Thermal Aware Design for Through-Silicon Via (TSV) based 3D Network-on-Chip (NoC) Architectures

Ujjwal Pasupulety, Bheemappa Halavar and Basavaraj Talawar
SPARK Lab, NITK Surathkal, Mangalore, Karnataka - 575025
{15it150.ujjwal, cs14f06.bheem, basavaraj}@nitk.edu.in

***Abstract*—Through-Silicon Vias(TSVs) are a type of on-chip interconnect used for communication between multiple layers of circuit elements in a 3D IC. Multiple TSVs form a vertical link connecting inter-layer elements in 3D Network-on-Chip(NoC) architectures. Microarchitectural parameters such as length, width, pitch, and operating frequency influence the total power consumed and heat dissipated by TSVs. Effective extraction of the heat between layers is a significant challenge in 3D NoCs. Modelling the power of the TSVs and the thermal profile of 3D NoCs accurately enable designers perform trade-off studies during the design phase. In this work, we evaluate the thermal behaviour of 2 layer 3D Mesh and CMesh NoC architectures. We extended HotSpot to provide support for the inclusion of a router-TSV circuit element as a part of the 3D NoC floorplan. For the 3D Mesh, the thermal behaviour was analyzed for the naive arrangement as well as a proposed thermally aware design of the router-TSV element. Additionally, the thermal effect of multiple cores sharing a single router-TSV in a CMesh architecture was investigated. Our experiments show that the average of the maximum temperatures of all the routers in the 4x8x2 thermal-aware 3D Mesh is lowered by 3% compared to the naive 3D Mesh design. Also, the average of the maximum temperatures of all the routers in a 3D CMesh is 7% more than the naive 3D Mesh and 9% more than the thermally aware 3D Mesh design.**

Keywords - 3D NoCs, Through-Silicon Via (TSV), Thermal profile,HotSpot, NoC Floorplan, Dynamic power consumption

I. INTRODUCTION

Through-Silicon Via(TSV) technology has enabled chip manufactures to stack more computational power into smaller packages. TSVs can be used as vertical interconnects in 3D NoC architectures. Since a single vertical link contains multiple TSVs, it is important to understand how power is consumed in TSV-based NoCs and the thermal consequences of the same. This knowledge will be useful in analyzing how traffic is routed across the circuit elements and how it can be re-routed via alternate pathways in case of a component failure due to overheating.

3D NoC architectures make use of a router to forward data packets for inter and intra-layer communication. The TSV behaves like an inter-layer conduit for transmitting the individual bits of the packets. The challenge is to minimize the number of TSVs used in the network while maintaining a low average network latency and good overall thermal distribution. This will pave the way for low-cost manufacturing of 3D ICs, occupying smaller areas.

Simulators help in visualizing the effect of changing various physical parameters of the TSV on the given performance attributes. HotSpot [1] is once such tool used for analyzing the thermal effect of on-chip elements. It uses a floorplan-based approach to map chip elements of given dimensions and position on the die to a power trace file.

978-1-5386-6575-6 /18/$31.00 © 2018 IEEE

The main focus of our work was to evaluate the thermal behaviour of 3D NoC architectures. The inclusion of a router-TSV element of standard dimensions into HotSpot's floorplan model gives a more accurate view of the heat distribution across the die in 3D NoCs. Using this modified floorplan, the thermal effect on two types of mesh architectures: 3D Mesh and 3D CMesh was analyzed. The power consumed by the TSVs is derived from an accurate power model which considers physical parameters.

The router is usually placed next to the Data cache, a circuit element which tends to get hot, for quick access to forward packets. Hence, a separate thermally aware design was considered under the basic 3D Mesh architecture where the router is slightly shifted away from the Data cache.

Our contributions in this work are outlined as follows:

- Design of a power optimal TSV configuration in 3D NoCs using state-of-the-art TSV power and delay models.
- Extended HotSpot for NoC architectures(3D Mesh and 3D CMesh) by adding support for a router-TSV element into the core layer floorplan.

This paper is structured as follows: Section II discusses the related work. Section III describes the model used for power estimation and the 3D Mesh architectures considered for analysis. Section IV provides information about the experimental setup. Section V demonstrates our results and analysis. The paper concludes in Section VI.

II. RELATED WORK

Kim et al. [2] proposed a scalable RLGC model(shown in Fig.1) considering multiple physical parameters(Table I) of the TSV to determine the capacitance of a single via. This is necessary in order to derive the power consumed under a given operating frequency and voltage. Their work in [3] used a simple equation to calculate the power consumed which involves the TSV capacitance values from [2] and the Activity Factor, a measure of the amount of work done by the underlying chip interconnects. The model is based on the via-last process since parasitic capacitances from elements like C_{imd} have to be neglected in the via-first or via-middle configurations. The work by Prabhat et al. [4] explored the Concentrated mesh(CMesh) 3D NoC as a low cost alternative to the naive 3D Mesh. It efficiently shares on-chip network

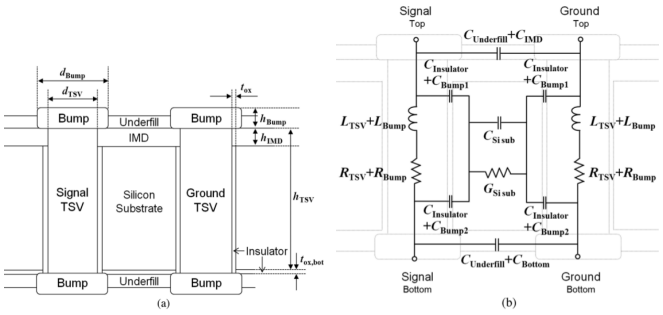

Fig. 1. (a) Structure of a signal TSV and a ground TSV with bumps with the via-last process and their structural parameters, and (b) the proposed scalable electrical model with labeled components. [2]

resources such as buffers and wire bandwidth. This architecture has been considered as a part of our analysis to verify whether there is an improvement in power consumption and thermal behavior.

Balfour et al. [5] and Psathakis et al. [6] made comprehensive studies on the internal elements of a router in order to optimize the Energy-Throughput Ratio of NoC architectures. Feero et al. [7] extensively evaluated and analyzed the throughput, latency, and energy dissipation performance in a variety of 3D NoC architectures.

Grot et.al [8] also made efforts in understanding of how network topologies(such as CMesh) scale with regard to cost, performance, and energy considering the advantages and limitations afforded on a die. The thermal behaviour of TSVs on these architectures is yet to be investigated.

For our analysis, we have considered a 2 layer NoC architecture with 32 cores per layer since manufacturing a 3D IC with more than 2 layers becomes unfeasible. Also, the thermal effect due to stacking many processing elements in a package will hinder the performance of the chip. We evaluated the thermal behaviour in 3D Mesh and 3D CMesh NoC architectures with two layers using an accurate power model and a modified floorplan-based approach. This new floorplan is a more accurate representation of NoC architectures than previous work. This is because the router is a fundamental element which is responsible for all intra and inter-layer communication taking place within the 3D IC. The power consumed and consequently, the heat generated by the router is non-negligible. Furthermore, the power consumption of the TSV, which performs the actual transmission of bits to other layers also can not be ignored. Therefore, including the router and TSV power as significant elements in a processing element using an accurate model brings the thermal distribution in 3D NoCs closer to real-world scenarios.

III. 3D TSV Power Estimation and NoC Floorplanning

A. TSV Dynamic Power Estimation Model

The electrical model from Fig. 1(b) was further condensed to Fig. 2, highlighting only the capacitance of important elements and configured as per Table II. Power consumption for each configuration was obtained using (1) where AF is the activity factor, C_{TSV} is the TSV capacitance calculated from

TABLE I
Electrical Model Parameters(Fig. 1(b))

Parameter Name	Inference
$C_{Underfill}$	Underfill capacitance
C_{Bump1}, C_{Bump2}	Bump capacitance
$C_{Insulator}$	Insulator capacitance
$C_{Si\ sub}$	Silicon substrate capacitance
C_{Bottom}	Bottom Oxide Layer capacitance
C_{imd}	Inter-metal dielectric Layer capacitance
R_{TSV}	TSV Resistance
R_{bump}	Bump Resistance
L_{TSV}	TSV Inductance
L_{bump}	Bump Inductance
$G_{Si\ sub}$	Silicon substrate conductance

TABLE II
Reduced Model Parameters(Fig. 2)

Parameter Name	Inference
C_{b1}	C_{ins} + C_{Bump1}
C_{b2}	C_{ins} + C_{Bump2}
C_1	$(C_{b1}*C_{b2})/(C_{b1}+C_{b2})$
C_2	$C_{Si\ sub}$
C_3	$C_{Underfill}$ + C_{Bottom}

the equation (2) and f and V are the operating frequency and voltage respectively [3].

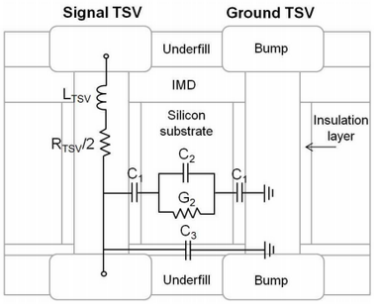

Fig. 2. Logical Layout of the TSV electrical model considered in the dynamic power model. [3]

$$Power_{TSV} = AF \,.\, C_{TSV} \,.\, V^2 \,.\, f \qquad (1)$$

$$C_{tsv} = C_3 + \frac{C_1 * C_2 * (1+\sigma_{Cu}/(\varepsilon_{si} * \omega))}{C_1 + 2 * C_2 * (1+\sigma_{Cu}/(\varepsilon_{si} * \omega))} \qquad (2)$$

The model was used to find an ideal TSV configuration with the least dynamic power within the safe limits of length, diameter and pitch given in [9] and remaining parameters from [10]. The ideal TSV has a length of 20μm, diameter of 20μm and pitch of 180μm, consuming 4.2 μWatts of power to transmit a bit.

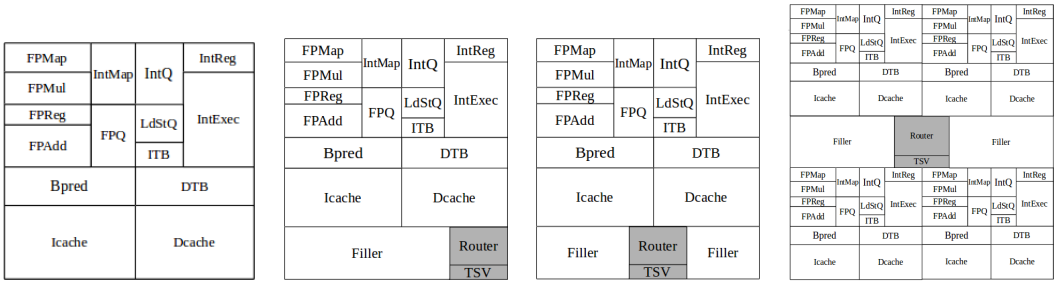

Fig. 3. Logical representations of (a) Default Alpha Ev6 processor layout in HotSpot (b) Modified layout with router next to the Data cache (c) Modified layout with router shifted away from the Data cache (d) CMesh architecture with one router shared between 4 cores (not to scale)

B. Floorplans of 3D Mesh and CMesh Topologies

The Alpha Ev6 processor architecture(Fig. 3(a)) is the base floorplan provided by HotSpot. It is yet to include support for an fundamental NoC elements such as routers and TSVs. A router-TSV element of standard dimensions was incorporated into the Ev6 floorplan.

Since the router is responsible for pushing packets across the network, an ideal placement would be a region where most of the data resides. Two separate configurations were considered under the naive 3D Mesh architecture: i) Router at the bottom-right corner, next to the Data cache(Fig. 3(b)) and ii) Router shifted to the centre, between the Instruction and Data cache(Fig. 3(c)). Another architecture considered was the 3D Concentrated Mesh. This involves one router being shared by 4 cores(Fig. 3(d)). HotSpot was extended to provide support for all the three mentioned 3D NoC architectures to analyze their respective thermal behaviours.

While placing the router next to the data cache may result in good performance due to lower link lengths, the data cache tends to be one of the hottest parts of the core. This high temperature from a neighboring element may affect the long-term performance of the router and, effectively the chip as a whole.

On the other hand, shifting the router away from the data cache, reduces the interfacial contact between the two elements. While the link length increases, the effect of thermal conduction from the hotter data cache is lowered, thereby extending the lifetime of the router. The expense of a slight drop in performance is compensated with an improved overall thermal distribution.

1) Mesh: A regular 3D NoC architecture(Fig. 4) where every processing element has a dedicated 6-port router element. The number of horizontal and vertical links are high, resulting in good performance but high power consumption. This serves as a baseline for studying the thermal effect of TSVs.

2) Concentrated Mesh: A low-cost extension of the mesh architecture(Fig. 5) where one router is shared between a set of processing elements. Such a set is referred to as a concentration(Ex. C=2,C=4,..). The Concentrated Mesh (CMesh) reduces the total number horizontal links at the expense of a slight but not significant degradation in performance. A router in a 3D CMesh setup is usually larger than a router used in 3D Mesh since it's responsible for routing packets from multiple cores. In order to handle this added complexity, CMesh routers consist of 9 ports. In a 64-core 4x8x2 NoC setup, only 16 such routers are present, which is 48 lesser routers as compared to the naive 3D Mesh architecture.

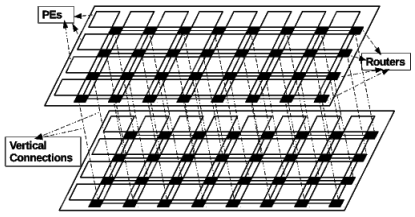

Fig. 4. 3D Mesh NoC architecture

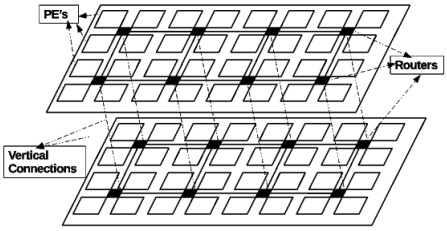

Fig. 5. 3D CMesh NoC architecture

IV. Experimental Setup

Router-TSV elements were added to the floorplans with dimensions specified in Table III. Adding this router element results in the final picture not being a perfect closed figure(a rectangle or square). For this, the Filler elements were added as dead space. All temperatures are in degree Kelvin(K) and the HotSpot simulations were run in the environment specified in Table IV. The power consumed by the router was taken from Orion [11]. It is assumed that a single vertical link contains 64 TSVs. Considering the power optimal configuration from the previous section, the power consumed by one link is roughly 0.5mWatts.

The process of generating floorplans and power trace files for both 3D Mesh and 3D CMesh NoC architectures was automated using scripts which take a small number of parameters(Fig. 6). Power trace files contain the power consumed by every element in the NoC floorplan at each time step. In our simulation, power values for 100 time steps were generated

TABLE III
MODIFIED FLOORPLAN AND POWER TRACE DETAILS

Architecture	Component	Dimension	Power
3D Mesh	Router	1.1mm x 1.1mm	0.42W
3D Mesh	TSV	0.18mm x 1.1mm	0.5mW
3D Mesh	Filler	1.28mm x 5.1mm	0W
3D Mesh(shifted)	Router	1.1mm x 1.1mm	0.42W
3D Mesh(shifted)	TSV	1.28mm x 1.1mm	0.5mW
3D Mesh(shifted)	Filler	2.18mm x 2.55mm	0
3D CMesh	Router	1.33mm x 1.33mm	0.7W
3D CMesh	TSV	0.18mm x 1.33mm	0.5mW
3D CMesh	Filler	1.51mm x 5.535mm	0

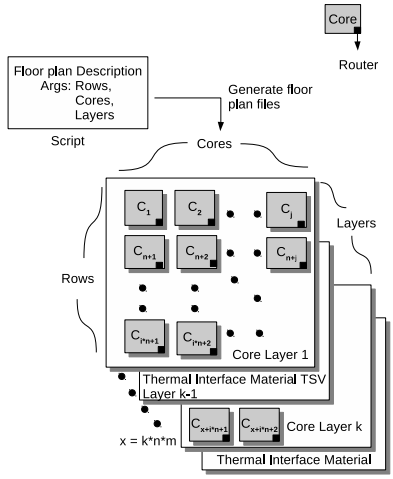

Fig. 6. Automated generation of floorplans in the extended version of HotSpot(with additional router-TSV element) for 3D NoC

from the existing power trace files in HotSpot based on the SPEC CPU2000 Benchmark [12]. Corresponding components on all cores consume exactly the same power in each time step. For example, a router associated with Core0 and another associated with Core1 consume 5 Watts of power at a given time step. All the layers are arranged in the layer configuration file.

TABLE IV
SIMULATION ENVIRONMENT

Simulator	HotSpot
Operating Frequency	2.5 GHz
Activity Factor	0.15
Clock Cycles(*CCycle*)	100000
Topology Simulated	3D 4x8x2 Mesh and CMesh
TSVs per Vertical Link(*TSVL*)	64
Power Consumed per TSV(*PCT*)	4.2 μW

V. RESULTS AND ANALYSIS

A general observation in all of the evaluated architectures is that the outer routers show lower temperature values than those located at the centre. This can be observed as a sharp increase in the router temperatures along the x-axis after the first few routers. This is because heat dissipation will be good at the boundary, but the elements in the middle of the die are surrounded from all sides. The only way heat can escape is through the the Thermal Interface Material below. Hence the elements at the centre get hotter than the boundary elements over the duration of the simulation.

A. Thermal behavior in 3D Mesh architectures

The position of the router in the floorplan of the chip has an effect on the heat distribution across its neighbours. The Data cache and Integer Registers are the primary regions where temperature is high due to their relatively higher power consumption in each time step. After R_31(in Fig. 7 (d) and (e)), the remaining routers correspond to the lower core floorplan layer, closer to the heatsink.

1) Router next to Data cache: The router is sandwiched between the hot Data cache and Integer Registers(Fig. 7(a)). As a consequence of thermal conduction, the router's temperature at the end of the simulation is also high(Fig. 7(d)). This setup may have a higher performance since the link latency between the cache and router is very low, but at the expense of poor overall heat distribution.

2) Router shifted away from Data cache: Our thermal aware architecture layout, overcomes the thermal shortcomings of the naive 3D mesh. By shifting the router towards the centre, contact with the Data cache is reduced and completely avoided with the Integer Registers. As a result, the (Fig. 7(b)) shows lower temperature in the router. Hence, by placing the router away from the elements that tend to heat up due to their higher power consumption, the router is less hotter than in the naive 3D mesh. However, this also means that the Data cache has lesser places to dissipate heat and hence, heats up more than it would if the router was directly next to it. The average of the maximum temperatures of all the routers in a 4x8x2 3D NoC is lowered by 3% in the thermal aware Mesh architecture(Fig. 7(e)).

B. Thermal behaviour of a 3D Concentrated Mesh

In a 3D CMesh architecture, one router is shared between 4 cores(Fig. 7(c)). After R_7, the remaining routers correspond to the lower core floorplan layer, closer to the heatsink. It can be observed that the only major hot regions in proximity to the router-TSV module are the upper and lower left cores. This results in the left half of the module having a higher temperature. The average of the maximum temperatures of all routers in 3D CMesh is 7% more than the naive 3D Mesh and 9% more than the thermally aware 3D Mesh design(Fig. 7(f)).This is due to the larger size and higher power consumption of the 9 port router. Also, the reduced area between surrounding cores contribute to the increased heat within the router.

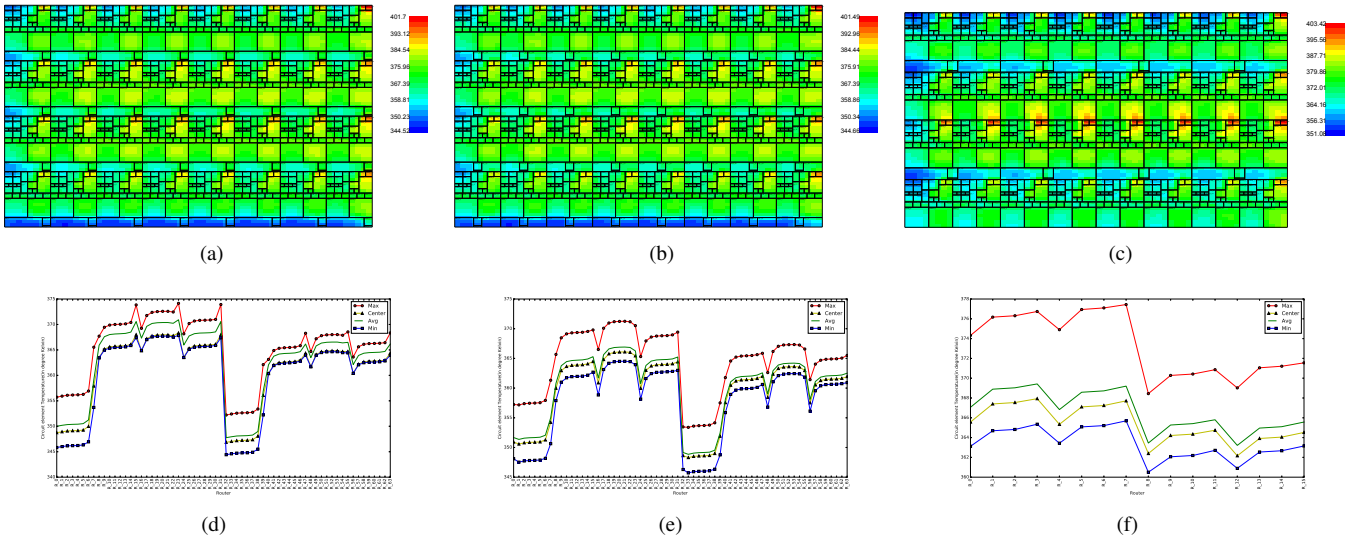

Fig. 7. Heatmaps in a 64-core (a) 3D Mesh architecture with the router next to the data cache, (b) 3D Mesh architecture with the router shifted away from the data cache, (c) 3D Concentrated Mesh Architecture and Temperature distribution across routers in (d) 3D Mesh architecture with the router next to the data cache, (e) 3D Mesh architecture with the router shifted away from the data cache and (f) 3D Concentrated Mesh Architecture.

VI. Conclusion and Future Work

In this work, we evaluated the thermal behaviour of 3D NoC architectures and proposed a thermal aware 3D Mesh NoC architecture. We made use of accurate power estimation models for fundamental 3D NoC elements, namely the router and TSV. The router power was obtained using Orion. Using this configuration the thermal effect of the TSV and router position on the chip floorplan was analyzed by modifying HotSpot for 3D Mesh and 3D CMesh NoC architectures. Considering a 4x8x2 3D Mesh setup, placing the router away from the data cache results in lower temperature across all routers in the 3D Mesh as compared to placing it directly next to it. The results of our experiments show that, the average of the maximum temperatures of all the routers in the 4x8x2 thermal-aware 3D Mesh is lowered by 3% compared to the naive 3D Mesh design. Also, the average of the maximum temperatures of all the routers in a 3D 4x8x2 CMesh is 7% more than the naive 3D Mesh and 9% more than the thermally aware 3D Mesh design.

In the future, we intend to use this information to build a thermal aware routing algorithm that takes into account any TSVs that may have overheated(after crossing a certain temperature threshold). In reality, the router contains some internal elements, each with its own power consumption. Extending HotSpot's floorplan to accommodate the micro-architectural elements of the router will give a more accurate picture on the heat distribution within the router as well its neighbouring elements.

References

[1] W. Huang *et al.*, "Differentiating the roles of ir measurement and simulation for power and temperature-aware design," in *2009 IEEE Intl. Symposium on Performance Analysis of Systems and Software*, April 2009, pp. 1–10.

[2] J. Kim *et al.*, "High-frequency scalable electrical model and analysis of a through silicon via (tsv)," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, no. 2, pp. 181–195, Feb 2011.

[3] J. K. et al., "I/o power estimation and analysis of high-speed channels in through-silicon via (tsv)-based 3d ic," in *19th Topical Meeting on Electrical Performance of Electronic Packaging and Systems*, Oct 2010, pp. 41–44.

[4] P. Kumar *et al.*, "Exploring concentration and channel slicing in on-chip network router," in *Proceedings - 2009 3rd ACM/IEEE Intl. Symposium on Networks-on-Chip,2009*, 10 2009, pp. 276–285.

[5] J. D. Balfour and W. J. Dally, "Design tradeoffs for tiled CMP on-chip networks," in *20th Annual Intl. Conf. on Supercomputing,Australia, 2006*, 2006, pp. 187–198.

[6] A. Psathakis *et al.*, "A systematic evaluation of emerging mesh-like cmp nocs," in *2015 ACM/IEEE Symposium on Architectures for Networking and Comm. Systems*, May 2015, pp. 159–170.

[7] B. S. Feero and P. P. Pande, "Networks-on-chip in a three-dimensional environment: A performance evaluation," *IEEE Transactions on Computers*, vol. 58, no. 1, pp. 32–45, Jan 2009.

[8] B. Grot *et al.*, "Express cube topologies for on-chip interconnects," in *2009 IEEE 15th Intl. Symposium on High Performance Computer Architecture*, Feb 2009, pp. 163–174.

[9] R. Weerasekera *et al.*, "Compact modelling of through-silicon vias (tsvs) in three-dimensional (3-d) integrated circuits," in *2009 IEEE Intl. Conf. on 3D System Integration*, Sept 2009, pp. 1–8.

[10] M. Lee *et al.*, *Electrical Design of Through Silicon Via*. Springer Netherlands, 2014.

[11] A. B. Kahng *et al.*, "Orion 2.0: a fast and accurate noc power and area model for early-stage design space exploration," in *Proc. of the Conf. on Design, Automation and Test in Europe*. European Design and Automation Association, 2009, pp. 423–428.

[12] J. L. Henning, "Spec cpu2000: Measuring cpu performance in the new millennium," *Computer*, vol. 33, no. 7, pp. 28–35, Jul. 2000.