

A Survey of Low Power Techniques for Efficient Network-on-Chip Design

Emmanuel Ofori-Attah, Michael Opoku Agyeman
Faculty of Art, Science and Technology
Univeristy of Northampton, UK.

Abstract—Power consumption continues to be a challenge for designers as the complexity of NoC increases. The scaling down of technology towards the deep nanometer era will only cause an increase in the amount of power NoC components will consume. Therefore, low power design solution is one of the essential requirements of future NoC-based System-on-Chip (SoC) applications. Several techniques have been proposed over the years to improve the performance of the NoCs, trading-off power efficiency; particularly power hungry elements in NoC routers. Power dissipation can be reduced by optimizing the router elements, applying architecture saving techniques and communication links. This paper presents recent contributions and efficient saving techniques at the router, NoC architecture and Communication link level.

I. INTRODUCTION

The poor scalability of bus-based architectures with technological growth has resulted in the emergence of the Network-on-Chip (NoC) paradigm as the communication standard for SoC [1]. The main idea of NoC is to allow simultaneous use of wires across a network, thus enabling parallelism, whereas in bus-based architectures, only one communication is achieved at a time. Figure 1 is an image of a typical 2D NoC-based shared-memory chip multiprocessor (CMPs) comprised of 36 nodes. Each node is comprised of a core, private level 1 instruction and data caches, a second level cache which could either be shared or made private, a router and logic. Connection is established between each node through the routers using links. Consequently, NoC's enrichment supply of parallelism impart high throughput, high bandwidth, and low latency [2]. However, switching activities and leakage power of the resources increases the on-chip power consumption. Furthermore, prior investigations suggest that as transistor size shrinks with technology, leakage power contributes to a substantial amount of NoC power consumption; particularly in NoC Routers [3], [4]. NoC router components consume a significant amount of power yet enhance the network performance. Nonetheless, reducing leakage power is proving to be a problem for designers as the complexity of NoC increases.

The growth in core count increases the number of on-chip resource. For this purpose, existing work trade-off network performance for power efficiency. Some have even resulted to the removal of power-hungry elements such as buffers and virtual channels. However, this results in impecunious performances.

Consequently, this rapid growth of technology does not only affect the power consumption in routers, but also impose

heavy delays on their implemented SoC design. Over the last few years, existing 2D NoC architectures have not been able to comprehend the demands of modern SoC design. As more resources are added, hop distance caused by additional wire length adds to the power consumption of the NoC, causing a negative effect on the performance. Therefore, efficient techniques are required to balance network and power performance. Particularly, to achieve this, we have categorized NoC into three main areas and have investigated power saving technique that can be applied in these areas. The rest of the paper is organised as follows. Section 3 discusses efficient techniques for power in NoC architectures. Section IV provides low-power techniques for communication links and finally, section V concludes the paper.

II. ROUTER ARCHITECTURE

Routers are the main components in NoC. A traditional single-stage router architecture consists of an arbiter, buffer, crossbar, virtual channels, input, and output ports (Figure 1c). As powerful and as effective as NoC is, research confirms that, it is responsible for consuming 40% of chip power [5], [6]. Majority of this consumption is caused by power hungry elements in routers such as the buffers and crossbar. Buffers are embedded into routers as temporal storage for packets while awaiting transmission. A study conducted by [7] reveal that 33% of dynamic power in routers are consumed by the buffers. Particularly, input buffers are considered to consume 45% of router power and occupies 15% of area [8]. On the other hand, the use of large buffer size expands the power consumption and area overhead whereas a reduction of it diminishes the overall performance of the network. Power consumption in the crossbar is solely dependent on the of the number of PEs employed. Recently, the increase of PEs in NoC has caused a major increase in the size of crossbars resulting in high power consumption, scalability issues, and large area [9]. Consequently, Intel's TeraFLOPS Processor [10] and MIT RAW [11] crossbars constitutes to a combined 40% of router power.

Therefore, to reduce the power consumption of the routers, existing work have utilized novel techniques to improve power efficiency, the crossbar size, arbiter, and buffer designs. For this purpose, this paper focuses on optimized power saving techniques at the buffer and crossbar level.

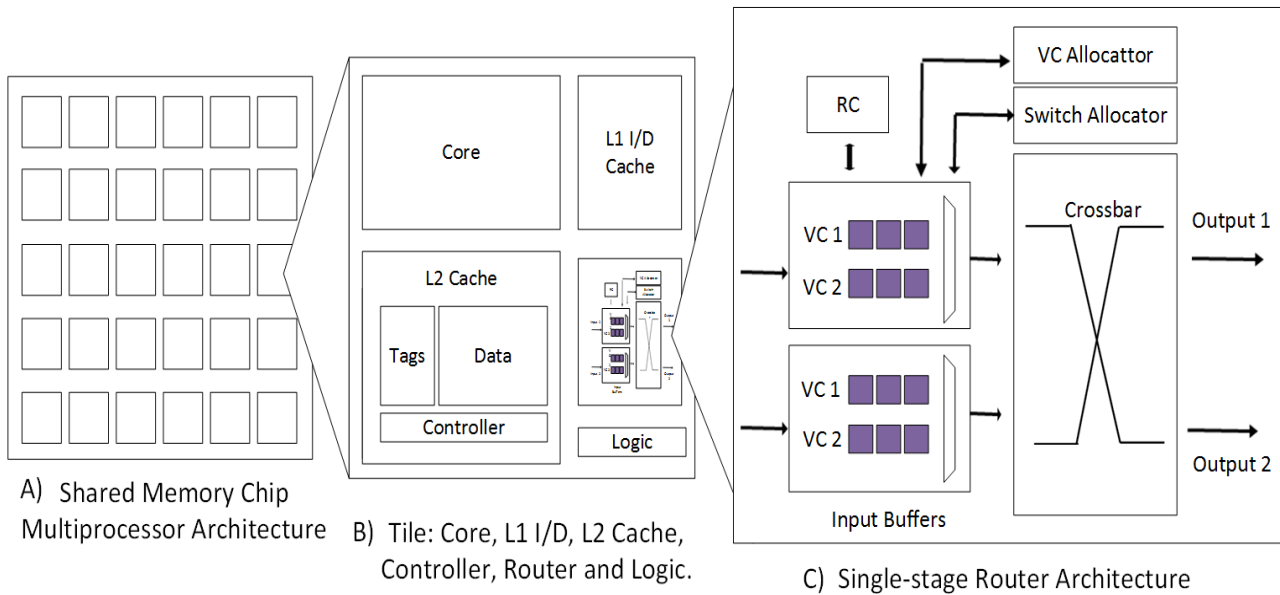


Fig. 1. 2D NoC-based CMP

A. Bufferless Routers and Buffered Routers

The power demand of modern SoC design is proving to be a struggle for designers. NoC is expected to provide high bandwidth, high throughput and low latency while still operating at a low power. This expectation however is proving to be design constraint for designers because of NoC components. Of all the power consuming elements there are in NoC routers, buffers and crossbars are the most consuming. Bufferless routers such as CHIPPER [12], SCEPTER[13] and Multi-ejection [14] ports have therefore been proposed forward as an alternative solution to combat this challenge. These two-different types of buffers are often disputed about their significance to the overall performance on NoC. While many prefer buffered routers, its high-power consumption and large area overhead often leads to the implementation of bufferless routers. In Bufferless routers, buffers are substituted with flow control deflection algorithms to transmit packets as soon as they arrive. Therefore, low power and low area pipeline registers are the only buffers employed in bufferless routers [15]. Consequently, the implementation of bufferless routers trade-off network performance for low power making it a high candidate to replace buffers.

On the other hand, many contests that bufferless routers at times contradicts its main target of providing low power. For example, CHIPPER [12] has its algorithm built based on a permutation tree. Permutation trees on the hand, are known to be high consumers of power [16].

Furthermore, because they are comprised of single ejection ports, bufferless routers suffer extensively from performance degradation when the network traffic reaches its peak. In addition, due to their inability to house packets, contention materialises when multiple flits arrive at the same time and contest to eject at the same node. Eventually, one flit will

be designated an output port while the other is diverted on a different route causing additional latency and eventually resulting in the packet to never reach its destination (Livelock). Therefore, Livelock is considered a major drawback and problem for bufferless routers because, at some point, the network will saturate and some packets will be deflected of course. This high deflection rates in the NoC contributes to bandwidth issues, high latency, and an increase in power consumption.

Consequently, various techniques have been proposed forward to resolve latency and livelock issues in bufferless routers. CHIPPER employs an algorithm which prevents contention by allocating output ports to packets based on their priority. Even though CHIPPER promises a reduction of 54% less power than conventional buffered routers, it is not evident that livelock is permanently eliminated. Whereas, input buffers with embedded virtual channels (VC) can be deployed to allow simultaneous transmission without a packet being thrown of course.

To tackle the deflection rates in Bufferless routers, Feng et al. proposed the multi-ejection port router [14] while Xiang et al. [17] proposed Deflection Containment (Dec). The bufferless routers used by Feng can be optimized to have four ejection ports which in effect will reduce the contention rates and latency. However, if more than four flits arrive, livelock can still materialise. Furthermore, an extension of the ejection ports will lead to an increase in the size of the crossbars resulting to more power being consumed in NoC. Xiang on the hand proposed an architecture which consists of virtual routers. These virtual routers allow the joining of sub routers. An extra link has been added to each virtual router to join sub networks together. The link allows packets which has been denied access in the current network to be transmitted to neighbouring sub networks to contest for an ejection port.

Moreover, bufferless routers present a great idea for reducing power however the employment of buffered routers trade-off area and power consumption to prevent deadlock, livelock, and high throughput. For this purpose, we focus on presenting techniques which can be applied in buffered routers to reduce the power consumption.

B. Low Power Buffered Router Techniques

1) *Power-Gating the resources in the Router Architecture:* Power-gating (PG) is an effective technique applied to many architectures to manage the amount of power generated by NoC resources. Their use at the router architecture level reduces the amount of static power dissipated in circuits which are rarely used. Many proposed architectures split resources into different parts and de-activate them based on the network traffic. The following authors have divided VCs into different groups; each group can be switched-off depending on the network traffic and performance.

Muhammad et al. [18] introduces the Traffic-Based Virtual Channel Algorithm. The algorithm divides the VCs in a switch port into three cells. Any one of these three cells can be activated and deactivated based on the network traffic and congestion. Thus, allowing resource power to be saved when they are not being used. Similarly, DimNoc is proposed by [19] to effectively manage PG operations. VCs are grouped and divided into levels. The lower level of the VCs is designed with SRAM and the higher level with SST-RAM. The lower level can be either be powered on or left in a drowsy state. The high-level VC's are utilized when there is heavy traffic. [20] employs a PG control unit which after completing a certain amount of cycles, it disables buffers from routers which are idle.

PG may be power efficient, however, an excessive use of it can have a diminishing impact on the performance of a network. Recently, many proposed architectures focus on saving power by using PG to disable resources however, they ignore the problems that arise when a packet encounters an idle router. According to [21], PG architectures are prone to deadlock. This is because, idle routers block all paths it intersects with, causing packet transmission to be halted resulting in performance degradation (wake-up latency). In addition to this, the constant turning on and off routers leads to non-negligible power overhead. In such an infrastructure, wake-up signals can be employed to either alert powered-off routers or routers which are scheduled to be powered-off of an incoming packet. For this purpose, Chen et al. proposed Power Punch, an optimized technique which sends wake-up signal 3 hops ahead, ensuring intersecting routers which are powered-off are activated in time to avert latency [22].

2) *Different types of Buffers:* Another novel technique is to use alternative buffers other than input buffers. Kodi et al. proposed iDeal [23], an architecture which employs the use of dual-function links. Unlike static buffer allocation, the proposed architecture uses a dynamic router buffer allocation to assign incoming flits to any free buffer. iDeal permits the reduction in buffer size to decrease power consumption by

employing existing repeaters to function as buffers during network congestion. Similarly, DiTomaso et al. proposed QORE [7], an architecture which improves power consumption using power-efficient Multi-Function channel buffers (MFC) and enhances the performance through reversible links. The use of MFC enables the channel buffers to be utilized instead of the buffers in the routers. [24] deals with power consumption by replacing the conventional SRAM with 3T₁N eDRAM. Significantly, buffer area was reduced by 52% and power by 43%.

3) *Reduction in the pipeline stages:* Packets traverse around the network through many pipeline stages: Buffer Write (BW), Route Computation (RC), Virtual Channel Allocation (VA), Switch Allocation (SA), and Switch Stage (ST).

Postman et al. [25] highlights that, buffers are not effectively optimized in existing architectures. Particularly, little emphasis is placed on path availability and network congestion which causes additional pipeline stages. The implementation of an algorithm which considers these challenges can help packets evade the buffering stage. Consequently, they proposed SWIFT NoC. The SWIFT NoC achieves low power consumption by allowing flits which bypass the buffering stage to do so in one cycle: Escaping the need for read/write power. Bypassing the buffering stage results in the use of fewer buffers and less power. Similarly, [26] proposes virtual circuit switching: A hybrid scheme which combines circuit and packet switching to allow flits to traverse through the network with only one stage. In contrast with the virtual point to point connections, 6.8% is the decrease in latency and 11.3% in power respectively.

C. Crossbar Switches

A crossbar switch is composed of individual switches arranged in a matrix form between several inputs and outputs. Crossbar switches can be categorized into two groups, single stage and multi stage. Figure 1.c depicts a typical $n \times n$ Crossbar switch in a single-stage router architecture.

1) *Crossbar size:* To achieve low power consumption and small area, existing work focuses on splitting large crossbars into smaller ones. Kim et al. [27] proposed a router architecture composed of two crossbars. In the proposed router architecture, Smaller crossbars are employed to reduce the size of the Virtual Channel Allocator (VA), Switch Arbiter units (SA) and shorter logic depth. Similarly, Park et al. propose an optimized crossbar [28] which combines decomposition and segmentation to effectively reduce power consumption by 35%. The crossbar has been disassembled in two small crossbars to reduce area and power. However, in large scale networks, there will be an increase in average latency.

Recently, multi-stage crossbars such as the Clos and Benes network [29], [30], [31] has been proposed. Multi-stage crossbars provide low power and smaller area. Yikun Jiang conducted a study on Circuit design [32] and concluded that the Clos networks outperform their counterparts (Benes[31] and Single-stage Crossbars) in several ways. In the Clos network, there is a reduction in the number of logic units used. The

Benes network suffers from 65% delay in timing and less power in Clos consumed because of the size of crossbars.

Naik et al. proposed a heterogeneous NoC [33] embedded with circuit switched routers composed of buffered and bufferless routers and a 3-stage CLOS network. In comparison to crossbar switch of the same size, the results of this is a reduction of 26% in power consumption and 32% in area. However, circuit switched network causes additional latency when a transmission is established between a source and its destination.

Bansal et al. proposed a power efficient 3-Dimensional Crossbar Switch with 7 input ports and 7 output ports. Although the 3D routers in the architecture consists of more ports compared to 2D routers (5 input ports and 5 Output ports), the crossbar consumes the same amount of power yet offers high throughput and low latency because of the extra two ports (up and down).

2) *Switching Algorithm:* In theory, there are two different types of routers; circuit switched routers and packet switched routers. In packet switched routers, data is encoded into packets and routed individually through the network. Circuit switching routers on the hand establishes a connecting between the source and destination and specifically allocates resources which will be used for transmission [34]. In Circuit switching routers, there is guaranteed throughput because all packets can be transmitted at the same time without delay in any router however, there is an increase in latency. This is because during the transmission process, the resources allocated cannot be accessed.

For this purpose, a switching mechanism has been proposed to effectively use the benefits of packet switching and circuit switching [35]. In this architecture, messages are split into different groups: High priority and Low priority. High priority messages are transmitted using circuit switching and low priority messages are transmitted using packet switching. The employment of these two mechanism allow power rails to be disconnected and PG to be used to disconnect parts of the router which are not used during a transmission.

III. LOW NETWORK ARCHITECTURE

Novel NoC architectures have been proposed to reduce the average packet latency while increasing the throughput. However, this usually at the expense of power consumption. To combat the challenges imposed by these power-hungry NoCs, various architectures have been proposed.

The exponential increase in the number of cores in multi-core over the last decade has resulted in the emergence of Three-Dimension (3D) NoC as the platform for on-chip communication. 3D NoC allows multiple silicon layers to be stacked together to not only enhance the throughput and latency, but also to reduce power consumption [36], [37]. In 3D NoC, the lengthy wires are replaced with short through silicon vias (TSVs) to minimize the number of hops it takes for a packet to traverse through the network. Particularly, the increase in the number of connectivity in 3D Integrated

Circuits allows the transmission of more messages around the network [38].

Debora Matos et al. proposed the 3D HiCIT, an architecture comprised of two hierarchical levels with a mesh topology at the top level. In comparison with the traditional 3D-SPIN and 3D Mesh topologies, the proposed architecture reduces the average latency to 50% and 54% respectively, with the 3D-SPIN been the latter [39]. In addition to this, the architecture is comprised of a crossbar and low-cost routers. Compared to the 3D-SPIN and 3D-BFT, the proposed architecture uses less TSVs.

However, limitations such as power density caused by the chip size, the cost of TSV and its defects [40], [39] prevents 3D NoC from reaching its potential. For this purpose, Nayak et al. [41] recommends the use of monolithic 3D. One approach to reduce power consumption is to use fewer buffers at the router port [42], [43], [44]. Similarly, Fang et al. proposed RRCIES, an architecture based on a mesh topology. RRCIES allows multiple cores to be connected between one router. As a result, hop distance is reduced. The use of fewer routers constitutes to a reduction in power hungry components such as buffers, crossbar, switches, and virtual channels [45].

Another alternative is to employ PG. The study of vertical slit field effect transistors led to the proposal of a 3D Hybrid architecture in [46]. PG and clock gating is employed in this architecture to enable different level of buffers to be deactivated. The proposed architecture splits the input buffers into three 3. Each input port is designed to access all three levels and permit any virtual channel destination to be chosen. In addition, the buffers from ports which are not being used are shared among busy ports.

For further enhancement in NoC architectures, Wireless NoC (WiNoC) has been proposed to overcome the bottleneck and limitations 3D NoCs. WiNoC reduces the hop-count between routers. The implementation of WiNoC allows the transmission of messages with a single hop, long range wireless links [47]. In such WiNoC architectures, communications over short distances materialise through the wired connections and long range communication occur through the wireless layer [48]. However, conventional WiNoC only permits one active wireless communication. During this period, the remaining wireless interfaces which are not being used dissipate static power. The proposed architecture in [49] categorizes the routers in the network into three zones. These zones are high utilization zone (UTZ), low utilization zone and rare utilization zone. The routers which are rarely utilized are power-gated and have their data rerouted. As a result, 88.76% of static power in the base router can be saved.

IV. COMMUNICATION LINKS

Although routers consume more power in NoCs, the communication links can be optimized to accommodate this. According to [50] and [51], routers and communication channels contribute to most of the power consumption in NoC. Therefore, novel techniques have been developed to reduce the amount of power consumed by the links.

1) *Voltage Scaling* : The voltage swing in the communication links can be optimized to reduce the amount of power it consumes. However, this is at a cost of a rise in bit-error-rate. For this purpose, Mineo et al. [52] proposed to reduce power consumption by using a technique which permits two working levels in a link. A flag is attached to each communication to identify their priority. Low prioritised communications (Body and Tail flit) can be transmitted on a low-level voltage swing while the others (Head flit) can be sent using a normal level voltage.

2) *Half-cycle Flits* : The longer flits traverse through the links in NoC, the more power is consumed. Therefore, decreasing the number of cycles of it takes from a flit to transmit routers would not only enhance the performance of the network but also save power. A. Psarras et al. proposed a technique which allows flits to only use a half cycle to hop between routers. By allowing flits to spend less time in the links, less power is consumed compared to single cycle routers where one cycle is used to execute all operations in the router and one is used to hop between routers [53].

TABLE I
NoC COMPONENTS & POWER SAVING DESIGN TECHNIQUES

Router Architecture	Bufferless [54], [14], [17], [55]	Buffered [25], [26], [18] [19], [20], [21] [22], [23], [7]
Network Architecture	3D NoC [45], [56], [38], [36], [40], [39], [41],	Wireless NoC [47], [49], [48]
Communication Links	Voltage Scaling [52]	Half-Cycle Flits [53]

V. CONCLUSION

In this paper, several NoC power saving techniques have been evaluated. Particularly, the effect of buffered and bufferless routers on power consumption has been presented. Moreover, a summary of these techniques have presented to compare and contrast their trade-off. The combination of some of the architectures presented, if employed, can help improve the amount of power consumed by NoC resources which can either be removed or switched-off. This could be the adjustment of the components (crossbar size, buffers, virtual channels) in the router architecture, modification of the architectures (resource management) and the amount of voltage used in the communication links. Also, we explore low power techniques used in emerging NoC Architectures (3D NoC and WiNOC). Based on our discussions, we can conclude power dissipation can be reduced in all areas of a network infrastructure.

REFERENCES

- [1] M. H. Neishaburi and Z. Zilic, "A fault tolerant hierarchical network on chip router architecture," in *IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems*, 2011, pp. 445–453.
- [2] A. B. Achballah, "A survey of network-on-chip tools," 2013.
- [3] L. Chen and T. M. Pinkston, "Nord: Node-router decoupling for effective power-gating of on-chip routers," in *45th Annual IEEE/ACM International Symposium on Microarchitecture*, 2012, pp. 270–281.
- [4] C. Sun, C. H. O. Chen, G. Kurian, L. Wei, J. Miller, A. Agarwal, L. S. Peh, and V. Stojanovic, "Dscent - a tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling," in *Networks on Chip (NoCS), 2012 Sixth IEEE/ACM International Symposium on*, 2012, pp. 201–210.
- [5] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar, "A 5-ghz mesh interconnect for a teraflops processor," *IEEE Micro*, vol. 27, no. 5, pp. 51–61, 2007.
- [6] M. B. Taylor, J. Psota, A. Saraf, N. Shnidman, V. Strumpfen, M. Frank, S. Amarasinghe, A. Agarwal, W. Lee, J. Miller, D. Wentzloff, I. Bratt, B. Greenwald, H. Hoffmann, P. Johnson, and J. Kim, "Evaluation of the raw microprocessor: an exposed-wire-delay architecture for ilp and streams," in *Computer Architecture, Proceedings. 31st Annual International Symposium on*, 2004, pp. 2–13.
- [7] D. DiTomaso, A. K. Kodi, A. Louri, and R. Bunescu, "Resilient and power-efficient multi-function channel buffers in network-on-chip architectures," *IEEE Transactions on Computers*, vol. 64, no. 12, pp. 3555–3568, 2015.
- [8] P. Kundu, "On-die interconnects for next generation cmpps, in workshop on on- and off-chip interconnection networks for multicore systems," 2006.
- [9] K. Sewell, R. G. Dreslinski, T. Manville, S. Satpathy, N. Pinckney, G. Blake, M. Cieslak, R. Das, T. F. Wenisch, D. Sylvester, D. Blaauw, and T. Mudge, "Swizzle-switch networks for many-core systems," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 2, pp. 278–294, 2012.
- [10] S. R. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntla, C. Roberts, Y. Hoskote, N. Borkar, and S. Borkar, "An 80-tile sub-100-w teraflops processor in 65-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 29–41, 2008.
- [11] M. B. Taylor, J. Kim, J. Miller, D. Wentzloff, F. Ghodrati, B. Greenwald, H. Hoffman, P. Johnson, J.-W. Lee, W. Lee, A. Ma, A. Saraf, M. Seneski, N. Shnidman, V. Strumpfen, M. Frank, S. Amarasinghe, and A. Agarwal, "The raw microprocessor: a computational fabric for software circuits and general-purpose programs," *IEEE Micro*, vol. 22, no. 2, pp. 25–35, 2002.
- [12] C. Fallin, C. Craik, and O. Mutlu, "Chipper: A low-complexity bufferless deflection router," in *IEEE 17th International Symposium on High Performance Computer Architecture*, 2011, pp. 144–155.
- [13] B. K. Daya, L. S. Peh, and A. P. Chandrakasan, "Towards high-performance bufferless nocs with sceptor," *IEEE Computer Architecture Letters*, vol. 15, no. 1, pp. 62–65, 2016.
- [14] C. Feng, Z. Liao, Z. Lu, A. Jantsch, and Z. Zhao, "Performance analysis of on-chip bufferless router with multi-ejection ports," in *IEEE 11th International Conference on ASIC (ASICON)*, 2015, pp. 1–4.
- [15] —, "Performance analysis of on-chip bufferless router with multi-ejection ports," in *IEEE 11th International Conference on ASIC (ASICON)*, 2015, pp. 1–4.
- [16] C.-K. Hsu, K.-L. Tsai, J.-F. Jheng, S.-J. Ruan, and C.-A. Shen, "A low power detection routing method for bufferless noc," in *Quality Electronic Design (ISQED), 14th International Symposium on*, 2013.
- [17] X. Y. Xiang and N. F. Tzeng, "Deflection containment for bufferless network-on-chips," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2016, pp. 113–122.
- [18] S. T. Muhammad, M. A. El-Moursy, A. A. El-Moursy, and A. M. Refaat, "Optimization for traffic-based virtual channel activation low-power noc," in *Energy Aware Computing Systems Applications (ICEAC), International Conference on*, 2015, pp. 1–4.
- [19] J. Zhan, J. Ouyang, F. Ge, J. Zhao, and Y. Xie, "Hybrid drowsy sram and stt-ram buffer designs for dark-silicon-aware noc," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 10, pp. 3041–3054, 2016.

- [20] N. Nasirian and M. Bayoumi, "Low-latency power-efficient adaptive router design for network-on-chip," in *28th IEEE International System-on-Chip Conference (SOCC)*, 2015, pp. 287–291.
- [21] H. Farrokhbakht, M. Taram, B. Khaleghi, and S. Hessabi, "Toot: an efficient and scalable power-gating method for noc routers," in *Tenth IEEE/ACM International Symposium on Networks-on-Chip (NOCS)*, 2016, pp. 1–8.
- [22] L. Chen, D. Zhu, M. Pedram, and T. M. Pinkston, "Power punch: Towards non-blocking power-gating of noc routers," in *IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*, 2015, pp. 378–389.
- [23] A. K. Kodi, A. Sarathy, A. Louri, and J. Wang, "Adaptive inter-router links for low-power, area-efficient and reliable network-on-chip (noc) architectures," in *Asia and South Pacific Design Automation Conference*, 2009, pp. 1–6.
- [24] C. Li and P. Ampadu, "A compact low-power edram-based noc buffer," in *Low Power Electronics and Design (ISLPED)*, *IEEE/ACM International Symposium on*, 2015, pp. 116–121.
- [25] J. Postman, T. Krishna, C. Edmonds, L. S. Peh, and P. Chiang, "Swift: A low-power network-on-chip implementing the token flow control router architecture with swing-reduced interconnects," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, pp. 1432–1446, 2013.
- [26] S. Shenbagavalli and S. Karthikeyan, "An efficient low power noc router architecture design," in *Online International Conference on Green Engineering and Technologies (IC-GET)*, 2015, pp. 1–8.
- [27] J. Kim, C. Nicopoulos, and D. Park, "A gracefully degrading and energy-efficient modular router architecture for on-chip networks," in *33rd International Symposium on Computer Architecture (ISCA'06)*, 2006, pp. 4–15.
- [28] D. Park, A. Vaidya, A. Kumar, and M. Azimi, "Mode-x: Microarchitecture of a layout-aware modular decoupled crossbar for on-chip interconnects," *IEEE Transactions on Computers*, vol. 63, no. 3, pp. 622–636, 2014.
- [29] Y. Xia, M. Hamdi, and H. J. Chao, "A practical large-capacity three-stage buffered clos-network switch architecture," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 317–328, 2016.
- [30] S. Yang, S. Xin, Z. Zhao, and B. Wu, "Minimizing packet delay via load balancing in clos switching networks for datacenters," in *International Conference on Networking and Network Applications (NaNA)*, 2016, pp. 23–28.
- [31] J. Zhang and H. Gu, "A partially adaptive routing algorithm for benes network on chip," in *2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 614–618.
- [32] Y. Jiang and M. Yang, "On circuit design of on-chip non-blocking interconnection networks," in *27th IEEE International System-on-Chip Conference (SOCC)*, 2014, pp. 192–197.
- [33] A. Naik and T. K. Ramesh, "Efficient network on chip (noc) using heterogeneous circuit switched routers," in *International Conference on VLSI Systems, Architectures, Technology and Applications (VLSI-SATA)*, 2016, pp. 1–6.
- [34] N. Chin-Ee and N. Soin, "A study on circuit switching merits in the design of network-on-chip," in *Computer and Communication Engineering (ICCCE)*, *2010 International Conference on*, 2010, pp. 1–5.
- [35] M. FallahRad, A. Patooghi, H. Ziaeeziabari, and E. Taheri, "Circket: A performance efficient hybrid switching mechanism for noc architectures," in *Euromicro Conference on Digital System Design (DSD)*, 2016, pp. 123–130.
- [36] C. Chao, "Traffic and thermal aware run time thermal management scheme for 3d noc systems," pp. 223 – 230, 2010.
- [37] W. R. Davis, "Application exploration for 3d integrated circuits: Tcam, fifo and fft case studies," pp. 496 – 506, 2009.
- [38] A. W. Yin, "Change function of 2d/3d network-on-chip," 2011.
- [39] D. Matos, M. Prass, M. Kreutz, L. Carro, and A. Susin, "Performance evaluation of hierarchical noc topologies for stacked 3d ics," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1961–1964.
- [40] D. e. a. Velenis, "Impact of 3d design choices on manufacturing cost," 2009.
- [41] D. K. Nayak, S. Banna, S. K. Samal, and S. K. Lim, "Power, performance, and cost comparisons of monolithic 3d ics and tsv-based 3d ics," in *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, *IEEE*, 2015.
- [42] A. T. Tran and B. M. Baas, "Roshaq: High-performance on-chip router with shared queues," in *Computer Design (ICCD)*, *2011 IEEE 29th International Conference on*, 2011, pp. 232–238.
- [43] K. Latif, A. M. Rahmani, L. Guang, T. Seceleanu, and H. Tenhunen, "Pvs-noc: Partial virtual channel sharing noc architecture," in *19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing*, 2011, pp. 470–477.
- [44] R. S. Ramanujam, V. Soteriou, B. Lin, and L. S. Peh, "Design of a high-throughput distributed shared-buffer noc router," in *Networks-on-Chip (NOCS)*, *2010 Fourth ACM/IEEE International Symposium on*, 2010, pp. 69–78.
- [45] J. Fang, J. Lu, and C. She, "Research on topology and policy for low power consumption of network-on-chip with multicore processors," in *International Conference on Computational Science and Computational Intelligence (CSCI)*, 2015, pp. 621–625.
- [46] V. S. Nandakumar and M. Marek-Sadowska, "A low energy network-on-chip fabric for 3-d multi-core architectures," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 2, pp. 266–277, 2012.
- [47] M. A. Wanas, M. A. A. E. Ghany, and K. Hofmann, "Hybrid mesh-ring wireless noc for multi-core system," in *Design and Diagnostics of Electronic Circuits Systems (DDECS)*, *IEEE 16th International Symposium on*, 2013, pp. 295–296.
- [48] P. P. Pande, R. G. Kim, W. Choi, Z. Chen, D. Marculescu, and R. Marculescu, "The (low) power of less wiring: Enabling energy efficiency in many-core platforms through wireless noc," in *Computer-Aided Design (ICCAD)*, *IEEE/ACM International Conference on*, 2015, pp. 165–169.
- [49] H. K. Mondal, S. H. Gade, R. Kishore, S. Kaushik, and S. Deb, "Power efficient router architecture for wireless network-on-chip," in *17th International Symposium on Quality Electronic Design (ISQED)*, 2016, pp. 227–233.
- [50] J. Balfour and W. J. Dally, "Design tradeoffs for tiled cmp on-chip networks in ics," 2006.
- [51] S. M. Hassan and S. Yalamanchili, "Centralized buffer router: A low latency, low power router for high radix nocs," in *Networks on Chip (NoCS)*, *Seventh IEEE/ACM International Symposium on*, 2013, pp. 1–8.
- [52] A. Mineo, M. Palesi, G. Ascia, and V. Catania, "Runtime online links voltage scaling for low energy networks on chip," in *Digital System Design (DSD)*, *Euromicro Conference on*, 2013, pp. 941–944.
- [53] A. Psarras, J. Lee, P. Mattheakis, C. Nicopoulos, and G. Dimitrakopoulos, "A low-power network-on-chip architecture for tile-based chip multi-processors," in *International Great Lakes Symposium on VLSI (GLSVLSI)*, 2016, pp. 335–340.
- [54] S. M. Hassan and S. Yalamanchili, "Centralized buffer router: A low latency, low power router for high radix nocs," in *Networks on Chip (NoCS)*, *Seventh IEEE/ACM International Symposium on*, 2013, pp. 1–8.
- [55] H. Kim, Y. Kim, and J. Kim, "Clumsy flow control for high-throughput bufferless on-chip networks," *IEEE Computer Architecture Letters*, vol. 12, no. 2, pp. 47–50, 2013.
- [56] R. Das, S. Eachempati, A. K. Mishra, V. Narayanan, and C. R. Das, "Design and evaluation of a hierarchical on-chip interconnect for next-generation cmps," in *IEEE 15th International Symposium on High Performance Computer Architecture*, 2009, pp. 175–186.