Design of Thermal Management Unit with Vertical Throttling Scheme for Proactive Thermal-aware 3D NoC Systems

Kun-Chih Chen[†], Shu-Yen Lin[‡], and An-Yeu (Andy) Wu[†]
[†] Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 10617, Taiwan
[‡] Information and Communications Research Laboratories, ITRI, Hsinchu, 31040, Taiwan

ABSTRACTION

The three-dimensional Network-on-Chip (3D NoC) has been proposed to solve the complex on-chip communication issues. However, the thermal problems become more exacerbated because of the larger power density and the heterogeneous thermal conductance in different silicon layer of 3D NoC. To regulate the system temperature, the Dynamic Thermal Management (*DTM*) techniques will be triggered when the device is thermal-emergent. However, these kinds of reactive *DTM* schemes result in significant system performance degradation. In this paper, we propose a proactive *DTM* with vertical throttling (*PDTM-VT*) scheme, which is controlled by the distributed Thermal Management Unit (*TMU*) of each NoC node. Based on the expected temperature resulted from the proposed thermal prediction model, the *TMU* can early control the temperature of the thermal-emergent device. The experimental results show that the proposed thermal prediction model has less than 0.25% prediction error against actual temperature measurement within 50ms. Besides, the *PDTM-VT* can reduce 11.84%~23.18% thermal-emergent nodes and improve 0.47%~47.90% network throughput.

1. INTRODUCTION

As the complexity of System-on-Chip (SoC) grows with Moore's law, the three-dimensional Network-on-Chip (3D NoC) has been proposed to provide larger interconnection bandwidth to achieve higher performance with lower power consumption [1]. However, the thermal issues become the main challenges of 3D NoC due to die stacking [1][4]. Besides, the routers have been shown as the sources generating thermal hotspots due to their higher switching activity, which leads to severer heat problem in 3D NoC systems [2].

In the modern thermal-aware 3D NoC systems, some dynamic thermal managements (*DTMs*) were proposed to regulate the system temperature, and they can be classified into temporal managements and spatial ones [3]. The temporal *DTMs* slow down the activities of the overheated nodes to perform cooling. The Dynamic Frequency Scaling (*DFS*), Dynamic Voltage Scaling (*DVS*), Dynamic Voltage and Frequency Scaling (*DVFS*), Clock Throttling *etc.* are the popular temporal *DTMs* [2][4]-[6]. Although the temporal *DTMs* can regulate the system temperature within short cooling time, they lead to significant performance overhead. On the other hand, the spatial *DTMs* can control the system temperature without frequency scaling through packet migration [7]. However, the spatial *DTMs* need long cooling time because of the thermal coupling between each node and the heterogeneous capability of thermal conductance in each silicon layer of a 3D NoC system [3]. Therefore, the temporal *DTMs* are adopted in this work due to the better cooling efficiency for the 3D NoC systems.

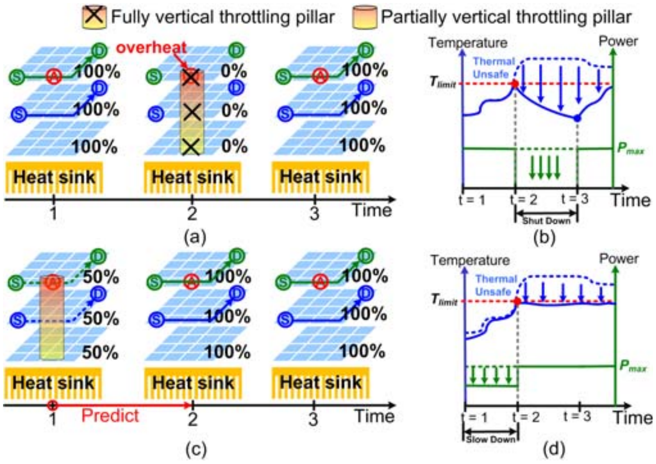

Fig. 1 (a)(b) Reactive *DTM-VT* leads to performance impact, and (c)(d) *PDTM-VT* improves the system performance.

To mitigate the performance impact caused by the temporal *DTMs*, the proactive *DTMs* were proved as an efficient way to avoid thermal emergency and performance degradation [3]. In [6], Wegner *et al.* proposed a central control scheme to proactively control the temperature of an NoC system. However, the central control results in serious communication overhead in a large scale NoC system. In [5], Zanini *et al.* proposed a proactive *DTM* to assign the feasible clock frequency of a multi-core system. However, the computation complexity is very large because of the iterative algorithm for the error reduction. In [8], Cochran *et al.* proposed a table-based two-phases thermal management. However, the large table requirement is not feasible to apply in an NoC system. In [9], Ayoub *et al.* proposed to use the thermal resistance and thermal capacitance (*Thermal RC*) model to predict the future temperature and perform temperature control for a memory subsystem. However, it still suffers from the large performance impact due to the memory gating.

In this paper, based on the *Thermal RC* model, we propose a *Thermal RC*-based thermal prediction model. Because of the applied *Thermal RC* model, the proposed thermal prediction model has a constant computation complexity. Based on the predictive temperature, a proactive *DTM* is proposed to maximize the system performance under a certain thermal limit. Because the previous proposed vertical throttling (*VT*) scheme has better cooling capability and smaller performance impact [2][4], the *VT* is involved as the *DTM* policy in this work. Fig. 1 illustrates an example. The performance is degraded rapidly, if the *DTM-VT* is triggered while node *A* is overheated at time 2, as shown in Fig. 1(a)(b). In the proposed proactive *DTM-VT* (*PDTM-VT*), the performance impact will be reduced, because we can early control the temperature with frequency scaling at time 1, as shown in Fig. 1(c)(d). The contributions of this paper are summarized as follows for clarity:

978-1-4673-4436-4/13/$31.00 ©2013 IEEE

1) A *Thermal RC*-based prediction model is proposed with low computation complexity.
2) A proactive *DTM-VT* is proposed for early temperature control.

We use the traffic-thermal mutual coupling co-simulation platform [10] to demonstrate our proposed *PDTM-VT*. The experimental results show that the prediction error of the proposed prediction model is less than 0.25% against actual temperature measurement within 50ms. Besides, the proposed *PDTM-VT* can improve 0.47%~47.90% network throughput due to early temperature control.

The rest of this paper is organized as follows. In Section 2, we introduce some related proactive dynamic thermal managements. In Section 3, the proposed *PDTM-VT* scheme is described. In Section 4, the experiments are shown and discussed. Finally, we conclude this paper in Section 5.

2. RELATED WORKS

A. *Central Thermal Management Unit for NoC System [6]*

To early control the temperature of the NoC system, Wegner *et al.* proposed a mechanism of central temperature control. Based on the *Thermal RC* model, each node of the NoC system can predict the future temperature. The central thermal unit will assign the clock frequency for each node based on the information of predictive temperature through sending the instruction packets. However, the method of central control is not feasible to apply in a large scale 3D NoC system due to the following two problems:

1) The extra instruction packets may increase the heavy traffic load.
2) The response time is increased with respect to the NoC scale.

B. *Approximate Explicit Model Predictive Thermal Control [5]*

For the temporal *DTM*, the frequency assignment of each device is important to determine the performance. Based on the thermal emergency state and the current operation frequency of each device, Zanini *et al.* obtains the feasible frequency through the proposed approximate explicit model. To reduce the approximation error, the authors proposed to use an iterative computation for error reduction. However, the iterative computation results in long response time.

C. *Phase-aware Thermal Prediction Methodology [8]*

Cochran *et al.* proposed to use *k*-means clustering to classify the global workload phases and the corresponding thermal models in offline analysis. By using the performance counter, the transient workload can be measured. Based on the analytic results in offline phase, the changes of thermal behavior can be predicted. Although the method can be applied in heterogeneous workload, the large table requirement results in significant area overhead.

D. *Thermal RC-based Prediction Model for Memory Subsystem [9]*

Ayoub *et al.* proposed to use *Thermal RC* model to predict the future temperature of a memory subsystem. Based on the results of the predictive temperature, the authors use memory gating and control the fan speed to achieve the energy efficiency under a certain thermal limit. The method has the benefit of low computational complexity and application independency. However, it still suffers from the significant performance overhead due to the memory gating.

3. PROACTIVE DYNAMIC THERMAL MANAGEMENT WITH VERTICAL THROTTLING (*PDTM-VT*) SCHEME

The framework of the proposed proactive dynamic thermal management with vertical throttling (*PDTM-VT*) is shown in Fig. 2, which is controlled by an embedded Thermal Management Unit (*TMU*). Because the embedded thermal sensor is a popular practice method in the thermal-aware architecture, we assume the *TMU* can be integrated in each node of the NoC system. The framework of the *PDTM-VT* contains two phases:

1) *Thermal Prediction Phase*: Based on the *Thermal RC* model, the future temperature ($T_{predict}$) can be predicted, which will be introduced in Section 3.B.
2) *Thermal Management Phase*: Based on the $T_{predict}$ and the current one ($T_{current}$), the embedded *TMU* can adaptively control the temperature with hybrid reactive control and proactive one for a pillar, which will be introduced in Section 3.C.

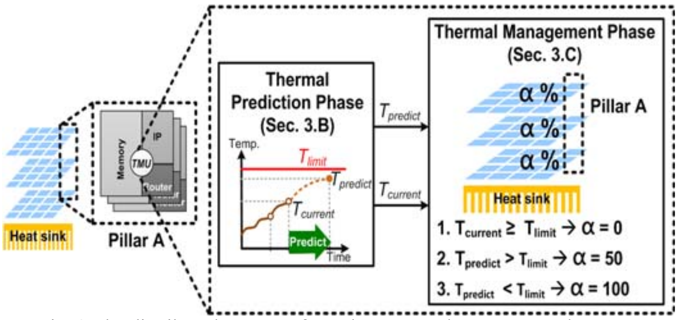

Fig. 2 The distributed *TMU* perform the proposed *PDTM-VT* scheme.

A. *Thermal RC-based Thermal Model*

To obtain the involved thermal model, we follow the derivation of the heat transfer equation in [11]. We assume that the *P(t)* and *T(t)* are the transient status of power consumption and temperature of a specific node in the network, respectively. By the Fourier's Law and the First Law of Thermodynamics, the change of temperature in a time unit can be formulated as:

$$\frac{dT(t)}{dt} = \frac{P(t)}{C} - \frac{T(t)}{RC}, \tag{1}$$

where *R* and *C* are the thermal resistance and thermal capacitance, respectively. To solve the linear equation (1), we assume that the initial temperature at t_0 is T_0 (*i.e.,* $T(t_0)=T_0$), and the steady-state temperature is T_{ss} (*i.e.,* $T(\infty)=T_{ss}$). Besides, to simplify the problem, the *P(t)* is assumed as a constant in usual practice. Therefore, the employed thermal model in this work can be written as:

$$T(t) = T_{ss} - (T_{ss} - T_0) \cdot e^{-bt}, \tag{2}$$

where *b* is a technology-dependent constant, which is equal to *1/RC*. In this work, we employed the default setting in the popular thermal simulator, *HotSpot* [12], and the *b* is equal to 1.98.

B. *Proposed Thermal RC-based Thermal Prediction Model*

To control the system temperature before thermal emergency, we present a thermal prediction scheme in this section. In the normal operation (*i.e.,* the operation period when the *DTM* is not triggered), the change of temperature is usually an exhaustive increasing trend, as shown in Fig. 3(a). The design goal of the thermal prediction model is to precisely predict the time when the temperature of the node achieves the temperature threshold in the normal operation.

If we assume the embedded thermal sensor provide an information of temperature every thermal sensing period Δt_s, the temperature at the time after $k\Delta t_s$ (as shown in Fig. 3(a)) can be described as:

$$T(t+k\Delta t_s) = T(t) + \Delta T(k\Delta t_s)\,, \quad (3)$$

where k is the thermal prediction distance (*i.e.,* the thermal sensing time far away from the current time).

To predict the $\Delta T(k\Delta t_s)$, similar to [9], we use the derivative analysis to extract the temperature difference within each Δt_s. Therefore, the first derivative of (2) can be shown as:

$$\frac{dT(t)}{dt} = b\cdot(T_{ss} - T_0)\cdot e^{-bt}\,, \quad (4)$$

which is the temperature slope between the current temperature at current time t and the one at the previous thermal sensing time $(t-\Delta t_s)$. With this equation, we can predict the temperature slope between the temperature at time $(t+k\Delta t_s)$ and $(t+k\Delta t_s-\Delta t_s)$ by using the following recursive equation:

$$\frac{dT(t+k\Delta t_s)}{dt} = \frac{dT(t)}{dt}\cdot e^{-b\cdot k\Delta t_s}\,. \quad (5)$$

Therefore, $\Delta T(t+k\Delta t_s)$, which is temperature difference between the temperature at time $(t+k\Delta t_s)$ and $(t+k\Delta t_s-\Delta t_s)$ can be derived as:

$$\Delta T(t+k\Delta t_s) = \Delta T(t)\cdot e^{-b\cdot k\Delta t_s}\,. \quad (6)$$

With (3) and (6), the $\Delta T(k\Delta t_s)$ is the accumulation of each change of temperature in each Δt_s from the time t to the time $(t+k\Delta t_s)$, which is shown as:

$$\Delta T(k\Delta t_s) = \sum_{i=1}^{k}\Delta T_i = \Delta T(t)\cdot\frac{e^{-b\cdot\Delta t_s}\cdot(1-e^{-b\cdot k\Delta t_s})}{1-e^{-b\cdot\Delta t_s}}\,, \quad (7)$$

which is shown in Fig. 3(b). Therefore, the temperature at the time after $k\Delta t_s$ can be predicted as:

$$T(t+k\Delta t_s) = T(t) + \Delta T(t)\cdot\frac{e^{-b\cdot\Delta t_s}\cdot(1-e^{-b\cdot k\Delta t_s})}{1-e^{-b\cdot\Delta t_s}}\,. \quad (8)$$

Note that the term of $\frac{e^{-b\cdot\Delta t_s}\cdot(1-e^{-b\cdot k\Delta t_s})}{1-e^{-b\cdot\Delta t_s}}$ in (8) will be a constant. If the thermal prediction distance k is determined, the computation complexity of the proposed prediction model is *O(1)*. Therefore, the proposed thermal prediction model is a low-cost approach. Besides, the proposed thermal model is application independent because of the involved *Thermal RC* model.

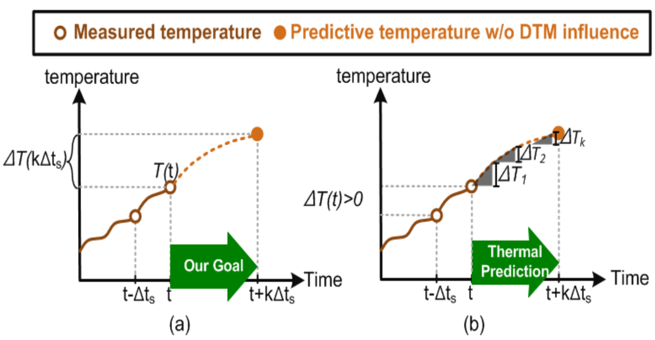

Fig. 3 (a) Temperature will increase in normal operation, and (b) The future temperature can be predicted through accumulation.

C. *Proactive Dynamic Thermal Management with Vertical Throttling (PDTM-VT) Scheme*

Fig. 4 shows the operation flow of the proposed *PDTM-VT* scheme, which is controlled by the distributed *TMU* of each NoC node. In *Thermal Management Phase*, the *TMU* will determine the *DTM* policy based on the predictive temperature and the current sensing temperature, which are resulted from the *Thermal Prediction Phase*. If the current temperature exceeds the triggering temperature T_{limit}, the fully vertical throttling scheme with be triggered for the consideration of emergent cooling. On the other hand, if the predictive temperature exceeds the triggering temperature T_{limit}, the partially vertical throttling scheme is performed to consider both thermal dissipation and performance maintenance.

It is important to implement the partially vertical throttling scheme. With the Fourier's Law in (1), in addition to the current temperature, the change of temperature depends on the current power consumption of the node. Therefore, we adopt the *DFS* to perform the vertical throttling scheme in this work, because the power consumption depends linearly on the clock frequency. However, the problem of frequency assignment was proved as an *NP-hard* problem [5]. To consider the implementation cost of partially throttling scheme, we only choose two throttled ratio to perform the *PDTM-VT*: (i) full off (*i.e.,* throttled ratio is 100%) and (ii) half off (*i.e.,* throttled ratio is 50%). Because of the *DTM* policy is decided without any computation overhead, the proposed *PDTM-VT* has the benefit of short response time.

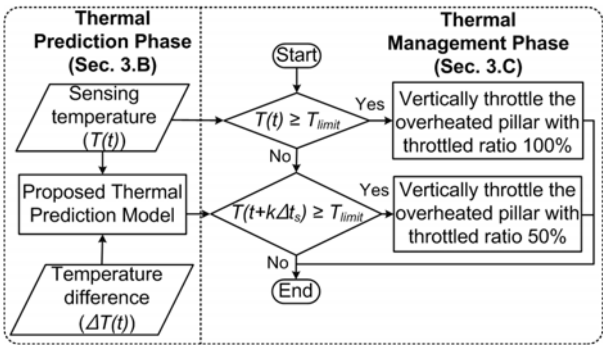

Fig. 4 The operation flow of the proposed *PDTM-VT* scheme.

4. EXPERIMENTS AND DISCUSSIONS

Through a traffic-thermal co-simulation platform [10], we compare the system performance and the temperature of an 8x8x4 3D NoC system by using (i) Reactive Vertical Throttling Scheme (*VT*) and (ii) Proactive Vertical Throttling Scheme with Prediction Distance *N* (VT_PD_N), respectively. For each router, the channel depth of the buffer is 4 flits without virtual channel. The network size is an 8x8x4 3D NoC, and the packet length is 8 flits. To simplify the problem, the XYZ routing algorithm is employed under the random traffic pattern.

A. *Accuracy Analysis of the Proposed Thermal Prediction Model*

Fig. 5(a) shows the result of temperature prediction while the prediction distance is equal to one (PD_1). The prediction error over different prediction distance is shown in Fig. 5(b). Obviously, the prediction error increases with respect to the prediction distance.

Because the prediction error affects the efficiency of *DTM*, we set the tolerated prediction error as 0.2˚C in this paper as a design example. For other applications, the designers can select the feasible prediction error constrain based on the design specification.

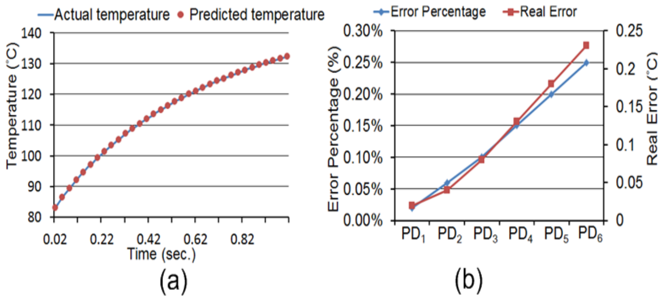

Fig. 5 (a) Temperature prediction result, and (b) Prediction error analysis.

B. Evaluation of the Proposed PDTM-VT

To prevent transient temperature from exceeding the hard thermal limit (*i.e.* 100˚C), the the triggering temperature for *VT*, VT_PD_1, VT_PD_2, VT_PD_3, VT_PD_4, and VT_PD_5 are 93.1˚C, 93.5˚C, 95˚C, 94˚C, 95.4˚C, and 94.2˚C, respectively. Fig. 6 shows the transient numbers of thermal-emergent node (*i.e.,* the temperature of the current node exceeds the triggering temperature) under random traffic when the *VT*, VT_PD_1, VT_PD_2, VT_PD_3, VT_PD_4, and VT_PD_5 are applied. Table 1 shows the comparison of system performance and the transient numbers of thermal-emergent node between the reactive *VT* and the proposed *PDTM-VT* scheme with different prediction distance. Because of early temperature control before thermal emergency, the proposed *PDTM-VT* can reduce the numbers of thermal-emergent node by around 11.84%~23.18% compared with the conventional *VT* [2][4]. Besides, the system performance can be improved by around 0.47%~47.90%.

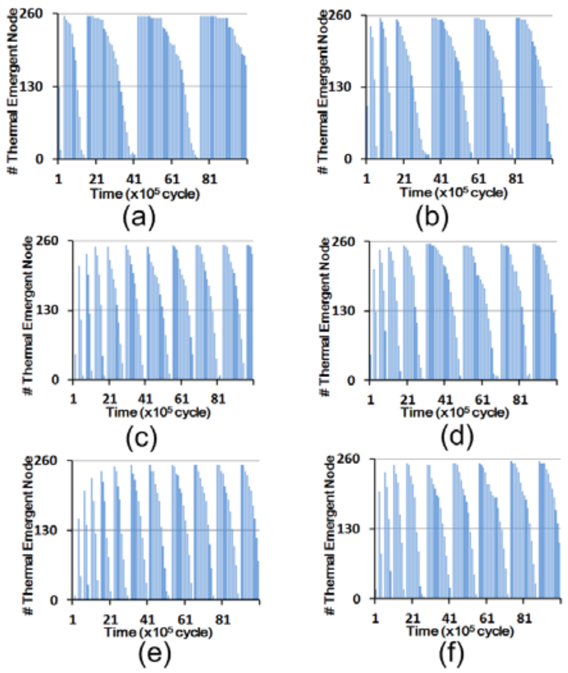

Fig. 6 The numbers of thermal-emergent node as applying (a) *VT*, (b) VT_PD_1, (c) VT_PD_2, (d) VT_PD_3, (e) VT_PD_4, (f) VT_PD_5 under random traffic.

Table 1 Comparison of each *DTM* under random traffic.

	# Thermal-emergent node	Throughput (flits/cycle)
VT [2][4]	18,272	3.92
VT_PD_1	15,688 (-14.14%)	3.94 (+0.47%)
VT_PD_2	14,464 (-20.84%)	5.71 (+45.54%)
VT_PD_3	16,108 (-11.84%)	4.52 (+15.20%)
VT_PD_4	14036 (-23.18%)	5.80 (+47.90%)
VT_PD_5	15,396 (-15.74%)	4.66 (+18.66%)

5. Conclusions

In this paper, to perform the early temperature control for a 3D NoC system, we propose a framework of proactive dynamic thermal management with vertical throttling (*PDTM-VT*) scheme, which is controlled by the distributed Thermal Management Unit (*TMU*). Based on the *Thermal RC*-based thermal prediction model, the future temperature of each node in 3D NoC system can be predicted with *O(1)* computation complexity. The experimental results show that the prediction error is less than 0.25% within 50ms. Besides, the *PDTM-VT* can reduce 11.84%~23.18% thermal-emergent nodes and improve 0.47%~47.90% network throughput.

Acknowledgement

This work was supported by the National Science Council under NSC 100-2220-E-002-013 and NSC 100-2220-E-002-016.

References

[1] B.S. Feero and P.P. Pande, “Networks-On-Chip in a Three Dimensional Environment: A Performance Evaluation,” *IEEE Trans. Comput.*, vol.58, no.1, pp.32-45, Jan. 2009.

[2] C.H. Chao *et al.*, “Traffic- and Thermal-Aware Run-Time Thermal Management Scheme for 3D NoC System” *IEEE Intl. Symp. Network-on-Chip (NOCS)*, pp.223-230, May 2010.

[3] I. Yeo *et al.*, “Predictive Dynamic Thermal Management for Multicore Systems,” *ACM/IEEE Design Automation Conference (DAC)*, pp.734-739, Jun. 2008.

[4] C.H. Chao *et al.*, “Transport Layer Assisted Routing for Run-Time Thermal Management of 3D NoC Systems,” to be appeared in ACM Transactions on Embedded Computing Systems.

[5] F. Zanini *et al.*, “Multicore thermal management using approximate explicit Model Predictive Control,” Proc. IEEE Int’l Symp. Circuits and Systems (ISCAS), pp.3321-3324, May 2010.

[6] T. Wegner *et al.*, “Impact of Proactive Temperature Management on Performance of Networks-on-Chip,” Int’l Symp. System on Chip (ISSOC), pp.116-121, Oct. 2011.

[7] Y. Ge *et al.*, “A Multi-Agent Framework for Thermal Aware Task Migration in Many-Core Systems,” on-line publication in *IEEE Trans. Very Large Scale Integr. Syst.*.

[8] R. Cochran *et al.*, “Consistent Runtime Thermal Prediction and Control Through Workload Phase Detection,” *ACM/IEEE Design Automation Conference (DAC)*, pp.62-67, Jun. 2010.

[9] R. Ayoub *et al.*, “Energy Efficient Proactive Thermal Management in Memory Subsystem,” *ACM/IEEE Int’l Symp. Low-Power Electronics and Design (ISLPED)*, pp.195-200, Aug. 2010.

[10] 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. IEEE Intl. Symp. on VLSI Design, Automation, and Test* (VLSI-DAT’10), pp.135-138, Apr. 2010.

[11] S. Wang and R. Bettati, “Reactive speed control in temperature-constrained real-time systems,” *Real-Time Systems,* vol.39, pp.73-95, 2008.

[12] W. Huang *et al.*, “HotSpot: A compact thermal modeling methodology for early-stage VLSI design,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol.14, no.5, pp.501-513, 2006.