Online Machine Learning-based Temperature Prediction for Thermal-aware NoC System

Kun-Chih (Jimmy) Chen and Yuan-Hou Liao
Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan

Abstract— **The Network-on-Chip (NoC) has been proposed to solve the communication problem in multicore systems, but it usually suffers from the serious thermal problem due to high power density. To solve this problem, the proactive thermal management (PDTM) has been proven as an efficient way to prevent the system from overheating and mitigate the performance impact during the temperature control period. Based on the predicted temperature results, the PDTM controls the system temperature in advance to make the system temperature under the thermal limit. However, due to the thermal-coupling effect on the chip, it is hard to have a precise thermal prediction model, which makes the PDTM cannot control the system temperature efficiently. In this paper, we propose a lightweight thermal prediction model based on the machine learning method accompany with an online training algorithm that can adapt the hyperplane of the temperature behavior of NoC system during the runtime. The proposed model can adapt varying situations of the temperature behavior of NoC systems on the fly. Compared with the traditional thermal prediction model, the proposed approach can reduce 40.6-51.5% average error and 39.4%-54.2% maximum error.**

Keywords: online learning; neural network; temperature prediction

I. INTRODUCTION

The complexity of multiprocessor system grows with respect to the advance in the semiconductor technology. The Network-on-Chip (NoC) has been proposed as an efficient solution to solve the interconnection among processing elements [1]. However, because of the high power density on the chip, the NoC system usually suffers from the severer thermal problem. The thermal issue leads to many negative impacts, such as lower system reliability and longer system latency. To prevent the NoC system from overheating, the Dynamic Thermal Management (DTM) is usually applied to keep the temperature of system under the thermal limit.

Conventional DTM is activated to throttle the nodes when the temperature of the nodes exceeds the thermal limit of system. However, as the thermal-emergent nodes are cooled down through DTM scheme, it will result in significant performance impact. Therefore, proactive dynamic thermal management (PDTM) which control temperature in advance by using predicted temperature is proposed to mitigate the performance impact during the temperature control period [2]. However, most of the temperature prediction models in the traditional PDTM techniques are based on certain prior-knowledge of the thermal behavior [3][4]. Hence, the conventional temperature prediction models are not always available for the various system workload. The varying workload would increase the error of temperature prediction, which affects the efficiency of the involved PDTM scheme significantly, as shown in Fig. 1(a).

In this paper, we propose a lightweight thermal prediction model based on machine learning for NoC systems. By the machine learning technique, the prediction model can learn the temperature behavior of system during training phase. We employ the Artificial Neural Network (ANN) method with the factors that causing temperature changes as input to build proposed prediction model. However, the weight of the model trained on pre-defined dataset could not adapt all situation of temperature changes. We thus further adopt Stochastic Gradient Descent (SGD) [5] to dynamically adjust the weight of model on runtime phase. The prediction error and performance impact would be reduced through the adjustment mechanism, as shown in Fig. 1(b). The contributions of this paper are

This work is supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2218-E-110-010.

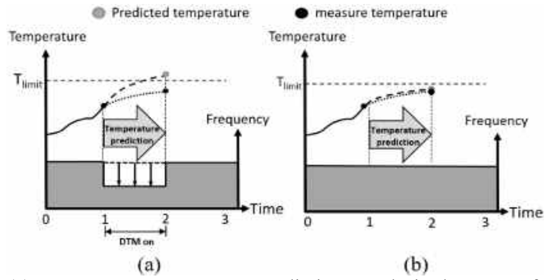

Fig. 1 (a) Inaccurate temperature prediction result in larger performance impact; (b) our goal is to predict more accurately to mitigate performance impact.

1) We adopt a lightweight ANN-based thermal prediction model to predict the system temperature.
2) We involve the weights training algorithm to propose an online learning approach to adapt the different temperature change during the runtime operation.

II. ONLINE MACHINE LEARNING-BASED TEMPERATURE PREDICTION MODEL

A. ANN-based Temperature Prediction Model

The conventional temperature prediction model is usually built based on physical knowledge. Zhang *et al.* proposed a thermal prediction model in [6], which is formulated as

$$T(k+N \mid k) = B^{N} T(k \mid k) + \sum_{i=k}^{q} B^{q-i} Du(i) . \qquad (1)$$

The B and D are both coefficients depending on the parameters of the certain thermal-RC circuit and $u(i)$ is the power vector. Because it is necessary to understand the physical phenomenon to build this model, it is difficult to find these detailed physical characteristics as the scale of system becomes large.

To solve the aforementioned problem, a method which can find the hyperplane of the temperature behavior of a system during the runtime is needed. Because the model in (1) is composed of a series of multiply and accumulate, the model can be seen as an operation of a neuron in ANN. The ANN is usually used to transform the input to proper expected target, which can be used to approximate a hyperplane of temperature behavior. Thus, we can rewrite the (1) as

$$T(k+N \mid k) = \sum_{n=1}^{k-q+2} W_n \cdot X_n , \qquad (2)$$

where W_n is equal to B and D in (1). X_n in (2) is the input data $T(k \mid k)$ and $u(i)$. Through ANN approach, we can calculate the similar computation in (1). Besides, based on the weight training (*i.e.*, W_n in (2)) approach, we can obtain the weight to build a model instead of certain physical coefficients.

Generally, the input of ANN computation affects the precision of prediction result significantly. The current temperature and the throttling activities of the local node and surrounding nodes are key factors causing the temperature change of the local node. Hence, in this work, we set the temperature and the throttling activities of the local node and surrounding nodes as the input data of the ANN neuron computing. Fig. 2 illustrates an

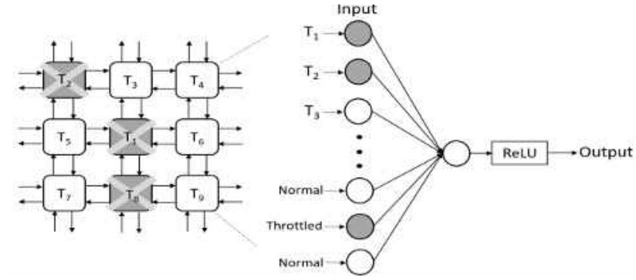

Fig. 2 Proposed ANN-based thermal prediction model

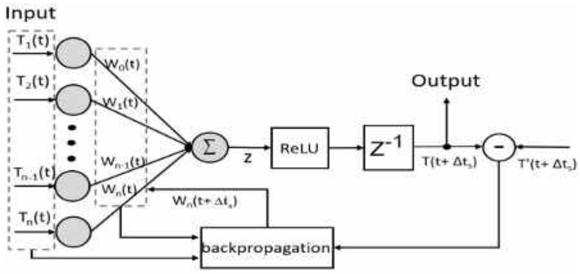

Fig. 3 Block diagram of proposed thermal prediction model with backpropagation algorithm.

example. By using the proposed ANN model, the temperature of the node1 can be calculated with the several inputs (*i.e.*, T_n and throttling information of nodes). On the other hand, to reduce the hardware overhead, we assume each NoC node has its own single-neuron ANN-based computing unit. In this work, we adopt the ReLU as the activation function because the temperature of the chip is positive, and the temperature change can be considered linear in a short time. After the proposed ANN-based temperature prediction computing, the information of the predicted temperature can be obtained.

B. Online learning Algorithm in the Model

In our proposed ANN-based temperature prediction model, the result of temperature prediction is obtained by using pre-trained weights. However, the pre-trained weights highly depend on pre-defined training dataset which cannot cover every situation of temperature change, making the prediction model unable to adapt to high-diverse temperature behavior of the system. Therefore, it is necessary to adjust the weights during the runtime based on the different temperature situation.

To adjust the weight during the runtime operation, we adopt the SGD [5], which is widely used backpropagation algorithm to train the weight of the ANN, to calculate gradient to find the minimum value of loss function and update the weight. The formulation of the gradient can be defined as

$$W_n(t+\Delta t_s) = W_n(t) - \eta \frac{\partial E(t)}{\partial W_n(t)}, \quad (3)$$

where

$$E(t) = (T'(t+\Delta t_s) - T(t+\Delta t_s))^2 . \quad (4)$$

The E is loss function and the η is the learning rate; the W_n is the trained weight of the ANN; the Δt_s is the thermal sensing period. In (3), the partial differential of $W_n(t)$ to $E(t)$ can be derived as

$$\frac{\partial E(t)}{\partial W_n(t)} = (-2)T_n(t)(T'(t+\Delta t_s) - T(t+\Delta t_s)) . \quad (5)$$

According to (5), we can rewrite (3) as

$$W_n(t+\Delta t_s) = W_n(t) - \eta T_n(t)(T'(t+\Delta t_s) - T(t+\Delta t_s)) , \quad (6)$$

where the $T_n(t)$ is the current input data. In each thermal sensing period, the backpropagation algorithm is used to dynamically adjust the weight based on the loss function. The $W_n(t)$ is adjusted based on prediction error ($T'(t+\Delta t_s)$ - $T(t+\Delta t_s)$) and current input data $T_n(t)$, as shown in Fig. 3.

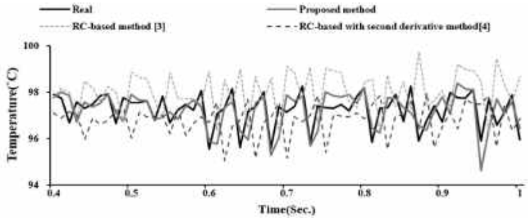

Fig. 4 Comparison of predictive temperature under random traffic.

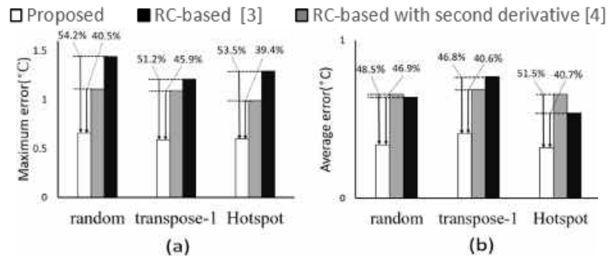

Fig. 5 The comparison of (a) the average and (b) maximum temperature prediction error under three traffic patterns.

III. EXPERIMENTAL RESULT AND ANALYSIS

To evaluate the proposed thermal prediction model, we use a traffic--thermal co-simulator [7] to simulate an 8x8 NoC system to simulate three difference synthetic traffic patterns. Fig. 4 shows the comparison between the actual temperature and predicted temperature by using different approaches. Compared with the conventional approaches in [3] and [4], the proposed approach can achieve more precise temperature prediction results. The reason is that the conventional approaches are executed based on thermal RC model and only consider the quasi-stationary case. In Fig. 5, because the proposed thermal prediction model learns the temperature behavior of the system, the prediction result is more accurate than [3] and [4]. The proposed approach can reduce 40.6%-51.5% average error and 39.4%-54.2% maximum error.

IV. CONCLUSION

We propose a novel temperature prediction method by using lite online learning model. We adopt ANN to predict the temperature of the local node in NoC system in the next temperature sensing period. Based on the prediction error information, we adopt an online learning model to adjust the weight during the runtime period. The experiment result shows that the proposed approach can reduce 40.6%-51.5% average error and 39.4%-54.2% maximum error.

REFERENCES

[1] Y. Hoskote, *et al.*, "A 5-GHz mesh interconnect for a teraflops processor." *IEEE Micro,* vol.27, no. 5, pp. 51-61, Sep. 2007.

[2] T. Wegner, M. Gag, and D. Timmermann, "Impact of proactive temperature management on performance of networks-on-chip," *Proc. Int. Symp. System on Chip*, Oct. 2011, pp. 116–121.

[3] K.-C. Chen, E.-J. Chang, H.-T. Li, A.-Y. Wu, "RC-based temperature prediction scheme for proactive dynamic thermal management in throttle-based 3D NoCs", *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 206-218, Jan. 2015.

[4] Z. Lei, *et al*. "A Predictive Thermal Model for Multiprocessor System-on-Chip," *Proc. of the World Congress on Engineering and Computer Science.* , Oct. 2015, pp. 27-32.

[5] S. Amari, "Backpropagation and stochastic gradient descent method", *Neurocomput.*, vol. 5, no. 5, pp. 185-196, Jun. 1993.

[6] Y. Zhang, A. Srivastava, "Accurate temperature estimation using noisy thermal sensors for Gaussian and Non-Gaussian cases", *IEEE Trans. Very Large Scale Integration (VLSI) Systems,* vol. 19, no.9, pp. 1617 – 1626, Sep. 2011.

[7] K.-Y. Jheng, C.-H. Chao, H.-Y. Wang, A.-Y. Wu, "Traffic-thermal mutual-coupling co-simulation platform for three-dimensional network-on-chip", *Proc. Int. Symp. VLSI-DAT*, Apr. 2010, pp. 135-138.