

# A Thermal Driven Genetic Algorithm for Three Dimensional Network-on-Chip Systems

**Ranjita Kumari Dash**  
National Institute of  
Technology  
Rourkela, India, 769008  
ranjita.rakhi@gmail.com

**José L. Risco-Martín**  
Complutense University of  
Madrid  
Madrid, Spain  
jlrisco@ucm.es

**Ashok Kumar Turuk**  
National Institute of  
Technology  
Rourkela, India, 769008  
akturuk@gmail.com

**José L. Ayala**  
Complutense University of  
Madrid  
Madrid, Spain  
jlayala@ucm.es

## ABSTRACT

3D network-on-chip (NoC) has emerged as a cutting edge technology that provides better performance by combining features of NoC and die-stacking IC technology. It is able to push the limits of Moores law by increasing the density of components in a chip resulting in higher functionality. The increasing packing density and power consumption of systems-on-chip (SoC) have made thermal effects one of the most important concern of chip designers. Increase in temperature degrades the performance, life time, reliability, and increases the maintenance cost. Addition of more layers in the z dimension has increased the length of heat conduction path and power density per unit area. Besides, the cooling capabilities of the 3D stacks are much less because of the adiabatic inter-layer materials. To ensure thermal safety, we propose a novel algorithm that uses Multi-Objective Genetic Algorithm to minimize the peak temperature of each layer in 3D NoC. Resistors model is used for thermal modeling. Three different architectures is considered in our experimental work and the results show a decrease in the peak temperature per layer. Experimental results have been validated with a well known thermal model named 3D ICE.

## Author Keywords

Genetic Algorithm; Hotspot; 3D NoC; Heat sink; Power density.

## INTRODUCTION

The 2D chip fabrication has given rise to various 2D NoC structures in recent years. Conventional 2D NoCs are facing lots of challenges due to exponential increase of PEs in a single layer. It leads to increase in number of hops from source to destination, ultimately increasing the latency to transmit

the messages. Implementation of heterogeneous technologies like high level scientific applications becomes critical in a single layer NoC . The increase in wire delay and power consumption made the chip designers to think adding IPs in the z-direction.

3D NoC is the combination of NoC and 3D IC. The NoC concept replaces design-specific global on-chip wires with a generic on-chip interconnection network realized by specialized routers that connect generic processing elements (*PE*) such as processors, ASICs, FPGAs , memories and so forth to the network and facilitate communications or links between them. The benefits of the NoC based systems include scalability, predictability and higher bandwidth with support for concurrent communications. Multiple silicon layers are stacked vertically in 3D NoC. It has emerged as an effective mechanism to overcome the physical constraints and communication delays found in 2D NoCs. 3D NoCs saves space as it results in smaller footprints in each layer. The dies can be placed few millimeters apart vertically. Vertical placement of dies reduces interconnection length notably [22] . The reduced interconnection length results in low latency and increase in performance. Data can be moved in both directions i.e. horizontally as well as vertically boosting the performance by 30% to 40%. More number of vertical vias between layers, increases the bandwidth. 3D NoCs consume 40-50% less power than its 2D counterpart [22]. As 3D NoC supports heterogeneous integration, the dies placed one above the other need not to be similar ones leading to optimization of chip components according to their diverse function [11, 12, 5].

3D NoC exacerbates the thermal problem in the inner layers because of longer heat dissipation path from chip surface to heat sink. With the increase in transistor density, the power density in the modern electronics devices has also grown exponentially, doubling in every three years [23]. The increase in power consumption results in a drastic increase in the chip operating temperature [3].

A major part of the power consumed by electronic devices is directly proportional to the heat generated, the exponential increase in power results in substantial increase in the chip temperature. Regulating temperature in advanced SoCs has become a major problem for computer system architects and designers. The soaring heat has forced high cooling costs in electronic devices (1-3 dollar/watt) [23]. Traditional methods of air-cooling is becoming inadequate to decrease the elevating on-chip temperature [20]. The seriousness of thermal problems is highlighted by Intel's cognizance of hitting a "thermal wall" [18]. The cooling cost in data centers has grown 400% in the last decade and it is expected to continue at the same rate in the near future. It is also pointed that for 1 watt of computation work, half or one watt will be consumed for cooling [2]. The exorbitant cooling methods are posing serious threat to the global economy, leading to the "economic melt down of Moore's law" [4].

The increase in power consumption as well as the elevated on-chip temperature creates enormous challenges to the future computing systems. Clock frequency of micro-processors nearly doubles in each generation. The scaling in supply voltage is unable to remunerate the resultant increase in power [10]. It has estimated the value of power density to reach 10,000 W/cm<sup>2</sup> by the end of 2016 [5]. Increase in temperature affects the performance and reliability of the chip.

This paper proposes a multi objective evolutionary algorithm (MOEA) based thermal-aware floorplanning for 3D NoCs. The previous works discussed in this section have experimented only on 3 layer architecture. The figures shown in the experimental part clearly shows the addition of two another layers and to study its effect on the system. The proposed floorplanner can manage the optimization of functional units, taking into account placement restrictions in the model. It can generate thermal profiles of the original and optimized floorplan, and peak temperature of individual layers. Results show that the inclusion of thermal-aware genetic algorithm decreases the peak temperature of individual layers to a great extent.

The paper is organized as follows: Section 2 describes previous research works carried out related to thermal management in 3D NoC. Section 3 illustrates the 3D NoC modeling, optimization and simulation platform. Section 4 details the experimental set up. Section 5 discusses the results obtained and finally, Section 6 concludes this paper.

## RELATED WORK

In this section, we discuss thermal-aware modeling during the design and optimization of 3D NoCs. One method to control the on chip temperature is via routing. Chao *et al.* proposed a thermal-aware adaptive routing using proactive downward routing [16] which ensures thermal safety for throttled NoCs. A hybrid adaptive routing algorithm to reduce the peak temperature, is proposed in [21].

Mapping is another method to control the on-chip peak temperature. Mapping algorithms assign various tasks to cores to optimize the thermal constraint [17]. Peak temperature can be reduced by minimizing the network communication rate, net-

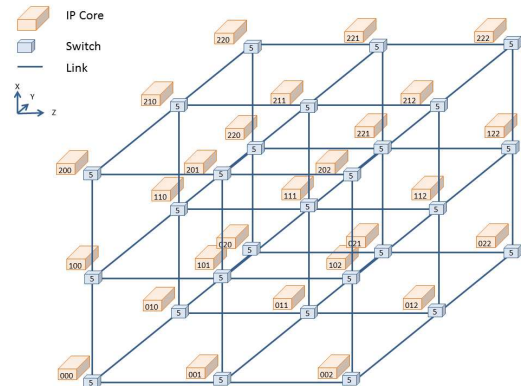


Figure 1: 3D Stacked-Mesh Architecture showing 3 layers

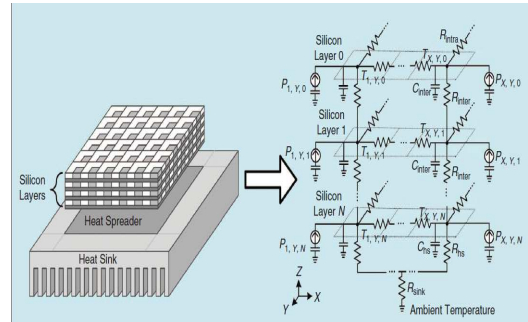


Figure 2: Thermal model of a 3D NoC system

work congestion, network communication energy and thermal balancing of the network [1, 25, 15, 6, 14]. Integer linear programming (ILP) approach is used in [12] to alleviate the on-chip temperature.

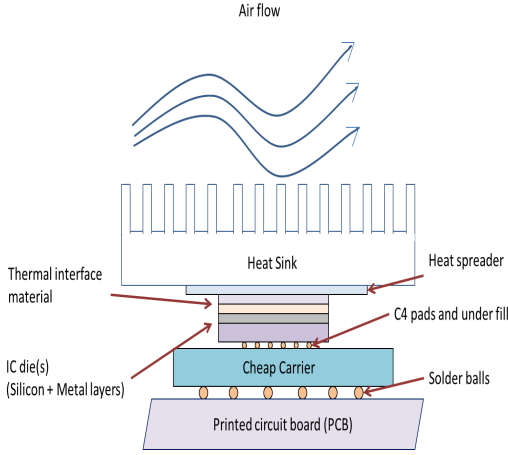
Thermal and communication aware mapping using genetic algorithm is used to decrease the peak temperature in 2D and 3D NoCs [15]. Skadron *et al.* [23] proposed to use equivalent electrical RC-circuits in order to model the thermal behavior of a chip [23]. Modeling can be done on various granularity levels, giving rise to different degrees of accuracy and speed. Several other works focused on 3D floor-planning, placement and thermal related issues [13, 19]. Coskum *et al.* [7] proposed dynamic thermal management methods for 3D NoCs.

Our approach, however, used Multi-Objective Evolutionary Algorithm, as it deals with a set of possible solutions simultaneously. It explores the solutions over the entire search space, which avoids trapping of the solution in local minima or maxima.

## THE 3D MODELING, OPTIMIZATION AND SIMULATION PROBLEM

### Mathematical Model

The topology used in our experimental work is 3D stacked mesh architecture. The stacked (hybrid NoC bus) mesh architecture presented in Figure 1 is a hybrid architecture between the packet switched network and the bus architecture to overcome the mentioned 3D symmetric NoC challenges.



**Figure 3:** Sideview of a typical 3D NoC System

It takes advantage of the short inter-layer distances. It integrates the multiple layers of 2D mesh networks by connecting them with a bus spanning the entire vertical distance of the chip. As the inter-layer distance for 3D ICs is small, the bus length will also be smaller. This makes the bus suitable for inter-layer communication in vertical direction. By using the stacked mesh architecture, six-port router is required instead of seven ports for typical 3D NoC router and vertical communication is just one hop away to any destination layer.

The side view of a typical 3D NoC is shown in Figure 2. The chip is usually mounted on the printed circuit board (PCB). Several heat conductive layers are present between the PCB up to the ambient. Here the ambient is represented by air flow. Individual layers are placed in a manner to maximize the heat flow from the active layers and to remove excess heat generated by the active layers. The heat sink is generally cooled by a fan. The package is having several heat flow paths like, primary heat flow path, secondary heat flow path and lateral heat flow paths. The primary heat flow path is made up by the following layers; the die, thermal-interface material, heat spreader and heat sink. The individual layers are made up of aluminum, copper and other material with high conductivity to maximize the heat flow. The secondary heat flow path is from the die to the PCB. It is designed in a way to minimize the heat flow die to PCB and to protect the PCB and other devices installed on it. It consists of the die, C4 pads and under fill, ceramic, solder balls and the PCB. Primary and secondary heat transfer path represents the vertical heat transfer, whereas there are other lateral heat transfer paths.

The thermal model of this system can be defined in terms of the RC model as depicted in Figure 2. The equation regulating heat diffusion via thermal conduction is,

$$C_m \cdot \frac{dT(t)}{dt} + A_m \cdot T(t) = \vec{P} \cdot u(t) \quad (1)$$

where  $A_m$  represents a  $M \times M$  sparse matrix representing the thermal conductivity matrix, the thermal capacitance matrix

$C_m$  is an  $M \times M$  diagonal matrix, whereas,  $\vec{T}(t)$  and  $\vec{P}$  represents  $M \times 1$  temperature and power vectors.

For steady state thermal analysis, the first term in Equation 1 can be eliminated, and rewritten as,

$$A_m T = P_m \Leftrightarrow T = A_m^{-1} P_m \quad (2)$$

From Equation 2, it is clear that given power vector  $P_m$  and thermal conductivity matrix  $A_m$ , we can calculate the steady state temperature. The difficulty in determining the steady state temperature lies with computing the inverse of  $A_m$ . Calculating  $A_m^{-1}$  is a difficult task as blocks change their location continuously in the optimization process in order to achieve better objective.

### Optimization Model

Several approaches have been proposed in the literature to calculate the temperature without evaluating the inverse of the thermal conductivity matrix. In this work, the hottest blocks are placed at maximum distance possible from each other at the end of the optimization process. This leads to the fact that hottest blocks are the main contributors of elevated temperature of the chip. The hottest blocks have the highest power densities among all the blocks. Therefore, the 3D stacked mesh NoC can be discretized in to small blocks. In this paper, the fitness function is organized as a weighted sum of two objectives. The first objective is defined by means of the topological relations among placed blocks. It represents the number of topological constraints violated (no overlapping between placed blocks as current volume  $\leq$  maximum volume). The second objective is a measure of the thermal impact, computed as follows. Given a block b, it's temperature can be modeled as follows:

$$T_b = \sum_i^N A_{bi}^{-1} \cdot P_i \approx A_{bb}^{-1} \cdot P_b \propto P_b \quad (3)$$

where  $P_b$  is the normalized power density of block b.

Therefore our objective can be stated as follows,

$$\text{Min.}(T) = \sum_{i < j} \frac{B_i \cdot B_j}{\sqrt{dx_{ij}^2 + dy_{ij}^2 + dz_{ij}^2}} \quad (4)$$

where  $B = \{PE_1, PE_2, \dots, PE_n, SW_1, SW_2, \dots, SW_n\}$ ;

PE is a processing element and SW represents a switch.

Our proposed algorithm will try to place the hottest processing elements and switches as far as possible. As described in [8], this method can reduce maximum temperature on the chip. For our 3D NoC stacked mesh scenario, we have used two objectives. The Performance and temperature objectives have an opposite nature: the better the performance, higher the temperature. In our proposed algorithm, our objective function is stated as follows:

$$\text{Min.}(F) = (n + r) \times \frac{T}{T^*} \quad (5)$$

where  $r$  is the set of topological constraints violated,  $T^*$  is the evaluation of Equation 4, in a stacked mesh topology. We have used the mesh topology as a baseline because of its high performance. For our future work, we are currently analyzing mechanisms to include communication as a third objective, making the mesh topology an excellent baseline for the fitness function.

The first objective described above, is related to topological relations among placed units. This objective verifies that the current area created by functional units is less than or equal to maximum area. It also ensures that there are no overlap between the placed units. If this objective is not taken into account, our floor-planner could find unfeasible solutions.

The use of MOEA is proposed in this work as it is an efficient method to solve NP-hard problems. MOEAs are stochastic optimization heuristics in nature. These approaches have been successfully applied to many NP-hard combinatorial optimization problems. In order to apply MOEAs to a problem, a genetic representation of each individual has to be found out. After this, an initial population has to be created and the cost function to measure the fitness of each solution should be defined. In the second step, the genetic operators to be applied have to be defined. These operators allow to produce a new population of thermal-aware floor-planning solutions from a previous one. It does this by capturing the inter-dependencies of the different topological constraints working concurrently. Then, the genetic operators are iteratively applied to the current population. After this, the fitness of the best individuals in the population converges to targeted solutions, according to the metrics to be optimized.

A genetic representation of design space of all the possible floor-planing alternatives needs to be defined in order to apply MOEA correctly. Our floor-planner guarantees that all the chromosomes, which are the codification of our final floor-plan, represent real and feasible solutions to the problem and ensures that the search space is covered in a continuous and optimal way. Permutation encoding is used in our approach, where every chromosome is a string of records, that represents a position in a sequence. These collect all the important information relative to the functional unit such as name of the functional unit, its length and width, connection map and power consumption of the functional units. The order in which the functional units appear in the chromosome determines their position in the final layout.

Figure 4 depicts two chromosomes used in our MOEA based floor-planning algorithm. A random pool of initial population is taken into account. As shown in the Figure 4, a chromosome is formed of four Processing Elements  $PE_i$  ( $i = 1, 2, 3, 4$ ) and four Switches  $SW_i$  ( $i = 1, 2, 3, 4$ ).

In each generation, two chromosomes are selected in tournament selection. Tournament selection runs several tournaments among few individuals chosen at random from the population. The winner of the tournament is the chromosome having the best fitness function. The winner chromosome is selected for crossover. Selection pressure is adjusted by changing the tournament size. Bigger tournament size dis-

cards the chance of weak individuals to be selected. The selection operator in our floor-planner selects two random chromosomes from the entire population and then as described above, the best of these are selected. This task is repeated twice in order to obtain two chromosomes or parents. In Figure 4, these chromosomes are selected as chromosome A and chromosome B. The best individuals from the tournament are the one having highest fitness value. The winners of the tournament are inserted in a mating pool. The mating pool consists of the tournament winners having higher fitness value. This method is more efficient and leads to a sub-optimal solution. Advantages of tournament selections are it is efficient to code, works on parallel architectures and it allows the selection pressure to be easily adjusted.

Figure 4. (b) demonstrates that the cycle crossover is applied. Crossover is the process of taking two parent solutions and producing a child from them. As seen in Figure 3, it starts with the first allele of chromosome A ( $PE_1$ ). In the next step, the allele at the same position in chromosome B is taken. In our case, it is  $SW_1$ . Next, the position with the same allele that was in chromosome B is taken in A and this allele is added to the cycle. The previous step is repeated until the first allele of A is taken again and breaks the loop. In the final step, the alleles of the cycle in the first child are put on the positions they have in the first parent, and take next cycle from the second parent. As can be seen in Figure 4, dark squares are unchanged units from the parents and gray squares represent when the children are swapped. This process guarantees that the offspring are feasible solutions to the problem. This process eliminates the chances of duplicity in the functional units.

The reproduction operator randomly selects a pair of two individual strings for the mating process. Selection of a cross site is done at random along the entire string length. In the last step, the position values are swapped between the two strings following the cross site. The main reason of applying crossover operator is that the newly generated chromosomes should contain good parts of the old chromosome and thus this process results in producing better off-springs.

After the crossover has created children, mutation procedure can occur in two different ways, with the same probability. The chromosome reveals the order of the functional units. Each gene inside the chromosome contains its position information. As seen in figure 3 (c), some blocks are chosen and swapped in the case of first chromosome, whereas in case second chromosome, the block  $PE_2$  is rotated  $90^\circ$ . In this case, the length and width of the block are interchanged.

Heat sources are those components of a circuit with greater power density. Our floor-planner helps to keep the heat sources as far as possible and it generally places them at the border of the chip. This technique helps in reducing the on chip temperature due to diffusion. Vertical heat spread is taken into account by not placing the heat sources one above the other. From the optimized floor-plans, it is clearly visible that most of the IPs are placed either in the first layer or in the last layer. Routers are placed at the border lines of the chip which yields to less heat.

In order to optimize the thermal deviation between nodes, we propose the use of this multi-objective function in a novel thermal-aware floor-planning algorithm for 3D NoCs. The algorithm consists of four phases. The population of chromosomes is modified to a new generation by applying three operators : selection, crossover and mutation. Selection takes good chromosomes based on the fitness function of each individual to apply. Crossover, which picks two chromosomes randomly with a probability of  $p_c$  and mixes information. The mutation operator swaps 1 to 0 and vice-verse with a probability of  $p_m$ . Genetic algorithm applies the above three operators in every generation until a stopping criteria is met. This algorithm can search a large solution space while ingoing the unsuitable regions. The pseudo code of the genetic algorithm is illustrated in Algorithm A.

## RESULTS AND DISCUSSION

After the optimization algorithm is executed using the approximated thermal model in Equation 4 , the results are validated against 3D-ICE [24], which uses the more accurate thermal model shown in Equation 1. Figure 5. represents the thermal distribution of our baseline scenario, where all the processing elements are labeled as  $PE_i$  and switches are labeled as  $SW_i$ , where  $i$  is an identifier. As seen in the figure, hot spots appear, specially in the first and second layer as they can not dissipate heat like the top layers.

The proposed algorithm is configured with a maximum population of 100 individuals, and a maximum number of 250 generations. The probability of mutation depends on the number of variables. In this case, it is the inverse of the number of blocks. Then, we set a cycle crossover with a probability of 0.90 and the tournament selection method, as mentioned in [9].

This section presents the thermal results obtained in the scenarios described in Section 3. The benchmarks used here is divided in to 3 categories , they are 3 layer , 4 layer, and 5 layer stacked 3D NoC mesh architecture. The floor-plans are modified in order to include an increased number of functional units in every layer. The inter-layer communication is carried out with a set of buses that route the communication signals from one layer to another. The base architecture is shown in Figure 1. All the layers are homogeneous in nature.

Figure 5 depicts the optimized thermal map for 3 layer , 4 layer and 5 layer benchmarks. As can be seen in Table 1, we achieve a reduction of 32.6588K in peak temperature when compared with the baseline scenario. However, as Table 2 and Table 3 shows the peak temperature of 4 layer and 5 layer benchmarks have gone up to 423.6995k and 453.5233k respectively. The high rise in temperature is because of the fact that the inner layers are not able to dissipate the heat to the ambient. After applying the optimization algorithm, the corresponding reduction in temperature are 44.3604K and 75.1354K respectively, for the layers showing highest temperature. The peak temperature of other layers can be referred from Table 1, Table 2 and Table 3.

From the results reflected in the tables, it can be derived that our floor-planner algorithm decreases the maximum temperature in all the active layers for all the designs. This is due to the homogeneous thermal distribution due to replacement of functional units. The reduction of on-chip temperature translates in to a reduced reliability risk and diminished leakage currents.

Figure 5 shows the temperatures of the active layers, in which the processing elements and switches are placed. A color map schema is used and it is shown in the legend. The functional units of each layer is represented by filling each of the cell with the color that corresponds with its temperature. In order to evaluate the improvements in the optimization process, the scale for each design is kept constant.

From the visual analysis of the Figure 5, it is clearly noticed that there are high temperature regions , where hotspots are created. The problem is more exacerbated in first layers, where the functional units are not able to dissipate heat to the environment. The thermal problems affects the performance and reliability of the chip due to their high temperature. As the power dissipated in the die increases, the temperature increases exponentially.

From the results reflected in the Table 1, Table and Table 3 , it can be found out that the optimizer decreases the peak temperature in all active layers. This is due to the fact that , the replacement of functional units ensured a more homogeneous

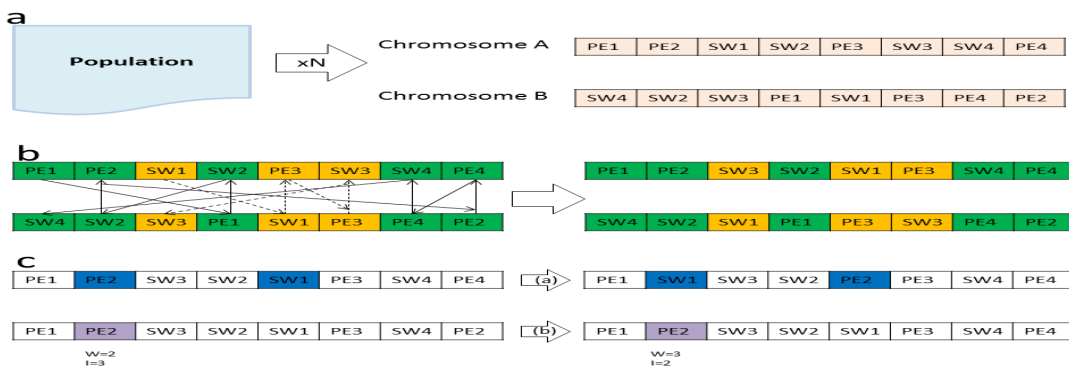
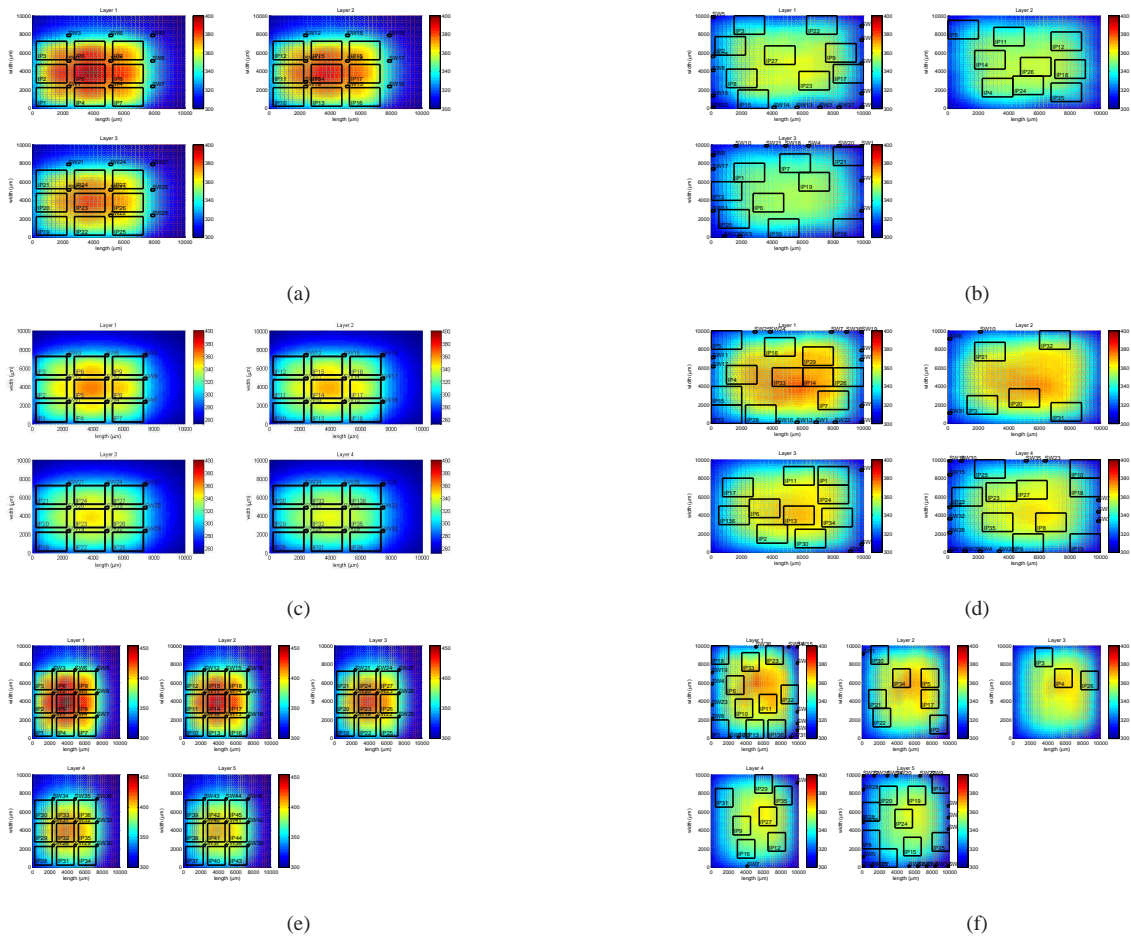
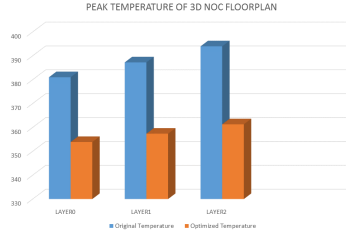


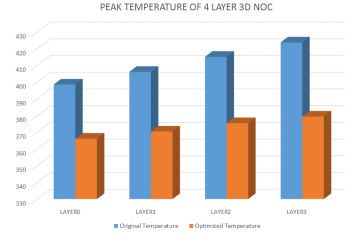
Figure 4: Steps showing operators of MOEA in thermal-aware floor-planning algorithm



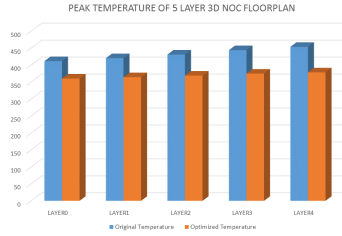
**Figure 5:** (a) Thermal map of original 3 layer floorplan (b) Thermal map of 3 layer floorplan after applying MOEA (c) Thermal map of original 4 layer floorplan (d) Thermal map of 4 layer floorplan after applying MOEA (e) Thermal map of original 5 layer floorplan (f) Thermal map of 4 layer floorplan after applying MOEA



(a)



(b)



(c)

**Figure 6:** (a) Original and optimized peak temperature of 3 layer homogeneous 3D NoC (b) Original and optimized peak temperature of 4 layer homogeneous 3D NoC (c) Original and optimized peak temperature of 5 layer homogeneous 3D NoC

Layer Number	Temp. of original floorplan	Temp. of optimized floorplan	Reduction of Temp.
Layer0	381.013	353.8788	27.1342
Layer1	387.2342	357.3654	29.8688
Layer2	393.9507	361.2919	32.6588

**Table 1:** Peak and optimized temperature of 3 layer floorplan

Layer Number	Temp. of original floorplan	Temp. of optimized floorplan	Reduction of Temp.
Layer0	398.5079	366.1133	32.3946
Layer1	406.0838	370.3774	35.7064
Layer2	415.0838	375.3961	39.6877
Layer3	423.6995	379.3391	44.3604

**Table 2:** Peak and optimized temperature of 4 layer floorplan

Layer Number	Temp. of original floorplan	Temp. of optimized floorplan	Reduction of Temp.
Layer0	411.5411	360.1468	51.3943
Layer1	420.2010	364.1730	56.028
Layer2	430.7982	368.9245	61.8737
Layer3	443.8383	374.6677	69.1706
Layer4	453.5233	378.3879	75.1354

**Table 3:** Peak and optimized temperature of 5 layer floorplan

thermal distribution. It results into a minimized reliability risk and reduced leakage currents.

Figure 5-a depicts the thermal distribution of our baseline scenario, where all processing elements are labeled as  $IP_{id}$  and

routers as  $SW_{id}$ , where id is an identifier. As can be seen in the figure, hot spots appear in the topmost layer as the heat conduction path is longer in topmost layers. Similarly, Figure 5-c and Figure 5-e depicts the original thermal diagram of 4 layer and 5 layer stack based 3-D NoC respectively.

## CONCLUSION

Our work proposes a novel optimization process for thermal management in 3D NoCs. Our algorithm is capable of placing the functional units, taking into account their thermal contribution to the entire stack. The optimization in the placement reduces the peak temperature of individual layers. The effectiveness of the proposed algorithm regarding reduction in temperature has been demonstrated by the reduction of on-chip temperature. Further, our simulations demonstrated peak temperature improvements compared to a typical stacked mesh 3D NoC.

## REFERENCES FORMAT

### REFERENCES

1. Addo-Quaye, C. Thermal-aware mapping and placement for 3-d noc designs. In *SOC Conference, 2005. Proceedings. IEEE International, IEEE (2005)*, 25–28.
2. Belady, C. In the data center, power and cooling costs more than the it equipment it supports 2007. URL <http://www.electronics-cooling.com/articles/2007/feb/a3>.
3. Borkar, S. Thousand core chips: a technology perspective. In *Proceedings of the 44th annual Design Automation Conference, ACM (2007)*, 746–749.

4. Brill, K. G. The invisible crisis in the data center: The economic meltdown of moores law. *white paper, Uptime Institute* (2007).
5. Chatterjee, D., and Manikas, T. Power-density aware floorplanning for reducing maximum on-chip temperature. In *18th IASTED Int. Conf. on Modelling and Simulation (ICMS)* (2007), 319–324.
6. Chou, C.-L., and Marculescu, R. Contention-aware application mapping for network-on-chip communication architectures. In *Computer Design, 2008. ICCD 2008. IEEE International Conference on*, IEEE (2008), 164–169.
7. Coskun, A. K., Ayala, J. L., Atienza, D., Rosing, T. S., and Leblebici, Y. Dynamic thermal management in 3d multicore architectures. In *Design, Automation & Test in Europe Conference & Exhibition, 2009. DATE'09.*, IEEE (2009), 1410–1415.
8. Cuesta, D., Risco-Martin, J. L., and Ayala, J. L. 3d thermal-aware floorplanner using a milp approximation. *Microprocessors and Microsystems* 36, 5 (2012), 344–354.
9. Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on* 6, 2 (2002), 182–197.
10. ESIA, J., and KSIA, T. Sia. international technology roadmap for semiconductors. Tech. rep., Technical report, ESIA and JEITIA and KSIA and TSIA and SIA, 2005. <http://public.itrs.net>, 2012.
11. Feero, B. S., and Pande, P. P. Networks-on-chip in a three-dimensional environment: A performance evaluation. *Computers, IEEE Transactions on* 58, 1 (2009), 32–45.
12. Hamedani, P. K., Hessabi, S., Sarbazi-Azad, H., and Jerger, N. E. Exploration of temperature constraints for thermal aware mapping of 3d networks on chip. In *Parallel, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on*, IEEE (2012), 499–506.
13. Hu, J., and Marculescu, R. Exploiting the routing flexibility for energy/performance aware mapping of regular noc architectures. In *Design, Automation and Test in Europe Conference and Exhibition, 2003*, IEEE (2003), 688–693.
14. Hu, J., and Marculescu, R. Energy-and performance-aware mapping for regular noc architectures. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 24, 4 (2005), 551–562.
15. Hung, W., Addo-Quaye, C., Theocharides, T., Xie, Y., Vijakrishnan, N., and Irwin, M. J. Thermal-aware ip virtualization and placement for networks-on-chip architecture. In *Computer Design: VLSI in Computers and Processors, 2004. ICCD 2004. Proceedings. IEEE International Conference on*, IEEE (2004), 430–437.
16. Lin, S.-Y., Yin, T.-C., Wang, H.-Y., and Wu, A.-Y. Traffic-and thermal-aware routing for throttled three-dimensional network-on-chip systems. In *VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on*, IEEE (2011), 1–4.
17. Marculescu, R., Ogras, U. Y., Peh, L.-S., Jerger, N. E., and Hoskote, Y. Outstanding research problems in noc design: system, microarchitecture, and circuit perspectives. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 28, 1 (2009), 3–21.
18. Markoff, J. Intels big shift after hitting technical wall. *New York Times* 17 (2004).
19. Murali, S., and De Micheli, G. Sunmap: a tool for automatic topology selection and generation for nocs. In *Proceedings of the 41st annual Design Automation Conference*, ACM (2004), 914–919.
20. Prasher, R. S., Chang, J.-Y., Sauciu, I., Narasimhan, S., Chau, D., Chrysler, G., Myers, A., Prstic, S., and Hu, C. Nano and micro technology-based next-generation package-level cooling solutions. *Intel Technology Journal* 9, 4 (2005).
21. Rahmani, A.-M., Liljeberg, P., Latif, K., Plosila, J., Vaddina, K. R., and Tenhunen, H. Congestion aware, fault tolerant, and thermally efficient inter-layer communication scheme for hybrid noc-bus 3d architectures. In *Networks on Chip (NoCS), 2011 Fifth IEEE/ACM International Symposium on*, IEEE (2011), 65–72.
22. Shang, L., Peh, L.-S., Kumar, A., and Jha, N. K. Thermal modeling, characterization and management of on-chip networks. In *Proceedings of the 37th annual IEEE/ACM International Symposium on Microarchitecture*, IEEE Computer Society (2004), 67–78.
23. Skadron, K., Stan, M. R., Huang, W., Velusamy, S., Sankaranarayanan, K., and Tarjan, D. Temperature-aware computer systems: Opportunities and challenges. *IEEE Micro*, 6 (2003), 52–61.
24. Sridhar, A., Vincenzi, A., Ruggiero, M., Brunschweiler, T., and Atienza, D. 3d-ice: Fast compact transient thermal modeling for 3d ics with inter-tier liquid cooling. In *Proceedings of the International Conference on Computer-Aided Design*, IEEE Press (2010), 463–470.
25. Zhou, W., Zhang, X., and Mao, Z. Pareto based multi-objective mapping ip cores onto noc architectures. In *Circuits and Systems, 2006. APCCAS 2006. IEEE Asia Pacific Conference on*, IEEE (2006), 331–334.