Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

Power Optimization Techniques for NOC

Abhinav Bijapur[1], Sumeet Siddappa Shirahatti[2], Dr. R. Jayagowri[3]
Department of ECE, BMS College of engineering
Bengaluru, India

***Abstract*— The on-chip network has become a significant solution for the communication limitation of SoC (System-on-chip). The demand for relative increment in bandwidth to facilitate high core utilization and the need for low power consumption as well as higher performance has increased. The major circuitry is the router in NoC, which barely affects on power dissipation, latency, and performance. The dynamic power consumption is one of the major components of total power consumption. This paper presents a detailed structure and verification of the router module and various power optimization techniques for NOC by restructuring the architecture. The design of the router is coded in Verilog, synthesized, and simulated in the Xilinx ISE Design Suite 19.1 tool.**

Keywords—: Network-on-Chip, SoC, Deadlock, Power Gating

I. INTRODUCTION

As technology is reaching a new height, memories and processors are becoming faster, smaller, cheaper, and energy efficient. This helps computer architects to incorporate a greater number of specifications in a single chip [1]. As Moore's Law is moving towards its limit, the intimation of simple cores is getting used to continue improving performance while reducing fabrication costs. Consequently, not only the computing power and memory access but also in inter-core communication are the bottleneck for performance.

As the Semiconductor industry is advancing towards a complex system on chip (SoC) design containing hundreds of IPs, the traditional bus architecture as shown in the figure.1, used for communication between the multiple cores. Bus architecture introduces a lot of latency [3] and area overhead in the design. As technology is scaling to a lower node, the need for high performance, higher efficiency, and low power design become the bottleneck. To satisfy these criteria of SOC, the bus-based communication lost its significance and there is a need for a better solution.

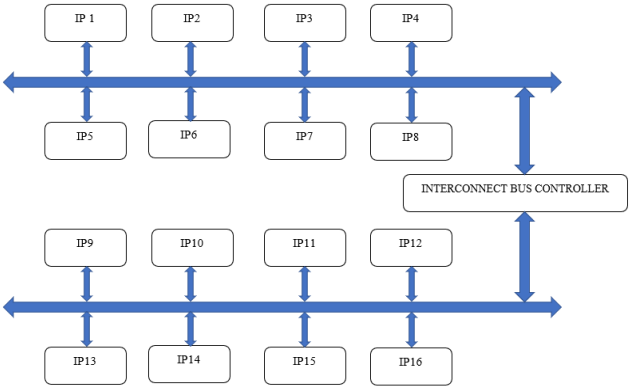

Figure-1: Classical bus-based architecture for multicore chip communication

As the technology is scaling down, the need for high performance, higher efficiency and low power design becomes the bottle neck. To satisfy these criteria of SOC, the bus-based communication lost its significance and there is a need for better solution.

The interconnected networks have become evident to replace traditional buses as the prevailing solution to provide fast, low cost, and scalable communications. They are the solution for successful future digital systems, both chip multiprocessors consist of identical cores and heterogeneous systems-on-chip. From the last ten years, there has been an in-depth effort towards optimizing the networks-on-chip (NoCs) from low-level physical aspects up to system-level and application-related problems. NoCs have currently reached a mature level of development with their integration as a fundamental component.

The Network on chip architecture differs based on the technology and application-specific. There are many topologies for NoC [2] like Chip level integration of communicating heterogeneous elements (CLICHE), Butterfly Fat Tree (BFT), Scalable programable integrated network (SPIN) and Mesh (Figure.2). Every topology has its own advantages and disadvantages. The mesh topology has the following advantages over other topologies.

a. Each node is connected to 4 neighbours, except for the ones on the boundary.
b. Easy to layout on-chip: regular and equal length links.
c. Path diversity: Many ways to get from one node to another.

The NoC consists of three main components and they are Routers, Interswitch links, and Repeaters. The wire linking two Router in NoC is called as interconnects. Routers are an important part of the NoC design. The performance of NoC depends on how effectively the router works.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

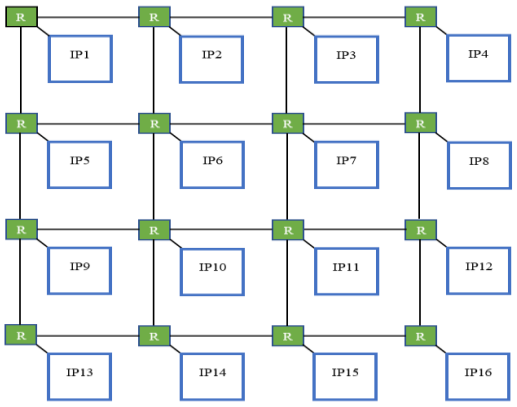

Figure-2: Typical Network on Chip (NoC) mesh topology

II. ROUTER MICROARCHITECTURE & FLOW CONTROL

A. *Flow Control*

Flow control is the method through which an upstream router will know about the availability of buffer in the downstream router. The router sending the data packets is called an upstream router, the router receiving the data packets is called a downstream router. This is done by exchanging credits between the two routers. If the buffer is available, a packet can be forwarded to the next router after the acknowledgment signal from the downstream router to the upstream router.

B. *Wramhole flow control*

Data packets are divided into smaller units called flits. The data packet consists of three flits, head, body, and tail. Flits are sent across the fabric in wormhole fashion. Body flit follows the head flit, tail follows the body flit and it happens in a pipelined manner. Figure 3 represents the wormhole flow control mechanism. If the head flit is blocked then the rest of the packet is stopped, since the information about the destination (Routing information) is given only to the head flit.

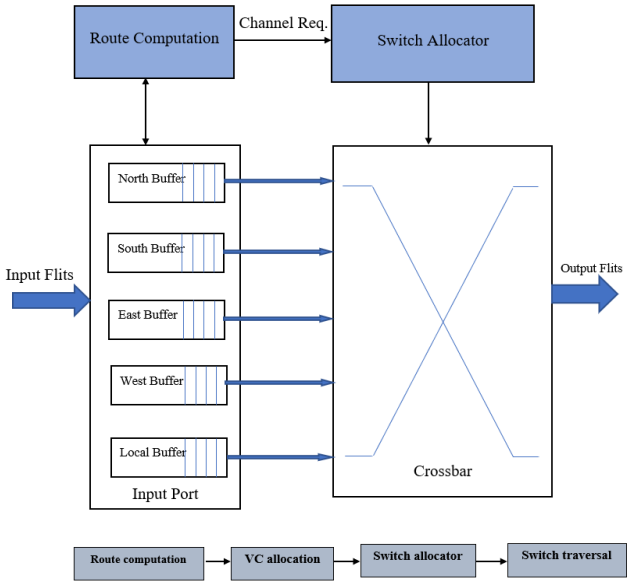

Figure-3: Wormhole flow control mechanism

This type of flow control mechanism has lower latency and since we are not reserving buffers for the entire packet, it has efficient buffer utilization. In this type of flow control, the packet can occupy resources across multiple routers. This wormhole routing has one problem called head of line blocking. If the head flit cannot move due to contention, another worm (flit) cannot proceed even though links are may be idle.

C. *Virtual Channel Flow Control*

As discussed above, to avoid the head of line blocking, the Virtual Channel Flow Control mechanism has been proposed. The one physical channel has been multiplexed with multiple virtual channels. Single FIFO buffer is replaced with multiple buffers. A single physical channel is terminating at multiple buffers. These buffers are known as virtual channels. This concept is known as Virtual Channel Flow Control. Virtual Channels (VC) number is allocated once at each router to the head flit and the rest of the remaining flits of the packet inherit the same VC number. Now the flits of different packets can be interleaved on the same physical channel.

D. *Router Microarchitecture*

As shown in figure 4, the routers consist of Input ports. There are five input ports North, East, West, South, and Local. Each port has virtual channels (depending on the design). The crossbar connects the input ports to the output. The control logic facilitates the smooth flow of packets from the input side to the output side.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

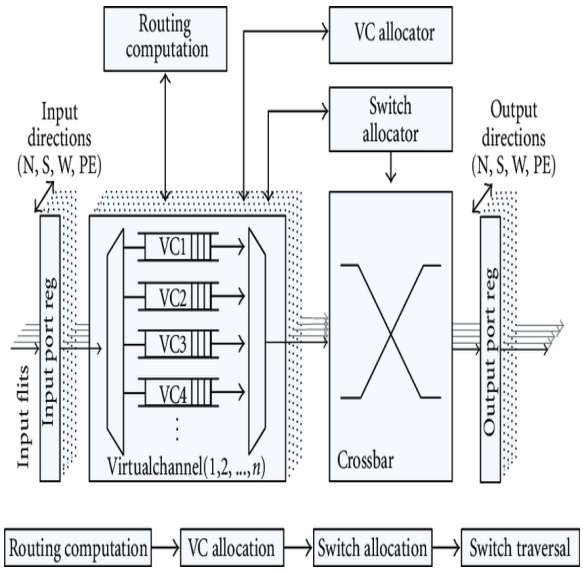

Figure-4: Virtual flow control mechanism[4]

The first function is the buffering of flits. Whenever a flit is coming to the channel, it occupies the buffer. The next task is route computation. For a flit that is residing inside a buffer, the route computation unit is going to assign the output port. The process of finding an output port for a packet residing inside a buffer is called route computing. Route computation is done for the head flit. The body and tail flit inherit the route assigned to the head flit. The next task is the virtual channel allocation. It is based upon handshaking between adjacent routers.

E. Router Algorythm

1. ΔPin = {ΔN,ΔS,ΔE,ΔW,ΔL}
2. ΔPout = {ΔN,ΔS,ΔE,ΔW,ΔL}
3. ΔD = {ΔH, ΔB, ΔT}
4. If ΔD == ΔH
5. RCUnit.NextΔP = ΔPout.AvoidDeadlock
6. If ΔBuffer == Δ(empty||~full)
7. VCAllocator = Assign VCNumber
8. If (ΔDh1, ΔDh2, ΔDh3,… ΔDhn) = ΔPout1
9. SWAllocator.HighestPriority(ΔD1, ΔD2, ΔD3,… ΔDn) = ΔPout1
10. Else ΔDh = ΔPout
11. Else RCUnit.NextΔP = ΔPin
12. If ΔD == (ΔB||ΔT)
13. Inherit VCNumber from ΔH
14. If (ΔDbt1, ΔDbt2, ΔDbt3,… ΔDbtn) = ΔPout1
15. SW Allocator. Highest Priority (ΔD1, ΔD2, ΔD3,. ΔDn) = ΔPout1
16. Else ΔDbt = ΔPout
17. Else RCUnit.NextΔP = ΔPin

The process of reserving buffer in the downstream router is called virtual channel allocation. The next task is the switch allocation. Whenever multiple flits require the same output port, the switch allocator chooses the flit. Once the flit has been chosen, the next task is to switch traversal. At most 5 flits can be traversed at any given clock cycle. These are the 5 logical cycles that happen inside the router.

III. IMPLEMENTATION OF POWER TECHNIQUES IN NOC

As technology is scaling down to deep Nanometer, gradually scales down threshold voltage which contributed enormously towards an increase in subthreshold current leakage, therefore, making high power dissipation. In Multi-core chips, transmission requires more bandwidth and speed in between IPs, due to an increase in more feature requirement. So NoC is one of the solutions which allows higher transmission speed in between core IPs, facilitates the latest technological trends like AI, Machine learning, etc. The development of NoC provides higher throughput, lower latency, and higher bandwidth. However, NoC suffers from power consumption caused by Leakage power and switching activity in multi-core circuitry.

To overcome this problem, the usage of Power gating in NoC will help. Buffers used as storage module for flits, which provides higher bandwidth and helps to avoid deadlock. In some recent research specifies [7] that buffer-less NoC helps in reducing power and area, but has trade-off for deadlock, low latency, and live lock. In the recent invention, power dissipation can be reduced with buffers included, by making some modification and adding circuitry in Integrated circuits.

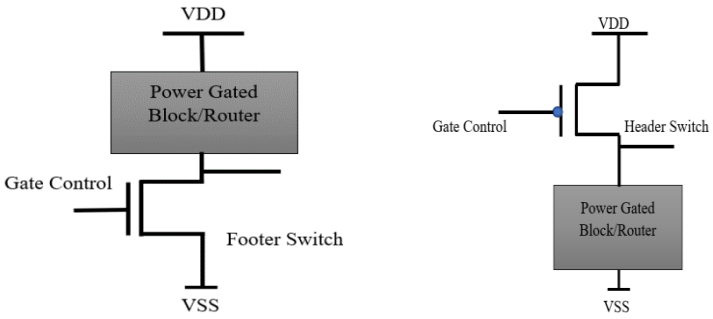

Figure-5: Power gating Block. a) footer switch b) Header switch

To reduce power dissipation in circuits, Power optimization techniques like Power gating, Clock gating, as well as multi vt (threshold voltage) and multi Vdd (Supply voltage), etc are suitable. Figure .5 represents traditional Power gating blocks. Power gating is a technique which shuts of the flow of current in Integrated circuits. From figure 5, the Gating of power can be done in two different ways, which is connecting NMOS and PMOS switch to the circuit. Footer switch contains NMOS connected between power gated block and Vss. Header switch power gating contains PMOS connected between Vdd and power gated block. Gate control signal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

helps, a suitable switch can be used in NoC. The power equation represented as:
Total power is given by the equation

P_{total} = Pswitching + Pshort-circuit + Pleakage (1)

IV. PROPOSED WORK

The present work determines the reduction of power consumption in core architecture as well as an increase in signal transmission-speed by implementing NoC. The proposed work contains several routers connected in mesh architecture and verified for its functionality. Each router is connected to different IP's, here IP's represents the processor, RAM, ROM, cache memory, sensor unit, DSP, Display Processor, WI-FI, GPU, etc.

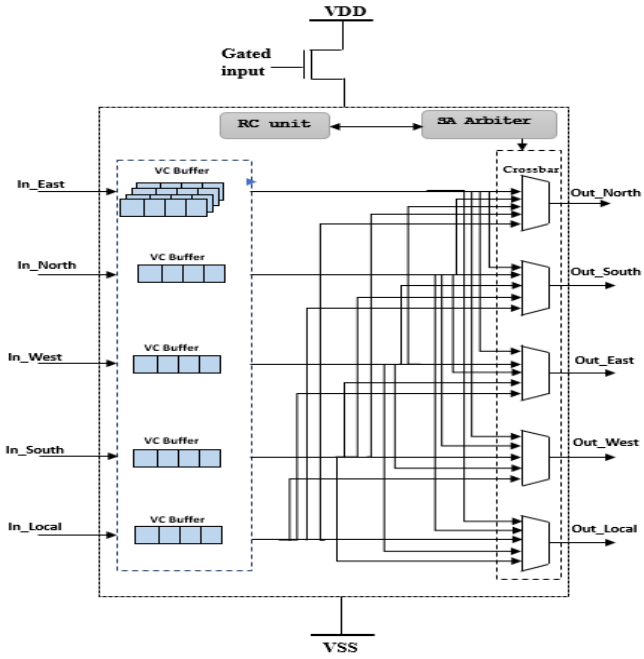

Figure-6: PG controller to Router

Figure 6. represents a detailed view of Power Gating applied to the router block. Power gating is connected between supply and enable pin of VC and switch allocator. which controls the active state and sleep state of the router. It works based on activating and deactivating router to save power.

The figure 7. represents the structure of 4*4 mesh topological NoC. In the below block all routers are active (ON) state, data flits are transmitting from source (upstream) router to destination (downstream) router. The transmission path will be a single path, path transmitting flits is the shortest path from the source router to the destination router. The flit traveling path shown in the dotted line figure 7

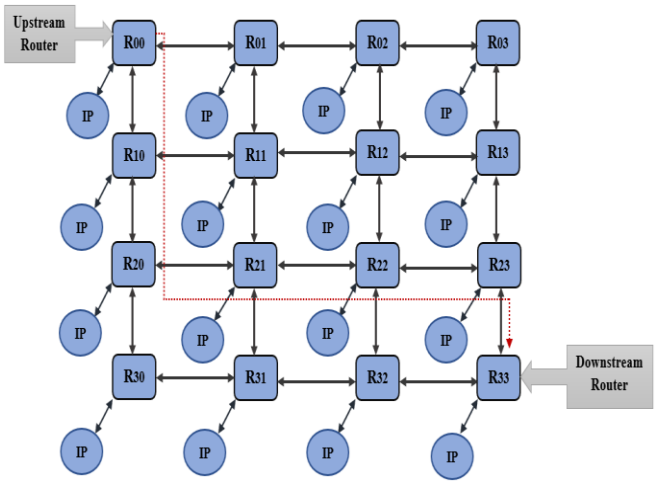

Figure-7: All routers in active-state

Except for the conducting routers, remaining routers are in idle-state until new data flits use idle router path. The routers in idle state dissipate power and increase device heating unnecessarily. For the betterment of device performance, the idle routers to be turned "off" or deactivate to save power. It is explained in the below section.

A. *Applying Power Gating to routers*

As mentioned in figure 8, while the data flits are transmitting from upstream router (transmitting router) to downstream router (receiving router), routers in this path will be inactive state by turning "off" PG. The destination router address is located at head flit of data, depending on the flit transmission path respective routers on that path will activate and few neighboring routers also turned “on” to avoid deadlock and latency problem.

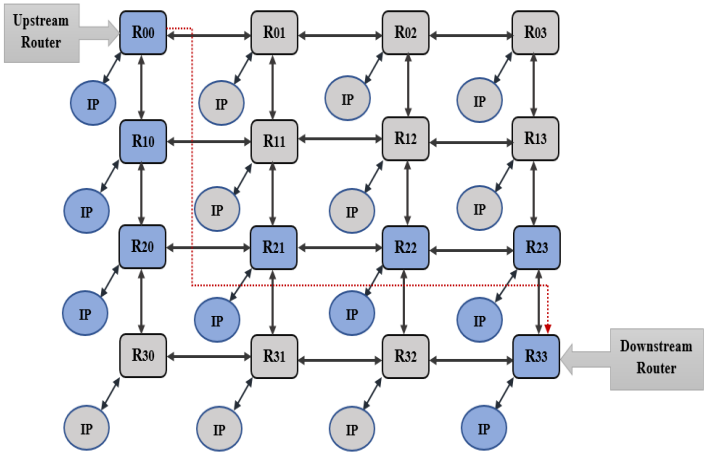

Figure 8: Selective Routers under active state

Figure 8. shows the Detailed structure of 4*4 mesh router. All routers can be turned “on” and “off” eventually using Power gating. As shown in figure 8 only selective routers are in an active state, where transmission of flits from source (upstream) router to destination (downstream) router is processing. The diagram shows R00 (router00) transmitting

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

data to R33 (router33) through routers R10, R20, R21, R22, R23, R33. The path allocated by route computation base on low latency and to avoid the deadlock. If already a few paths are transporting data flits, the deactivated routers made "on" if any new data flit arrives. Eventually, this methodology saves power consumption exceptionally but keeping the routers continuously in an active state consumes power.

The activation of PG is controlled by VC and data flits. After receiving all flits, if there is no further data transmission for a few cycles, then routers will go in sleep mode. A quick incoming flit during router is in the sleep state can lose the data packet or can cause latency in the transmission of data because turning on the power gating at the same cycle time will increase the latency. To avoid this problem, there is a methodology that generates an early notification for the routers.

B. *Early Notification to PG*

To avoid the problem of missing data due to the router off (sleep state), the proposed technique works in order to turn on (active state) the router before the flit is reached. The upcoming router in a path will get notified 5 cycles earlier to wake up to capture data. If the upstream router starts transmitting data to the downstream router, the downstream router gets notified to be an inactive state.

C. *Slotting Routers*

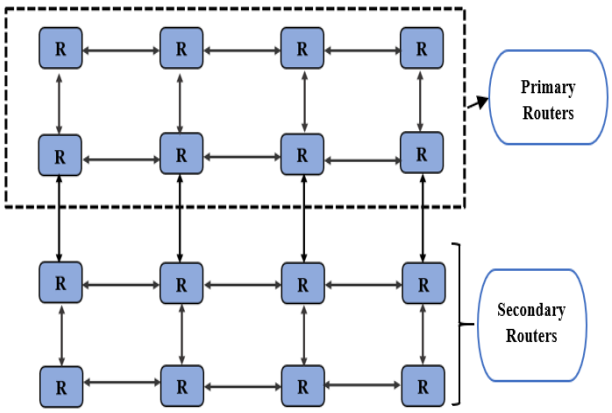

Figure 9: Router Slotting block

To overcome the drawback of keeping unused routers on, we propose a method to overcome the drawbacks of keeping unused routers ON. The routers are divided into Different slots like Primary routers and Secondary routers. Microprocessor core has few main regular IPs like CPU, memory, Execution units which do not need to be turned off for regular iterations. Other units like GPU, DSP, Sensor processing can be turned OFF. Major power loss happens by supplying power to units that are non-functional. So, the main routers will act as primary routers and remaining can be slotted as secondary routers. Primary routers are continuously in an active state and secondary routers in the sleep state. If a router fails to conduct completely, then one of the secondary routers acts as a primary router and provides complete functionality. We also propose a memory storage table, so it can be easy to router identify which neighboring router is in ON state and which router has failed in its functionality. The table has a router transmission path memory status from the upstream router to the downstream router.

D. *Route path Memory table*

The memory table stores the information about the shortest path information of the source router to the destination router. For on-chip Networks in the core, there are possibilities of router failures, in that case, memory table stores the failed router information and avoids other routers to consider the failed outer path. The memory table will store the shortest flit transmission path to avoid latency and improve efficiency. In case of a router failure, the automatically upstream router finds a new path to reach the downstream router, the updated path will be stored in the memory table.

V. EXPERIMENTAL RESULTS

For the proposed method simulation and synthesis conducted using Xilinx Vivado version V.2019.1. The power experiments with the proposed methodology are conducted on 2*2, 2*3, 4*4, 8*8 mesh topology routers. NoC model implemented using System Verilog. Experiments have been conducted on mesh NoC topology by applying Power gating and without power gating. Total power is the addition of dynamic power & static power.

Table.1 Power dissipation comparison

Sl No	Router	Dynamic Power(µW)		Static power(µW)		Total power(µW)	
		Without Power Gating	With Power Gating	Without Power Gating	With Power Gating	Without Power Gating	With Power Gating
1	2*2	47.603	40.912	0.102	0.091	47.705	41.003
2	2*3	93.779	82.293	0.131	0.121	107.3	93.91
3	4*4	215.34	174.4	1.735	1.09	217.09	175.498
4	8*8	871.36	697.6	6.98	4.06	878.34	701.66

The comparison between the without power gating model and with power gating model, of different size Mesh topologies with and without power gating for dynamic and static. By the experimental results, a number of routers increase the buffer count in each router increases, a high number of buffers consume more power. Routers are kept in an active state for a period of cycle time during packet transmission between them. once transmission completes routers wait for 5-6 cycles ideally, then go into sleep mode.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

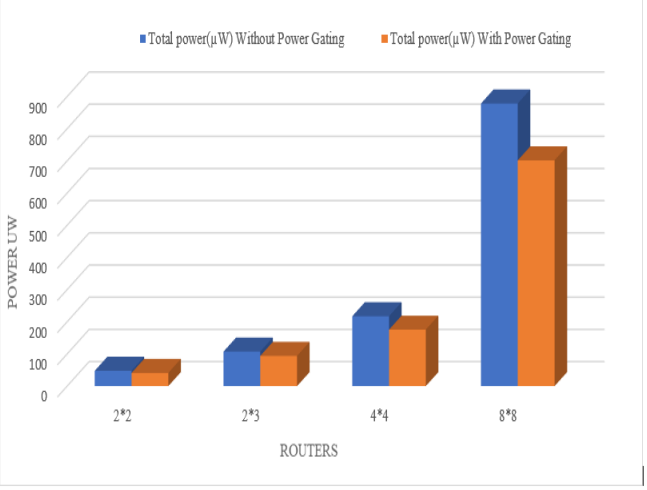

Figure-10. Total power comparison graph

The figure 10 is a graphical representation of total power consumed by the routers. The graph shows the difference between without power gating and with applying power gating at routers, for deterministic input path. Here an increase in router count increases power dissipation, measured in uW (microwatts). But comparative applied methodology helps in the reduction of power.

VI. CONCLUSION

The high performance and less power consuming NoC for multicore Integrated circuits have become important. We did a detailed study of NoC blocks, like router VC allocator, input buffer, route computation, Switch arbiter. Router with 2*2, 2*3, 4*4, 8*8 mesh topologies has been implemented in this paper, the total power consumption in routers without power gating and the power reduction in addition to power gating. Buffer is a major power-consuming part of the router. Power optimization techniques improve the efficiency of the circuit and reduce the power consumption. Also slotting the routers as a primary and secondary use for high-end applications that needs continuity of power with turning "off" them, in case of a router failure, the system considers alternative router by excluding the old router from path memory list. It contributes to avoiding the deadlock problem and reduces latency in interconnects.

REFERENCES

[1] Marta Ortín Obón, "Network On-Chip – From the optimization of traditional NOC's to design of design of Emerging optical NOC's" from webdiis.unizar.es, 2016

[2] G. Reehal and M. Ismail, "A Systematic Design Methodology for Low-Power NoCs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 22, no. 12, pp. 2585-2595, Dec. 2014.

[3] Juyeob Kim, Miyoung Lee, Wonjong Kim, Junyoung Chang, Younghwan Bae and Hanjin Cho, "Performance analysis of NoC structure based on Star-Mesh Topology," 2008 International SoC Design Conference, Busan, 2008, pp. II-162-II-165.

[4] Wen-Chung Tsai, Ying-Cherng Lan, Yu-Hen Hu, and Sao-Jie Chen "Networks on Chips: Structure and Design Methodologies," Journal of Electrical and Computer Engineering, 2012, 1–15.

[5] LizhongChena, DiZhub, MassoudPedramb, TimothyM.Pinkstonb, "Simulation of NoC power-gating: Requirements, optimizations, and the Agate simulator" 2016 Elsevier 0743-7315

[6] Ofori-Attah and M. O. Agyeman, "A survey of recent contributions on low power NoC architectures," *2017 Computing Conference*, London, 2017, pp. 1086-1090.

[7] X. Xiang and N. Tzeng, "Deflection Containment for Bufferless Network-on-Chips," 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Chicago, IL, 2016, pp. 113-122, doi: 10.1109/IPDPS.2016.17.

[8] Chung-Kai Hsu, Kun-Lin Tsai, Jing-Fu Jheng, Shanq-Jang Ruan, and Chung-An Shen, "A low power detection routing method for bufferless NoC," International Symposium on Quality Electronic Design (ISQED), Santa Clara, CA, 2013, pp. 364-367, doi: 10.1109/ISQED.2013.6523636.

[9] LizhongChena,DiZhub,MassoudPedramb,TimothyM.Pinkstonb,"SimulationofNoCpower-gating:Requirements,optimizations,and the Agate simulator "J. Parallel Distrib. Comput. 2016 Elsevier

[10] Feng Wang, Xiantuo Tang, and Zuocheng Xing Journal of Electrical and Computer Engineering Volume 2015, Article ID 862387 link http://dx.doi.org/10.1155/2015/862387

[11] Emmanuel Ofori-Attah, Michael Opoku Agyeman "A Survey of Recent Contributions on Low Power NoC Architecture" Computing Conference 2017 18-20 July 2017 London, UK

[12] 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 Link http://download.intel.com/pressroom/kits/Teraflops/Teraflops_Research_Chip_Overview.pd

[13] FengWang, XiantuoTang, and ZuochengXing "Applying Partial Power-Gating to Direction-Sliced Network-on-Chip" Hindawi Publishing Corporation Journal of Electrical and Computer Engineering Volume 2015, Article ID 862387, 16 pages