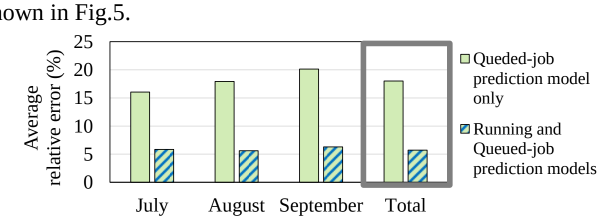

Fig. 4: Prediction results of total job power. Tow predication results are shown for comparison.

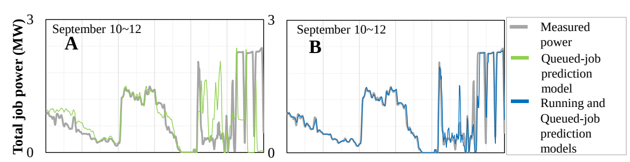

Fig. 5: Examples of total job power predict. A: Queued-job prediction model (September 10 to 12). B: Running-job and Queued-job prediction model (Same period as A).

4. Conclusion

We developed a highly accurate total job power predicting system to control the cooling units of a large-scale HPC system. By adopting the proposed method, we achieved an average relative error of 5.7% in a K computer as compared to the 18.0% obtained using the conventional one. This improvement in prediction accuracy contributes to reduce the cooling-unit power. For example, assuming 3-MW cooling units, relative errors of 18.0% and 5.7% are corresponding to the control margins of 0.54 MW and 0.17 MW, respectively. Therefore, it is possible to reduce cooling-units power by 0.37 MW (i.e., 12%).

5. REFERNCES

[1] TOP500 Lists. https://www.top500.org/lists/2018/11/

[2] The race to exascale: A story of superpowers and supercomputers. https://www.datacenterdynamics.com/analysis/superpowers-supercomputers-and-race-exascale/

[3] Norman P Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, et al. In-datacenter performance analysis of a tensor processing unit. In Proceedings of the 44th Annual International Symposium on Computer Architecture, pp. 1-12. ACM, 2017.

[4] 2018 Data Center Industry Survey Results - Uptime Institute LLC https://datacenter.com/wp-content/uploads/2018/11/2018-data-center-industry-survey.pdf

[5] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini. Predictive Modeling for Job Power Consumption in HPC Systems. In High Performance Computing: 31st International Conference, ISC High Performance 2016, Frankfurt, Germany, June 19-23, 2016.

[6] Y. Tarutani, K. Hashimoto, G. Hasegawa, Y. Nakamura, T. Tamura, K. Matsuda, and M. Matsuoka, "Temperature distribution prediction in data centers for decreasing power consumption by machine learning," December 2015.

[7] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, A. Bemporad Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities Energies, 11 (2018), p. 631

[8] Y. Chai, A. Wu, N. Dong, Y. Wang and Y. Li: "Dynamic Operation and Control Strategy of Abcorption Chiller under different working Conditions" In proceedings of the 13th World Congress on Intelligent control and Automation, 2018

[9] S. Wallace, X. Yang, V. Vishwanath, W. E. Allcock, S. Coghlan, M. E. Papka, and Z. Lan, "A data driven scheduling approach for power management on hpc systems," SC16: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 656–666, Nov 2016.

[10] Storlie, C., Sexton, J., Pakin, S., et al. "Modeling and predicting power consumption of high performance computing jobs", arXiv preprint arXiv:1412.5247, 2014.

[11] S. Wallace, X. Yang, V. Vishwanath, W. E. Allcock, S. Coghlan, M. E. Papka, and Z. Lan, "A data driven scheduling approach for power management on hpc systems," SC16: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 656-666, Nov 2016.

[12] S Suzuki, M Hiraoka, T Shiraishi, H Fukuda, T Yamamoto, S Matsui, A Uno, "Power prediction with probabilistic topic modeling for HPC", ISC2019 HPC RESEARCH POSTER, Frankfurt, Germany ,2019