

NoD: Network-on-Die as a Standalone NoC for Heterogeneous Many-core Systems in 2.5D ICs

Masoumeh Ebrahimi*, Awet Yemane Weldezion[†], Masoud Daneshtalab[‡]

*KTH Royal Institute of Technology, Sweden; mebr@kth.se

[†]Hangofay AB, Sweden; awet@hangofay.se

[‡]Mälardalen University, Sweden; masoud.daneshtalab@mdh.se

Abstract—Due to a high cost of 3D IC process technology, the semiconductor industry is targeting 2.5D ICs with interposer as a fast and low-cost alternative to integrate dissimilar technologies. In this paper, we propose an independent network-on-chip die, called Network-on-Die (NoD), for 2.5D ICs that operates as a communication backbone for heterogeneous many-core systems on interposer. NoD is responsible for routing packets from a source router to a destination router, and the connections between routers and cores pass through the interposer. This technique eliminates the complexity of the routing algorithms in heterogeneous systems by turning the irregular form of NoC in 2.5D ICs into a regular/optimized one in NoD. The performance evaluation is verified through RTL simulations for a heterogeneous many-core system of varying die sizes and with asymmetric shapes. We provide the theoretical justification for our simulation results.

Index Terms—Network-on-Die; Heterogeneous Integration; 2.5D Many-core System; Network-on-Chip; Silicon Interposer.

I. INTRODUCTION

System-on-Chip (SoC) design is moving towards the integration of tens to hundreds of intellectual property (IP) cores on a single chip. Networks-on-Chip (NoC) have been proposed as a promising solution for designing the interconnect fabric of these IP cores [1][2].

To overcome the interconnect delay and area limitations of a planar NoC, technology is moving rapidly towards the concept of integration in the third dimension (3D IC). In 3D ICs (Figure 1(a)), multiple active silicon layers are vertically stacked using fast interconnects, called Through-Silicon-Vias (TSV). The physical characteristics of TSVs differ fundamentally from the on-chip wires. For example the resistance of TSVs is much lower than on-chip horizontal links due to the shorter length and larger cross-section. Thereby, the RC delay constant is relatively smaller [3]. NoC can be extended to 3D systems in a straightforward way and shown to be an efficient platform for the network sizes up to hundreds of nodes.

While the main advantages of 2D NoCs are simplicity and ease of design, 3D NoCs have a number of advantages over 2D designs as: (1) lower interconnect delay; (2) shorter global interconnects; (3) higher performance; (4) lower power consumption; (5) higher packing density and smaller footprint; and (6) support for the implementation of mixed-technology

chips [4][5][6][7]. However, there are three main design challenges regarding 3D NoCs as follows:

- 1) **Area overhead:** Each TSV requires a pad (around $5\mu \times 5\mu$) with the pitch of around 8μ for bonding to a wafer [8][9], indicating that the area overhead of TSVs imposes constraints on the number of TSVs.
- 2) **Thermal problem:** 3D die stacking leads to a high power density which in turn increases the risk of thermal heating. In addition temperature is trapped in middle layers and cannot be easily dissipated, resulting in accumulated temperature and consequently serious damages to the device.
- 3) **Expensive manufacturing process:** The current TSV process flow uses copper as a main conducting material to fill VIA cavities in a silicon substrate. The process involves three levels:
 - In the first level, to avoid contacts between the copper and silicon substrate, a barrier is formed using SiO₂ and other similar materials. This process is repeated for every TSV made on a wafer. Accordingly, the cost of 3D integration is dominated by the cost of manufacturing TSVs.
 - In the second level, TSV filling is made for each die stack. Yield varies based on the size and number of TSVs. If the yield is low in one die stack, the whole package is wasted. Thereby, to keep the manufacturing cost low, the yield of TSV filling should be kept very high. On account of this, only the known good 3D stacks (KG3DS) are then ready for final 3D integration.
 - In the third level, wafer bumping and packaging is done with the KG3DS. If the yield of packaging is low, all TSVs made in the previous steps together with the stacked dies are wasted.

Considering these challenges, chip integration on interposer (2.5D IC) can be seen as a suitable alternative to 3D ICs, offering a low-cost approach while keeping most advantages of 3D ICs. In 2.5D IC, as shown in Figure 1(b), dies are placed on an interposer linked with interconnects through the interposer. As an example, a 2.5D IC approach used by Xilinx is based on a stacked silicon interposer (SSI) technology where multiple FPGA dies are placed on a silicon interposer and connected to each other through the interposer [10][11]. The

*The first two authors contributed equally to this work.

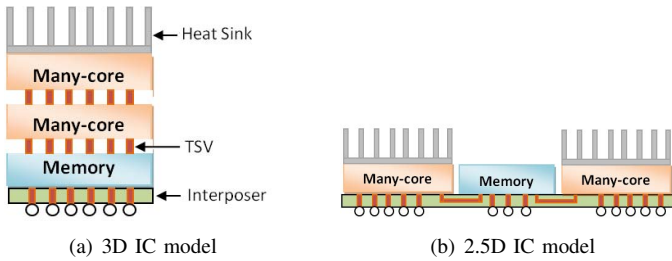


Fig. 1. Models of (a) 3D IC and (b) 2.5D IC on interposer

FPGA dies are first fabricated individually using the existing process technologies and then bonded to the silicon interposer with micro-bumps.

We introduce the idea of Network-on-Die (NoD) for 2.5D ICs, suggesting a standalone die where the NoC die is fabricated independent of the cores. Figure 2(a) shows a SoC with a NoD in the central part. The NoD is responsible for routing packets from a source router to a destination router. Each core in SoC is connected to a router in NoD and the connections pass through the interposer as shown in Figure 2(b). The proposed NoD design combines the benefits of 2D and 3D designs where the main advantages are as: (1) design simplicity similar to 2D designs with broad architectural and design choices; (2) shorter global interconnects by decoupling cores from NoC; (3) diminishing the thermal issues of 3D designs; (4) low manufacturing cost; and (5) ease of heterogeneous integration than both 2D and 3D designs. On the other hand, the floor-planning choices of 2.5D ICs are limited as compared to 3D ICs but allowing a better integration than 2D ICs.

The need for the proposed NoD model is driven by the design challenges that arise with the technology trends. System-on-Chip designs are going toward heterogeneous systems where the cores supplied by third parties have different characteristics and sizes. Such trends demand a network infrastructure and communication backbone that could efficiently accommodate all these cores on a single chip. Proper irregular network designs for 2D NoC and 3D NoC are very limited and at a cost of a degraded performance, increased logic-circuitry and architectural complexity. However, the idea of Network-on-Die suggests several unique features as:

- NoD can be fabricated independently which allows accommodating cores with irregular sizes using a regular network.
- NoD can be easily optimized for performance and power due to its independence from the cores.
- NoD can be used as a plug-and-play type of die with different sizes, topologies, etc.
- Heterogeneous integration would be easily possible where cores can be fabricated with different technologies, enabling dissimilar technology integration.

The remainder of this paper is organized as follows: Section II discusses the related work. The Network-on-Die and the comparative analysis to the equivalent 2D, 3D, and traditional

2.5D NoC counterparts are provided in Section III. Section IV describes the link delay and energy modelling. The experimental results are reported in Section V and finally we conclude the paper in Section VI.

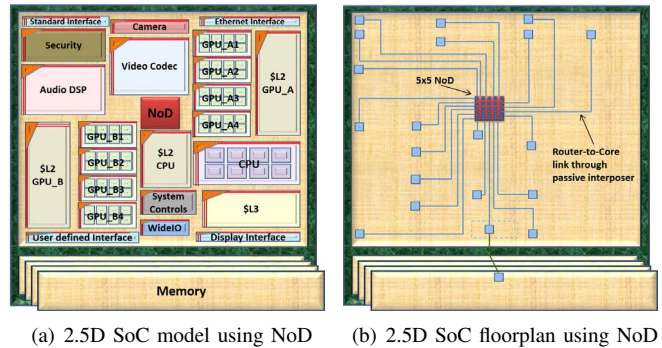


Fig. 2. A 2.5D mobile application processor with memory stacks on interposer

II. RELATED WORK

Mesh-based 2D NoC architectures have been commonly accepted as a solution for the complex on-chip communication problems [1][2][12]. Many studies have been carried out over the years proposing various architectures, topologies, and routing algorithms for both regular and irregular 2D NoCs. A typical 2D router consisting of 5 ports is shown in Figure 3(a).

3D NoC, on the other hand, can be considered as a new area of research where limited amount of work has been conducted so far. Design complexities and technological challenges have slow down the progress in this domain. A comprehensive comparison between 2D and 3D NoC designs has been performed in [6], including speed, power, and latency analysis. The work in [4] describes a detailed study on router architectures for 3D NoC designs. Various on-chip network topologies have been studied for 3D NoCs [8]. Mesh-based structures are popularly used in 3D systems as their grid-based regular architecture is intuitively considered to be matched to the 2D VLSI layout for each stack layer. Nevertheless, if the number of IP-cores and memories increases in a layer, more TSVs are necessitated to handle the inter-layer communication. Inasmuch as each TSV employs a pad for bonding, the area footprint of TSVs in a layer is augmented significantly [4]. A typical 3D router consisting of 7 ports is shown in Figure 3(b).

A few works have discussed NoCs considering the silicon interposer technology (2.5D stacking) [13][14]. An approach to efficiently employ the NoC architecture in the silicon interposer is proposed in [13]. This paper argues that the current design approaches only utilize the interposer for chip-to-chip routing and vertical connections to the package substrate for power, ground, and I/O. Thereby, a 2.5D design is proposed to make a better use of the abundant and unused routing resources on the interposer layer. NoC uses a topology that spans both the multi-core die and the interposer. Although this work discusses different aspects of the integration of NoC in 2.5D ICs in details, the integration is done in a traditional manner. In the other words, the NoC platform is closely coupled with the

processing elements as shown in Figure 4(a) with the floorplan illustrated in Figure 4(b). This type of integration limits the design choices and flexibility. The authors in [14] provide the cost and yield analysis of interposer technologies in multicore processors. A typical 2.5D router consists of 9 ports, four of which connected to the east, west, north and south routers through the interposer as shown in Figure 3(c). The number of ports can be reduced by the concentration method, discussed in [13].

In this paper, we introduce NoD for 2.5D ICs that offers a new perspective to NoC designs which is radically different from traditional NoC platforms in 2D, 3D, and 2.5D IC designs. A router in this architecture consists of 5 ports as shown in Figure 3(d) where the connection between a core and its corresponding router is made through passive interposer.

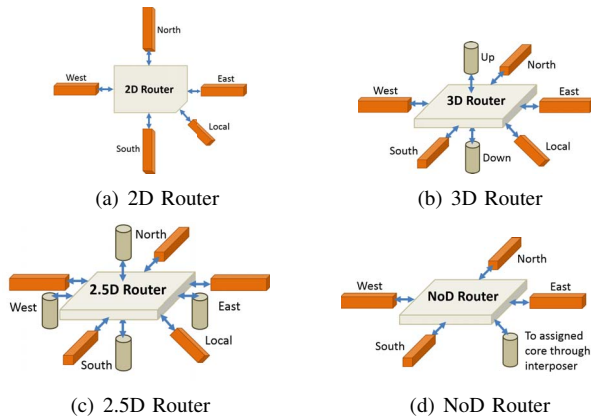


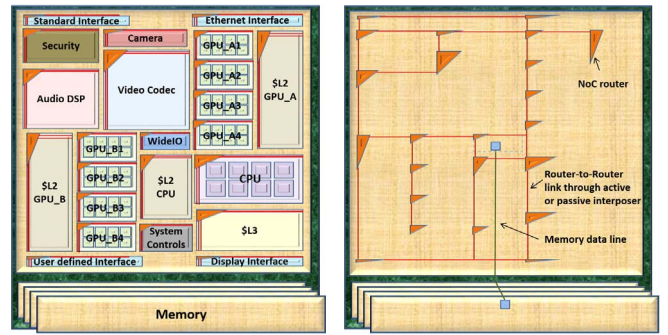
Fig. 3. Router models for 2D, 3D, 2.5D, and NoD router

III. NETWORK-ON-DIE

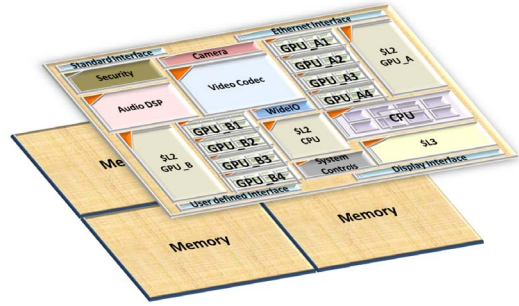
A 2D NoC architecture based on the mesh topology has been widely used in regular networks [1][2][12]. A 2D router is used to facilitate the flow of packets to different destinations in the network. In irregular networks shown in Figure 4(a), due to the differences in the shape and size of the cores, the length of the links connecting the routers is not the same. Such heterogeneous system configuration brings challenges in achieving optimal performance and power consumption due to the imposed network irregularity.

Routing algorithms for regular 2D NoCs are well-studied. However, routing algorithms are becoming very challenging in irregular networks such as the one shown in Figure 4(a). Thereby, even though 2D NoC is known to be a simple design option, it is not the case when network irregularity is taken into consideration. The recently proposed theory on interconnection networks, EbDa [15], helps to reach an optimal routing algorithm for a given topology. However, the optimization space is generally limited in irregular networks.

The complexity increases even further in 3D systems. In a heterogeneous system design (Figure 4(c)), the topology might be different in various layers. For example, one layer may have a mesh-based NoC while another layer may take a connectivity form of ring-based NoC. Despite the necessity



(a) 2.5D SoC model using NoC (b) 2.5D SoC floorplan using NoC



(c) 3D SoC

Fig. 4. A mobile application processor (a) typical 2.5D SoC model with NoC and WideIO connectivity; (b) 2.5D SoC floorplan; (c) 3D SoC with logic layer stacked on top of memory layer

of heterogeneous 3D NoC designs, the main issue is that the network cannot be easily optimized regarding the performance and power. Usually heuristic approaches are applied for each topology configuration. On the other hand, proposing a heuristic approach for different configurations of heterogeneous 3D NoCs imposes huge costs and engineering efforts.

Due to the aforementioned inherent limitations of 2D, 3D, and traditional 2.5D NoC designs in heterogeneous many-core systems, we suggest NoD as a standalone die with an optimized network (Figure 2). In the proposed design, cores can be connected to a router in the NoC die using Through Silicon Interposer (TSI). This allows a general purpose NoC platform providing a high-performance communication for a heterogeneous system (i.e. dies are supplied as IP cores designed by different vendors with varying shapes, sizes, and technology nodes).

2D, 3D, or traditional 2.5D NoCs are lacking efficient network solutions in heterogeneous systems as the network should be handled in an irregular manner. As was already discussed, an irregular network considerably limits design choices and also puts strict constraints on mapping cores on the platform. These issues are perfectly relaxed in the NoD architecture. The standalone NoD can be designed in a regular or optimized way regardless of the underlying system heterogeneity. The router architecture, routing algorithm, and topology can be flexibly designed for a better performance and a lower power consumption.

IV. NOC LINK DELAY AND ENERGY MODELLING

To compare delay and energy consumption of NoCs in 2D, 3D and 2.5D technologies, we model the interconnect for each case considering the latest technology parameters.

In 2D NoCs only horizontal links are available and the length of each link depends on the local core size attached to a router. Generally, the links are long and a distributed RC circuit is used to model the delay. In order to minimize the delay in such resistive wires, the common practice is to insert repeaters, breaking the wire into several segments as shown in Figure 5(a).

3D NoCs consists of both horizontal and vertical links. The properties of the horizontal interconnects in 3D NOC are the same as in 2D NOC. The new addition is the vertical interconnect known as Through-Silicon-Via (TSV). TSVs have unique properties as compared to horizontal interconnects. The first visible difference is that whereas the horizontal interconnects are long and thin, TSVs are short and fat. As a result, delay in long horizontal interconnects worsens and requires several stage repeaters in between to fasten the speed. With TSVs, repeaters are not implemented because TSVs are completely immersed within the silicon substrate. However when two TSVs are close to each other, the spacing between them is relatively wide. Thereby, when several TSVs are placed in parallel array, they consume a large area.

From the fact that a TSV behaves like a short wire, a lumped RC TSV model with a load and driver (RLC model) is sufficient to accurately model the TSV as shown in Figure 5(b).

2.5D NoD consists of on-chip wires and interposer interconnects. Through Silicon Interposer (TSI) technology with thick RDL layers has been considered in this comparison [16]. The RDL wires are significantly wide and thick and therefore their resistance is two orders of magnitude smaller than that of the 2D NoC wires. Since active devices are not presented in the silicon interposer, no repeaters can be presented to minimize the delay. However, the wire resistance is very low, and thus there is no need of inserting repeaters. An electrical model of a wire in TSI is shown in Figure 5(c).

V. RESULTS AND DISCUSSION

A. Evaluation of Link Delay and Energy

To analyse the physical properties of 2D NoCs, we assume a CMOS process technology of $32nm$ [17]. The physical parameters of wires in this technology and the extracted parasitics are reported in the 1st row of Table I using ITRS roadmap [17]. The required number of repeaters is estimated using the models in *Predictive Technology Model* [18]. Based on the Elmore delay model of Figure 5(a) and deriving with respect to the number of repeater stages, the optimal propagation delay (t_{pd}) per unit length can be expressed by $2.13\sqrt{R_w C_w FO1}$, where $FO1$ is the delay of an inverter deriving an identical copy of itself that is equal to $4.2ps$ in the $35nm$ technology. The required number of repeater stages is estimated by $\sqrt{\frac{2FO1}{R_w C_w}}$. Based on the wire parasitic,

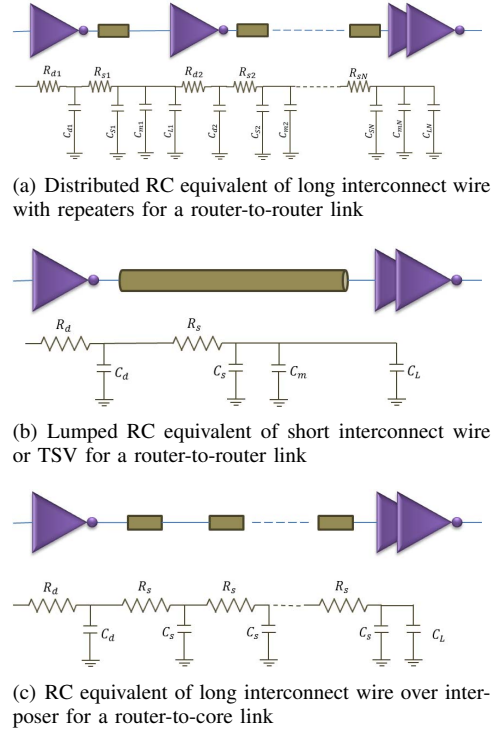


Fig. 5. RC equivalent of on-chip wires, TSV, and interposer interconnects

repeaters must be inserted at $0.34mm$ distances. Delay values for different wire lengths are listed in the 1st row of Table I.

The energy consumption of links is estimated separately for wires and repeaters. The energy consumption of a wire is measured by $0.5C_w l_W V_{dd}^2$ while the energy consumption of a repeater is calculated by $kHC_g V_{dd}^2$ where k and H are the number of repeaters and the repeater size, respectively and C_g is the gate capacitance of a repeater. Accordingly the energy consumption for different wire lengths is calculated (1st row of Table I).

The physical characteristics, delay and energy values of links in 3D NoCs (Figure 5(b)) are reported in the 2nd row of Table I. For the sake of simplicity we assume that the driver and receiver of TSVs are very close to each other so that no horizontal wire is required. Therefore, the delay model of TSV is half of the product of its resistance and capacitance. The energy consumption per TSV is given by the typical equation $0.5C_{tsv} V_{dd}^2$. TSV with the diameter of $5\mu m$ and the height of $50\mu m$ has been simulated in Synopsys Device. The capacitance and resistance are extracted as $24.5fF$ and 909Ω in the fully depleted region.

Finally, the 3rd row of Table I is dedicated to the 2.5D NoD design based on the link model of Figure 5(c). The reported parasitics of RDL wires are estimated based on *Predictive Technology Model* [18]. The delay and energy consumption of a wire per unit length are estimated using $0.5R_w C_w$ and $0.5C_w V_{dd}^2$, respectively.

$$T = \ln(2) \sum RC \quad (1)$$

TABLE I
WIRE DELAY AND ENERGY PER BIT BASED ON 35NM TECHNOLOGY

	Case	Parameters	Parasitics	Delay	Delay (in ps) for Different Wire lengths (mm)					Energy per bit per length (pJ/mm)	Energy (pJ/bit/wire) for Different Wire lengths (mm)			
					1	3	5	10	1		3	5	10	
Ref [19]	2D NoC	W=S=140nm H=308nm	R [Ohm/mm] 510 C [fF/mm] 154	repeated [in ps/mm]	36.00	36.0	108.0	180.00	360.00	1.54	1.54	4.62	23.10	231.00
	3D NoC (TSV)	Diameter=5μm Height=50μm	R [Ohm] 0.043 C [fF] 24.5	in ps	5E-4	5E-4	5E-4	5E-4	5E-4		0.01	0.01	0.01	0.01
Ref [16]	2.5D NoD	W=S=3μm H=3μm	R [Ohm/mm] 2.44 C [fF/mm] 116.6	in ps/mm ²	0.14	0.14	1.28	3.56	14.23	0.06	0.06	0.17	0.87	8.75

For the equivalent RC circuit shown in Figure 5(b), the time delay is given by:

$$T = (R_d[C_d + C_s + C_m + C_L] + R_s[C_s + C_m + C_L]) \ln(2) \quad (2)$$

Equation 2 contains two terms, delay from the drive throughout the wire and delay due to the total wire impedance. With further rearrangement, Equation 3 is obtained:

$$T = \ln(2)(R_d[C_d + C_s + C_L] + R_s[C_s + C_L] + (R_d + R_s)C_m) \quad (3)$$

Except the C_m , all the other variables are not affected by the crosstalk and can be represented as fixed values. C_m is the sum of individual mutual capacitance with the middle TSV.

The interesting point in this table is the delay and energy values of links in a 2D NoC vs. 2.5D NoD. The delay of a 1mm and 3mm wire on active devices is around 36ps and 108ps, respectively while the link delay over the interposer is around 0:14ps and 1:28ps for the same wire length. Similarly the link energy of a 1mm and 3mm wire in the 2D NoC design is around 1:54 and 4:62 pJ/bit/wire while the equivalent values in 2.5D NoD is around 0:06 and 0:17 pJ/bit/wire. The delay and energy values are very small in TSVs. These analyses indicate the fact that even though the local links (i.e. connecting a core to a router) are relatively long in 2.5D NoD but on the other hand the delay is very small. Therefore, packets can traverse the local link over the interposer within one clock cycle, e.g. by assuming a 2mm × 2mm SoC where the longest local link is less than 2mm, and the clock frequency is 1GHz. Based on these observations, for the evaluation of the next section we assume that the packet traversal over the local link takes only one clock cycle in 2D NoC, 3D NoC, and 2.5D NoD.

B. Experimental Evaluation

For the performance evaluation, we setup RTL-level simulation of a network with an injection rate of up to 0.9 packets per node per cycle. Without affecting the generality of NoD, for this study, a packet is considered as one flit long. The implemented network does not drop packets. The packet header generated by the core contains final destination address, and routers make routing decisions on fly based on this information. The data samples are extracted from the output files following the warm-up phase of the network and preceding the cool-down phase to ensure reliable results. Both uniform random traffic (URT) and hotspot traffic (HS) are

studied. The metrics used to quantify the performance of the network in each case are average latency and throughput.

We created two many-core models for the mobile application processor (MAP). The first model is a 3D IC shown in Figure 4(c) that uses the standard 2D NoC or 3D NoC routers attached to each core forming a tile. However, since the cores are of different sizes, the network follows an irregular-type routing. The logic plane is placed on top of the memory layer. Wide-IO standard is used as the interface between the cores and the memory stack. In Figures 6, 7, 8, and 9, the irregular 3D NoC refers to this model.

The second model is the standalone NoC in 2.5D ICs shown in Figure 2(a) that uses NoD as a communication backbone placed in the center of an interposer. Each core is connected to its router by the connections passing through the passive interposer. In addition, the memory dies are placed in the same interposer surrounding the logic cores. Wide-IO standard is used to interface the cores with the memory stacks connected through the metal wires in the interposer. In Figures 6, 7, 8, and 9, this model is called 2.5D NoD.

Figures 6 and 7 show the average latency of irregular 3D NoC and the proposed 2.5D NoD for different injection rates under uniform random and hotspot traffic patterns, respectively. It can be seen that the irregular 3D NoC is saturated earlier than 2.5D NoD despite the fact that the average distance is shorter in the 3D NoC. The reason behind this observation is the regularity of 2.5D NoD for the heterogeneous system architecture whereas an irregular topology structure has to be employed in the 3D NoC. The network irregularity in 3D NoC leads to an asymmetric packet flow and thus a higher congestion and an early saturation. However, since 2.5D NoD utilizes a regular network, the load is better balanced over the network which results in a superior performance of 2.5D NoD over 3D NoC. Figures 8 and 9 show the average throughput of irregular 3D NoC and the proposed 2.5D NoD under uniform random and hotspot traffic patterns, respectively. Due to similar reasons, 2.5D NoD offers a better throughput than irregular 3D NoC.

VI. CONCLUSION

In this paper, we proposed the idea of Network-on-Die (NoD), a standalone die where the NoC die is fabricated independent of the cores and thus can be easily optimized. We showed the viability of the proposed NoD in many-core systems with three main features: First, the model ensures the scalability of NoC in 2.5D ICs that has become the low cost

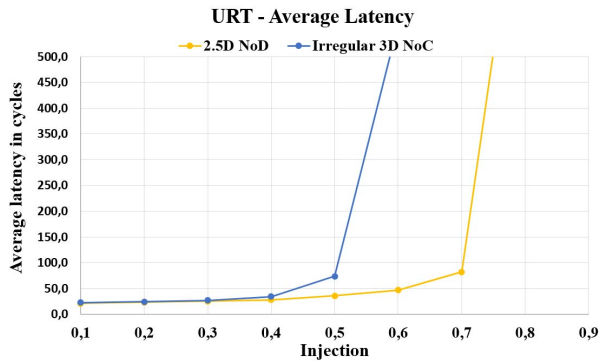


Fig. 6. Average latency in uniform random traffic - URT

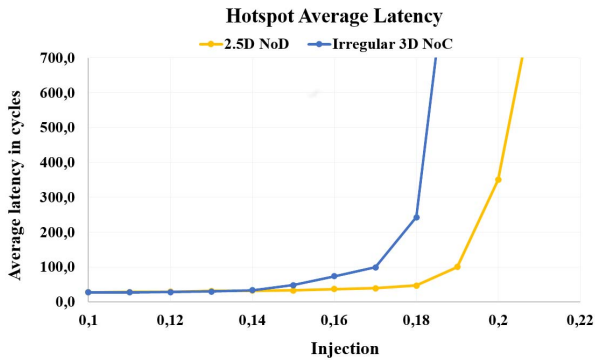


Fig. 7. Average latency in hot-spot traffic - HS

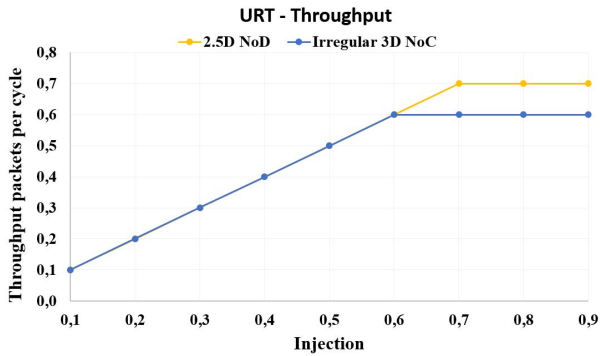


Fig. 8. Average throughput in uniform random traffic - URT

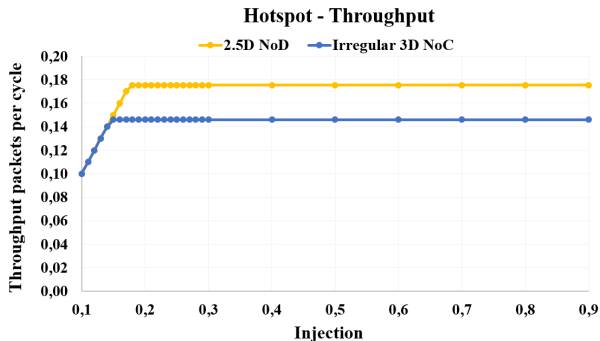


Fig. 9. Average throughput in hot-spot traffic - HS

alternative to 3D ICs. Second, by using a regular/optimized NoD, the challenges of designing high-performance routing algorithms for irregular networks are addressed. Third, the proposed model offers an integration backbone for heterogeneous cores fabricated with different technologies and process nodes. We analyzed the proposed model in terms of throughput and latency. The results show the superior performance and throughput of 2.5D NoD over its 3D counterpart.

REFERENCES

- [1] L. Benini and G. De Micheli, "Networks on chips: a new soc paradigm," *Computer*, vol. 35, no. 1, pp. 70–78, 2002.
- [2] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Millberg, and D. Lindqvist, "Network on chip: An architecture for billion transistor era," in *Proceeding of the IEEE NorChip Conference*, vol. 31, 2000.
- [3] R. Weerasekera, M. Grange, D. Pamunuwa, H. Tenhunen, and L. Zheng, "Compact modelling of through-silicon vias (tsvs) in three-dimensional (3-d) integrated circuits," in *3DIC*, 2009, pp. 1–8.
- [4] B. S. Feero and P. P. Pande, "Networks-on-chip in a three-dimensional environment: A performance evaluation," *Computers, IEEE Transactions on*, vol. 58, no. 1, pp. 32–45, 2009.
- [5] H. Matsutani, M. Koibuchi, and H. Amano, "Tightly-coupled multi-layer topologies for 3-d nocs," in *Parallel Processing, 2007. ICPP 2007. International Conference on*. IEEE, 2007, pp. 75–75.
- [6] V. F. Pavlidis and E. G. Friedma, "3-d topologies for networks-on-chip," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 15, no. 10, pp. 1081–1090, 2007.
- [7] C. Seiculescu, S. Murali, L. Benini, and G. De Micheli, "Sunfloor 3d: a tool for networks on chip topology synthesis for 3-d systems on chips," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 29, no. 12, pp. 1987–2000, 2010.
- [8] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir, "Design and management of 3d chip multiprocessors using network-in-memory," *ACM SIGARCH Computer Architecture News*, vol. 34, no. 2, pp. 130–141, 2006.
- [9] "Cluster-based topologies for 3d networks-on-chip using advanced inter-layer bus architecture," *Journal of Computer and System Sciences*, vol. 79, no. 4, pp. 475 – 491, 2013, jCSS CADs 2010.
- [10] K. Saban, "Xilinx stacked silicon interconnect technology delivers breakthrough fpga capacity, bandwidth, and power efficiency," *Xilinx White Paper: Virtex-7 FPGAs*, pp. 1–10, 2012.
- [11] M. Santarini, "Stacked & loaded: Xilinx ssi, 28-gbps i/o yield amazing fpgas," *Xcell Journal*, vol. 74, pp. 8–13, 2011.
- [12] S. Kumar, A. Jantsch, J. P. Soeninen, M. Forsell, M. Millberg, J. Öberg, K. Tiensyrjä, and A. Hemani, "A network on chip architecture and design methodology," in *VLSI, 2002. Proceedings. IEEE Computer Society Annual Symposium on*. IEEE, 2002, pp. 105–112.
- [13] N. E. Jerger, A. Kannan, Z. Li, and G. H. Loh, "Noc architectures for silicon interposer systems: Why pay for more wires when you can get them (from your interposer) for free?" in *Microarchitecture (MICRO), 2014 47th Annual IEEE/ACM International Symposium on*. IEEE, 2014, pp. 458–470.
- [14] A. Kannan, N. E. Jerger, and G. H. Loh, "Exploiting interposer technologies to disintegrate and reintegrate multicore processors," *IEEE Micro*, vol. 36, no. 3, pp. 84–93, May 2016.
- [15] M. Ebrahimi and M. Daneshalab, "Ebda: A new theory on design and verification of deadlock-free interconnection networks," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA '17. New York, NY, USA: ACM, 2017, pp. 703–715.
- [16] G. Katti, S. W. Ho, L. H. Yu, S. Zhang, R. Dutta, R. Weerasekera, K. F. Chang, J. K. Lin, S. R. Vempati, and S. Bhattacharya, "Fabrication and assembly of cu-rdl-based 2.5-d low-cost through silicon interposer," *IEEE Design Test*, vol. 32, no. 4, pp. 23–31, Aug 2015.
- [17] *International Technology Roadmap for Semiconductors (ITRS)*.
- [18] *Predictive technology model (PTM)*.
- [19] M. Grange, A. Jantsch, R. Weerasekera, and D. Pamunuwa, "Modeling the computational efficiency of 2-d and 3-d silicon processors for early-chip planning," in *2011 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2011, pp. 310–317.