Thermal-aware Dynamic Buffer Allocation for Proactive Routing Algorithm on 3D Network-on-Chip Systems

Yuan-Sheng Lee[†], Hsien-Kai Hsin[‡], Kun-Chih Chen[‡], En-Jui Chang[‡], and An-Yeu (Andy) Wu[†]
[†]Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 10617, Taiwan

***Abstract*—The thermal problems of three-dimensional Network-on-Chip (*3D NoC*) systems become more serious because of die stacking and different thermal conductance between layers. Up to now, most previous works cannot further achieve thermal balance of the 3D NoC systems since they consider either only temperature or only traffic information. We propose a Proactive Thermal-Dynamic-Buffer Allocation (*PTDBA*) scheme to constrain the routing resource around overheated regions. In addition, we reduce the frequency of packets switching in overheated router regions. By doing so, we can slow down the rate of temperature increment. Based on the proposed *PTDBA*, we can redistribute traffic load by means of buffer occupancy. The experimental results show that the proposed scheme can reduce the deviation of temperature distribution by 25.6% and help to improve network throughput in non-stationary irregular mesh by 74.8% compared with *PTB^{3}R*.**

Keywords- Proactive; Buffer Allocation; 3D NoC; 3D IC

INTRODUCTION

As the complexity of System-on-Chip (*SoC*) grows with the advance technology scaling, three-dimensional Network-on-Chip (*3D NoC*) emerges as a scalable on-chip communication paradigm for integrating higher amount of intellectual property (*IP*) cores [1]. 3D NoC reduces the distance of global interconnects and provides higher bandwidth with lower power consumption [2]. However, because of the die stacking and location of the heat sink, longer heat dissipation path and higher power density result in more serious thermal problems in 3D NoC. Thermal issues degrade the system performance and increases leakage power, which further causes the thermal runaway [15]. To solve the thermal issue on 3D NoC, two major approaches are employed: One is run-time thermal managements (*RTM*) [11][13] and the other is routing based thermal balance algorithm [5][12].

To keep the system temperature below a certain thermal limit, the throttling mechanisms of *RTM* are triggered as the system temperature reaches the alarming level [13]. However, these reactive *RTM* mechanisms usually result in huge performance impact under Non Stationary Irregular (*NSI*) mesh [16]. To mitigate the performance degradation, several proactive *RTMs* were proposed to take proper actions in advance based on the information of predictive thermal-emergency level [3][11]. However, it causes unbalance traffic due to the sudden traffic changes in the *NSI* mesh. Consequently, the system using proactive *RTM* (*PRTM*) still suffers from dramatic performance degradation in *NSI* mesh resulting from throttling mechanisms [13]. Hence we need to adjust routing resource (*e.g.*, buffer depth) through the use of thermal information to move traffic toward cooler and non-congested region in temporal and spatial domain.

On the other hand, Chen *et al.* proposed a Traffic- and Thermal-aware Adaptive Beltway Routing (*TTABR*) to balance the temperature distribution through balancing the traffic load [5]. The assumption is that traffic distribution can influence spatial thermal distribution. The *TTABR* provides an extra non-minimal path. Therefore, the packets have a chance to detour the congested minimal path region. Based on the traffic information, the *TTABR* adaptively selects a path between the minimal path and the non-minimal beltway path. However, the *TTABR* still has high probability to deliver the packets through potential overheated region due to the lack of thermal information.

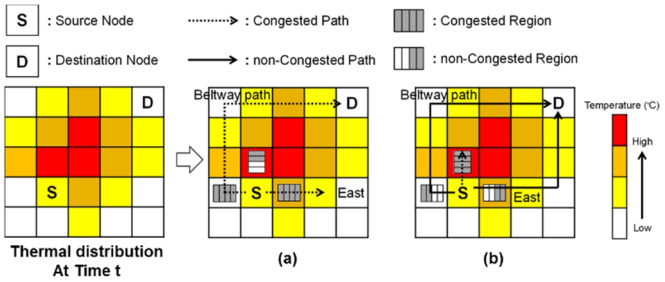

Fig. 1. (a) *PTB^{3}R* always choose the cooler path so it lets the cooler eastern path and beltway path become congested, and (b) *PTDBA* makes the overheated region relatively congested and selection strategy can focus on balancing traffic to choose the cooler beltway or eastern path.

Based on *TTABR*, Kuo *et al.* uses a novel thermal-aware routing index, Mean Time to Throttle (*MTTT*), to represent the remaining active time of the router before the temperature reaches the alarming level [12]. They proposed the Proactive Thermal-budget-based Beltway Routing (*PTB^{3}R*) for further balancing thermal distribution. However, the thermal information cannot be obtained from thermal sensors at any time. In Fig. 1(a), based on the last temperature sampling, *PTB^{3}R* always choose the cooler path so it makes the cooler eastern path and beltway path congested. To resolve the problem, the random selection according to the distribution of the normalized *MTTT* is needed. Although cooler routers can be achieved by handling fewer switched packets, the cooler region may suffer from two traffic conditions:

1) Cooler region caused by seldom choice of packet transmitting.
2) Cooler region caused by congestion. The routers in the region cannot switch more packets.

The *PTB^{3}R* in [12] does not take the second traffic condition into account since it transmits packets to the cooler and possible congested region only depending on thermal information.

Hence, to resolve the problem, we propose a *Proactive Thermal Dynamic Buffer Allocation (PTDBA)* scheme to constrain routing resource around the overheated region. We use the congestion-aware proactive routing to balance thermal distribution through bypassing the congested and overheated region as shown in Fig. 1(b). The proposed method also selects the path between the minimal path and the non-minimal path like *TTABR* and *PTB^{3}R*. The network can achieve more balanced temperature distribution. The contributions of this paper are summarized as follows:

1) *The PTDBA scheme considers both spatial and temporal thermal information:* Because thermal hotpots result from switching excessive packets, the adjustment of buffer depth according to the spatial difference from temporal predicted thermal information can constrain routing resource around overheated regions in order to slow down the rate of temperature increment.

2) We apply congestion-aware proactive routing algorithm to balance thermal distribution through bypassing the congested and overheated regions because *PTDBA* makes the overheated region relatively congested. The path selection strategy uses the buffer information of neighbor routers [14] to choose a path toward the cooler and non-congested region.

To evaluate the performance of the proposed method, the traffic-thermal mutual coupling co-simulation platform [10] is employed. The

978-1-4799-2776-0/14/$31.00 ©2014 IEEE

experimental results show that the proposed method can improve the system throughput by 74.8% and reduce the standard deviation of temperature among routers by 25.6% compared with *PTB*^{3}*R*.

The rest of this paper is organized as follows. In Section II, we introduce some related routing schemes for balancing temperature distribution. In Section III, the proposed *PTDBA* for proactive routing algorithm is described. In Section IV, the experiments are shown and discussed. Finally, we conclude this paper in Section V.

II. Related Works

A. Traffic- and Thermal-aware Adaptive Beltway Routing (TTABR) [5]

To solve the problem of unbalanced temperature distribution in 3D NoC systems, Chen *et al.* proposed the *TTABR*, which aims to balance the on-chip temperature profile by balancing traffic distribution. The *TTABR* provides multiple non-minimal paths (*i.e.*, beltway paths) to increase the lateral path diversity. Based on the traffic information of the network, the *TTABR* can adaptively select the minimal path or non-minimal beltway path. Because of higher lateral path diversities, the *TTABR* can balance the traffic load in the network system, which makes the temperature distribution become more balanced. However, the *TTABR* cannot prevent packet from routing through the potential hotspot region because it only refers to the traffic information. If the routers only consider traffic information and excessively propagate packets toward the non-congested region, the thermal hotspot will appear at the next sampling time of temperature.

B. Proactive Thermal-Budget-Based Beltway Routing (PTB^{3}R) Algorithm[12]

Many thermal-aware adaptive routings were proposed to dynamically select the path based on the traffic information to prevent from thermal hotspot [4-6]. Kuo *et al.* proposed Proactive Thermal-Budget-Based Beltway Routing (*PTB*^{3}*R*). The author uses an index called as Mean Time to Throttle (*MTTT*), obtained from thermal information, to solve thermal imbalance problem in 3D NoC. However, sensors get the thermal information every once in tens millisecond. Hence temperature is a long term and static information with regards to traffic activity between temperature sampling. Also, the cooler region represents either the region where packets seldom transmitted or congestion due to contention for a long time. This method results in severe congestion and performance degradation. For this reason, the author normalized the *MTTT* to get the distribution ratio in each direction. The pseudo random number generator is needed to randomize the selecting of directions for each packet to match the normalized distributed ratio. This method achieves thermal balance within each layer and achieves higher throughput in regular mesh at the cost of hardware overhead.

III. Thermal-aware Dynamic Buffer Allocation for Proactive Routing Algorithm

As mentioned before, thermal-aware adaptive routings use traffic information regardless of thermal condition can easily result in thermal imbalance and thermal hotspot. To balance traffic distribution, they generally try to solve the problem of congestion. However, heat generation results from steady flow traffic condition. In this condition, the routers may switch excessive packets and then overheat. On the other hand, the routing algorithm only using thermal information or the distribution of normalized *MTTT* to choose a cooler path can result in severe congestion or extra overhead. Besides, it cannot distinguish non-congested paths from the cooler region.

To solve these problems, we proposed a *Proactive Thermal Dynamic Buffer Allocation (PTDBA)* scheme to constrain routing resource around the overheated region. Then, the proactive routing algorithm redistributes traffic load to achieve thermal balance. Our proposed method takes not only thermal information but also traffic condition into account. All detail will be described in follows.

A. Proposed Proactive Thermal Dynamic Buffer Allocation (PTDBA) Scheme

Temperature changes on temporal and spatial domain. We always want to transmit packets to the region with lower temperature or slower rate of temperature increment to balance thermal profile in *3D NoC* system. The *Rate of Temperature Increment* (*RTI*) is defined as the temporally thermal information. If unit of time is set as one temperature sample time, rate of temperature increment from thermal prediction model [11] can be written as:

$$RTI = T(t + \Delta t_s) - T(t) \quad (1)$$

where Δt_s is the sampling period and $T(t+\Delta t_s)$ is the predicted temperature from thermal prediction model. *RTI* represents the amount of temperature increment per sampling period of temperature. *RTI* is related to the volume of packet switching on a router per sampling of temperature period. We can adjust the traffic on the overheated router by adjusting the maximum number of flits written into the buffer of its neighbor routers depending on spatial difference of the *RTI* value. Therefore, the *PTDBA* scheme takes not only temporal predicted thermal information but also spatial difference of *RTI* into account.

In Fig. 2, there is an overheated router in the middle. If the *RTI* of router *A*, *B* or *D* is lower than that of the overheated router, the maximum number of flits pushed into the buffer on the direction toward the overheated router will be adjusted. We reduce the maximum number of flits into these buffers by two. Based on wormhole routing, head flit (*H*) and body flits (F) of the flow *1* are pushed into the eastern input buffer of the router *A*, and the remaining flits stay in the buffers of the overheated router (*Hot*) and router *D*. If the head flit (*H*) of the flow *1* does not get the grant of output port at the router *A*, the grant of output port at the hot router and router *D* cannot be released to other packets. Hence the situation easily induces the switch contention and even congestion in the overheated region if the packet at the source node is transmitted toward north under Flow *1*, *2* and *3* existing. In this scenario, the packet transmitted by the source node can easily bypass the overheated and congested region.

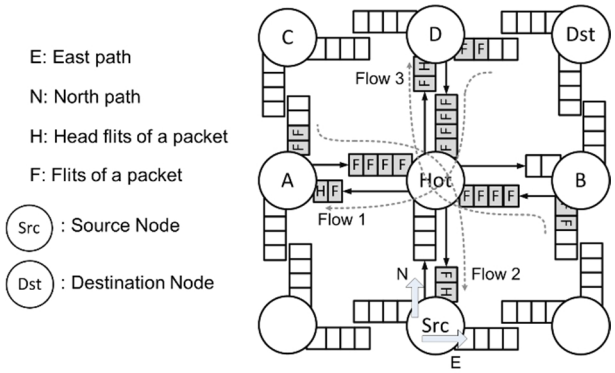

Fig. 2. Congestion caused by *PTDBA*.

We can further simplify the situation. For example in Fig. 3, the default channel depth of buffer is 8 flits on hardware. In our work, we adjust the maximum allowable flits in the buffers.

Step 1: Initially the maximum allowable flits in the buffers of the router *A* and *B* is set to 4.

Step 2: If *RTI* of the router *A* is higher than that of the router *B* (*RTI A – RTI B>0*), the maximum flits into buffer of the router *B* shrinks by two and that of the router *A* increases by two through simple buffer monitor logic controlled by *PRTM*.

Step 3: We can easily observe that the router *A* cannot transmit any other packets toward the eastern direction, because all western buffer space at router *B* has been occupied as shown in step 3 of Fig. 3.

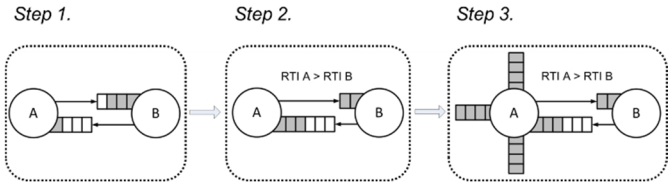

Fig. 3. Procedure of *PTDBA*: Step 1: Initial status, Step 2: If *RTI* of router *A* > *RTI* of router *B*, the maximum flits into buffer of the router *B* is reduced by two and that of the router *A* increases by two, and Step 3: Based on wormhole routing, congestion easily occurs at router *A*.

The increment in the eastern input buffer of router *A* can receive more flits of packets so that router *B* has better ability to transmit packets from one cooler region to another in addition to the direction toward router *A*. The shortening of the western input buffer of the router *B* can prevent packets coming from overheating region (router *A* in this scenario) so that switch contention easily occurs at router *A*.

The scheme of *PTDBA* reduces the maximum number of flits pushed into the buffer on the direction toward the overheated router to reduce frequency of packets switching in the overheated region. In this way, the scheme can reduce heat production and diffusion because the traffic condition is no longer steady flow case in the hotspots region and hence routers in this region switch fewer packets. We turn overheated regions into congested regions. Consequently, we can achieve thermal balance as long as packets can bypass or detour away from the congested regions.

B. Proactive Routing based on PTDBA

We use a congestion-aware proactive routing algorithm to balance the temperature distribution in this section. To achieve traffic balance, we adopt beltway routing [5] as a routing function to a routing function to provide higher path diversity. Because of the difference of thermal conductance between layers, we add the downward direction into routing function, instead of lateral first and then downward routing. By providing the downward direction in routing phase, packets have a chance to bypass the congested routing region which results from *PTDBA* and is also a thermal hotspot region.

If we can bypass the possible congested region resulting from constraining routing resource by proposed *PTDBA* scheme, we achieve not only thermal balance but also traffic balance. In order to achieve the goal, we use the information of buffer occupancy from all directions of the downstream routers to estimate the router delay based on [14]. The router delay is composed of channel transfer delay and output contention delay. The channel transfer delay can be simplified to buffer depth as the buffer is empty and buffer occupancy as the buffer is not empty. The output contention delay is composed of buffer occupancy of competing channels in the downstream router.

If a packet in the local router has two candidate channels of east and north paths, we choose a path with smallest router delay to transmit the packet into the downstream router. We start to compute the delay of downstream routers for candidate channels. For example in Fig. 4 (*e.g.*, eastern candidate channel), we average the sum of the buffer occupancy of competing channels in the downstream router of Fig 4(a) and that of Fig 4(b) to compute the output contention delay. Then, we add the contention delay and buffer occupancy of input buffer in the downstream router. We can get the router delay of one of candidate channels. As all router delays of candidate channels have been computed, we choose the smallest router delay to transmit packets into downstream routers.

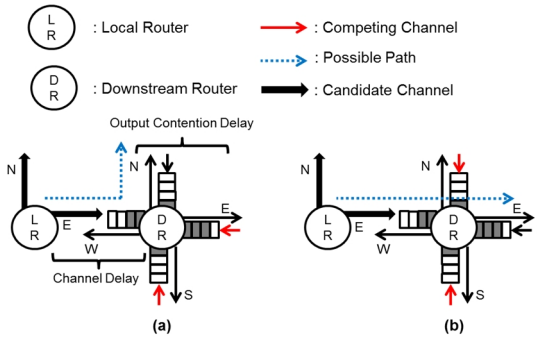

Fig. 4. (a) The competing channels in the case of packets choosing the northern path, and (b) packets choosing the eastern path.

The flow chart of *PTDBA* based proactive routing is shown in Fig. 5. Our proactive routing selects the non-congested path and hence bypassing the overheated regions.

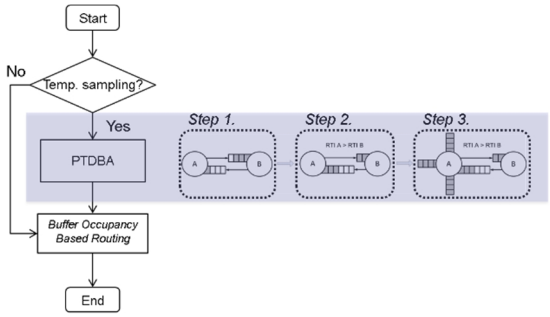

Fig. 5. Flow chart of the proposed *PTDBA* based proactive routing.

IV. Experiment and Discussion

In this section, the performance of the proposed method is evaluated through the traffic-thermal co-simulation platform [10]. For each router, the default channel depth of buffer is 8 flits without virtual channel, and each packet size is 8 flits. Besides, the maximum number of flits pushed into buffer is 8, the default number of flits pushed into buffer is 4 and the minimum number of flits pushed into buffer is 2. The network size is 8x8x4, and uniform random traffic pattern is employed.

A. Analysis of Temperature Distribution and Statistical Traffic Load Distribution (STLD)

Fig. 6(a) shows the comparison of temperature distribution of *TTABR* [7], *PTB³R* [5] and the proposed method, respectively. Because we use *PTDBA* to make thermal hotspot region potentially congested, the routing algorithm can transmit fewer packets to potentially congested and overheated region and take the non-congested path. The proposed method results in a more balanced temperature distribution.

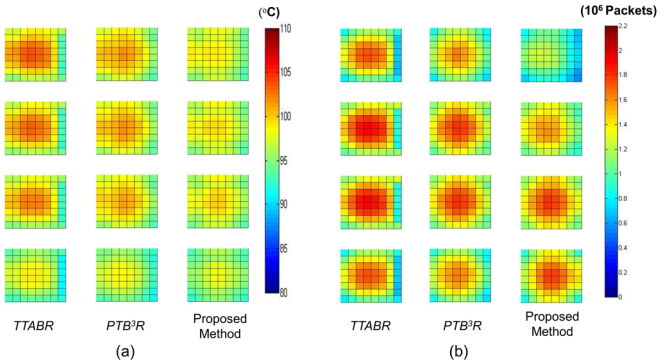

Fig. 6. (a) Temperature distribution, and (b) STLD of regular mesh.

As mentioned in [8], the temperature behavior is the long-term accumulation of packet switch activities. Our proposed method can improve thermal balance and traffic balance within layers as shown in Fig. 6(a) and (b). In summary, as shown in Fig. 7, the proposed method can help to reduce the standard deviation (σ) of temperature distribution and traffic distribution within layers by 25.6% and 14.4% respectively compared with *PTB³R*. In Fig. 8, the result further shows that our proposed method can lower the standard deviation of temperature over time before throttling within layers. However, because of different thermal conductance between layers, balanced thermal distribution balanced cannot imply more balanced traffic between layers as shown in Fig. 6(b).

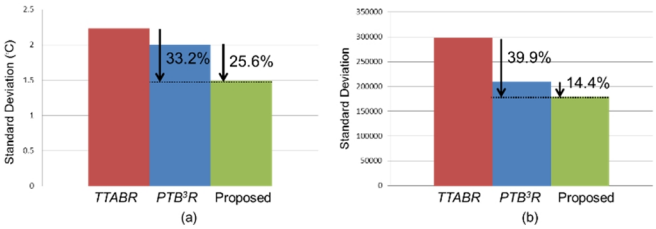

Fig. 7. σ of (a) Temperature distribution and (b) Traffic load distribution

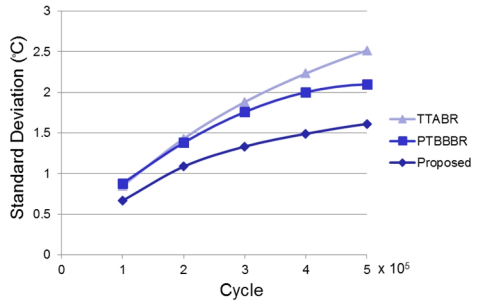

Fig. 8. σ of Temperature distribution over time

B. *Analysis of System Throughput in NSI Mesh*

As temperature increases to thermal limit, the regular mesh turns into Non-Stationary Irregular mesh (*NSI*). *NSI* mesh is a mesh changing with time because of throttling. Throttling causing *NSI* mesh is triggered by thermal alarm. Thermal distribution is affected by traffic load, and traffic distribution is dramatically affected by throttling. Therefore, the throttling mechanics causes dramatic performance degradation. Besides, the closed loop cannot be easily broken to do an experiment on a fixed irregular mesh because the initial traffic status affected by previous irregular mesh cannot be easily defined. Hence, we have *PTB³R* and our proposed method run in the closed loop. We observe the performance in *NSI* mesh. In Fig. 9 (a) and (b), the average delay rises slower, and the throughput is improved by 74.8% compared with *PTB³R* because our proposed method can gradually in advance move traffic to other cooler region before throttling.

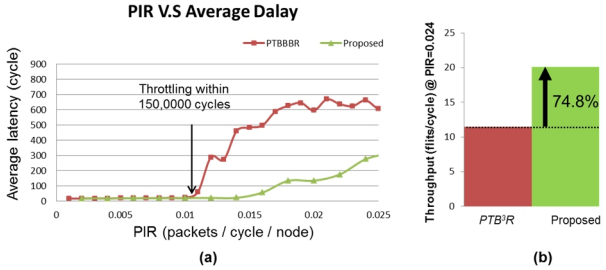

Fig. 9. (a) packets injection rate V.S. average delay, and (b) Performance improvement in *NSI* mesh.

V. CONCLUSIONS

In this paper, to balance the temperature distribution and to achieve high throughput under *NSI* mesh in 3D NoC systems, we propose a *PTDBA* scheme. The scheme can constrain routing resource (*i.e.*, buffer depth) around overheated regions in order to redistribute traffic load. We can easily bypass the congested and overheated regions only according to buffer information of neighbor routers. Consequently the system can sustain thermal balance by redistributing traffic load. Compared with the previous works, the proposed method reduces the standard deviation of temperature among routers by 33.2% and 25.6% compared with *TTABR* and *PTB³R* respectively. In *NSI* mesh, the system improves system throughput by 74.8% compared with *PTB³R*.

ACKNOWLEDGEMENT

This work was supported by the National Science Council under NSC 100-2221-E-002-091-MY3 and NSC 102-2220-E-002-001.

REFERENCES

[1] V. Pavlidis and E. Friedman, "3-D Topologies for Networks-on-Chip," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 15, no. 10, pp. 1081-1090, Oct. 2007.

[2] 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.

[3] I. Yeo, C.C. Liu, and E.J. Kim "Predicitve Dynamic Thermal Management for Multicore Systems," *in Proc. Design Automation Conference (DAC)*, pp.734-739, Jun. 2008

[4] Z. Qian and C.Y. Tsui., "A Thermal-aware Application Specific Routing Algorithm for Network-on-Chip Design," in *Proc. of the Asia and South Pacific Design Automation Conference*, pp. 449-454, Jan. 2011.

[5] K.C. Chen, C.C. Kuo, H.S. Hung, and A.Y. Wu, "Traffic- and Thermal-aware Adaptive Beltway Routing for three dimensional Network-on-Chip Systems," in *Proc. IEEE Int. Sym. Circuits and Systems (ISCAS)*, pp. 1660-1663, May 2013.

[6] F. Liu, H. Gu, and Y. Yang, "DTBR: A dynamic thermal balance routing algorithm for Network-on-Chip," *Computers & Electrical Engineering*, vol. 38, pp. 270-281, 2012

[7] C.J. Glass and M.N. Ni, "The turn model for adaptive routing", in *Proc. of the International Symposium on Computer Architecture*, pp. 278–287, May 1992

[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. IEEE Intl. Symp. on VLSI Design, Automation, and Test (VLSI-DAT)*, Apr. 2010.

[9] Mobile Intel Pentium 4 processor – M datasheet. http://www.intel.com

[10] L. Shang, L.S. Peh, A. Kumar, and R.P. Dice, "Thermal modeling, characterization and management of on-chip networks," in *Proc. Int. Symp. Microarch.*, pp.67-68, Dec. 2004

[11] K.C. Chen, S.Y. Lin, and A.Y. Wu, "Design of thermal management unit with vertical throttling scheme for proactive thermal-aware 3D NoC systems," in *Proc. IEEE Intl. Symp. on VLSI Design, Automation, and Test (VLSI-DAT)*, pp.118-121, Apr. 2013

[12] C.-C. Kuo, K.C. Chen, E.J. Chang, and A.Y. Wu, "Proactive Thermal-Budget-Based Beltway Routing Algorithm for Thermal-Aware 3D NoC Systems," in *Proc. IEEE int. Symp. Network-on-Chip(SoC)*, pp, 20-24, Oct. 2013

[13] C.H. Chao, K.Y. Jheng, H.Y. Wnag, J.C. Wu , and A.Y. Wu, "Traffic- and Thermal-Aware Run-Time Thermal Management Scheme for 3D NoC Systems," in *Proc. IEEE int. Symp. Network-on-Chip(SoC)*, pp. 223-230, May 2010

[14] G. Ascia, V. Catania, M. Palesi, and D. Patti "Implementation and Analysis of a New Selection Strategy for Adaptive Routing in Networks-on-Chip," *IEEE Trans. Comput.*, vol. 57, no. 6, pp. 809-820, June 2008

[15] A. Vassighi, and M. Sachdev, "Thermal Runaway in Integrated Circuits," *IEEE Trans. Device and Materials Reliability*, vol. 6, no. 2, pp.300-305, June 2006.

[16] Kun-Chih Chen, Shu-Yen Lin, Hui-Shun Hung, and An-Yeu (Andy) Wu, "Topology-Aware Adaptive Routing for Non-Stationary Irregular Mesh in Throttled 3D NoC Systems," *IEEE Trans. Parallel and Distributed Systems*, vol.24, no.10, pp. 2109-2120, Oct. 2013.