LSTM-based Temperature Prediction and Hotspot Tracking for Thermal-aware 3D NoC System

Tong Cheng, Haoyu Du, Li Li*, *Member, IEEE* Yuxiang Fu*, *Member, IEEE*
School of Electronic Science and Engineering,
Nanjing University
Nanjing, China
Email: yuxiangfu@nju.edu.cn

2021 18th International SoC Design Conference (ISOCC) | 978-1-6654-0174-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISOCC53507.2021.9613862

Abstract— **As the increasingly large energy consumption and diverse workload distribution in many-core systems, thermal problems, especially hotspots, have become a significant restriction for the performance of the Network-on-Chip system. Therefore, several proactive thermal management methods based on thermal prediction data were proposed. However, previous temperature prediction methods did not consider the impact of hotspots, and their prediction accuracy is not satisfying. In this paper, we propose a Long Short-Term Memory (LSTM)-based temperature prediction and hotspot tracking model in a thermal-aware 3D NoC system. As the experiment indicates, the Mean-Square-Error (MSE) of the 6-step temperature prediction of the proposed model is 0.411°C and the response time for tracking hotspot transfer is less than 0.075 ms.**

Keywords: On-Chip Network; temperature prediction; Long Short-Term Memory

I. INTRODUCTION

Multi-core system such as Systems-on-Chips (SoCs) is an effective way to implement high performance computing and integrates different types of cores into one chip. Following the development of the multi-core system, Network-on-Chip (NoC) [1] was proposed as a pivotal method for optimizing internal communication. Meanwhile, continually growing energy consumption coupled with uneven workload distribution in the NoC system can result in severe thermal issues, especially in three-dimensional Network-on-Chip (3D NoC) [2]. Hotspots are regional overheating nodes in the NoC system which receive vast quantities of packages spatially or temporally and always lead to long latency, performance decrease, and system unreliability [3].

To cope with the aforementioned challenges, Proactive Dynamic Thermal Management (PDTM) [4], [5]was proposed as a critical component in the NoC system to facilitate more proper thermal distribution. PDTM can control the temperature in advance based on the predicted temperature which highly relies on the prediction accuracy and the prediction step length of the temperature prediction model.

Recently proposed temperature prediction methods such as AutoRegressive Moving Average (ARMA)-based model [6] and Linear Regression-based model [7] both neglect the importance of the long-step prediction model and not considering challenges brought in by hotspots including the random drift and rapid temperature change of hotspots. Peculiarities of hotspots mentioned above place higher requirements for an efficient temperature prediction method which needs to predict potential hotspot far ahead of schedule to guarantee a moderate adjustment in PDTM.

In this paper, we propose a temperature prediction model based on LSTM in thermal-aware 3D NoC systems which can precisely do a multi-step-ahead prediction with outstanding accuracy. Besides, under the hotspot drift scenario, the proposed model can track hotspots with low response time.

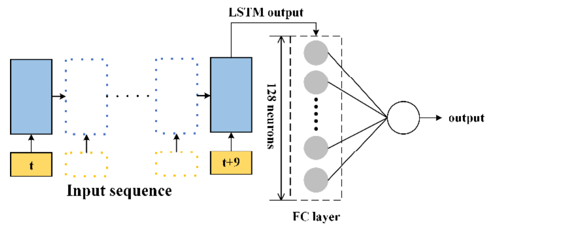

Fig. 1 Proposed neural network structure for temperature prediction

II. LSTM-BASED TEMPERATURE PREDICTION AND HOTSPOT TRACKING MODEL

LSTM is an improved version of the traditional Recurrent Neural Network (RNN), which is now pervasively used in time series prediction problems. LSTM and RNN both use the former time step output information and state information as the input of the current model. But the hidden layer in the LSTM model is more complicated than that in the RNN model, functioning as different gates to control data delivery.

Fig. 1 presents our neural network structure with an LSTM layer of 10-time-step input followed by a full connection layer with 128 neurons. To train the network, we adopt Relu and Sigmoid as state activation and gate activation, Adam as the optimizer, and a dropout layer with a 0.3 dropout probability to reduce overfitting. Furthermore, we set the batch size as 128 and epochs as 500 by trial and error and the loss function is MSE.

As shown in Fig. 2, the proposed model can be integrated in the NoC system as an independent global processing unit. Thermal sensors collect temperature data of each node in the 3D NoC mesh. By receiving temperature data and doing spatial and temporal analysis, the proposed model can have a better grasp on the overall temperature distribution and temperature changes of nodes of concern. Furthermore, PDTM will control temperature rise of the predicted potential hotspot.

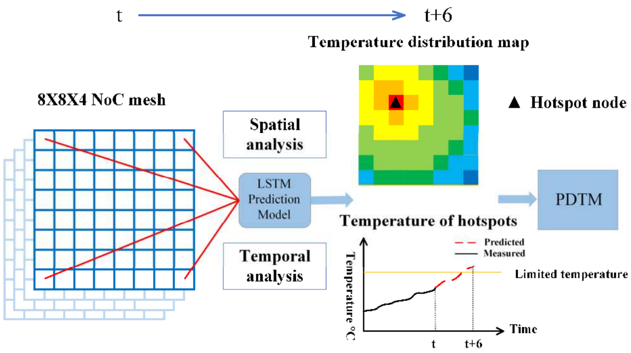

Fig. 2 Overall of the proposed prediction scheme for PDTM

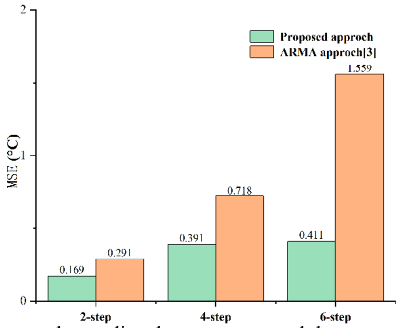

Fig. 3 MSE between the predicted temperature and the measured temperature by using the proposed model and the ARMA model [6]

III. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed thermal prediction and hotspot tracking model and compare it with an ARMA model [6], which is a traditional regression temperature prediction method. To establish the data set for building and evaluating the proposed LSTM model, we constructed a 3D NoC system on the AccessNoxim [8], which is an accurate NoC traffic and temperature simulator. AccessNoxim integrates traffic simulator--Noxim and thermal simulator Hotspot to realize a traffic-thermal mutual-coupling co-simulation platform. We built an 8×8×4 3D NoC system whose buffer depth in every node is 8 flits, and adopted the XYZ routing algorithm with the modified hotspot traffic pattern which can create several dynamic changing hotspot nodes in the system. To strengthen our temperature prediction model's robustness, we also added random fluctuation to the package injection rate to simulate the data flow fluctuation in real physical scenes.

The ARMA predictor proposed in [6] assumes that the temperature changing is a stationary stochastic process, which is not suitable for a random shift of hotspots in many practical cases. As shown in Fig. 3, the proposed model archives higher accuracy than the ARMA model [6] as the prediction step increases, improving by 41.92% to 73.63%.

In the AccessNoxim, the temperature sampling interval is set to 0.025 ms. Fig. 4 demonstrates that a node with normal temperature becomes a hotspot node at the sampling time t1 and its temperature then starts to rise rapidly. The proposed temperature prediction model can keep up with the temperature change of potential hotspot and locate the new hotspot, then makes an alarm after three intervals in less than 0.075 ms.

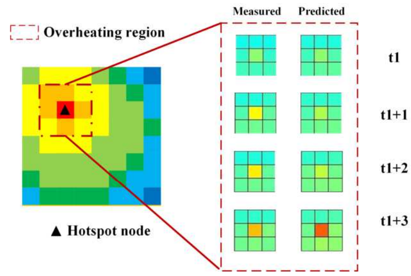

Fig. 4 Rapidly locating the new hotspot node using the proposed model

IV. CONCLUSION

In this paper, we focus on hotspots and propose a LSTM-based temperature prediction and hotspot tracking model for thermal-aware 3D NoC system. Experiment demonstrates that our model can locate the new hotspot within 0.075 ms. Besides, the average prediction accuracy is improved by 41.92% to 73.63% compared with ARMA-based temperature prediction method [6].

This work was supported in part by the National Nature Science Foundation of China under Grant No. 62104098 and in part by the Natural Science Foundation of Jiangsu Province for Youth.

REFERENCES

[1] W. Dally and B. Towles, “Route packets, not wires: on-chip interconnection networks,” in *Proc. DAC*, 2001, pp. 684–689.

[2] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, D. McCaule, P. Morrow, D. W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. Shen, and C. Webb, “Die stacking (3d) microarchitecture,” in *2006 39th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO’06)*, 2006, pp. 469–479.

[3] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. Stan, “Hotspot: a compact thermal modeling methodology for early stage vlsi design,” *IEEE Trans. VLSI Syst.*, vol. 14, no. 5, pp. 501–513, 2006.

[4] T. Wegner, M. Gag, and D. Timmermann, “Impact of proactive temperature management on performance of networks-on-chip,” in *2011 International Symposium on System on Chip (SoC)*, 2011, pp. 116–121.

[5] Y. Fu, L. Li, K. Wang, and C. Zhang, “Kalman predictor-based proactive dynamic thermal management for 3-d noc systems with noisy thermal sensors,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 11, pp. 1869–1882, 2017.

[6] A. K. Coskun, T. S. Rosing, and K. C. Gross, “Utilizing predictors for efficient thermal management in multiprocessor socs,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 10, pp. 1503–1516, 2009.

[7] E. W. Wachter, C. de Bellefroid, K. R. Basireddy, A. K. Singh, B. M. ¨ Al-Hashimi, and G. Merrett, “Predictive thermal management for energy efficient execution of concurrent applications on heterogeneous multicores,” *IEEE Trans. VLSI Syst.*, vol. 27, no. 6, pp. 1404–1415, 2019.

[8] K.-Y. Jheng, C.-H. Chao, H.-Y. Wang, and A.-Y. Wu, “Traffic-thermal mutual-coupling co-simulation platform for three-dimensional network-on-chip,” in *Proc. VDAT*, 2010, pp. 135–138.