

Design and FPGA Implementation of FREDO-3D-NoC for Low Power and High Throughput and its Protocol Interfacing

¹Sujata S. B. and ²Anuradha M. Sandi

Submitted: 20/10/2023

Revised: 20/12/2023

Accepted: 29/12/2023

Abstract: In the past 10 years, Network-On-Chips (NOC) is likely offered solution for future systems on chip design. It gives huge scalability compared to the shared bus-based interconnection and permits more processors to run concurrently since NOC has a custom built wires, we can able to foretell the performance.

Objectives: Within this frame of reference, using Flexible Routing for exact direction order (FREDO) for 5x5 with three layers and 3x3 with three layers we recommended a 3D- NoC for buffer and bufferless. Using Wormhole switching and a Stall-and-Go flow control scheme for buffer and bufferless, the mesh topology is used by the recommended design. Despite of having the advantages of FREDO-NoC over the shared bus system, it has got few constraints, like low throughput, high-cost communication and high-power consumption. To avoid these constraints, we recommend a 3D-NoC (3D FREDO-NoC), the expansion of our 3D FREDO-NoC.

Methods: In this document, the 3D FREDO -NoC architecture is narrated in detail and the output of preliminary evaluation. To meet the prerequisites of performance, power and area, it is required to design routers. This is the core of the network on chips. Few techniques lead to improve the number of buffers to upgrade performance, but it is also liable for a higher portion of the router power and area. The performance of the Router is improved in terms of upgrading the saturation rate by maintaining the constant amount of buffers as the Base Router in the Flexible Router architecture. Findings: Furthermore, it is evident that the flexible Router outperforms the base router in throughput by 23.1%, latency by 31.5% and 9% increase in saturation point for uniform random traffic at higher injection rates.

Novelty: In this document, to prove the functionality and judge the performance compared to the base router it focuses on hardware implementation and evaluation of flexible Router. Here we utilize the Verilog HDL on the considering uniform, hotspot traffic patterns and neighbour to get the cycle precise NoC Simulation system.

Keywords: Wormhole switching algorithm, 3D-NoC, for exact direction order algorithm Flexible Routing, and Buffer and Bufferless, Virtual Cut through (VCT)

1. Introduction

The consistent lowering of transistor geometry is critical to the development of a workable communications fabric for the latest multi-processor system-on-chip (MPSoC) designs in NoC [1]. Following Moore's Law, chip complexity and capacity have steadily increased over the previous few decades. The functions that over the previous ten years have made up a board-level system can now be smoothly merged into a single chip in modern system-on-chip (SoC) designs. Furthermore, it is impractical to start from scratch because to the extraordinary levels of complexity

reached by SoCs. Time-to-market demands demand the quick and dependable integration of several reusable intellectual property (IP) building elements. In place of conventional hierarchical point-to-point connections and bus systems, the on-chip network architecture offers an integrated interface that makes it simple to add new IP blocks to a system. Essentially, an on-chip network communication fabric houses a modern MPSoC, which is essentially a communication-centric system [2]. Almost all NoCs have synchronous networks since their constituent parts are synchronized by global or equivalent clocks. Global clocks and sophisticated electrical design automation (EDA) tools that enable timing assumptions make this synchronization easier. Despite ongoing design issues, these synchronous NoCs are renowned for their speed and area efficiency [3]. But diverse networks need assistance. An MPSoC is a heterogeneous system in which network nodes represent IP blocks with different hardware structures and functions, in

¹Research Scholar, Department of Electronics and Communication, GNDEC, Bidar, VTU Belgaum Karnataka, sujata.bidreddy@gmail.com

²Professor, Department of Electronics and Communication, GNDEC, Bidar Karnataka,

anu29975@gmail.com

contrast to multi-processor (CMP) systems, in which all network nodes are homogeneous processor units. Different working voltages, area sizes, and clock frequencies are provided for the IP blocks, which are then tested. The latency performance of synchronous networks is affected by these changes, which make it more difficult to define the network topology and achieve chip timing closure [4].

Transistor geometries that are continuously reducing are essential for the construction of a workable communications fabric in state-of-the-art multi-processor system-on-chip (MPSoC) designs in network-on-chip (NoC) [1]. Following Moore's Law, chip complexity and capacity have increased steadily over the previous few decades. The functions that over the previous 10 years have made up a board-level system can now be smoothly merged into a single chip in modern system-on-chip (SoC) designs. Furthermore, it is impractical to start from scratch because of the extraordinary levels of complexity reached by SoCs. It becomes imperative to integrate a wide range of reusable intellectual property (IP) blocks quickly and reliably in order to meet time-to-market requirements. In place of conventional hierarchical point-to-point connections and bus systems, the on-chip network architecture offers an integrated interface that makes it simple to add new IP blocks to a system. Based on an on-chip network communication fabric, a contemporary MPSoC is essentially a communication-centric system [2]. Nearly all NoCs belong to the category of synchronous networks since all of their components are synchronized by global or comparable clocks. Global clocks and sophisticated electrical design automation (EDA) tools that enable timing assumptions are responsible for this synchronization. These synchronous NoCs have a reputation for being fast and requiring little space, even with all of the ongoing design issues [3]. Support for heterogeneous networks, however, becomes essential. An MPSoC is a heterogeneous system where network nodes are thought of as IP blocks with different functionalities and hardware structures, in contrast to multi-processor (CMP) systems, where all network nodes are homogeneous processing units. Different working voltages, area sizes, and clock frequencies are provided and tested for IP blocks. These variances complicate the process of creating the network architecture, which in turn hinders the latency performance of synchronous networks and makes chip timing closure challenging [4].

- ❖ Less power utilization. Since SoC determines the maximum standby time of a handset device, it is important to reduce the power utilization. The significant amount of energy is utilized by a clock tree of synchronous on-chip networks [79] and by shrinking transistor geometry it is getting aggravated.
- ❖ The future sub-micron VLSI designs can be affected significantly by tolerance to variations, temperature, process and voltage fluctuation [72,74].

By 2024, 32% of the delay uncertainty resulting from divergence in the sign-off timing closure will have occurred [60], according to the international semiconductor technology plan. Traditional static timing analysis takes the place of statistical timing analysis [14] in order to cope with declining yield rates and unduly conservative timing estimates. Synchronous on-chip networks mitigate this effect by taking task-mapping procedure volatility into account [74]. However, this is only effective in networks that are homogeneous; otherwise, the routers will continue to operate at the poorly anticipated speed. Despite synchronous on-chip networks, asynchronous networks are ideal for addressing the aforementioned problems. Clock-less asynchronous circuits are used in the construction of the communication components of the asynchronous on-chip network. Data transmission occurs in accordance with certain handshake protocols, which may not be responsive to delays [5]. The interface between all IP blocks to the global asynchronous on-chip network is merged by the identical synchronous to/from the asynchronous interface due to this delay insensitivity. The asynchronous network isolates all synchronous blocks which makes easier for chip-level timing closure. Also, an asynchronous on-chip network is naturally tolerant to all the fluctuations and appreciations to the delay insensitivity. The function of these handshake protocols cannot affect by the deviation caused by the delay uncertainty. At least, asynchronous circuits not required clock, When no data is transmitted the asynchronous on-chip network utilizes zero dynamic power. Although, many of the asynchronous networks are slower than synchronous on-chip networks having uniform resources and structures [8]. However the synchronous circuits of the global clock is power consuming, for synchronizing combinational operations it uses an approach of speed and area efficient method. To control data circulation asynchronous courses depends on handshake protocols. To ensure the delay insensitivity the combinational functions are explicitly identified and

guarded [9]. The circuits utilized in identifying combinational functions presents the area and speed overhead. Delay insensitive asynchronous courses are basically torpid [10]. In asynchronous on-chip networks, implementing TDM structures without any use results in additional completion identification circuits and speed penalties. It is impossible to prevent the complete identification in speed penalty, though. The potential applications of asynchronous courses stem from the delay-insensitive handshake protocols. In asynchronous circuits the scale of the synchronization is limited to lower transmission units, like single pipeline. So only the speed penalty is alleviated. The question is that with such controlled synchronization how we can build asynchronous networks. The solution projected in this document is spatial parallelism. In asynchronous circuits the TDM is not a promising approach since it will bring additional synchronization and compromises speed. If the synchronization is restricted to lower scale, these pipelines are controlled and distributed. Put differently, the unsynchronized low-level components are partially spitted in the communication resources. Also the speed penalty of synchronization is minimized [11].

2. Related Work

The provided NoC is a thorough design method for identifying energy-efficient (near-field inductive coupling) NFICs in order to expand the applications of NICs within the framework of a trustworthy and effective NoC architecture. Statistical link analysis is employed in the provided design framework to determine the optimal NFIC-link arrangement. It is far more efficient in terms of energy efficiency and area overhead. We project that 3-D NoC's with NFIC-enabled linkages outperforms TSV equivalents in performance. Moreover, to encounter electro-migration and workload causes stress problems, all of the reliability of NFIC and TSV is enabled. NoC is an interconnect network in which cores exchange packets using routers as part of a packet-switching mechanism. A strong foundation for design techniques in interconnection networks has been established by extensive research conducted over the course of more than two decades. This includes the operating system services, software, hardware communication infrastructure, CAD tools for NoC synthesis, NoC testing, and a number of other aspects within the growing field of NoC research and development. Key components of

NoC design, including application mapping to NoC, communication infrastructure design, communication methodology, and assessment framework, are covered in detail in this lesson. It also draws attention to new issues in the field of NoC, such as 3D-NoC design, testing, reconfiguration, and synthesis [2] [3]. Compared to planar connections, the use of Through-Silicon Vias (TSVs) in symmetric 3-D mesh NoCs shows minimal bandwidth usage and hardly ever causes contention sites. In response, we suggest the TSV sharing (TS) technique, which allows neighboring routers to share vertical channels in a time division multiplexing fashion, saving TSVs in 3D network overlay controllers. Through a design space search, this study investigates different choices for TS implementation and shows how TS improves TSV-effectiveness (TE) in multicore CPUs. Experiments verify that the suggested technique improves TE with low performance overhead and provide a thorough study of TS's effect on all system levels [3].

In [4] mitigate the negative impacts of process variation, the intra-router stages and inter-router linkages are distributed throughout the tiers in the suggested design and optimization technique. The evaluated NoC architecture raises the EDP by an average of 27.4% across all benchmarks when compared to the process-oblivious design [4]. Frequency-aware adaptive routing (FAAR) and thermal-aware frequency scaling policy (TFSP) are two of the novel temporal and spatial management techniques that were presented. Based on anticipated temperature variations, TFSP dynamically modifies the Data Flow Switch Box (DFSB) frequency, proactively controlling data flow for effective heat dissipation. To demonstrate the success of this strategy, it is compared to global throttling and downward routing thermal management schemes in a 4x4x4 3-D NoC-bus architecture. According to experimental results, our suggested method satisfies a temperature constraint of 378.15 K while outperforming the other two systems by 24 and 56.2 percent in throughput and 33.1 and 45.7 percent in latency, respectively [5]. It is important to note that there could be large variances in the 3-D on-chip thermal profile and the IR-drop distribution in the Power Delivery Network (PDN).

Our innovative approach to system-level application-specific co-synthesis at the design time is based on the clever allocation of communication

and computing resources on a die to suit a given workload. The main goal is to improve the 3-D Power Delivery Network (PDN) architecture and reduce NoC and chip-cooling power in the context of a microfluidic cooling-based application-specific 3-D Multiprocessor System-on-Chip (MPSoC). This approach strives to meet thermal constraints and optimal performance goals. Our results show that the proposed 3-D NoC-PDN co-synthesis framework not only meets PDN design goals but also exhibits superior overall optimality, with an up to 35.4 percent improvement in solution quality over a probabilistic metaheuristic-based co-optimization approach [6]. This is in contrast to previous 3-D NoC synthesis approaches. 3D-NoC is becoming more and more popular among designers because of its scalability, increased bandwidth, fault tolerance, and dependability. The suggested method is modeled through the Access Noxim simulator, with simulation parameters supplied by Tezzaron Semiconductors and Global Foundries. Applications based on Application Traffic-Aware Routing (ATAR) enhance the temperature distribution and on-chip traffic uniformity, according to modeling and experimental results. In order to validate simulations, different traffic patterns are investigated, and specific 3-D design platforms are created to demonstrate thermal optimization in high-performance 3-D NoC-based parallel processing systems [7]. A thorough floor-planning method is necessary when using 3-D stacking of electrical and optical layers in the context of optical ONoCs. The physical mapping approach, which takes into consideration floorplanning, placement, and routing constraints in a 3-D-stacked environment, is supported in this article for wavelength-routed ONoC topologies. Therefore, based on their physical design flexibility, the research compares the signal-to-noise ratio and power efficiency of ring-based vs. filter-based wavelength-routed topologies [8].

An isolation-and-check process is suggested in [9] to improve the method's localization skills. TSV-OCT can find errors up to five times faster. TSV-OCT on a 3-D Network-on-Chip (NoC) router exhibits minimal performance degradation during testing, with a reaction time restricted to less than 65,000 cycles, making integration into real-time applications possible despite a significant area overhead [9]. There is an increasing need for heterogeneous, manycore, and multicore processors as mainstream design alternatives. One major problem has been inter-core communication, which

affects the power consumption and bandwidth of multicore processors. Network-on-Chips (NoCs) offer applications requiring ubiquitous high-performance computing a scalable connection solution. The cutting-edge 3D Network-on-Chip Octagon for Ubiquitous Computing (OUC), created for Embedded Ubiquitous Computing Systems, is presented in this paper. OUC shows a significant reduction in latency and an average throughput improvement of 21.54% under hotspots and 12.89% under uniform traffic patterns, respectively, with a modest network diameter and adequate route variety [10].

By enabling single-cycle multiple-hop transmission via bypass channels, SMART NoCs are able to achieve extremely low latency. Performance can be impacted by congestion on these bypass channels, though. It is demonstrated through experiments that the Scalable Mapping Technique (SMT) framework is more scalable than Integer Linear Programming (ILP). Shorter application schedule lengths are achieved across a range of workloads on 2D and 3D SMART NoCs thanks to the SMT framework's faster runtimes and enhanced scalability, as well as its 2D and 3D extensions with mixed dimension-order routing [11]. Spiking Neural Networks (SNN) have been successfully integrated in 3D-ICs with 3D-NoC interconnects in neuromorphic devices to minimize power consumption. The fault-tolerant 3D-NoC-based neuromorphic system NASH makes use of on-chip learning and lightweight spiking neuron processing cores (SNPCs). With 65k synapses and 256 leaky integrate-and-fire (LIF) neurons per SNPC, NASH tackles the reliability issue in extremely dense neuromorphic systems [12]. In order to optimize performance and cost metrics like power, dependability, area, heat distribution, and latency, the Sailfish Optimization Method addresses the mapping of application task graphs on Intellectual Property (IP) cores into Network-on-Chip (NoC) and presents an advanced mapping technique called SFOA.

An innovative method is shown in [14] to achieve quicker mapping across six conventional benchmarks while lowering NoC power dissipation. This method makes use of a shared k-nearest neighbor clustering mechanism with an empirical basis. The results of the experiments show that the suggested techniques perform better than other nature-inspired metaheuristic approaches, especially when dealing with big application task graphs. Even

though 3D NoC is thought to be a viable choice for next Chip Multiprocessors (CMPs), there are still some difficult design decisions to be made, like choosing an effective routing algorithm. In order to solve this issue, technology is used in [13] to forecast which routing algorithm will operate at the maximum throughput and lowest power consumption. Based on the traffic load rate for the NoC system, the suggested system dynamically switches between existing 3D routing algorithms, demonstrating great accuracy in anticipating optimal routing methods. The effectiveness of the 3D NoC throughput, energy consumption, and the NN-prediction technique are demonstrated by the experimental results, which were carried out with the 3D NOXIM simulator and confirmed with PARSEC workloads [15].

Hot-Cluster, a hotspot-aware self-correction platform for clustering flaws in 3D-NoCs, is introduced in [16]. In the case of a medium fault rate, this platform decreases redundancies by approximately 60% when compared to uniformly distributed redundancies. To fix problematic TSV clusters, HotCluster combines offline (max-flow min-cut offline approach) and online (weight-based) mapping algorithms. According to experimental data, under a 50% failure rate, less than 1% of routers are disabled while using the max-flow min-cut offline approach and the weight-based online mode with a redundancy of 0.25. Introduced in [17], the Optical-NoC Router uses a 3D NoC design approach and is built to automatically prevent network deadlock and livelock. The router in 3D NoC provides fast speed and low power consumption by using adaptive multicast routing. In comparison to existing approaches like EDXY and FADyAD during traffic situations, the 3D NoC's mesh network of 808 (64 nodes) shows a significant reduction in average latency (36.08% and 28.5%, respectively) and total power consumption (74.4% and 66.2%, respectively) for each layer.

Various hardware design elements, memory consumption, and timing parameters including minimum and maximum periods, frequency support, etc. are considered in the present study focused on Network-on-Chip (NoC) design. To forecast design accuracy and performance, machine learning approaches such as multiple linear regression, decision tree regression, and random forest regression are used. The Virtex-5 FPGA is used to verify interprocess communication between nodes,

where data is transferred in packets with different bit widths. Modelsim 10.1b is used for simulations, and Xilinx ISE 14.2 is used for design generation using the VHDL programming language. Robust performance of the constructed model is demonstrated on independent test data sources and it is verified [18]. NoC network performance could be greatly improved by implementing 3D IC technology, which consists of stacking several active NoC layers via Through Silicon Via (TSV) vertical linkages. In order to enhance bufferless mesh NoC performance, an asymmetrical routing algorithm and a special flit priority unit are combined in [19] to provide an interleaved vertical edge routing design strategy for 3D NoC. According to experimental findings, when compared to standard bufferless networks with an equal number of routers, the suggested Router offers better network performance with less hardware overhead.

In [20] examines performance trade-offs in 3D NoC, such as higher router space and power consumption in exchange for better network performance and lower latency. Using an asymmetrical routing algorithm and a special flit priority unit, an interleaved vertical edge routing design technique is created. In comparison to normal bufferless networks, experimental results show that the suggested Router delivers greater network performance with lower hardware overhead. Power optimization in NoC design is addressed in [21], introducing an evolutionary algorithm-based method for mapping Intellectual Property (IP) cores to 3D NoC. Simulated annealing is incorporated in the crossover operation stage of the genetic algorithm to enhance global optimization capability. Experimental results show improved convergence, lower power consumption, and quicker search for a better solution, with an average power consumption reduction of 42.2 percent when dealing with a large number of cores (124 IP cores). In [21] discusses power optimization in NoC design and presents an evolutionary algorithm-based technique for mapping 3D NoC to Intellectual Property (IP) cores. The genetic algorithm's crossover operation stage incorporates simulated annealing to improve its capacity for global optimization. When working with a large number of cores (124 IP cores), the experimental findings reveal improved convergence, reduced power consumption, and faster search for a better solution, with an average power consumption decrease of 42.1%.

The requirement for more processing power to support these algorithms is rising as deep learning technology becomes more and more commonplace in a variety of applications. High-performance, 3D heterogeneous manycore systems have become a viable option. Deep learning implementation on these systems, however, presents a number of architectural difficulties. In addition to efficiently managing the communication traffic between CPUs and GPUs, NoC needs to solve the thermal problems that arise from the high power density of 3D system designs. This study proposes a design methodology for a heterogeneous 3D NoC architecture that addresses thermal hotspot mitigation in addition to meeting the communication requirements of both CPUs and GPUs. LeNet and CIFAR, two popular convolutional neural networks (CNNs), are trained in order to accomplish this. Our suggested combination of performance-thermal optimization techniques lowers the maximum temperature by 22% compared to a solo performance-optimized 3D NoC, with only a 5% degradation in the full-system energy-delay product during CNN training.

In order to decrease network diameter, the study presents a novel design strategy that incorporates NLCA. Deadlock-free routing is proposed using this topology, and guidelines for the main node as a cluster head (CH) are given. When the suggested design is used, simulation results show a 10% decrease in energy consumption, a 5.3% improvement in network latency, and a 20% increase in throughput at a lower cost than competitors. A alternative for bus architecture, known as NoCs, has been proposed in response to the growing complexity of systems-on-chip (SoCs). NoCs provide effective performance utilization and scalability; nonetheless, an important factor in NoC design is the tradeoff between area/power and performance. This paper presents the Flexible Router, a novel architecture that improves network performance overall without adding virtual channels (VCs) or larger buffer sizes. When it comes to increasing the saturation rate for different traffic patterns, the Flexible Router performs better than conventional routers. The NoC has become the industry standard fabric architecture for chip multiprocessor (CMP) design. An effective multicast routing approach is required due to the increase in multicast traffic for barrier synchronization, multithreading, and cache coherence protocols. By making use of geographical variety in the input buffer, the proposed Multicast

Router with Buffer Sharing (MRBS) guarantees deadlock-free multicast routing. MRBS accomplishes minimal path routing without requiring huge buffers or extra virtual channels, and it offers an average 41.5% development in ADP for a range of network sizes over a standard tree-based router.

In recent years, the evolution of 3D NoCs has solidified their status as a mature multicore interconnection architecture. However, the limitations of traditional electrical lines, characterized by minimal bandwidth and high energy consumption, have prompted consideration of photonic interconnection for future 3D Optical NoCs (ONoCs). Addressing the need for fault-tolerant capabilities in 3D ONoCs, this paper introduces a robust optical router (OR) structure that prioritizes minimal redundancy to maximize restore paths. Additionally, a fault-tolerant routing algorithm is proposed, allowing for restore path identification within disabled ORs under deadlock-free conditions, known as fault-node reuse. The 3D NoCs have established themselves as a well-developed multicore connectivity architecture in recent years thanks to their progress. For upcoming 3D Optical NoCs (ONoCs), photonic interconnection has been considered due to the drawbacks of conventional electrical connections, which are marked by low bandwidth and high energy consumption. This work presents a resilient optical router (OR) structure that prioritizes minimal redundancy to maximize restore pathways, addressing the demand for fault-tolerant capabilities in 3D ONoCs. Furthermore, a fault-tolerant routing technique called as fault-node reuse is proposed that permits path identification restoration within disabled ORs without causing a deadlock.

The results of the experiments show that the suggested strategy outperforms earlier efforts, with throughput performance improvements of up to 81.1% and an average of 33.0% under a variety of synthetic and actual traffic patterns[27]. The multilayer architecture of 3-D NoC results in heat imbalances between layers, which affect performance and reliability. In order to mitigate these concerns and guarantee thermal safety, efficient cooling techniques are essential. In order to address path selection and thermal distribution, this research suggests a unique thermally aware routing technique that makes use of the Q-learning algorithm (Q-Thermal). When compared to cutting-

edge methods, Q-Thermal dramatically lowers heat and temperature throughout the various layers of 3-D NoCs, resulting in a balanced thermal distribution and notable gains in network performance [28]. The high-density integration, energy economy, and performance of monolithic 3-D (M3D) technology come with some temperature and process fluctuation difficulties [29]. The influence of inter-tier process variation on M3D-enabled NoCs is highlighted in this research, which also reveals that traditional designs overestimate the energy-delay product (EDP) by 50.8%. In order to reduce negative effects, a process variation-aware design strategy is suggested, dividing intra-router stages and inter-router links among tiers. When compared to designs that are unaware of processes, experimental results show a 27.4% improvement in EDP across benchmarks [30]. This research blends one or two heterogeneous floor planning layers with homogeneous regular mesh networks on a separate layer to introduce innovative 3D 2-layer and 3-layer NoC architectures. A thorough design process that incorporates cycle-accurate NoC simulation, router assignment, and floor planning is suggested. In comparison to its 2-layer sibling, the 3-layer architecture exhibits better network performance and provides more flexibility to improve performance through the augmentation of virtual channels, buffer size, or mesh size [31].

3. Main Contribution of Work On 3-D Noc

Examining spatial parallelism in asynchronous on-chip routers is the main objective of this work. By 2024, handshake protocols are predicted to power almost half (49%) of all communications worldwide, and through 2014, asynchronous signaling latency improvements are anticipated [60]. Since routers are essential parts of on-chip networks, optimizing the performance of asynchronous routers via spatial division techniques provides a workable way to satisfy future chip designs' speed requirements. It is envisaged that the methods developed in this work might be applied to conventional asynchronous circuits and may find use beyond asynchronous on-chip networks. Exploration of spatial parallelism will be the main focus at various layers. Although there isn't a single agreement on what constitutes a layer in on-chip networks, the routing layer, switching layer, and physical layer are typically used to categorize the lower communication structure [41]. There are three types of data that are communicated over a network:

frames, flits, and flits. Important communication resources like buffers and channels are included in the physical layer, which makes it easier for flits to move from one pad to another. In order to improve the overall performance of asynchronous on-chip routers, the research attempts to explore spatial parallelism within each of these levels.

A flit, which can consist of one or several flits, plays a crucial role in the communication structure. The switching layer dynamically manages the allocation of communication resources in the physical layer to different flits. This allocation process relies on the control flow method, involving both hardware structure and algorithms. In this context, a frame is identified as the smallest self-explanatory data unit for a network node and contains one or more flits. The routing layer determines the path a frame takes within the network. This research primarily delves into the exploration of spatial parallelism within the lowest two layers: the physical layer and the switching layer. In the physical layer, state-of-the-art routers employ synchronized multi-bit pipelines as buffer stages, akin to latches on buses in synchronous circuits. While this pipeline style simplifies control logic, it introduces notable speed overhead. The research scrutinizes the impact of speed degradation caused by synchronization and proposes techniques to mitigate this effect. The research introduces a new spatial division multiplexing (SDM) flow control method, which will be compared with the traditional virtual channel flow control method across various router implementations. This comparative analysis will encompass speed performance, area consumption, and power dissipation under different operational scenarios.

The contributions of this research are:

- In the physical layer, an examination was conducted to assess the speed and area overhead associated with the synchronization of multiple low-level pipelines. This analysis delved into understanding the impact of synchronization on performance and resource utilization during the physical implementation of these pipelines.
- In the switching layer, various analyses and methodologies were explored to enhance efficiency and reduce overhead:
- Overhead analysis of the VC flow control method: The study involved a comprehensive examination of the overhead associated with the virtual channel flow control method. This analysis focused on

understanding the impact of this method on performance, resource utilization, and overall efficiency within the switching layer.

- Utilization of SDM in asynchronous routers: The exploration involved the integration and application of spatial division multiplexing within asynchronous routers. This included assessing the advantages and challenges of incorporating SDM into asynchronous router architectures.
- A novel asynchronous SDM router: A pioneering asynchronous Spatial Division Multiplexing (SDM) router was introduced. This innovative router design leverages asynchronous communication principles while incorporating spatial division multiplexing, aiming to enhance overall efficiency and performance in on-chip routing systems.

The introduction of the Small graph WiNoC construction served as a defense against DoS attacks. On the other hand, a hash-based verification technique has been put out to prevent Error Detection (ED) events and improve WiNoC architecture security. This method not only prevents DoS, ED, and spoofing, but it also uses the OS to actively prevent DoS attacks in a WiNoC with contention-free channel access—the particular kind that is examined in this research. While there are a number of methods for detecting attacks, like event- and signature-based detection, which are usually implemented in software because of hardware overheads, their applicability in a multicore NoC is constrained by high latencies and processing overheads. Furthermore, a workable solution for low-complexity attack detection is threshold-based attack detection. However, since unrecoverable burst mistakes are not always related to jamming, using the recoverable error rate as a criterion to distinguish between burst and jamming-induced errors may lead to a large false negative rate. By addressing the detection and defense against jamming and ED assaults directly within the WiNoC architecture, this article adopts a unique method and provides a full solution incorporated into the Network on Chip itself.

3. Proposed 3-D Buffer And Bufferless Noc Using Fredo And Wormhole Ruting And Switching Algorithms

A mesh topology network with 3x3 and 5x5 dimensions using wormhole switching in the

OASIS-NoC is proposed in [9] and same is depicted in Fig. 2, the five ports of each routers are local, north, south, east, and west directions. Because of the switch's placement inside the network, the actual number of ports may differ. We give each control in the network an XY coordinate (x-addr and y-addr). Each of these three-bit address is merged into each 76-bit input flit. The flit includes several important information, including the 64-bit payload, the direction of the next port (Next-Port, 5 bits), and a one-bit tail signal for the packet's termination. For more information, see Fig. 3. Three pipeline phases are used to carry out the routing operation, as shown in Fig. 1. Destination addresses are taken out of the flit and then decoded in the first step, which is the routing calculation stage. By comparing the decoded address with the processing switch address, one can determine the direction of the next port. The switch allocation stage is then formed by sending this data to the switch allocator. For each input, the switch allocator determines when and to whom to assign the flits' output direction. When there is a conflict and many input ports want the same output, the arbiter uses the round-robin strategy to try to service each request fairly [10]. Then, in the crossbar traversal stage, the crossbar uses the switching control signals that the switch allocator provides to direct the transfer of flits to the computed output. Further information about the FREDO-NoC architecture is available in [9]. We do experiments using the JPEG codec application to evaluate the performance of the FREDO-NoC design [11] as shown in Fig. 2 and set the network size to 3x3. As shown in Table.1, task assignments were mapped onto the FREDO-NoC and were assigned at random. Fig. 4 presents the simulation results for the FREDO-NoC design, showing the total clock cycles in relation to the buffer size. Conversely, the 3D Oasis-NoC, depicted in Figure 3, utilizes a simple 2x2x4 mesh architecture. Each router is assigned an X, Y, and Z coordinate via x-addr, y-addr, and z-addr, in that order. Interestingly, each switch has a maximum capacity of seven input ports; the first five (input local, west, east, south, and north) are used for intra-layer connections, while the other two (up and down) are used for inter-layer communication. Depending on where the switch is placed in the design, the actual number of ports changes. This reduces power consumption by allowing unnecessary links to be removed that aren't connected to any other regulators.

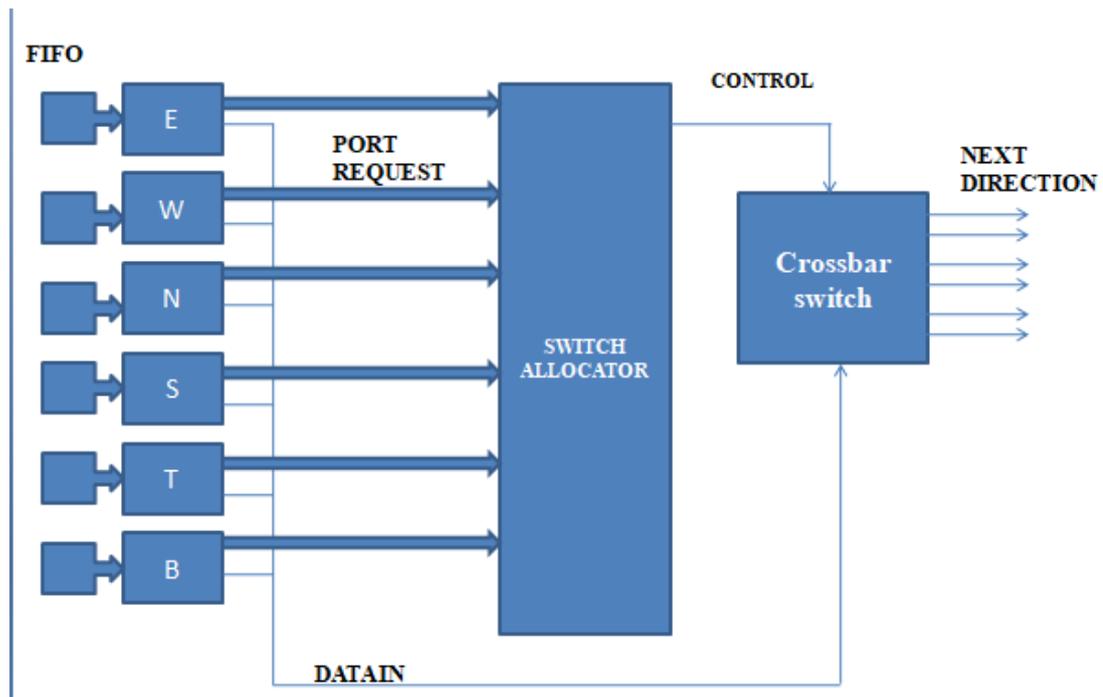


Fig1. Proposed 3D single NOC routing model

The packets that are transmitted for every clock cycles is np_{ck} and size of each packet is $packet_{size} = 19$ bits, flit size is indicated by $flit_{size}$ which is 4 bits, the total latency of packet transmission per cycle is $nsim_{pkt}$ and the total period of the clock is represented by T. The proposed 3D NoC includes three layers, each having 25 routers so $3 \times 25 = 75$ routers, and each layer is connected through a z-axis port to access routers in the middle and top layers. The True Random Number Generator (TRNG) is an entropy source to generate non-deterministic information and each piece of information is the size of 22 bits it includes a 4-bits destination router ID

and 2-bits to select layers dynamically and 16-bit data. The TRNG-generated data is passing through AHB to APB bridge to AHB interconnect at speed of 100MHz and is scaling down to 10MHz through APB protocol, which includes idle, setup, and access stated. In the setup state, the frequency scaling will take place to the required frequency. The low-speed data which is transmitted from APB is stored in transmitter FIFO as shown in Fig.2 and this storage is avoid loss of data and increases throughput and decreases latency. The FIFO depth is pre-calculated for the depth of 100, so 100 packets can receive from the bridge without any interference between write and read.

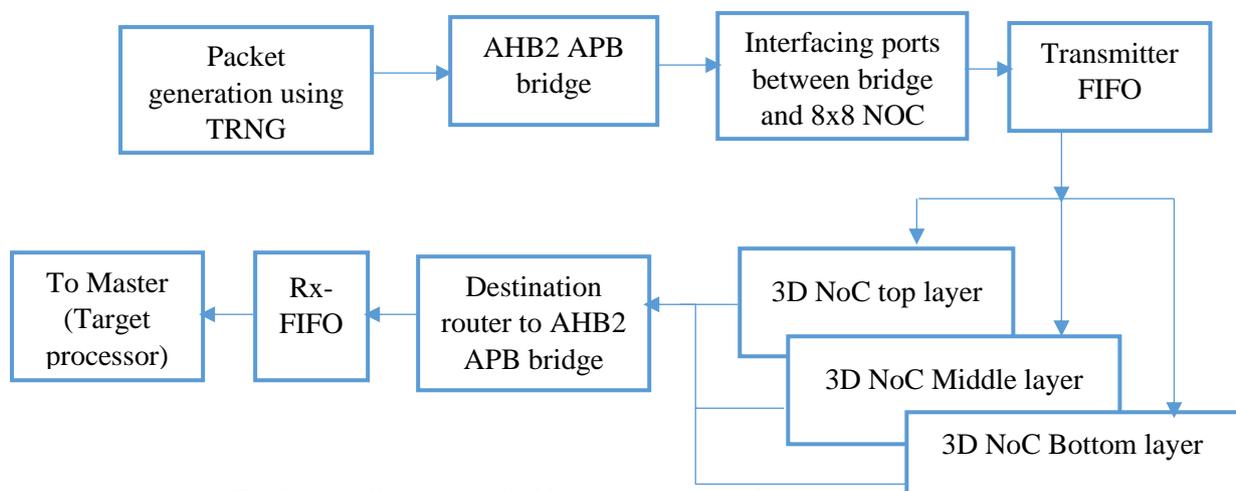


Fig 2. Overall proposed 3D NoC subsystem level architecture with bridge.

The 3D-NoC has been designed using FREDO which is a fast routing and switching algorithm, after transmission of the packet through all intermediate routers with help of FREDO algorithms, the received packets will write into a text file for easy monitoring, and every router has maintained one text file for writing all received packets. The destination router data packet is transmitted to the bridge and the same packet write into receiver FIFO (Rx-FIFO) and

the AHB2APB bridge reads the same packet and sends it to the processor. In order to minimize latency and hardware resources, the complete 3D NoC and other modules are designed with help of combinational logic, Table.1 shows a summary of the complete design in terms of slice registers, delay, number slice LUTs, and power and all these parameters are compared with existing results.

Table.1: Comparison between proposed and existing results

Parameter	Buffer less		Buffered	
	Existing results [64, 71,43]	Present results	Existing results	Present results
Slices	873	289	895	383
Area	672	288	873	403
Latency (ns)	21.42	1.154	21.54	1.541
Power (W)	0.4281	0.082	0.381	0.093
Slices+LUT's	15731	4901	13401	4529
Operating frequency (MHz)	64.7	357.015	82.2	386

The system that is being presented makes use of Xilinx Vivado, a hardware platform that has been reliably used in earlier systems to construct NoC. Notably, the suggested system adds unique elements, and Table 1 offers a condensed comparison of the suggested strategy with the current works. To provide a comprehensive visual aid and in-depth examination of the NoC performance, Fig. 2 shows relative plots that compare the system with and without buffers. Upon comparison with the existing works, the proposed system demonstrates superior performance across all parameters, whether with or without the application of buffers. This improvement is evident in factors such as speed, efficiency, and other relevant metrics. The comprehensive analysis

presented in Table 1 and Fig. 2 highlights the advancements achieved by the proposed system, reinforcing its efficacy and superiority over previous approaches.

4. Results And Discussion

This study, which emphasizes on-chip communication and makes use of the Xilinx tool, offers an overview of both the suggested system and the current system. A thorough examination of different current topologies serving as on-chip communication routers is provided. Although the research points to potential future applications to networks of NxN, the current network topological approaches are restricted to an 8x8 torus network. A major component of this study is the creation of a

DRNoC using FDOR systems with XY routing. This approach considers variables like slice registers and LUTs, leading to a significant improvement with low resource use. The overall performance is mostly

dependent on the number of flip-flops and delay factors; for systems with and without buffers, the suggested approach delivers notable gains in delay reduction and packet transmission.

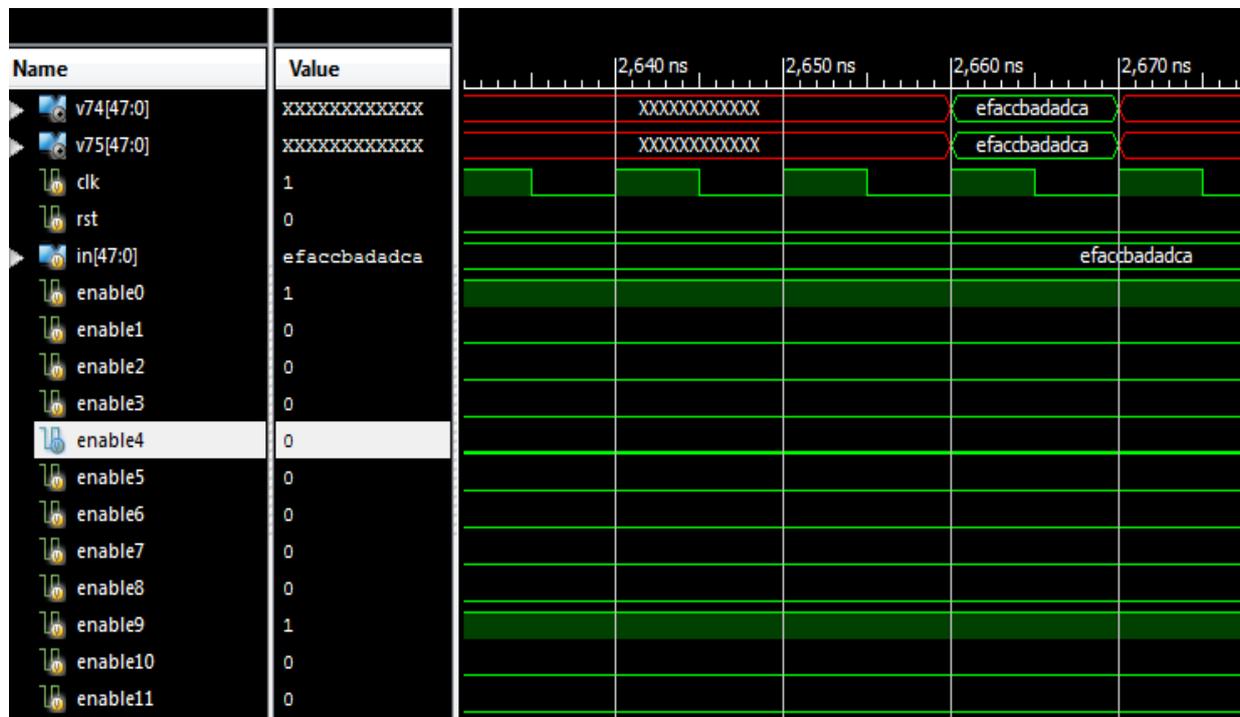


Fig 3. Simulated results of sourcing of packet at source router

When it comes to VLSI, factors like size, power, and frequency usually have an impact on SoC devices that follow Moore's Law. Nevertheless, the suggested approach distinguishes itself by lowering these parameters, demonstrating improved performance, and potentially being applicable to a wide range of future uses. The study highlights throughput as a crucial factor for NoC devices and shows that, in comparison to current methods, the suggested system has an overall throughput improvement. The application scenarios are covered in detail, with a focus on how NoC implementations without buffers work well for device-to-device applications, particularly in optical installations. On the other hand, NoC with buffers is the

recommended option in scenarios involving dynamic traffic routing and real-time data packet transfer to handle traffic. Thus, the research offers a comprehensive examination of the benefits of the suggested system, both in terms of performance and adaptability to various application scenarios. Channel routing method: Based on the architectural arrangement of linked sub-modules, this algorithm determines the best path between them. The outcomes show that it greatly improves stability and efficiency in the NoC (Network-on-Chip) network congestion. Screenshots of the suggested algorithm are shown in Figs. 3 and 4, which offer a visual explanation of the labor done.



Fig 4. Simulated results of packet delivered at destination router

Routing topology: this is a unique kind of strategy for the control operation and different accessible calculations for the topology advancements; in the routing topology kind, stream control application modules are done in the module oriented among the variable size and existing data transfer rate. The uniqueness of the NoC in the reconfigurable style can be employed to minimize the number of development support of the developers, which will reduce the effort of one chip and less power design utilization. At the same time, the typical challenges are expected from the delays caused due to the connecting wires among the channel for transmission using predefined. The routing can be designed and implemented in two divisions: free steering and definite routing, in which initial step and secondary steps, respectively.

In this research work, an attempt was made to realize and implement of generic data routing strategy with evaluation criteria for the demonstration of the virtual cut and wormhole techniques through the different data switching architecture for the non-congestion control by the insight upgradation of packet switching for the two-dimensional topology of the kind of mesh with the network arranged in 8X8 matrix style. In the view of successful traffic data handling in the already existing buffer with the operational principle of FIFO, where the preamble cause and the prime concern of the article discussion are to attain the optimal solution for the challenges

in the system architecture in view of packet delivery ratio, latency, and area in spite of available resources it was attained with the utilization of 24.85% BRAM and its interpretation is done in Fig.3 &4, the prime focus in the paper is to achieve the minimal utilization of the LUT of around 3% in the available list of resources from the designer perspective with the predefined applications.

5. Conclusion

An obvious continuation of our group's previous 3D-NoC design effort is the 3D FREDO-NoC. This study presents the hardware design for 3D NoC and presents preliminary evaluation findings. When compared to the 2D OASIS-NoC architecture, the 3D FREDO-NoC performs less quickly and has a marginally higher power overhead and space usage penalty. Even with more complicated technology, the 3D FREDO-NoC manages to lower latency; in comparison to the 2D FREDO-NoC, delay is reduced by 22%. We will be concentrating on routing algorithm optimization in the future to improve overall design performance. To assess the usable performance of our system, we also want to simulate 3D FREDO-NoC using actual workloads, such as the JPEG application that was previously tested with 2D FREDO-NoC in part II. The DRNoC with FDOR strategies are required to fulfill the demands of data handling on an FPGA device over various types of topologies as well as computational

methods for memory or based on the buffer and without memory or buffer less data transmission. Several techniques that are based on packet switching in on-chip networks assume local or internal routing strategies to address a variety of aspects like area along with power reduction. It is also taken into account other factors like minimizing delay without changing the frequency of operation for scenarios like buffer as well as buffer-less routing strategies. When compared to the conventional strategies, the proposed buffer-less routing algorithm was found to be more efficient at the prototyping level, whereas this algorithm is not suitable for real-time applications. Buffer-based routing on a practical FPGA as well as other application-specific devices plays a vital role in the operating topology. We are limited to an 8 X 8 switch with mesh topology under a two-dimensional structure under design in this research work. The power, as well as the area used by router buffers in NoC, seem to be a major issue in the deep submicron domain. Bufferless routing is been developing as a good alternative for achieving power as well as area efficiencies for NoC with the elimination of buffers. A unique buffer-less routing algorithm is employed with any topology in this work. The recommended routing technique mainly depends on the principle of making-a-stop (MaS) intending to achieve deadlock as well as livelock freedom in wormhole-switched NoC. With the help of a flit-level, cycle-accurate network simulator, the performance of synthetic traffic situations can be assessed. When compared to another conventional bufferless routing algorithm, the computational results demonstrate that the designed routing algorithm optimizes average latency by 24%, power consumption by 19%, as well as area overhead by 74%.

References

- [1] Thomas Moscibroda.at.al, “A Case for Bufferless Routing in On-Chip Networks”, *ISCA'09*, June 20–24, 2009, Austin, Texas, USA. Copyright 2009 ACM 978-1-60558-526-0/09/06
- [2] Juan Fang.at.al, “Hybrid Network-on-Chip: An Application-Aware Framework for Big Data”, *Hindawi, Complexity Volume 2018*, Article ID 1040869, 11 pages,
- [3] <https://doi.org/10.1155/2018/1040869>
- [4] Jing Lin.at.al, “Making-a-stop: A new bufferless routing algorithm for on-chip network”, *J. Parallel Distrib. Comput.* 72 (2012) 515–524, 2012 Elsevier Inc.
- [5] Rose George Kunthara.at.al, “ReDC: Reduced Deflection CHIPPER Router for Bufferless NoCs”, 978-1-5386-6575-6 /18, 2018 IEEE
- [6] Alexander Shpiner.at.al, “On the Capacity of Bufferless Networks-on-Chip”, *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, VOL. 26, NO. 2, FEBRUARY 2015
- [7] Yu-Hsiang Kao.at.al, “Design of A Bufferless Photonic Clos Network-on-Chip Architecture”, *IEEE TRANSACTIONS ON COMPUTERS*, Digital Object Identifier 10.1109/TC.2012.2500018-9340/12, 2012.
- [8] Chaochao Feng.at.al, “Addressing Transient and Permanent Faults in NoC With Efficient Fault-Tolerant Deflection Router”, *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS*, Digital Object Identifier 10.1109/TVLSI.2012.2204909, 1063–8210, 2012 IEEE
- [9] Dominic DiTomaso.at.al, “Resilient and Power-Efficient Multi-Function Channel Buffers in Network-on-Chip Architectures”, *IEEE TRANSACTIONS ON COMPUTERS*, information: DOI 10.1109/TC.2015.2401013, *IEEE Transactions on Computers*, 0018-9340 , 2015 IEEE.
- [10] Anh T. Tran.at.al, “Achieving High-Performance On-Chip Networks With Shared-Buffer Routers”, *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS*, 0.1109/TVLSI.2013.2268548, 1063-8210, 2013 IEEE.
- [11] 1063-8210/\$31.00 © 2013 IEEE.at.al, “System-Level Buffer Allocation for Application-Specific Networks-on-Chip Router Design”, *IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS*, VOL. 25, NO. 12, DECEMBER 2006.
- [12] Chih-Hao Chao.at.al, “Routing-Based Traffic Migration and Buffer Allocation Schemes for 3-D Network-on-Chip Systems With Thermal Limit”, *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS*, VOL. 21, NO. 11, NOVEMBER 2013.
- [13] Mohammad Arjomand and Hamid Sarbazi-Azad, “Power-Performance Analysis of Networks-on-Chip with Arbitrary Buffer Allocation Schemes”, *IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS*, VOL. 29, NO. 10, OCTOBER 2010.

- [14] Cunlu Li.at.al, "RoB-Router : A Reorder Buffer Enabled Low Latency Network-on-Chip Router", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 1045-9219, 2018 IEEE.
- [15] Chun-Wei Wu.at.al, " A Hybrid Multicast Routing Approach with Enhanced Methods for Mesh-Based Networks-on-Chip", IEEE TRANSACTIONS ON COMPUTERS, MANUSCRIPT ID: TC-2017-11-0610-R1, 0018-9340, 2018 IEEE.
- [16] Davide Zoni.at.al, "CUTBUF: Buffer Management and Router Design for Traffic Mixing in VNET-based NoCs", 10.1109/TPDS.2015.2468716, IEEE Transactions on Parallel and Distributed Systems, 1045-9219, 2015 IEEE
- [17] Yu-Yin Chen, at.al, "Path-Diversity-Aware Fault-Tolerant Routing Algorithm for NOC systems", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 10.1109/TPDS.2016.2588482, 1045-9219, 2016 IEEE.
- [18] Anastasios Psarras.at.al, "Networks-on-Chip With Double-Data-Rate Links",IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I: REGULAR PAPERS, 10.1109/TCSI.2017.2734689, 1549-8328, 2017 IEEE.
- [19] Zhiliang Qian.at.al, "FSNoC: A Flit-Level Speedup Scheme for Network-on-Chips Using Self-Reconfigurable Bidirectional Channels", IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, 10.1109/TVLSI.2014.2351833, 1063-8210, 2014 IEEE
- [20] Lei Yang.at.al, "Optimal Application Mapping and Scheduling for Network-on-Chips with Computation in STT-RAM based Router", IEEE TRANSACTIONS ON COMPUTERS, 10.1109/TC.2018.2864749, 0018-9340, 2018 IEEE
- [21] Feiyang Liu.at.al, "Wavelength-Reused Hierarchical Optical Network on Chip Architecture for Manycore Processors", DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing, 2377-3782, 2017 IEEE
- [22] Michael Opoku Agyeman.at.al, "Performance and Energy Aware Inhomogeneous 3D Networks-on-Chip Architecture Generation", 10.1109/TPDS.2015.2457444, IEEE Transactions on Parallel and Distributed Systems, 2015.
- [23] Alexandre Coelho.at.al, "FL-RuNS: A High-Performance and Runtime Reconfigurable Fault-Tolerant Routing Scheme for Partially Connected Three-Dimensional Networks on Chip", IEEE TRANSACTIONS ON NANOTECHNOLOGY, 1536-125X, Volume 18, 2019.
- [24] Haseeb Bokhari and Sri Parameswaran, Network-on-Chip Design, Springer Science+Business Media Dordrecht 2017 S. Ha, J. Teich (eds.), Handbook of Hardware/Software Codesign, DOI 10.1007/978-94-017-7267-9_16.
- [25] Assad Abbas.at.al, "A survey on energy-efficient methodologies and architectures of network-on-chip", Computers and Electrical Engineering, <http://dx.doi.org/10.1016/j.compeleceng.2014.07.012>, 0045-7906/2014 Elsevier Ltd.
- [26] Ahmed Aldammas.at.al, "The efficiency of buffer and buffer-less data-flow control schemes for congestion avoidance in Networks on Chip", Journal of King Saud University Computer and Information Sciences, 2016, 28, 184–198, <http://dx.doi.org/10.1016/j.jksuci.2015.11.002>, 1319-1578, Production and hosting by Elsevier B.V.
- [27] Armin Runge, "FaFNoC: a Fault-tolerant and Bufferless Network-on-chip", Procedia Computer Science 56 (2015) 397 – 402, 2015.
- [28] Chris Fallin.at.al, "MinBD: Minimally-Buffered Deflection Routing for Energy-Efficient Interconnect".
- [29] Gnaneswara Rao Jonna.at.al, "Minimally Buffered Single-Cycle Deflection Router", 978-3-9815370-2-4/DATE14/2014 EDAA.
- [30] George Nychis.at.al, On-Chip Networks from a Networking Perspective: Congestion and Scalability in Many-Core Interconnects, SIGCOMM'12, August 13–17, 2012, Helsinki, Finland. Copyright 2012 ACM 978-1-4503-1419-0/12/08
- [31] Marwa Shaheen.at.al, "Modified CONNECT: New Bufferless Router for NoC-Based FPGAs", 978-1-5386-7392-8/18, 2018 IEEE
- [32] Michael Opoku Agyeman.at.al, "Extending the Performance of Hybrid NoCs beyond the Limitations of Network Heterogeneity", Journal of Low Power Electronics and Applications, www.mdpi.com/journal/jlpea, Appl. 2017, 7, 8; doi:10.3390/jlpea7020008.