# Area, Power and Performance analysis of Optimal 3D BFT NoC Architecture

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

*Bheemappa Halavar, Basavaraj Talawar*

*SPARK Lab, Computer Science and Engineering, National Institute of Technology Karnataka, India.*
** E-mail: {cs14f06.bheem,basavaraj}@nitk.edu.in*

**Abstract:**
Network-on-Chips in 3D stacked Chip Multiprocessors and SoCs have improved communication latencies and better power-performance characteristics. In this paper, we explore power and performance tradeoffs in two, 2-layer 3D Butterfly Fat Tree (BFT) variants using a floorplan driven approach. The first 3D BFT variant analyzed is a standard stacked BFT (3DBFT) derived from a 2D BFT topology. A power-performance optimal 3D BFT (OP3DBFT) is evolved from the standard 3DBFT using overall performance, link and TSV minimization, and power-performance trade-offs. The OP3DBFT has symmetric link lengths and 75% lesser TSVs compared to the 3DBFT. Both the BFT variants are analyzed under the Random and Round robin based deflection routing algorithms. The Booksim cycle accurate NoC simulator was augmented with TSV power and delay models, 3D BFT topology generation and simulation capabilities for this study. Accurate wire and TSV lengths were derived from floorplans and plugged into the models for tradeoff studies. The OP3DBFT with round-robin deflection routing delivers up to 44% higher performance and consumes up to 23% lesser power compared to the 3DBFT over various synthetic traffic patters. Over uniform random, transpose and bit-reversal synthetic traffic, OP3DBFT delivers 1.44×, 1.38× and 1.37× better performance compared to 3DBFT. From an Energy perspective, OP3DBFT has an average 23% decrease in Flits-per-Joule, and up to 46% improvement in Energy-Delay-Product compared to the 3DBFT. The BFT variants were synthesized on Xilinx Artix-7 FPGAs for area comparison. OP3DBFT consumes 12% lesser area compared to 3DBFT.

## 1 Introduction

Three-dimensional integrated circuits(3D ICs) are an attractive solution for scalable CMPs and SoCs with the potential to achieve high performance and low power usage. 3D ICs distribute logic and memory in stacked layers and use Through Silicon Vias (TSVs) as vertical interconnects[1]. The Processing Elements (PEs) inside the chip are interconnected using micro-networks called Network-on-Chips (NoCs) [2]. The topology of the NoC defines the arrangement of PEs, through routers, interconnected by links in a latency, bandwidth constrained structure. Topology selection depends on the bandwidth required, resources available, area budget, power and performance constraints of the SoCs. Topologies are classified as regular and irregular. Topologies are evaluated through performance, energy, and cost measures[3].

The most common topology in CMPs and SoCs is the Mesh, as it symmetric in nature and has the shortest wire lengths between routers compared to other topologies[4]. Butterfly Fat Tree (BFT) is a hierarchical, tree based topology having 2.3× less routers compared to the Mesh for a given number of Processing Elements(PEs). In BFT, PEs are placed at the leaves and routers are placed at the top and intermediate levels(Fig. 1). Each router at level $l$, connects 4 nodes (PEs or routers) in the next level($(l+1)$) and 2 routers in the previous level$(l-1)$. Router-4 in level-2 connects to routers 12-15 in level-3 and router 0 and 2 in level-1[5]. In each level of routers (except the last-which connected to PEs), a pair of router connects to 4 router in the next level. Routers(0, 2) connects to (4, 6, 8, 10) and routers (4, 5) connects to (12-15) Once the network topology is selected, the floorplan determines the physical placement of PEs and routers physical dimensions. Floorplan influences the overall area and the length of physical links. Link lengths are driven by the physical dimensions of the components on the chip. Link latency plays a significant role in the overall performance of a NoC. The floorplan based link delay requires the length, the operating frequency, and delay of the interconnect. Using floorplan-based link lengths and corresponding simulation delays in 2D BFT topology leads to an increase in average network latency up to 8× (Fig. 2). Thus, considering physical dimensions during simulation leads to accurate NoC architecture evaluation[6]. Accurate modeling of link delay is necessary in early stage design trade-off studies. State of the-art NoC simulators[7] model the link as a fixed delay component. Incorporating microarchitecture link delay and TSV delay models will enable accurate performance evaluation 3D NoC Topologies.

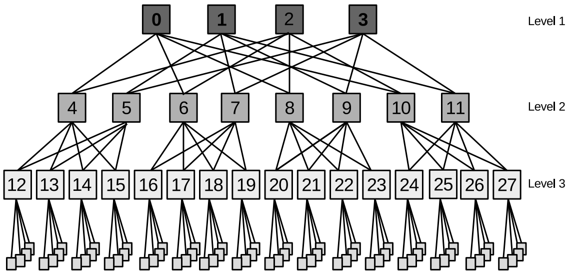


**Fig. 1**: 2D BFT topology

In this work, a low cost, performance optimal 3D BFT(OP3DBFT) is proposed. Firstly, a 64 node 3D BFT is evaluated for traffic flow using two output path selection routing mechanisms: Random Output Deflection(ROD) and Round Robin Output Deflection (RROD). Based on the performance constraints, link and TSV minimization, and power-performance trade-offs, the OP3DBFT is derived. The BookSim simulator has been extended with micro-architectural link delay models, and power models of the TSVs for the evaluation of 3D BFT NoC architecture. OP3DBFT contains 75% lesser TSVs compared to the conventional 3DBFT while having identical performance. The analysis of the BFT variants have been performed in the presence of uniform, transpose, and tornado traffic patterns with low level horizontal link and

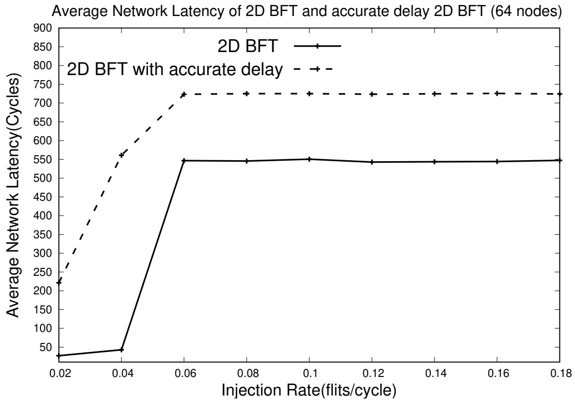


**Fig. 2**: 2D BFT topology network latency comparison of constant and floorplan based link delay.

TSVs delay and power models. The 3D BFT variants FPGA area costs have been evaluated using the HDL output from CONNECT[8]. The syntheses results show the area utilization of the OP3DBFT 12% less compared to 3DBFT on the Vertex-7 FPGA.

This paper is organized as follows. Section 2 discusses the related work and state-of-the-art TSVs delay and power models. Analysis of 2D and 3D BFT topologies with microarchitectural, data serialization, routing and link delay estimation details are presented in Section 3. Section 4 discusses the TSV count minimization. The design of the optimal 3DBFT including the floorplan is proposed in Section 5. Section 6 discusses the experimental setup. Section 7 presents the Results. The paper concludes in Section 8.

## 2 Related Work

This section presents the state-of-the-art in work related to BFT topology design and the TSV latency and power modelling.

### *2.1 BFT Topology Design and Characterization*

In [9], author presented the performance model for wormhole routed interconnection network. Further the performance model is applied Butterfly fat tree topology and compared with simulation results which predicted accurate performance. P Pande[10] proposed switch-based network-centric architecture using butterfly fat tree architecture to connect multiple heterogeneous IP blocks. The switch-based network-centric architecture shows low hardware overhead and latency. Latency, energy dissipation, and wire area overhead of Mesh and BFT variants of 3D NoC architectures are compared with 2D NoC architectures in Feero et al.[11] 3D Mesh topology shows better network latency with a small area overhead over 2D NOC topology. In [12], a new 3D topology derived from the BFT topology. The derived BFT is compared with standard NoC topologies such as the Mesh, torus, butterfly, flattened butterfly for performance. Constant wire delays have been considered in the simulation. Rahmani et al. [13] evaluate a 3 x 3 x 3 3D NoC by replacing a unidirectional TSVs pair into the bidirectional channel, and there is significant inter-layer communication with little performance overhead.

In [14], proposed a 3D Hierarchical Crossbar-based Interconnection Topology (3D-HiCIT) and compared with other hierarchical topologies in terms of flexibility, scalability and performance. The evaluation methodology to compare the performance of the 2D NoC topology such as SPIN, Torus, Folded torus, Octagon, BFT NoC architectures are disused in [2]. In [15], authors have proposed an ultra-optimized interlayer communication for 3D NoC Bus Hybrid architecture to achieve power and performance improvement. In [16], 3D partitioning and floorplanning approach is considered in Clos NoC (CNOC) 3D integration to avoid very long global wires.

Various architectural design space is explored to optimize both performance and energy-efficiency of 2D and 3D NoC architectures. Architectural choices for NoCs include partitioning using multiple sub-networks, concentration and express physical links. Various NoC architectures have been proposed based on design options like Multiple physical networks (P), cores concentration (C), express channels (X), flit widths (W), and virtual channels(V) and evaluated almost all combinations of these design option for power and performance.[17, 18].

BookSim[7] cycle-accurate simulator is configurable and used for performance evaluation of 2D and cube-based NoC architectures. It supports for evaluation of Mesh, CMesh, torus and Fat tree topologies. ORION[19] is a power estimation tool for a network of routers. It has energy model and area models which are used for 2D NoCs power and area evaluation. However, On-chip simulators consider constant delay on the communications, length and the delay of links vary according to the floorplan of the NoC.

### *2.2 Through-silicon via (TSV)*

Various TSV delay and power models have been proposed by considering the microarchitectural details of coupling capacitance(C), resistance(R) and inductance(L) between the vias[20–22]. Weerasekara et al. [20] modelled TSV bundle by deriving reduced electrical circuit models(R, L, C). Ahmed et al. [21] proposed TSVs delay aware floorplanning. It considers the coupling capacitance between adjacent TSVs($C_{TT}$) and the horizontal wire and TSV($C_{TW}$). The propagation delay depends on the driver resistance and TSV capacitance. You et al. [22] used approximate ring oscillator model to characterise TSVs. TSVs detailed microarchitectural parameters, regression method and approximation techniques, RC parasitics and dimensional analysis methods are carried out for evaluation accurate TSVs power[23–25]. RC parasitics of TSVs are modelled as a 3D interconnect with buffers with little delay overhead[24]. The model is created in Khalil et al.[25, 26] is based on TSVs dimensional analysis which takes input parameters namely the TSV length, radius and pitch parameter for accurate TSVs power evaluation.

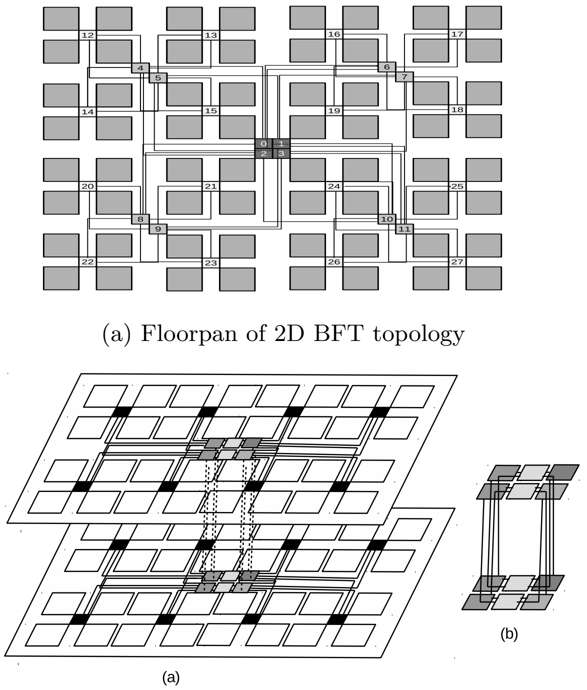
The current work incorporates from Kim et al. model (TSV power and delay calculations) into BookSim[7] as vertical interconnects power(dynamic power estimation), delay model and TSVs data serialisation for evaluation 3D NoCs. i.e. ORION and TSVs delay and power models are incorporated in BookSim for accurate evaluation of 3D BFT NoC architecture The power, performance and FPGA based area utilization of 2D, 3DBFT and proposed 3DBFT topology is characterised in the presence of uniform, transpose and bit-compliment traffic.

## 3 Analysis of 2D and 3DBFT topology

An equivalent 3D BFT topology is constructed from the 2D BFT topology. Floorplans of both the 2D and 3D BFT topologies are shown. TSV Serialization and routing options in the BFT topologies are analyzed.

### *3.1 Floorplanning*

The conventional 64 PE BFT topology is shown in Fig. 1. Except for the top level routers(which have 4 ports each), all routers contains 6 ports. The ports of all the 6-port routers are connected as follows: 4 ports connected to all 4 child



(a) Floorpan of 2D BFT topology

(b) (a) Floorplan of 3DBFT (two-stacked layer) BFT connected using TSVs($8 \times 4 \times 2$). (b) Inter-layer connections.

**Fig. 3**: Floorplan of 2D and 3DBFT topology variants.

nodes, remaining 2-port connect to two parent nodes. Fig. 3 (a) shows the floorplan of the 2D BFT topology. The floorplan consists of a system with a tiled Chip Multiprocessor containing 64 Sun-SPARC cores(area of each core is 3.4mm$^2$) [27]. Router area is estimated from the ORION area module. Table 1 lists some of the microarchitectural parameters used to derive the floorplan. Based on the floorplan, the 2D BFT has five different links lengths. The link lengths are used to estimate the delay of the link for performance evaluation (shown in Table 3).

The 3D BFT floorplan is derived from the 2D BFT floorplan by equally distributing the PEs and the associated routers in 2-layers (Fig. 3 (b)). The level 2 routers are moved closer to level 1 routers to reduce the link length. Eight vertical links, made up of TSV bundles, are shown in the 3DBFT floorplan.

**Table 1** Floorplan parameters

| Parameter | Value |
|---|---|
| PEs area | 3.4mm$^2$ |
| 3-port router area | 0.69mm$^2$ |
| 5-port router area | 1.25mm$^2$ |
| 6-port router area | 1.57mm$^2$ |
| 7-port router area | 1.91 mm$^2$ |
| Channel width | 128 bit |

### *3.2 Through Silicon Via Link Delay Model*

Each signal TSV is paired with a ground TSV as shown in Fig. 4 (a). Fig. 4 (b) shows the electrical model of the signal and ground TSV pair [28]. The propagation delay depends on the dimensions(TSV length, radius and pitch) of TSVs. Khalil et al.[25] uses an analytical model of TSVs to get the TSVs delay, power and valid TSV configuration. An analytical model of the propagation delay of the TSV is shown in Algorithm 1 and TSV delay is depends on the from Height/Length($l$), Diameter($d$) and Pitch/Separation($s$). To avoid the manufacturer complexity during the fabrication process safe limits (safe limits are from [20, 29]) each microarchitectural considered. The safe limits of each parameter are considered to the generated set of TSVs configuration. The TSVs configuration with 20$\mu$m height, the diameter of 20$\mu$m and pitch of 40$\mu$m yields lowest TSV power.

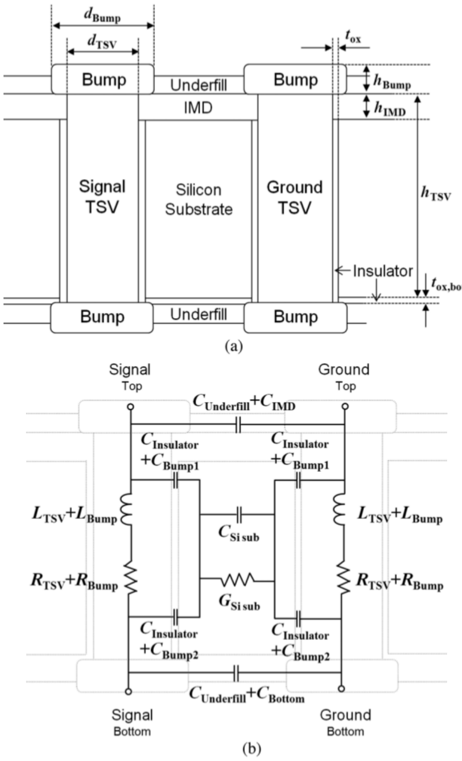


**Fig. 4**: (a) Signal and ground TSV structure( via-last process), and (b) Electrical model of Signal and ground TSV. [28]

**Algorithm 1:** TSV Delay Estimation

```
START
$r = d/2$
$l_o = \frac{\sigma_{Cu} * r^2 * \sqrt{(\mu_o/\varepsilon_{si})} * acosh(s/d)}{0.693*(1+0.617*(r/s))}$
if($l \geq lo$)
    $delay = \sqrt{(\mu_o/\varepsilon_{si})} * l * l/l_o$
else
    $delay = \sqrt{(\mu_o/\varepsilon_{si})} * l$
END
```

### *3.3 Data serialization over TSVs*

Data serialization is used to reduce the area footprint of the TSVs with an additional power overhead. The channel size used in this work is 64 bits. The area footprint of 64-bit TSVs connection adversely effects the overall area of the chip. The lowest yield TSV configuration (h=20$\mu$m, d=20$\mu$m and p=40$\mu$m) has a higher pitch. The area of TSVs is directly proportional to the pitch size. We consider the 2:1 data serialization where the footprint area of TSVs is decreased to half

(64 TSVs to 32 TSVs). Table 2 shows the TSV count, Keep Out Zone (KOZ) and dimensions for non-serialized and 2:1 serialized TSVs. The TSV array dimension reduces to half in the case of 2:1 serialized TSVs. The delay of the TSV is 50ps which is much smaller than 0.4ns (2.5GHz). The TSV delay is considered as one clock cycle throughout out the paper.

**Table 2** Parallel and serial case with the TSVs design parameters and TSV count

| | 1:1 TSVs(per channel) | 2:1 Serialisation TSVs (per channel) |
|---|---|---|
| TSVs count | 64 | 32 |
| KOZ ($\mu$m) | 5 | 5 |
| TSV array dimension ($\mu$m) (TSV+KOZ) | 270 x 640 | 135 x 640 |

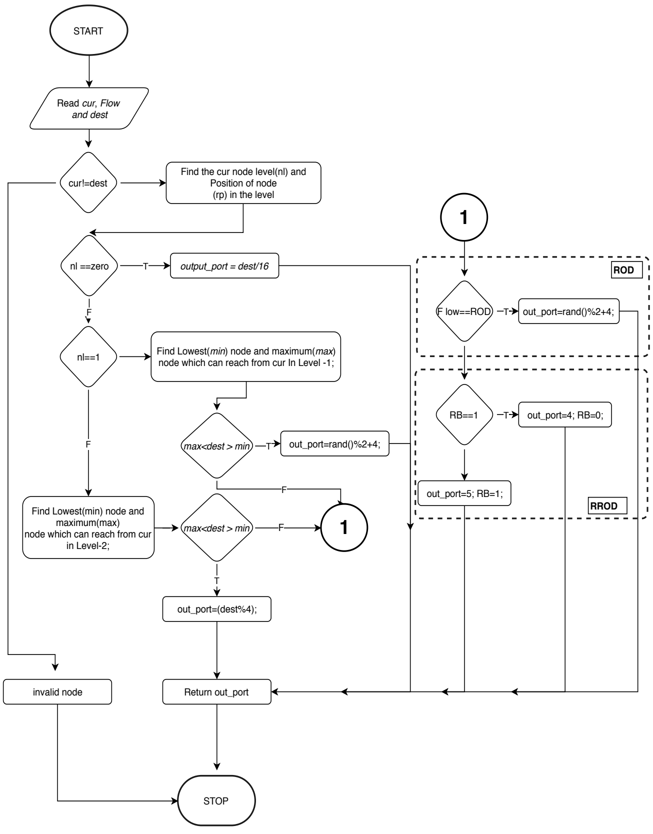


**Fig. 5**: NCA Routing flowchart of BFT topology with ROD and RROD

### 3.4 Nearest Common Ancestor(NCA) Routing in BFT Topology

Fig. 6 (a) shows the example of possible routing paths from source(**node-0**) to destination (**node-32**). The destination (node 32) can be reached from source ( **node-0** ) from two different output paths. Similarly there is always two paths for each packets which ejected from any source. The two different paths are available in Level 3 and Level 2, so alternative paths can be chosen. We analyse random output path selection and round robin output path selection in the NCA algorithm flowchart(Figure 5). Random Output Deflection (ROD) routing is illustrated in Fig. 6(b). ROD is a selection of random output path while sending flits from source to destination(Figure 5). In ROD, selecting same output port for different packets leads to additional latency due to contention. In the Round Robin Deflection (RROD), the output path is selected in round-robin order (Fig. 6 (c)). The alternative output path selection helps in balanced traffic on links. Figure 5 depicts NCA flowchart algorithm with both ROD and RROD path selection mechanisms.

### 3.5 Link Delay Estimation

The floorplan based link delays are estimated using ORION RC delay models (2.5GHz frequency) for wires and TSVs delay, power and valid TSV configuration using Khalil et al.[25] model, which takes three parameters, namely TSV length, radius and pitch. We generated an ideal TSV configuration by combining these models by considering safe limits for each parameter to avoid the manufacturer complexity during the fabrication process. TSV delay depends on the Height/Length($l$), Diameter($d$) and Pitch/Separation($s$).

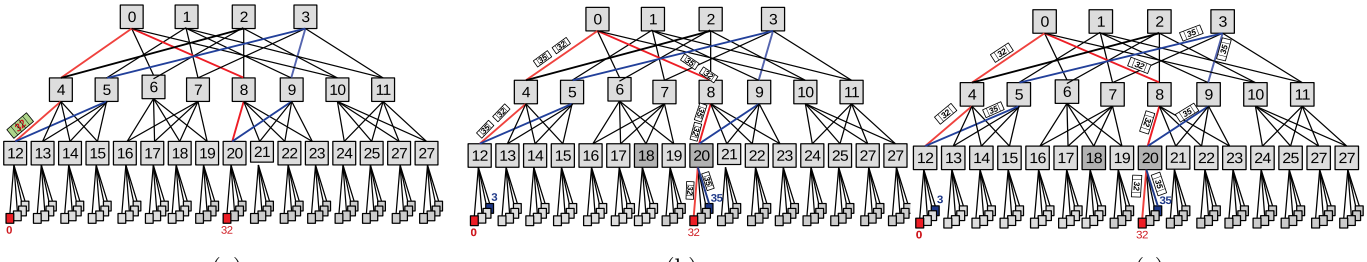
Table 3 shows the delays of the horizontal wires and the TSVs in 2D and 3DBFT based on the flooplans (Fig. 3). Each vertical link contains 32-TSVs (2:1 serialisation) TSVs count and TSVs delay of each vertical link is depicted in last two columns of Table 3.

## 4 Power and Performance Optimal OP3DBFT

The utilization of TSVs in a conventional 3DBFT are analyzed under synthetic traffic patterns. The TSVs with the least utilization are removed under performance constraints. The optimal 3DBFT (OP3DBFT) topology and floorplan are presented.

### 4.1 TSV Count Minimisation

The vertical links from Fig. 3(b) are marked L1 - L8 in Fig. 6. The channel width of each link is 64-bit. A 64-bit TSV channel, with a 64 pairs of signal and ground TSVs result in a prohibitive area of 0.1728$\mu$m$^2$.The reduction in the TSV area through 2:1 serialisation was presented in Section 3.3.



(a) (b) (c)

**Fig. 6**: 2D BFT topology (a) Two possible output paths for from **node 0** to **node 32** with red and blue colour links. (b) ROD routing for from **node 0** to **node 32** and **node 3** to **node 35** with red and blue colour links respectively (c) RROD for from **node 0** to **node 32** and **node 3** to **node 35** with red and blue colour links respectively

**Table 3** Link length and delay details of BFT topologies variants.

| Topology | Wire (mm) | Delay (clock cycle) | Number of TSVs (32-TSVs per link) | Delay (Clock Cycle) |
|---|---|---|---|---|
| 2D BFT | 9.376 | 92 | - | - |
| | 8.976 | 85 | - | - |
| | 4.4889 | 21 | - | - |
| | 4.088 | 18 | - | - |
| 3DBFT | 8.176 | 68 | 256 | 1 |
| | 7.776 | 63 | | |
| | 4.088 | 18 | | |
| | 3.688 | 14 | | |
| | 1mm | 1 | | |

Table 4 depicts the 3DBFT topology vertical links(L1-L8) utilisation for uniform, transpose and bit-reversal traffic.

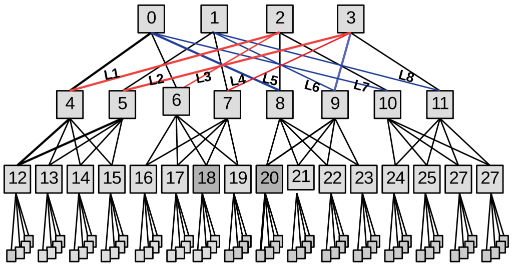


**Fig. 7**: 2D BFT L1 to L8 links are vertical interconnect for 3DBFT topology with red and blue colour links.

**Table 4** Links utilisation (Injection rate=0.018) of 3DBFT (8-vertical links (TSVs)) and OP3DBFT (2-vertical links(TSVs)) for uniform , transpose and bit-reversal. Average utilisation of 3DBFTis 40% and OP3DBFT is 80%.

| Traffic | 3DBFT | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| pattern | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
| Uniform | 0 | 43.062 | 21.619 | 21.463 | 43.673 | 43.032 | 43.679 | 43.742 |
| Transpose | 65.128 | 65.092 | 43.402 | 43.524 | 43.648 | 43.509 | 43.242 | 43.737 |
| bit-reversal | 65.128 | 65.092 | 43.402 | 43.524 | 43.648 | 43.509 | 43.242 | 43.737 |

**Table 5** Links utilisation (Injection rate=0.018) of OP3DBFT (2-vertical links (TSVs)) for uniform , transpose and bit-reversal.

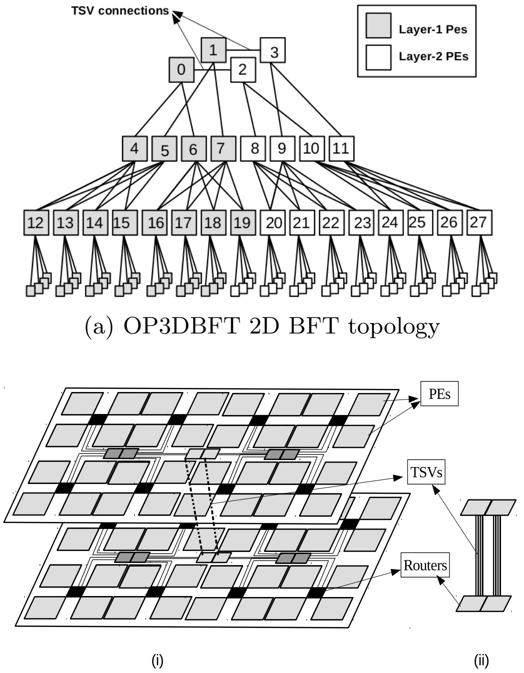
| Traffic | OP3DBFT | |
|---|---|---|
| pattern | L1 | L2 |
| Uniform | 83.759 | 84.419 |
| Transpose | 84.303 | 84.581 |
| bit-reversal | 83.968 | 83.968 |

The average utilisation for 3DBFT is 32%, 47%,48% for uniform, transpose and bit-reversal traffic respectively. An average 50% links(TSVs) are under-utilised in the 3DBFT topology. We attempt to reduce the number of vertical links without affecting the overall performance of the BFT.

Fig. 8 (a) shows the modified 3DBFT topology, with only 2 vertical links (reduced from 8). One TSV link is removed from Level-1 routers, thereby reducing the degree of the routers from 4 to 3. The overall connectivity has not been altered. Applying TSV serialisation and TSV count minimisation, the OP3DBFT is proposed. The topology and floorplans are in Fig. 8. Table 5 lists the link utilisation of L1 and L2 links for uniform, transpose and bit-reversal traffic patterns. The average link utilisation for OP3DBFT is 84%, 85%, 85% for uniform, transpose and bit-reversal traffic respectively.

### *4.2 OP3DBFT - Topology and Floorplan*

Fig. 8 (a) shows modified 3DBFT topology where each layer consists of 32 PEs each. Level-1 routers have a degree of 3 - one output port of each router connects the next odd router in level-0 (vertical interconnection) and two output ports are connected to level-1 routers (horizontal links). Fig. 8(b) shows the floorplan of the OP3DBFT. In modified 3DBFT, six links(vertical interconnect) reduced to two as compare to 3DBFT(Fig. 3 (b)). Based on OP3DBFT floorplan, there only two different link lengths(4.088mm and 3.688mm) and the delay of both links is 13 clock cycles. The OP3DBFT has up to 80% lesser TSVs, 75% lesser TSV area compared to the regular 3DBFT.



(a) OP3DBFT 2D BFT topology

(b) (i) 8 x 4 x 2 2-layer OP3DBFT with two stacked layers. (ii) Inter-layer(TSVs) connections.

**Fig. 8**: OP3DBFT - Topology and Floorplan.

## 5 Experimental Setup

The BookSim simulator was extended to support 3D NoC by adding (a) TSV delay and power modules for vertical links, (b) Orion power and delay modules for horizontal links as shown in Fig. 8. The floorplan module takes as input the topology, PE size and router area to output the lengths of the links. These parameters are passed to link delay and power module. Link delay module calculates the delay of individual horizontal and vertical links. The horizontal link ($T_{D\_H}$) delay is calculated using ORION, and vertical link

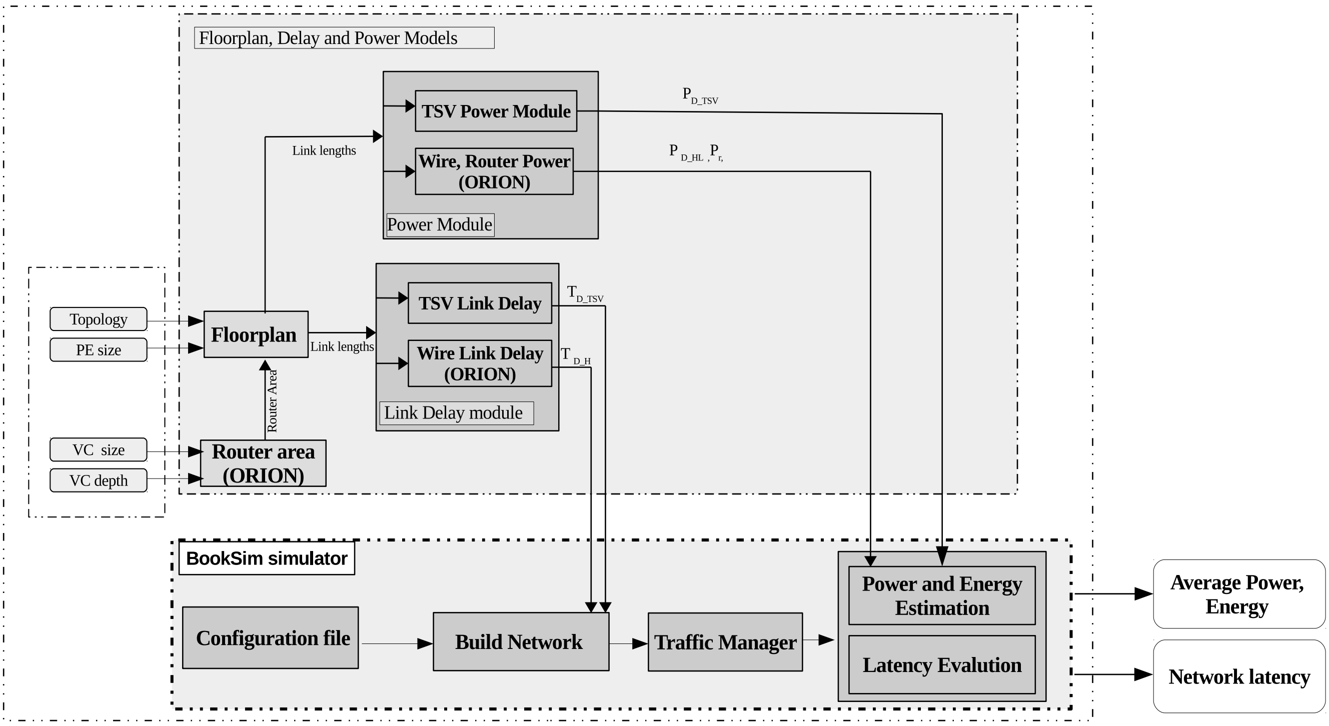


**Fig. 9**: Simulation framework for evaluating power and performance. BookSim was extended with 3D TSV delay, power and link delay modules.

delay($\mathrm{T_{D\_TSV}}$) is calculated from TSV delay module. The delay of individual links is passed to the simulator to create topology(build network). Links (horizontal wire and verticals) delay were modelled(as described in the Section 3.5). The 3DBFT topologies as shown in Fig. 3 and 8 are implemented in simulator. Nearest common ancestor algorithm with ROD and RROD routing implemented for 64 nodes. The routing functionality and BFT network topology tested and implemented in BookSim simulator. Power module takes the links length and router details to calculate the accurate power details. The vertical links power($\mathrm{T_{D\_TSV}}$) is calculated using the TSV power module and the router ($\mathrm{P_r}$) and horizontal links power($\mathrm{P_{D\_H}}$) are calculated using ORION. The accurate power details are used when the transfer of traffic starts. The topologies simulated in BookSim are 2D BFT, 3DBFT and OP3DBFT with network size 64-nodes. There are 28 routers with a 8 VCs per port with a VC buffer depth of 16. The simulation time is of $10^5$ cycles.

## 6 Results and Discussion

### *6.1 Performance Analysis*

The performance comparison of 3DBFT and OP3DBFT for uniform, transpose and bit-reversal traffic is shown in Fig. 10. The OP3DBFT has a performance improvement of up to 1.54×, 1.38×, and 1.37× compared to 3DBFT uniform, transpose and bit-reversal traffic respectively. The improved performance because there is reduction of up to 75% of TSV count i.e 6 vertical links have been reduced compared to the regular 3DBFT.

### *6.2 Energy Analysis*

Energy(Joules) per flits (JpF) is calculated using Equation 1, $\mathrm{F_t}$ is the total number flits delivered throughout the simulation and T is the total simulation in the cycle.

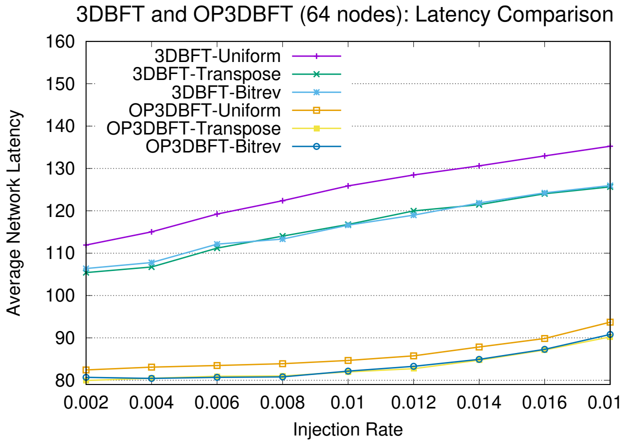


**Fig. 10**: Latency comparison of 2-layer and OP3DBFT topology for uniform, transpose and bit-reversal traffic.

$$JpF = \frac{(P_{\mathrm{t}} * T)}{F_{\mathrm{t}}} \qquad (1)$$

Total Power consumption is calculated as the sum of the powers of links and routers. Equation 2 depicts the total power consumption. $\mathrm{P_t}$ is the total power, $\mathrm{P_r}$, $\mathrm{P_l}$, $\mathrm{P_{tsv}}$ are powers of the routers, links and TSVs respectively.

$$P_t = P_r + P_l + P_{tsv} \qquad (2)$$

$\mathrm{P_{tsv}}$ was obtained using Equation 3, where AF is the activity factor, $\mathrm{C_{TSV}}$ is the TSV capacitance calculated from the Equation 4. V and f are the Voltage and the operating frequency respectively[24].[26].

$$P_{\mathrm{tsv}} = AF \,.\, C_{\mathrm{tsv}} \,.\, V^2 \,.\, f \qquad (3)$$

In Equation 4, The electrical permittivity of the silicon substrate is $\epsilon$, and conductivity of the silicon substrate, $\sigma$. From Fig. 4 (b), sum of insulator and bump2 capacitance is $C_1$, $C_2$ is the silicon substrate capacitance and $C_3$ is the under-fill capacitance.

$$C_{\text{tsv}} = C_3 + \frac{C_1 * C_2 * (1 + \sigma_{\text{Cu}} / (\varepsilon_{\text{si}} * \omega))}{C_1 + 2 * C_2 * (1 + \sigma_{\text{Cu}} / (\varepsilon_{\text{si}} * \omega))} \tag{4}$$

Fig. 11 shows the Joules per flit of OP3DBFT and regular 3DBFT variants for uniform, transpose and bit-reversal traffic. From the results,OP3DBFT has average 23% decrease in JPF compared to regular 3DBFT. The JPF in OP3DBFT has decreased up-to 23%, 22% and 21% in uniform, transpose and bit-reversal traffic respectively.

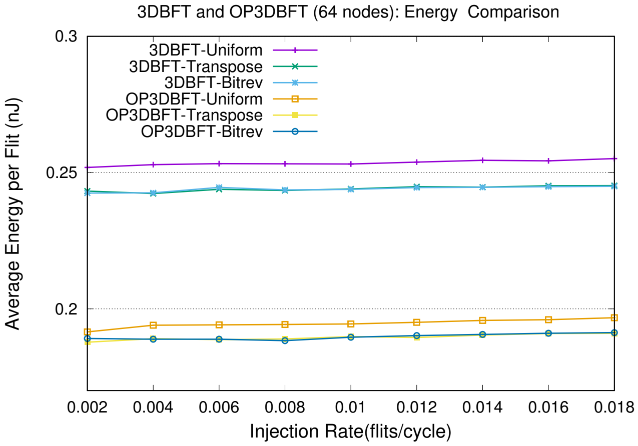


**Fig. 11**: Energy per flit comparison of 2-layer and OP3DBFT topology for uniform, transpose and bit-reversal traffic.

### *6.3 Energy Delay Product (EDP)*

Fig. 12 shows the Normalised EDP of OP3DBFT and regular 3DBFT variants for uniform and transpose traffic pattern. The transpose and bit-reversal traffic has average reduction of EDP up to 10% and 11% compared to uniform traffic pattern in 3DBFT topology throughout the simulation(Fig. 12)as BFT is suited for localised traffic rather than uniformly distributed traffic. The EDP is the product of average network latency and average Energy per flit. The OP3DBFT shows 46%, 44% and 44% reduction in EDP compared 3DBFT uniform, transpose and bit-reversal traffic. Overall, the OP3DBFT's EDP is lower than 3DBFT because there is a reduction in OP3DBFT's latency (Fig. 10) and energy (Fig. 11) compared to 3DBFT.

### *6.4 Area Utilization*

The BFT topologies were implemented using CONNECT, a web-based NoC generator tool[8]. The HDL models of OP3DBFT was obtained from modifying the BFT HDL models. The synthesis results were obtained using Xilinx Vivado. Xilinx Artix-7 XC7A200T FPGA board has used to analyse the FPGA resource utilization and Table 6 shows the detailed synthesis results. From Table 6, it can be seen that the regular 3DBFT topology consumes 1.12% more LUTs than Optimal Power and performance 3DBFT topology. The proposed topology has 12% reduction in area compared to regular BFT topology without compromising in performance.

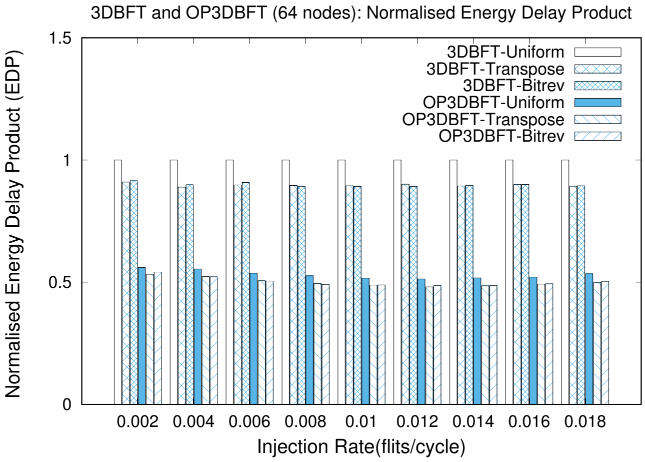


**Fig. 12**: Normalised EDP of regular 3DBFT and OP3DBFT for uniform, transpose and bit reversal traffic.

**Table 6** Synthesis results of 3DBFT Topology variants

| H/W utilisation (%) | 3DBFT | OPP3DBFT |
|---|---|---|
| LUTs | 54.6 | 50.04 |
| FFs | 10.47 | 9.72 |
| Freq | 100 MHz | 100 MHz |

## 7 Conclusion

A novel, low cost, power-performance optimal 3D BFT topology (OP3DBFT) is proposed. OP3DBFT is evolved from the standard 3D BFT after eliminating extraneous TSV links under a performance constraint. The utilization of links in 3DBFT is analyzed under the uniform, transpose and bit-reversal traffic. The regular 3D BFT and the OP3DBFT employ 2:1 serialization to reduce the area footprint of the TSVs links. Two path selection schemes, the round-robin output(RROD) and the random(ROD) selection, based on the Nearest Common Ancestor routing are used to evaluate the performance of the BFT topologies. State-of-the-art TSV delay and power models have been incorporated into the BookSim simulator. The delays of horizontal wires are derived from ORION delay models

OP3DBFT with RROD is optimal area, power and performance 3DBFT architecture compared to regular 3DBFT. Using RROD results in a more balanced distribution of traffic among links. The OP3DBFT shows up to 1.44×, 1.38× and 1.37× performance improvement compare to 3DBFT uniform, transpose and bit-reversal traffic respectively. OP3DBFT performs better due to its modified structure(75% of TSV count reduction). The Joules Per Flit(JPF) in OP3DBFT has decreased up-to 23%, 22% and 21% in uniform, transpose and bit-reversal traffic respectively. The EDP of OP3DBFT shows 46%, 44% and 44% reduction compared 3DBFT uniform, transpose and bit-reversal traffic respectively. Based on the synthesis results, OP3DBFT consumes 12% lower area compared regular 3D BFT topology.

The OP3DBFT shows up to 1.44×, 1.38× and 1.37× performance improvement compare to 3DBFT uniform, transpose and bit-reversal traffic respectively. OP3DBFT performs better due to its modified structure(75% of TSV count reduction). The Joules Per Flit(JPF) in OP3DBFT has decreased up-to 23%, 22% and 21% in uniform, transpose and bit-reversal traffic respectively. The transpose and bit-reversal traffic has average reduction of EDP up to 10% and 11% compared to uniform traffic pattern in 3DBFT topology as

BFT is suited for localised traffic rather than uniformly distributed traffic. The EDP of OP3DBFT shows 46%, 44% and 44% reduction compared 3DBFT uniform, transpose and bit-reversal traffic respectively. Based on the synthesis results optimal power and performance 2-layer 3D topology consumed 0.12% lower area compared regular 3D BFT topology.

## 8 References

1 D. H. Kim, K. Athikulwongse, and S. K. Lim, "A study of through-silicon-via impact on the 3d stacked ic layout," in *Proc. of the 2009 International Conference on Computer-Aided Design*. ACM, 2009, pp. 674–680.

2 P. P. Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh, "Performance evaluation and design trade-offs for network-on-chip interconnect architectures," *IEEE Transactions on Computers*, vol. 54, no. 8, pp. 1025–1040, 2005.

3 W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Design Automation Conference, 2001. Proceedings*. IEEE, 2001, pp. 684–689.

4 S. Kumar, A. Jantsch, J. P. Soininen *et al.*, "A network on chip architecture and design methodology," in *Proc. IEEE Computer Society Annual Symp. on VLSI. ISVLSI 2002*, 2002, pp. 105–112.

5 C. Grecu, P. P. Pande, A. Ivanov, and R. Saleh, "A scalable communication-centric soc interconnect architecture," in *International Symposium on Signals, Circuits and Systems. Proc., SCS 2003. (Cat. No.03EX720)*, 2004, pp. 343–348.

6 B. Halavar, U. Pasupulety, and B. Talawar, "Extending booksim2.0 and hotspot6.0 for power, performance and thermal evaluation of 3d noc architectures," *Simulation Modelling Practice and Theory*, p. 101929, 2019.

7 N. Jiang, D. U. Becker *et al.*, "A detailed and flexible cycle-accurate network-on-chip simulator," in *Performance Analysis of Systems and Software (ISPASS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 86–96.

8 M. K. Papamichael, "Fast scalable fpga-based network-on-chip simulation models," in *Ninth ACM/IEEE International Conference on Formal Methods and Models for Codesign (MEMPCODE2011)*, July 2011, pp. 77–82.

9 R. I. Greenberg and L. Guan, "An improved analytical model for wormhole routed networks with application to butterfly fat-trees," in *Proceedings of the International Conference on Parallel Processing*, ser. ICPP '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 44–48. [Online]. Available: http://dl.acm.org/citation.cfm?id=645533.656511

10 P. P. Pande, C. Grecu, A. Ivanov, and R. Saleh, "Design of a switch for network on chip applications," in *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03.*, vol. 5, May 2003, pp. V–V.

11 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.

12 A. Bose, P. Ghosal, and S. P. Mohanty, "A low latency scalable 3d noc using bft topology with table based uniform routing," in *2014 IEEE Computer Society Annual Symposium on VLSI*, July 2014, pp. 136–141.

13 A. Rahmani, P. Liljeberg, J. Plosila, and H. Tenhunen, "Bbvc-3d-noc: An efficient 3d noc architecture using bidirectional bisynchronous vertical channels," in *2010 IEEE Computer Society Annual Symposium on VLSI*, July 2010, pp. 452–453.

14 M. Debora, P. Max, K. Marcio, C. Luigi, and S. Altamiro, "Performance evaluation of hierarchical NoC topologies for stacked 3D ICs," *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2015-July, pp. 1961–1964, 2015.

15 A. M. Rahmani, P. Liljeberg, J. Plosila, and H. Tenhunen, "Lastz: An ultra optimized 3d networks-on-chip architecture," in *2011 14th Euromicro Conference on Digital System Design*, Aug 2011, pp. 173–180.

16 A. Zia, S. Kannan, H. Jonathan Chao, and G. S. Rose, "3d noc for many-core processors," *Microelectron. J.*, vol. 42, no. 12, pp. 1380–1390, Dec. 2011.

17 A. Psathakis, V. Papaefstathiou *et al.*, "A systematic evaluation of emerging mesh-like CMP NoCs," in *Architectures for Networking and Communications Systems (ANCS), 2015 ACM/IEEE Symp. on*. IEEE, 2015, pp. 159–170.

18 A. Psathakis, V. Papaefstathiou, M. Katevenis, and D. Pnevmatikatos, "Design space exploration for fair resource-allocated noc architectures," in *2014 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XIV)*. IEEE, 2014, pp. 141–148.

19 A. B. Kahng, B. Li, L.-S. Peh, and K. Samadi, "Orion 2.0: a fast and accurate noc power and area model for early-stage design space exploration," in *Proc. of the conference on Design, Automation and Test in Europe*. European Design and Automation Association, 2009, pp. 423–428.

20 R. Weerasekera, M. Grange, D. Pamunuwa, H. Tenhunen, and L.-R. Zheng, "Compact modelling of through-silicon vias (tsvs) in three-dimensional (3-d) integrated circuits," in *2009 IEEE International Conference on 3D System Integration*. IEEE, 2009, pp. 1–8.

21 M. A. Ahmed, S. Mohapatra, and M. Chrzanowska-Jeske, "Tsv- and delay-aware 3d-ic floorplanning," *Analog Integr. Circuits Signal Process.*, vol. 87, no. 2, pp. 235–248, May 2016.

22 J. You, S. Huang, Y. Lin, M. Tsai, D. Kwai, Y. Chou, and C. Wu, "In-situ method for tsv delay testing and characterization using input sensitivity analysis," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 3, pp. 443–453, March 2013.

23 C. Jueping, J. Peng, Y. Lei, H. Yue, and L. Zan, "Through-silicon via (tsv) capacitance modeling for 3d noc energy consumption estimation," in *2010 10th IEEE International Conference on Solid-State and Integrated Circuit Technology*, Nov 2010, pp. 815–817.

24 D. H. Kim and S. K. Lim, "Through-silicon-via-aware delay and power prediction model for buffered interconnects in 3d ics," in *Proceedings of the 12th ACM/IEEE International Workshop on System Level Interconnect Prediction*, ser. SLIP '10. New York, NY, USA: ACM, 2010, pp. 25–32.

25 D. Khalil, "Analytical model for the propagation delay of through silicon vias," in *9th International Symposium on Quality Electronic Design (isqed 2008)*, March 2008, pp. 553–556.

26 J. Kim, J. Cho, J. S. Pak, T. Song, J. Kim, H. Lee, J. Lee, and K. Park, "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.

27 T. C. Xu, P. Liljeberg, and H. Tenhunen, *A Greedy Heuristic Approximation Scheduling Algorithm for 3D Multicore Processors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

28 J. Kim, J. S. Pak, J. Cho *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.

29 M. Lee, J. S. Pak, and J. Kim, *Electrical Design of Through Silicon Via*. Springer Publishing Company, Incorporated, 2014.