A survey of recent contributions of high performance NoC Architectures

***Abstract*— The Network-on-Chip (NoC) paradigm has been herald as the solution to the communication limitation that System-On-Chip (SoC) poses. However, power Consumption is one of its major defects. To ensure that a high performance architecture is constructed, analyzing how power can be reduced in each area of the network is essential. Power dissipation can be reduced by adjustments to the routers, the architecture itself and the communication Links. In this paper, a survey is conducted on recent contributions and techniques employed by researchers towards the reduction of power in the router architecture, network architecture and communication links.**

Keywords*— *Network-on-chip; formatting; low power router architecture; communication links NoC; Virtual channel sharing.

I. INTRODUCTION

The integration of multiple cores on a single chip has escalated since the concept of Moore's law was implemented into the development of technology. Moore's law depicts that as technology evolves, more transistors will be integrated on a single chip to improve the performance [1]. This advancement will no longer permit the traditional bus architecture to be implemented because of its limitation of providing higher bandwidth, and a scalable network. Although Buses are easy to configure, time management provides difficulties as more Processing Elements(PE) are added. The introduction of network-on-chip (NoC), a packet-switched-network communication paradigm allows packets to be exchanged between the PE's using routers and links. With NoC, as the network scales so does bandwidth which is the opposite to the busses [2].

Demand for high performance has seen the emergence of different architectures, routing algorithms and approaches been undertaken to reduce power consumptions in the router. In this paper, we analyse different techniques used to enhance the performance of the router architecture, network architecture and the communication links in NoC.

II. ROUTER ARCHITECTURE

Many have argued against the idea of using buffers in NoC routers since they contribute to power consumption in the network. This section of the paper presents techniques others have undertaken to reduce the amount of power consumed in buffers.

A. Virtual channel sharing

The use of virtual channels aids network architectures to share the physical link on a network. The physical link is divided into several virtual ports for packets to be forwarded. The impact of this to improve latency, throughput and network congestion. The rule of thumb is that, the more virtual channels used in the router, the higher the throughput and latency. However, this is at a cost of power dissipation. Although the use of virtual channel can be of an advantage, one may argue one major drawback of its use is, high power consumption. Many have proposed the idea of using bufferless routers however this is at cost of deadlock, livelock and network congestion when the packet injection rises. According to [3], recent architectures which has implemented virtual channel do not make effective use of the buffers. Packets go through different stages before it reaches its destination which consumes power; in addition to this, they are transmitted without regard to the network congestion and availability of the path. This causes packets to traverse through more stages increasing latency and power. Adjustments to the network and appropriate utilization of routing algorithm can help packets to evade the buffering stage.

The authors in [4] propose a Traffic-Based Virtual Channel Activation Algorithm (TVA). This technique groups the virtual channels in the switch port in to three cells. These three grouped cells can be deactivated by TVA depending on the network traffic or network congestion. The purpose of TVA is to disable cell groups which are not in use. In addition to this, when traffic arrives and free channels are not available for packets to be stored, the idle channel will be activated and allocated.

Catnap architecture is proposed by [5]. Multiple network design is power gated and grouped into a number of subnets. Routers with empty input buffers are automatically turned off.

The authors in [6] propose the partial virtual-channel sharing NoC architecture which shares free buffers between a limited number of ports depending on the network requirements. This is because sharing resources among all input ports increases the size of the crossbar. In this architecture, the virtual channel allocator (VC) is not in charge of allocating free buffers but rather an implemented control login and input crossbar is; VC enters in sleep mode when a connection is established for a packet to be forward to the next router thus saving power.

B. Buffer stages reduction

Another technique widely used to improve power consumption is circuit switching. Packets when injected into the network goes through several stages before traversing to the next router. These stages are buffer write(BW), route computation (RC), virtual channel allocation (VC), switch allocation (SA) and the switch traversal stage. When circuit switching is implemented, all stages other than the switch allocation is used. This helps reduce power as other stages are not used however flexibility is an issue. To solve this issue, the authors in [7] propose virtual circuit switching. Virtual circuit combines virtual channels and circuit switching together to allow flits to be sent with just one stage. The results show that 33.2% power can be saved.

The SWIFT NoC proposed by [3] 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. In this architecture heavy traffic are able to bypass the buffering stage which means less buffers are used resulting in less power consumption.

C. Power management

The amount of power consumed by buffers can also be managed. Dynamic power management units can be employed in the routers to reduce the amount of power or voltage used. The study of replacing conventional buffers by drowsy SRAM have been studied for the authors in [8]. The authors propose DimNoc, which combines SRAM and STT-RAM to reduce unnecessary power-gating operations. Rather than turning off idle buffers, routers are placed in sleep mode. How power consumption is achieved with the proposed technique is that, virtual channels are grouped into levels. The lower level of the virtual channels is designed with SRAM; the higher level is designed using SST-RAM. The use of SRAM allows the lower level to either be powered on or left in drowsy state. The high SST-RAM Virtual channels are activated when there is heavy traffic.

Similar to [4], [9] propose a power control unit which turns idle buffers off which are not active for a certain amount of cycles. The algorithm uses the status of neighboring routers, and a routing block which implements minimal routing, to find two different paths a packet can sent from source to destination. When the two different paths are concluded, the input ports of the neighbours port will be analysed and allocated a score. The packet will then be sent through the port with the highest score and the port with the lowest will be not be activated which will help save power.

Although all the proposed solutions help save power consumption when the traffic on the network is low, one may argue it does not fully solve the problem as more virtual channels will be activated when the network traffic rises. In addition to this, the deactivating of virtual channels prevents the use of adaptive routing when there is congestion in the network.

III. NETWORK ARCHITECTURE

A. *3D Architectures*

Many approach to low power consumption has seen the implementation of different architectures, first with 2D, hierarchical NoC and the emergence of 3D. 3D NoC is solely down to the limitation that 2D NoC imposes. As more cores are integrated in a 2D NoC, a rise in latency occurs. This is caused when more processing elements are implemented. More cores mean more hops counts for a packet to reach its destination. 3D NoC provides a solution by allowing multiple silicon layers to be stacked together. With lengthy wires reduced to short wires through silicon vias (TSVs), the number of hops are minimised however due to size of the chip, they experience a surge in power. TSV is relatively new so more research is being conducted to reduce the failures they suffer in the architecture. According to [10], the increase in number of links in 3D ICs allows the transmission of more messages around the network. One may argue the introduction of 3D IC's allow the combination of different type of technology such as digital, analogue, memory etc. to be merged together thus enhancing the performance of the chip, however consumes more power [11] [12]. The power consumption can however be reduced according to the concluded results in [13]. The reduced in wire length can save up to 37% of power if monolithic 3D is implemented.

In [9], a thorough investigation is conducted into analysing the performance of 2D and 3D NoC. Based on the results concluded, 3D NoC has a percentage increase of 21.43% meaning the network is able to contain more flits allowing an increase in the transmission of messages across the network. Because routers consume a lot of power, many proposed architecture looks to use less routers and more cores. Debora Matos et al. proposed the 3D HiCIT architecture, which in comparison with the traditional 3D-SPIN and 3D Mesh topologies reduces the average latency to 50% and 54% respectively, with the 3D-SPIN been the latter [11].

The architecture is comprised of two hierarchical levels with a mesh topology at the top level; The routing algorithm used is XYZ routing. Each router consists of seven (7) ports which are Local, North, West, East, South, Up and Down. In addition to this, the architecture is comprised of less routers and TSVs. What has been utilized can reduce power consumption of the network. The less routers used in a network the less power is consumed.

B. Utilization of less routers

Juan Fang et al. proposed RRCIES which is based on a mesh topology. RRCIES allows many cores to be connected to the network through one router thus minimizes hop distance and the utilization of less buffers. One router allows four cores to be connected to the network reducing the power in the

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

network since less routers are used [14]. In addition to this, because less routers are used, this means a reduction in number of buffers, crossbar switches and virtual channels which consumes a lot of power.

The authors in [15], conducted a study of vertical slit field effect transistors (VeSFETs) and propose a 3D Hybrid architecture which improves power consumptions in 3D networks. Buffers from ports which are not activated are shared among busy ports. The proposed architecture splits the input buffers into three (3) levels. Each input port is designed to access all three levels and permits any virtual channel destination to be chosen. The architecture is designed in a way that a level can be switched off by applying power and clock gating.

The architecture of a network has a massive effect on the performance of the network. To gain high performance, analysis has to be on carried on the most effective way to develop a low powered NoC. The presented architectures provide different ways researches have undertaken to improve power consumption. Whether it been the utilization of less routers or power-gating techniques, power consumption can be achieved through ways.

IV. COMMUNICATION LINK

A. *Power managing in links*

Although power consumption highly takes place in the router, the power the links consumes in the network can be reduced to accommodate this. Several techniques have been proposed which adopt different techniques to reduce power consumption.

According to [16], the voltage swing in the communication links have severe impact of the power consumed in the network which in fact can be reduced. However, reducing the voltage swing will result a surge in error bit rate. The proposed technique allows the link to work at two different levels. Communications which do not need to be secured because the error bit rate is low can be sent using low level voltage swing. The head flit of a packet is normally sent using normal voltage swing because of the data it carries and the rest of the flits are sent using low voltage.

Other architectures have employed the use of algorithms which can predict the traffic between routers. According to the authors in [17] adjustment of the link leads to low power consumption. To do this, they have developed the adaptive ATPT. ATPT allows the voltage of a link to be adjusted.

Photonic links are emerging as the default links used for transmission of data. With low loss in optical waveguides and bit rate transparency, if carefully researched into, they could replace the traditional metallic interconnects [18] [19]. Their ability to provide high bandwidth has seen a lot of research into its area. Laser power dissipation however it's an issue with architectures developed with them. The authors in [20] propose a power management to save the amount of power loss in the laser source by reducing the number of L2 cache banks utilized. This allows the photonic links associated L2 cache banks to be turned off.

B. *Half cycle flits*

The authors in [21] utilized a technique which allows flits to only use a half cycle to hop between routers. The effect of this is that the less a flit spends in the link, the more power is used. 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.

Based on the techniques presented, the power dissipated by the communication links can be reduced by allowing flits to traverse a lot quicker, by turning off links associated with idle resources and also employing dynamic techniques which can manage the amount of voltage used to transmit a flit.

V. CONCLUSION

In this paper, several techniques have been presented which can help reduce power dissipation in NoC. The combination of some of the architectures presented if employed can help improve the amount of power consumed by resources which can either be removed or switch off. The communication links of a network can be adjusted to only transmit flits with half the voltage it normally utilizes thus reducing power; the architecture can also be employed with less routers yet many cores can still be connected in the network and also the use of control units to share virtual channels can help reduce the amount of power used for buffering. Based on our discussions, we can conclude that 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," *2011 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems*, Vancouver, BC, 2011, pp. 445-453.

[2] Ahmed Ben Achballah, "A Survey of Network-On-Chip Tools", 2013.

[3] 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," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 8, pp. 1432-1446, Aug. 2013.

[4] 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," *Energy Aware Computing Systems & Applications (ICEAC), 2015 International Conference on*, Cairo, 2015, pp. 1-4.

[5] R. Das, S. Narayanasamy, S. K. Satpathy, and R. G. Dreslinski, "Catnap: energy proportional multiple network-on-chip," in Computer Architecture (ISCA), Proceedings of the 2013 40th Annual International Symposium on. ACM, 2013, pp. 320–331.

[6] K. Latif, A. M. Rahmani, K. R. Vaddina, T. Seceleanu, P. Liljeberg and H. Tenhunen, "Enhancing Performance of NoC-Based Architectures Using Heuristic Virtual-Channel Sharing Approach,"*2011 IEEE 35th Annual Computer Software and Applications Conference*, Munich, 2011, pp. 442-447.

[7] S. Shenbagavalli and S. Karthikeyan, "An efficient low power NoC router architecture design," *2015 Online International Conference on*

Green Engineering and Technologies (IC-GET), Coimbatore, 2015, pp. 1-8.

[8] J. Zhan, J. Ouyang, F. Ge, J. Zhao and Y. Xie, "Hybrid Drowsy SRAM and STT-RAM Buffer Designs for Dark-Silicon-Aware NoC," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 10, pp. 3041-3054, Oct. 2016.

[9] N. Nasirian and M. Bayoumi, "Low-latency power-efficient adaptive router design for network-on-chip," *2015 28th IEEE International System-on-Chip Conference (SOCC)*, Beijing, 2015, pp. 287-291.

[10] Alexander Wei Yin, "Change Function of 2D/3D Network-on-Chip", 2011.

[11] Deborah Matos, "Performance Evaluation of Hierarchical NoC Topologies for Stacked 3D ICs", 2015.

[12] A. Bose, P. Ghosal and S. P. Mohanty, "A Low Latency Scalable 3D NoC Using BFT Topology with Table Based Uniform Routing," *2014 IEEE Computer Society Annual Symposium on VLSI*, Tampa, FL, 2014, pp. 136-141.

[13] Deepak Kumar Nayak, Svrinivasa Banna, "Power, Performance and Cost Comparisons of Monolithic 3D ICs and TSV-based 3D ICs.

[14] Juan Fang, "Research on Topology and Policy for Low Power Consumption of Network-on-chip" with Multicore Processors, 2015.

[15] V. S. Nandakumar and M. Marek-Sadowska, "A Low Energy Network-on-Chip Fabric for 3-D Multi-Core Architectures," in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 2, pp. 266-277, June 2012.

[16] A. Mineo, M. Palesi, G. Ascia and V. Catania, "Runtime Online Links Voltage Scaling for Low Energy Networks on Chip," *Digital System Design (DSD), 2013 Euromicro Conference on*, Los Alamitos, CA, 2013, pp. 941-944.

[17] H. K. Mondal, S. H. Gade, R. Kishore and S. Deb, "Adaptive multi-voltage scaling in wireless NoC for high performance low power applications," *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, 2016, pp. 1315-1320.

[18] The international technology roadmap for semiconductors (ITRS), http://www.itrs.net/, interconnections, 2011.

[19] R. Ramaswami and K.N. Sivarajan, Optical Networks: A Practical Perspective 2002 edition.

[20] R. Ramaswami and K.N. Sivarajan, Optical Networks: A Practical Perspective 2002 edition. H.X. Gu, K.H. Mo, J. Xu, and W. Zhang, A low-cost optical router for optical networks-on-chip in mulitprocessor systems-on-chip, 2009 IEEE Computer Society Annual Symposium on VISI 1924(2009).

[21] A. Psarras, J. Lee, P. Mattheakis, C. Nicopoulos and G. Dimitrakopoulos, "A low-power network-on-chip architecture for tile-based chip multi-processors," *2016 International Great Lakes Symposium on VLSI (GLSVLSI)*, Boston, MA, 2016, pp. 335-340.