VNU Journal of Science:
Computer Science and Communication Engineering

Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

Thermal Distribution and Reliability Prediction for 3D Networks-on-Chip

Khanh N. Dang[1,*], Akram Ben Ahmed[2], Abderazek Ben Abdallah[3], Xuan-Tu Tran[1]

[1]*VNU University of Engineering and Technology, Vietnam National University, Hanoi, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*
[2]*National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8568, Japan*
[3]*University of Aizu, Aizu-Wakamatsu, Japan*

Received 02 April 2020
Revised 02 June 2020; Accepted 06 June 2020

Abstract: As one of the most promising technologies to reduce footprint, power consumption and wire latency, Three Dimensional Integrated Circuits (3D-ICs) is considered as the near future for VLSI system. Combining with the Network-on-Chip infrastructure to obtain 3D Networks-on-Chip (3D-NoCs), the new on-chip communication paradigm brings several advantages. However, thermal dissipation is one of the most critical challenges for 3D-ICs, where the heat cannot easily transfer through several layers of silicon. Consequently, the high-temperature area also confronts the reliability threat as the Mean Time to Failure (MTTF) decreases exponentially with the operating temperature as in Black's model. Apparently, 3D-NoCs and 3D ICs must tackle this fundamental problem in order to be widely used. However, the thermal analyses usually require complicated simulation and might cost an enormous execution time. As a closed-loop design flow, designers may take several times to optimize their designs which significantly increase the thermal analyzing time. Furthermore, reliability prediction also requires both completed design and thermal prediction, and designer can use the result as a feedback for their optimization. As we can observe two big gaps in the design flow, it is difficult to obtain both of them which put 3D-NoCs under thermal throttling and reliability threats. Therefore, in this work, we investigate the thermal distribution and reliability prediction of 3D-NoCs. We first propose a new method to help simulate the temperature (both steady and transient) using traffic values from realistic and synthetic benchmarks and the power consumption from standard VLSI design flow. Then, based on the proposed method, we further predict the relative reliability between different parts of the network. Experimental results show that the method has an extremely fast execution time in comparison to the acceleration lifetime test. Furthermore, we compare the thermal behavior and reliability between Monolithic design and TSV (Through-Silicon-Via) based design. We also explore the ability to implement the thermal via a mechanism to help reduce the operating temperature.

Keywords: Thermal dissipation, Reliability, Through-Silicon-Via, 3D-ICs, 3D-NoCs.

* Corresponding author.
E-mail address: khanh.n.dang@vnu.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.245

1. Introduction

3D Networks-on-Chip (3D-NoCs), as a result of combining Networks-on-Chip (NoCs) [1] with 3D Integrated Circuit (3D-ICs) [2], is considered as one the most promising technologies for IC design [3]. By providing parallelism and scalability of the NoCs to 3D-ICs, we even obtain lower power consumption, shorter wire length while reducing the design area cost by several times. Among several 3D-ICs, Through-Silicon-Via which constitutes as inter-layer wire is one of the near-future technologies. Monolithic 3D ICs is another method to implement the 3D-ICs [4, 5]. With both technologies, we expect to have multiple layers of the system. To support communication within the system, 3D-NoCs offer a router-based infrastructure where the 3D mesh topology is used.

Despite several advantages, 3D-ICs and 3D-NoCs have to confront the thermal dissipation issue. The temperature variation between the two layers has been reported to reach up to 10°C [6]. *Cuesta et al.* [7] also conducted an experiment of four-layer and 48 cores which gives the temperature variation up to 10°C between a single layer. The main reason for thermal dissipation difficulty in 3D-ICs is the top layers act as obstacles that prevent the heat could be dissipated by the heatsink. To solve this problem, fluid cooling [7] or thermal cooling TSV [8] has been proposed.

By having higher operating temperatures, it is apparent that 3D-NoCs easily encounter thermal throttling. Moreover, in terms of reliability, there is an expected acceleration in the failure rate (or a reduction in Mean-time-to-Failure). For semiconductor devices, one of the most well-known models of thermal impact in reliability is the Black's model [9] where the fault rate acceleration π_T is:

$$\pi_T = A(J)^n \times e^{-\frac{E_a}{k_B T}} \quad (1)$$

where A is constant, J is the energy, k_B is Boltzmann constant, $E_a$is activation energy and T is the temperature in Kelvin. Here, we would like to note that the activation energy of Copper is much higher than CMOS material which makes TSV more vulnerable than the normal gates. Since TSV can act as a cooling device, TSV-based NoC has a lower operating temperature than Monolithic; however, TSV also has lower reliability. Therefore, the reliability differences between Monolithic and TSV-based 3D-ICs need to be investigated.

While the thermal behavior could be extracted by performing the real-chip, reliability cannot be directly measured. Most industrial methods are based on Black's model [9] in Equation 1 by baking the chip under high temperature to accelerate the failure [10-12].

In this work, we have investigated the impact of the thermal dissipation difficulty of Network on Chip based 3D-ICs by proposing a method to predict the temperature and MTTF of each region of the targeted system. We first use commercial EDA tools to design and analyze the power and energy per data bit of 3D-NoC router. Then, we extract the number of bits and the operating time of synthetic and PARSEC benchmarks to obtain the average power consumption of each router inside the network. We then use a thermal emulation tool named Hotspot 6.0 [13] to obtain the steady grid temperature of the system. By adopting the Black's model of reliability, the tool follows up with a reliability prediction of the system. By following the method, designers can fast extract the potential hotspots inside the 3D-ICs and predict the potential of the vulnerable regions due to high operating temperatures. The results also suggest the possible mapping of fluid cooling or thermal TSV insertion [7]. The contribution of this work is as follows:

- A platform to model the power, temperature, and reliability of any NoC systems. Here, we specify for 3D-NoCs but the technique is general and can be applied for the traditional planar NoC systems.
- The reliability analyses of Monolithic and TSV-based NoCs. While TSV-based NoCs have a lower operating temperature, TSV's material (Copper) has lower reliability.

- Exploration and comparison between different layout strategies and cooling methods.

The remaining part of this paper is organized as follows. Section 2 surveys the existing works. Section 3 describes the proposed method in detail. Experimental results are discussed in Section 4. Finally, Section 5 concludes this work.

2. Related Works

In this section, we summarize the literatures related to our proposed method. We start with the power model and then present the work on thermal estimation. Finally, the reliability estimations for 3D-NoCs are presented.

2.1. Power Modeling for 3D Network-on-Chip

To measure the power consumption of a 3D-IC, the straight forward method is to fabricate and set up a measuring system [16]. However, it is difficult to obtain such a system, especially designing and fabricating the chip are expensive, time-consuming and designers want to estimate the value before sending to production. Therefore, modeling the power consumption is a necessary step.

To model the power of any digital IC system, two major parts which are static and dynamic power are considered as follows:

$$P = P_{dynamic} + P_{static} = s f_c C_L V_{DD}^2 + I_{off} V_{DD} \quad (2)$$

where s is the switching probability (or activity ratio), f_c is the clock frequency, C_L is the load capacitance, I_{off} is the leakage current and V_{DD} is the supply voltage. Based on Equation 2, common EDA tools can estimate the power consumption based on the parameter of the library and the switching activity. In fact, power estimation tool such as PrimeTime requires switching activity to obtain the most accurate result.

Using Equation 2 can estimate the power consumption of any circuit; however, for a fast prediction, the power consumption of NoCs can be obtained by its switching activity. By obtaining the number of flits went through the router during simulation, it can estimate the dynamic power consumption. Meanwhile, the static power consumption is constant for the same configuration (voltage, frequency, design). For instance, ORION 2.0 [17] models power consumption as dynamic and static power. Physical parameters such as wire length and leakage current are calculated to estimate the static power. In [18], the authors use regression to estimate the power consumption of the system based on the existing values. Other works in [19][20] also consider dynamic voltage frequency scaling in power consumption.

While these works can help estimate the power consumption of our system, we observe it is not the most accurate one because of the differences in design choice and library. Therefore, in this work, we propose our power extraction method. We use the EDA tools to estimate the dynamic and static power and then combine with the switching of the routers in the used benchmarks.

2.2. Thermal Behavior Prediction for 3D Network-on-Chip

Once we obtain the power consumption of modules within a system, we can estimate the temperature of the chip. HotSpot [13] is one of the ealier tools to help estimate the temperature grid. The 6th version of HotSpot now can estimate the temperature of 3D-ICs. There are also different tools such as 3D-ICE [14] and MTA [15]. While MTA performs a similar task as Hotspot by using the finite element method, 3D-ICE focuses on the potential of liquid cooling. *Cuesta et al.* [7] also explored different layout strategies and liquid cooling for 3D-ICs.

2.3. Reliability Prediction for 3D Network-on-Chip

By having the temperature of the system, we now can estimate the potential reliability. As we previously have metioned, Black's model [9] in Equation 1 is one of the first models for CMOS designs. MIL-HDBK-217F of the US Military [22] also released its own

model of reliability acceleration related to temperature. HRD4 from industry [23] and RAMP from academics [24] are the other two models to estimate the reliability of the system.

Among these models, HRD4 consider the reliability as the same for the chip bellow 70°C. The rest of the models follows the exponential acceleration with operation temperature (in Kelvin).

On the other hand, industrial approaches on reliability prediction [10-12] are to bake the chip to high temperature and measure the average time to failure of the samples. By using Black's model, they can estimate the potential lifetime reliability under normal temperature.

3. Proposed Method

Figure 1 shows the proposed method for the thermal and reliability prediction of 3D-NoCs. We first built Verilog HDL of 3D-NoC. Then, synthesis and place & route are the following steps to obtain the layout, netlist file, wire length, and physical parameters.

We then perform post-layout simulation and use Synopsys PrimeTime to extract the power consumption of the system. Based on the number of data-bit, we further extract the energy per data bit. Then, we now can estimate the power consumption of all benchmarks by multiplying the obtained value with the number of bits per router per time. The power consumption of each router is taken to the temperature estimator tool (Hotspot 6.0) to obtain the temperature map. At the end of this step, we obtain all temperature maps of all benchmarks.

One notable thing in 3D-NoCs is the possibility to have redundant Through-Silicon-Vias (TSVs). TSVs are usually made out of Copper and have a larger size than normal wire which can dissipate heat faster than normal silicon. Monolithic 3D-ICs fails to have the same feature since the via is extremely small. Consequently, we take the redundancy mapping into the hotspot prediction.

Once we can predict the temperature, we can obtain the reliability prediction using the Black's model in Equation 1. Note that the activation energy also varies among materials. The output of reliability can also affect redundancies mapping as a close loop. Consequently, designers can further optimize the system to have the most balancing point of temperature, reliability, and area overhead. In the following part, we explained in detail each part of the proposed method.

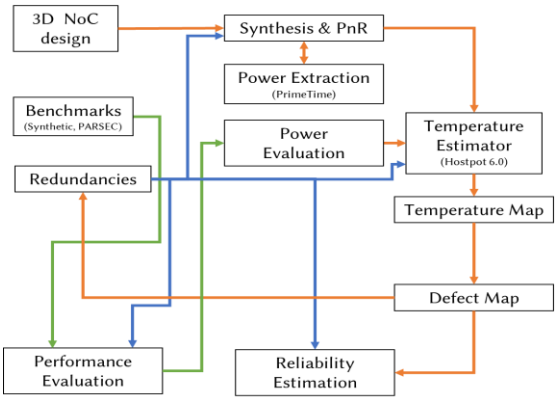

Figure 1. Thermal and reliability prediction method of 3D Networks-on-Chip.

We would like to note that our method reuses and follows the principle of existing works in academic and industrial approaches [10-12, 22-24].

3.1. Design of 3D Network-on-Chip

Here, we adopted our previous work in [3] with some modifications where the TSVs of a router are divided into four groups and placed in four directions (west, east, north, south) of the router to support sharing and fault tolerance. However, we here provide more flexibility in the design since fault tolerance is not our objective of this work. Figure 4 shows the architecture of our 3×3×3 Network on Chip. Each router can connect to at most six neighboring routers in six directions and one local connection to its attached processing element. The inter-layer connections are TSVs and we support optional the redundant TSV group (yellow TSVs) which can be used to repair a faulty group in the router. Borrowing and sharing mechanisms are another features

we support to have high reliability in our system. More details on the fault tolerance method can be seen in our previous work [3].

Each router receives a header flit of packet and support routing inside the network. Based on the destination, it forwards the header flit and the following flits (body and tail flits) to the desired port. Once the tail flit completes its transmission, the router starts to route a new packet.

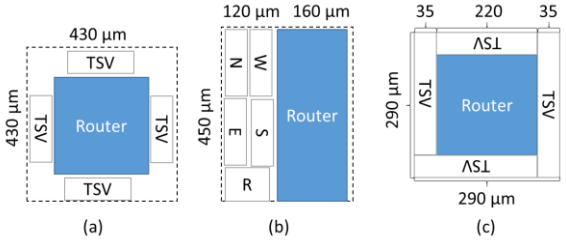

Figure 2. Layout option for 3D-NoC router: (a) Previous work in [21]; (b) Separated TSV region; (c) Surround TSV region.

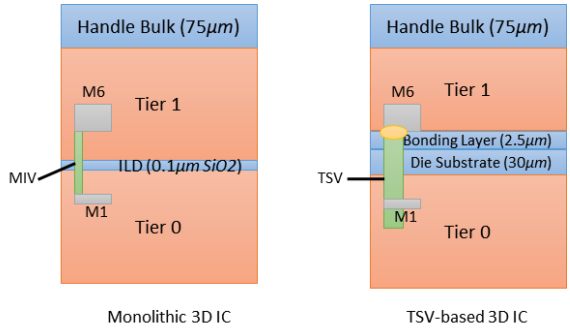

Figure 3. 3D IC layer structure (heat sink on top) of Monolithic 3D IC vs TSV-based 3D IC.

In the router layout of [3], the design is not well optimized since it leases space between routers in layout. Figure 2(a) shows the layout of [3]. In order to optimize it, we use two different floorplans in this work. We first place TSVs and router logics in separated regions as in Figure 2 (b). Then, we place TSVs surrounding the router logics as in Figure 2 (c). We can notice that we reduce the size of the router significantly by removing the empty space.

Among the two new layouts, Figure 2(c) provides the best thermal balance because it isolates the logic of a router to the nearby module. Since routers are usually hotspots inside the system, placing them near a hot area can raise its temperature significantly. Here, by surrounding by TSVs, we create isolation for the router. Furthermore, Copper has low thermal resistivity which can dissipate the heat from the router to the upper layers. By doing so, we can transfer then heat to the top layer and the heatsink. In the evaluation section, we then discuss the efficiency and cost of inserting thermal via in our design.

Figure 3 shows the different between Monolithic and TSV-based 3D-ICs. While TSV is made out of Copper that dissipate thermal faster than Silicon layers. However, there are bonding layers between stacking using TSVs which creates an isolation of thermal disspation between them.

3.2. EDA tools and Power Extraction

The following part of the method is to use EDA tool to extract the power consumption. Apparently, we can use any supported EDA to obtain power consumption. For our experiment, we use Synopsys Design Compiler, ICC and PrimeTime to do the physical design and extract the power consumption.

To extract the power, we perform a heuristic transmission benchmark of a single router. Here, we generate two packets of ten flits in all possible directions. Because our router supports returning the flit from it sending ports, we have 7×7=49 possible directions. By using PrimeTime, we can obtain the dynamic and static power.

Here, we also classify the energy into static and dynamic. While static power consumption is stable, we keep the value as it is. For the dynamic power, we calculate the total energy and the energy per data bit.

3.3. Power and Temperature Estimation

Once we obtain the energy per data-bit, we can obtain the overall power consumption as follows:

$$P = P_{static} + P_{dynamic} = P_{static} + E_{dynamic} \times \frac{N_{bit}}{time} \quad (3)$$

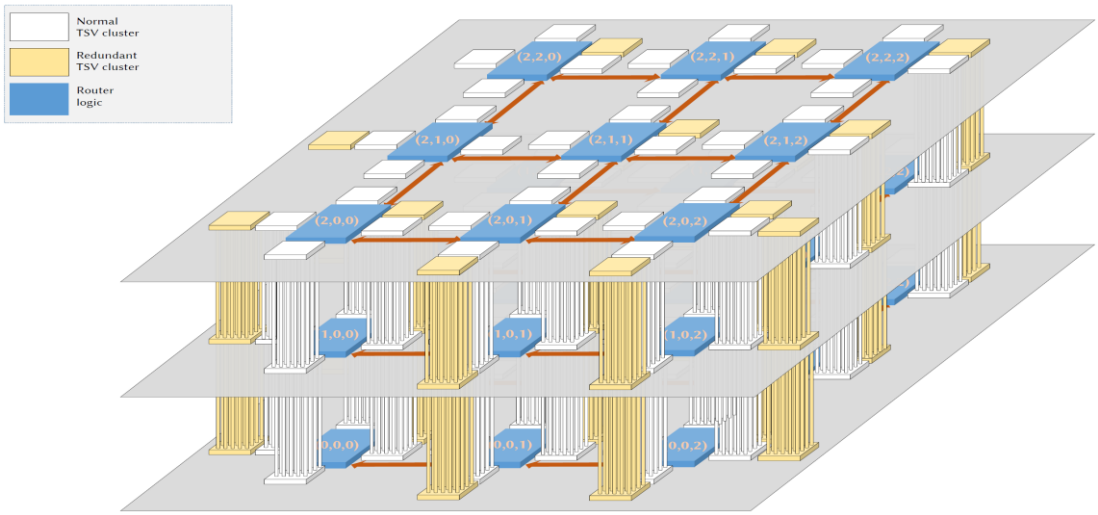

Figure 4. Architecture of our 3D Network-on-Chip with the size of 3x3x3.

where N_{bit} is the number of a data bits in the benchmark. We can also scale the power with the dynamic frequency and voltage if needed. Here, we also support dynamic scaling for voltage and frequency by using Equation 2 where different voltage and frequency can be converted using the following equations:

$$\frac{P_{static}^{V_1F_1}}{P_{static}^{V_2F_2}} \approx \frac{V_1}{V_2} \quad (4)$$

$$\frac{P_{dynamic}^{V_1F_1}}{P_{dynamic}^{V_2F_2}} \approx \frac{f_1V_1^2}{f_2V_2^2} \quad (5)$$

where V_1,f_1 and V_2,f_2 are two pairs of supply voltage and frequency.

The power trace and floorplan are taken into Hotspot 6.0 to obtain the thermal map of the design. The results of Hotspot 6.0 are the steady temperature of each router and its TSVs. We can also support transient power and temperature. However, since we consider reliability as the major target, the steady temperature is the most important value.

3.4. Defect Mapping

After getting the thermal map, we can extract the reliability to obtain the defect map. Figure 6 shows the normalized thermal acceleration model in academics and industry. We illustrate the MIL-HDBK-217F of the US Military[22], HRD4 from industry [23] and RAMP from academics [24]. Notably, we used the Black's model [9] in our work. However, we could also adopt the existing model if needed as in Figure 6. One common between the model is the exponential curve of acceleration of the fault rate with the temperature. Note that HRD4 uses 70°C as the threshold of reliability concern.

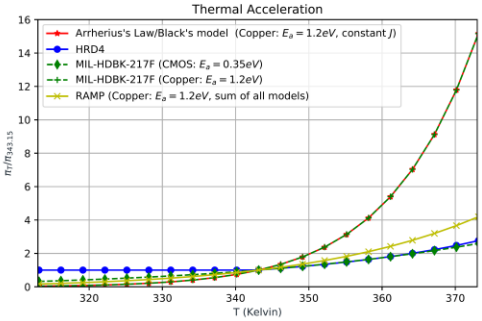

Figure 6. Normalized thermal acceleration of fault rate.

Table 1 shows the fault rate mapping obtained by Black's model [9]. At 30°C, the fault rate is less than 2% at 70°C (343.15K). However, once the IC operates at 80°C (353.15K), its fault rate is 2.6× at 70°C

(343.15K) and 220× at 30°C (303.15K). By mapping to fault rates, we can find the critical part of the 3D-NoCs in terms of reliability.

Table 1. Normalize fault rate of Copper TSV mapping using Black's model [9]

Temperature (K)	Normalize fault rate to 70°C
303.15	0.011537
313.15	0.039174
323.15	0.123317
333.15	0.362371
343.15	1
353.15	2.605435
363.15	6.439561
373.15	13.94691

4. Experimental Results

In this section, we evaluate the 3D Network on Chip [3] using the proposed platform. Furthermore, we explore the idea of the different floorplan and cooling strategies. At first, we extract the power consumption from the synthetic benchmark of a router. Then, we estimate the power consumption of the 3D-NoC system under various benchmarks. Then, temperature and reliability prediction are illustrated. In the final part, we compare different strategies for layout and cooling.

4.1. 3D-NoC Router Power Estimation

We used the router model in our previous work [3] to estimate the power consumption and the energy. Note that we modified the router with some optimizations and further fault tolerances. We use NANGATE 45nm library [25] and NCSU FreePDK TSV [26]. The hardware complexity of the router is shown in Table 2. We perform a heuristic benchmark for this router by sending each port to all possible ports two packets of ten flits of 32 bits. The number of bits is 7×7×2×10×32= 31360 bits. The desired injection rate is 1 flit/port/cycle. The final results for static power and energy per data bit are 7.66e-4 W and 9.246e-13 J/bit, respectively.

Table 2. Hardware complexity of our 3D-NoC router

Parameter	Value
Area cost	38,838 μm^2
Maximum Frequency	537.63 *MHz*
Operating Frequency	500 *MHz*
Technology	45*nm* (NANGATE 45)
Voltage	1.1 *V*
Static Power (at 500MHz)	7.64e-4 *Watt*
Dynamic Power (at 500MHz)	1.028e-2 *Watt*
Simulation time	2.823200e-6 *second*
Energy	2.9022496e-8 *Joule*
Energy per data bit	9.2546e-13 *Joule/bit*

4.2. 3D-NoC System Power Estimation

To estimate the power of 3D-NoC system, we use Equation 3 with the scaling Equation 4 and 5 for different voltage and frequency pairs if needed. Apparently, we need to obtain the number of the bits through the routing during its operation. Here, we perform both synthetic benchmarks (Matrix, HotSpot, Uniform, and Transpose) from [3], and we design a 3D-NoC version of garnet 2.0 in gem5 [27] then perform the PARSEC benchmarks suite [28]. PARSEC is one of the most well-known benchmarks for multi-core computing systems. Here, we use 64 core x64 processors as the processing elements of the PARSEC benchmarks. Here, we only extract the number of flits that went through the routers to estimate the power consumption. The power consumption of the processing elements can be obtained by using McPAT [29]; however, it is out-of-scope of this work.

Figure 7 shows the power consumption of our 3D-NoC under PARSEC benchmark. Here, we scale the frequency to 2GHz to fit with the configuration of gem5 using Equation 4 and 5. Among these benchmarks, we observe the benchmark *cannel* has the highest power consumption and also the highest variation (between the minimum and maximum power of router).

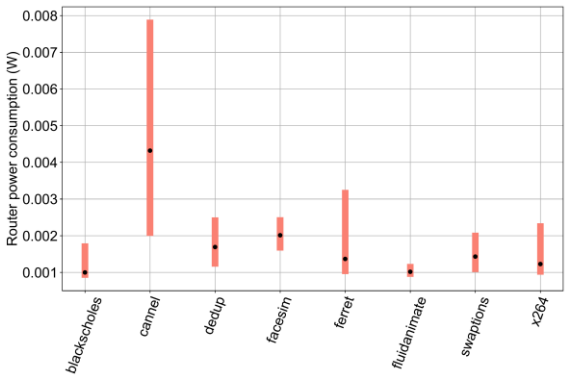

Figure 7. Power consumption of our 3D-NoC under PARSEC benchmarks.

Figure 8 shows the power consumption of the 3D-NoC system under synthetic benchmarks. We keep the frequency as of 500MHz and inject the flit with a maximum inject rate. Note that we perform two Hotspot benchmarks where two nodes are the destination of 5% and 10% of total flits. We can easily observe the significant drop when increasing the number of flits to the hotspot nodes. This can be explained by the congestion created due more flits coming to these nodes which extend the execution time of the system. On the other hand, the matrix benchmark has the lowest router power consumption. We also notice that the synthetic benchmarks have much higher power consumption than the PARSEC benchmarks since no computation is taken in this benchmarks. As a consequence, the execution time is shorter, which makes the power consumption higher than PARSEC.

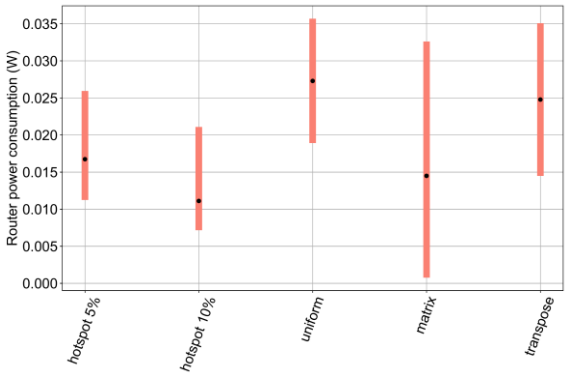

Figure 8. Power consumption of our 3D-NoC under synthetic benchmarks.

4.2. 3D-NoC Thermal Estimation

By using the power estimation of the previous section, we conduct the thermal estimation using Hotspot 6.0 [13]. Table 3 shows the configurations for thermal estimation using Hotspot 6.0. We modify the thermal resistivity corresponding to our designed TSV (Copper with the size of $4.06\mu m \times 4.06\mu m$) using the following equation [30]:

$$R_{joint} = \frac{Area}{\frac{Area - Area_{TSV}}{R_{TIM}} \_ \frac{Area_{TSV}}{R_{Copper}}} \quad (6)$$

where TIM is the thermal interface material. The result of the thermal resistivity of the layout in Figure 2(c) can be found in Table 3. The final TSV area thermal resistivity is 0.0226mK/W.

Table 3. Configurations for thermal estimation

Parameter	Value
Router floor-plan	290 μm ×290 μm
Floorplan	Figure 2(c)
One TSV area	4.06μm×4.06μm
Router logic area	220 μm ×220 μm
Router logic utilization	80%
TSV area/utilization	35,700 μm^2 / 10.16%
Copper thermal resistivity	0.0025*mK/W*
TIM thermal resistivity	0.25*mK/W*
TSV area thermal resistivity	0.0226*mK/W*

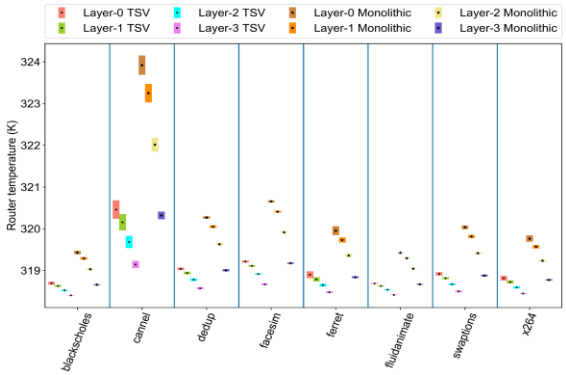

Figure 9. Temperature of our 3D-NoC under PARSEC benchmarks.

To compare with Monolithic 3D-IC, we also adopt the method in [32] where we remove the bonding layers between silicon layers. We keep the thickness of the silicon layer as it is for a fair comparison. Obviously, if we thin the layer, the transfer of heat is much faster.

Figure 9 shows the router temperature under the PARSEC benchmark. Here, we also compare with the monolithic technology where no TSV needed [32]. As we can observe in Figure 9, the TSV-based system has lower operating temperature thanks to the ability to transfer the heat of Copper TSVs. The difference in temperature is around 1K at the bottom layer and even reach 3.5K in the cannel benchmark.

Figure 10 shows the operating temperature under synthetic benchmarks of our 3D-NoC. We can easily notice that the operating temperature of Monolithic systems is much higher than TSV ones since we stress the system under its saturation points. The highest temperature of Monolithic 3D-NoC even reaches 351.64 K (78.49°C). The hottest layer of the TSV-based system has a similar temperature as the coolest layer of Monolithic 3D-NoC.

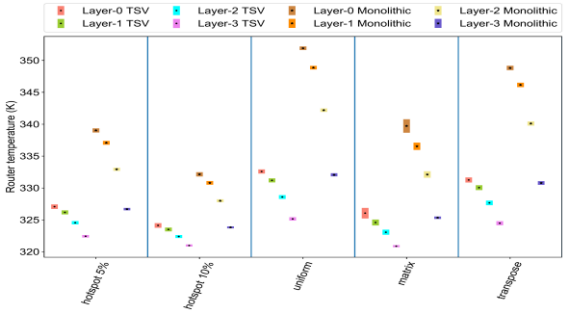

Figure 10. Temperature of our 3D-NoC under synthetic benchmarks.

4.2. 3D-NoC Reliability Estimation

In this section, we use the Black's model to evaluate the MTTF of 3D-NoC. Figure 11 and Figure 12 show the normalized MTTF of each layer to 323.15K (50°C) under PARSEC and synthetic benchmarks. Here, we can observe the TSV-based 3D-NoC dominates Monolithic in the PARSEC benchmark. With synthetic benchmarks, TSV-based 3D-NoC is slightly better than Monolithic ones.

4.4. Exploring Different Layout and Thermal Dissipation Method

In this section, we explore different layouts and their thermal dissipation behaviors for our 3D-NoC. First, we perform thermal and reliability prediction for our layout in Figure 2(b). Then, we insert four thermal TSVs with the size 15 μm ×15 μm in four corners of the router floorplan in Figure 2(c). This size of TSV is still feasible in the existing manufacture process [7]. We also add 10 μm Keep-out-Zone distance this thermal TSV to avoid mechanical stress. The thermal TSV went through all layers of TSVs but did not contact with the heatsink. The heatsink and thermal TSV are separated by a layer of thermal interface material.

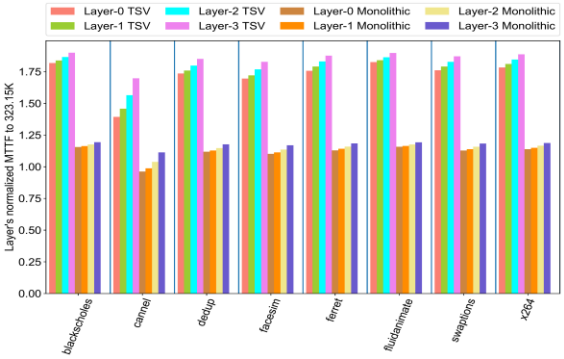

Figure 11. Normalized MTTF of our 3D-NoC under PARSEC benchmarks.

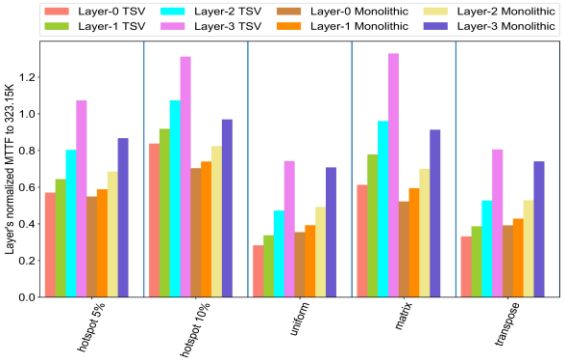

Figure 12. Normalized MTTF of our 3D-NoC under synthetic benchmarks.

Figure 13 and Figure 14 show the thermal behaviors under PARSEC and synthetic benchmarks for different layouts and cooling. We can notice that the layout in Figure 2(b) has the worst thermal behavior among the TSV designs. On the other hand, adding thermal TSV can help reduce the operating temperature significantly. By adding four TSVs, we can even reduce the temperature by nearly 1K at the bottom layer in the uniform benchmark which is the most stressed benchmark. Other benchmarks' results also show a slight improvement in thermal behaviors.

One thing we can easily notice the top layer's temperatures do not change. This is due to the fact it is already cool down by the heatsink and adding TSV cannot help it reduces the temperature. Also, the heatsink temperature is raised near the top layer temperature which reduces the ability to transfer heat. If the thermal TSV can contact the heatsink, it can significantly cool down the bottom layer. Also, liquid cooling could be extremely helpful in this situation.

In comparison to the traditional 2D-ICs, we observe that the TSV-based ICs have higher operating temperatures. The 2D-based 3D-NoCs operate under 319K and 322K with PARSEC and synthetic benchmarks, respectively. On the other hand, TSV-based system increases at most 10K in maximum temperature with the layout in Figure 2(b).

In summary, different layouts can make different thermal behaviors. The layout in Figure 2(b) does not surround the router by TSV area, therefore, the router could heat up each other and reach a higher temperature. On the other hand, adding thermal TSV to cool down the bottom layer is helpful since it can reduce nearly 1 Kelvin in the worst case. By mapping to the reliability, we can easily obtain a 2×~3× improvement of MTTF.

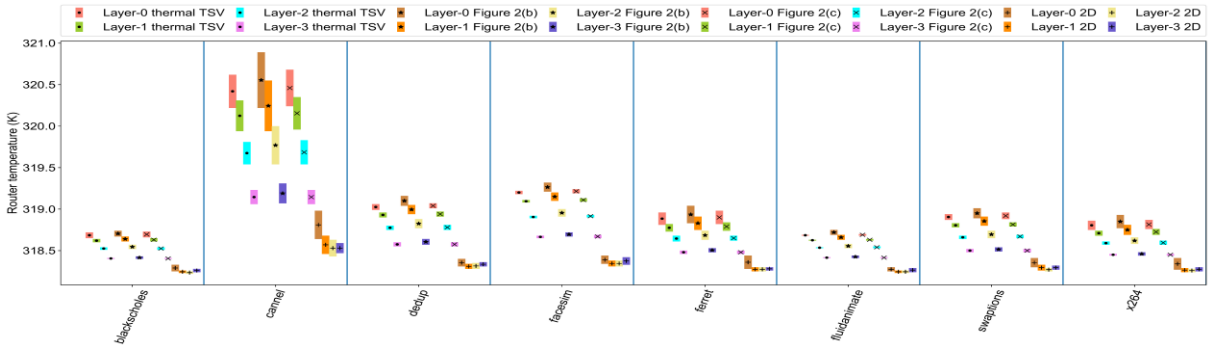

Figure 13. Thermal behavior of different layouts and cooling methods under the PARSEC benchmark.

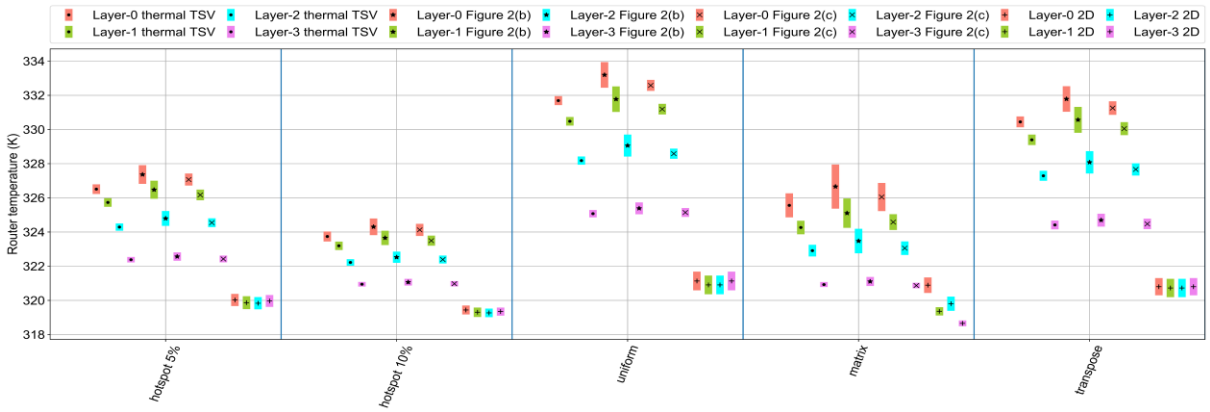

Figure 14. Thermal behavior of different layouts and cooling methods under the synthetic benchmarks.

4.5. Execution Time

In this work, we evaluate the proposed method using a system with Xeon E5-2620 8 cores 2.1GHz, 16GB RAM and Linux Subsystem and PowerShell under Windows 10. The platform is written under C++, Python, and Bash. The execution time is measured using command time under Linux and Measure-Command under Windows PowerShell. Here, the simulation time of PARSEC and synthetic benchmarks are not considered because they are separated from our flow. As shown in Table 4, all steps in our flow perform under two seconds. Our method easily outperforms in terms of execution time the fabrication-based methods which usually take hours regardless of designing, fabrication and assembly time [10-12].

Table 4. Execution time of the proposed flow

Work	Step	Time
Ours	Power extraction (one benchmark)	1.22 s
	Floorplan generate	0.095 s
	Temperature estimation (one benchmark)	81 s
	Reliability estimation (12 benchmarks)	1.12 s
[10]	Reliability test	96h
[11]	The longest step in reliability test	1000h
[12]	Lifetime acceleration test	100-5000h

Although our approach is fater than real-chip testing [10-12], it cannot as accurate as the baking tests due to the deviations during simulation and the potential of manufacturing variation. However, as the close-loop design flow, having an understand of the potential reliability threat is helpful for designers.

4.6. Discussion

In this section, we would like to discuss some technical details of our methods. Advantages and drawbacks are also mentioned in this part.

In our evaluation, we point out that Monolithic has a higher temperature than TSV-based 3D-NoCs due to two major reasons: i) TSVs act like thermal conduct devices and ii) Monolithic 3D-ICs has a higher density than TSV-based system. However, we would like to note that Monolithic 3D-ICs have lower area cost than TSV-based systems.

Fluid cooling [7] is one of the most advanced methods to reduce the operating temperature of the system. Although we have not explored the ability of this method, it has shown promising efficiency for 3D-ICs [7]. With a fast velocity of the fluid, we expect the system can be cooled down significantly. However, we would like to note that fluid cooling has unknown reliability which needs to be carefully investigated for being widely used.

5. Conclusion

In this work, we proposed a platform to quickly estimate the power, thermal behavior, and reliability of 3D-NoC systems. The method has shown extremely short execution time. We also analyze and simulate the reliability of TSV and Monolithic 3D-ICs. Furthermore, we explore and compare different layout strategies and cooling methods.

From our experiments with 3D-NoC, we can realize that lower index layers have higher operating temperatures and are more critical in terms of reliability. Although this conclusion cannot cover all possible cases; this is a consensus of the tested benchmark Based on these experiments, designers can decide their fault-tolerance or thermal dissipation up on their required specification.

In the future, advanced cooling techniques such as liquid could be investigated. The impact of DVFS and fault tolerance on performance and thermal behavior also could be studied.

Acknowledgments

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2018.312.

References

[1] Khanh N. Dang, Akram Ben Ahmed, Xuan Tu Tran, Yuichi Okuyama, Abderazek Ben Abdallah, "A Comprehensive Reliability Assessment of Fault-Resilient Network-on-Chip Using Analytical Model," IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 25(11) (2017) 3099-3112. https://doi.org/10.1109/TVLSI.2017.2736004.

[2] K. Banerjee K. Banerjee, S.J. Souri, P. Kapur and K.C. Saraswat, "3-D ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration," Proc. IEEE. 89(5) (201) 602-633. https://doi.org/10.1109/5.929647.

[3] Khanh N. Dang, Akram Ben Ahmed, Yuichi Okuyama, Abderazek Ben Abdallah, "Scalable design methodology and online algorithm for TSV-cluster defects recovery in highly reliable 3D-NoC systems", IEEE Transactions on Emerging Topics in Computing, 2017, pp. 1-14 (in-press). https://doi.org/10.1109/TETC.2017.2762407.

[4] Wong, Simon, et al. "Monolithic 3D integrated circuits" International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA), IEEE, 2007.

[5] Y.J. Park et al., "Thermal Analysis for 3D Multi-core Processors with Dynamic Frequency Scaling", in IEEE/ACIS 9th Int, Conf, on Computer and Information Science, Aug 2010, pp. 69-74.

[6] Van der Plas, Geert, et al., "Design issues and considerations for low-cost 3-D TSV IC technology". IEEE Journal of Solid-State Circuits 46(1) (2010) 293-307.

[7] D. Cuesta et al., "Thermal-aware floorplanner for 3D IC, including TSVs, liquid microchannels and thermal domains optimization," Applied Soft Computing 34 (2015) 164-177. https://doi.org/10.1016/j.asoc.2015.04.052.

[8] Park, Changyok, "Dummy TSV to improve process uniformity and heat dissipation", U.S. Patent 10, 181, 454, 15 Jan, 2019. https://patents.google.com/patent/US20110215457A1/en (access 16 March 2020).

[9] J.R. Black, "Mass transport of aluminum by momentum exchange with conducting electrons", in 6th Annual Reliability Physics Symposium (IEEE), IEEE, 1967, pp. 148-159.

[10] Hamada, M. Dorothy June, J. William, Roesch, "Evaluating device reliability using wafer-level methodology", CS Mantech Conference, 2008.

[11] Renesas's Semiconductor Reliability Handbook https://www.renesas.com/us/en/doc/products/others/r51zz0001ej0250.pdf/, 2017 (access 17 March 2020).

[12] Toshiba's Reliability Handbook https://toshiba.semicon-storage.com/content/dam/toshiba-ss/shared/docs/design-support/reliability/reliability-handbook-tdsc-en.pdf /, 2018 (access 17 March 2020).

[13] Zhang, Runjie, Mircea R. Stan, Kevin Skadron, "Hotspot 6.0: Validation, acceleration and extension", University of Virginia, Tech, Rep, 2015.

[14] Sridhar, Arvind, et al., "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling", 2010 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), IEEE, 2010.

[15] Scott Ladenheim, Yi-Chung Chen, Milan Mihajlović, Vasilis F. Pavlidis, "The MTA: An Advanced and Versatile Thermal Simulator for Integrated Systems", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 37(12) (2018) 3123-3136. https://doi.org/10.1109/TCAD.2018.2789729.

[16] Erdmann, Christophe, et al., "A heterogeneous 3D-IC consisting of two 28 nm FPGA die and 32 reconfigurable high-performance data converters", IEEE Journal of Solid-State Circuits 50(1) (2014) 258-269. https://doi.org/10.1109/JSSC.2014.2357432.

[17] Kahng, B. Andrew, et al., "ORION 2.0: A fast and accurate NoC power and area model for early-stage design space exploration", Design, Automation & Test in Europe Conference & Exhibition, IEEE, 2009.

[18] Lee, Seung Eun, and Nader Bagherzadeh, "A high level power model for Network-on-Chip (NoC) router", Computers & Electrical Engineering 35(6) (2009) 837-845. https://doi.org/10.1016/j.compeleceng.2008.11.023.

[19] Lee, Seung Eun, Nader Bagherzadeh, "A variable frequency link for a power-aware network-on-chip (NoC)", Integration 42(4) (2009) 479-485. https://doi.org/10.1016/j.vlsi.2009.01.002.

[20] Lebreton, Hugo, Pascal Vivet, "Power modeling in SystemC at transaction level, application to a DVFS architecture", 2008 IEEE Computer Society Annual Symposium on VLSI, IEEE, 2008.

[21] Khanh N. Dang Akram Ben Ahmed, Abderazek Ben Abdallah, Xuan-Tu Tran, "TSV-OCT: A

Scalable Online Multiple-TSV Defects Localization for Real-Time 3-D-IC systems" IEEE Transactions on Very Large Scale Integration Systems 28(3) (2020) 672 - 685. https://doi.org/10.1109/TVLSI.2019.2948878.

[22] United States of America: Department of Defense, Military Handbook: Reliability Prediction of Electronic Equipment: MIL-HDBK-217F, 1991.

[23] J.B. Bowles, "A survey of reliability-prediction procedures for microelectronic devices", IEEE Trans, Rel. 41(1) (1992) 2-12. https://doi.org/10.1109/24.126662.

[24] J. Srinivasan et al., "Lifetime reliability: Toward an architectural solution", IEEE Micro. 25(3) (2005) 70-80. https://doi.org/10.1109/MM.2005.54.

[25] NanGate Inc., "Nangate Open Cell Library 45nm" http://www.nangate.com/, 2016 (accessed 16 June 2016).

[26] NCSU Electronic Design Automation, "FreePDK3D45 3D-IC process design kit", http://www.eda.ncsu.edu/wiki/FreePDK3D45:Contents/, 2016 (accessed 16 June 2016).

[27] Binkert, Nathan, et al., "The gem5 simulator", ACM SIGARCH computer architecture news 39(2) (2011) 1-7.

[28] Bienia, Christian, et al., "The PARSEC benchmark suite: Characterization and architectural implications", Proceedings of the 17th international conference on Parallel architectures and compilation techniques, 2008.

[29] Li, Sheng, et al., "McPAT: an integrated power, area and timing modeling framework for multicore and manycore architectures", Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture, 2009.

[30] J. Meng, K. Kawakami, A.K. Coskun, "Optimizing energy efficiency of 3-d multicore systems with stacked dram under power and thermal constraints", in DAC Design Automation Conference 2012, IEEE, 2012, pp. 648-655.

[31] Khanh N. Dang, Akram Ben Ahmed, Abderazek Ben Abdallah, Michael Corad Meyer, Xuan-Tu Tran, "2D Parity Product Code for TSV online fault correction and detection", REV Journal on Electronics and Communications (in-press). http://dx.doi.org/10.21553/rev-jec.242.

[32] Samal, Sandeep Kumar, et al., "Fast and accurate thermal modeling and optimization for monolithic 3D ICs", 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC), IEEE, 2014.