Segmented Buffer for NOC Router

Biman Debbarma and S. N. Pradhan

***Abstract*—As the number of transistors in a single chip is increasing day by day, SoC is a very popular choice among the developers as it gives more speed and functionality. The IPs in the SoCs can be from different vendors and may be used for a third party for manufacturing a chip. For effective communication among the IPs or more commonly Cores, an in built network structure called Network on Chip (NOC) has been developed over the years and different versions of NOCs are available in the present day. This paper proposes a segmented buffer structure for NOC router for effective utilization of the NOC buffers.**

***Index Terms*—Router, ViCHAR, segmented buffer, segmented control logic, power gating.**

I. Introduction

An NOC router has many variations in this day depending upon their routing techniques, communication methods etc. In this paper wormhole routing is taken into consideration which consists of Virtual Channels. Generally each router is connected to five ports. Four in the cardinal directions which are in turn connected to other routers and the fifth is connected to its own IP. The job of the router is to deliver a packet to its correct destination depending upon the routing table (either static routing table or dynamic routing table). In this paper a segmented buffer structure for NOC has been introduced. The rest of the paper is as follows- in Section II, a summary of the generic router architecture is given. in Section III different routing techniques. Section IV explains Power Gating technique. Concept of Wake up latency is explained in Section V. Section VI emphasizes on ViChaR (a dynamic virtual channel regulator) architecture. In Section VII a brief literature survey is given and finally the proposed architecture is explained in Section VIII.

II. NOC Router

An SoC chip is a collection of homogeneous or heterogeneous modules called IP cores. Each IP is connected to a router and all the routers are interconnected to form a network of routers. When an IP module wants to send a packet to another IP module, first the packet goes to the router associated with the originating module. A routing algorithm is applied on the packet and it is transferred to another adjacent router. That router does the same operation and the packet traverses through various routers until it reaches the destination.

A generic NOC router consists of the following components: Virtual Channels, Virtual Channel Arbiter, Switch Allocator, Crossbar, routing computation unit and I/O ports. Fig. 1 shows a complete NOC router structure [1] and Fig. 2 shows a generic router buffer [2].

A. Virtual Channels

Virtual channels are collection of buffers which provide alternate path for flits which may not proceed further due to congestion in some route though its destined route is not busy. It gives flexibility to the structure and throughput is greatly increased.

B. Virtual Channel Arbiter

VCA assigns a Virtual Channel to particular flit if more than one flit are requesting for the same VC.

C. Switch Allocator

SA grants permission to a certain VC to use the crossbar switch among other VCs.

D. Crossbar Switch

Crossbar Switch is an N×N switch matrix which transfers the flits from a VC to the destined output port.

E. Routing Computation Unit

The RC unit transfers the header flit to its destination I/O port. Other flits of that packet follow the header flit.

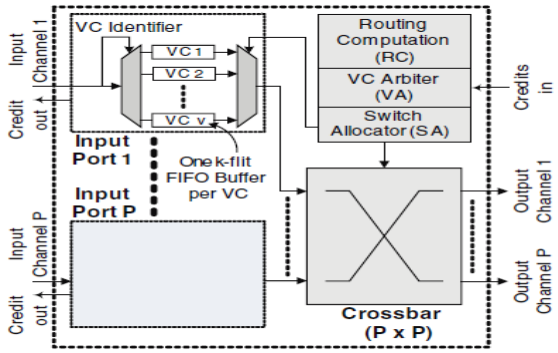

Fig. 1. An NOC router [1].

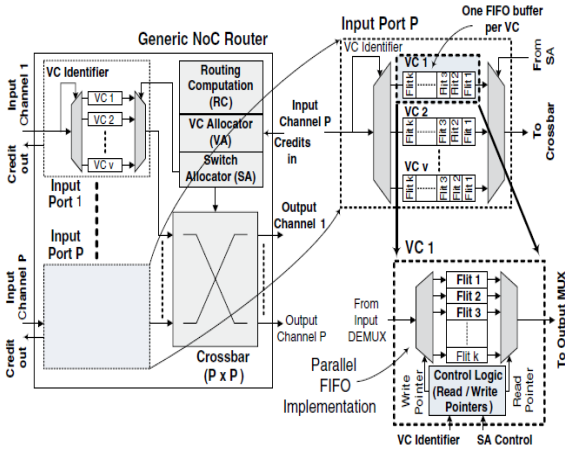

Fig. 2. A generic NOC buffer (courtesy nicopoulos *et al*. [2]).

Manuscript received June 27, 2014; revised September 23, 2014.

Biman Debbarma and S. N. Pradhan are with the National Institute of Technology Agartala, India (e-mail: bimandebbarma75@gmail.com).

DOI: 10.7763/IJMO.2015.V5.431

III. Various Routing Techniques

A. XY Routing

This is a type of routing algorithm in which flits first moves in x-direction and then in y- direction to the receiver. In this the addresses of the routers are given in XY-coordinates. One of the advantages of this routing algorithm is that it never runs into deadlock or live-lock.

B. West First Routing

In this type of routing, flits move to the west direction first to reach its destination router. And if the flits take any other direction then further west turn is prohibited.

C. Negative First Routing

In this type of routing algorithm, flits move in negative direction i.e. either west or south first. These turns are not allowed in future.

D. Odd Even Routing

In this type of routing algorithm no flit is allowed to make in East-North turns & East-South turns if origination router is in even column. In case of odd column North-South turns & South-West turns are prohibited.

E. Dyad Routing

This routing algorithm is totally dependent upon network traffic condition. It is the combination of deterministic & adaptive routing algorithm. Basically it switches between adaptive & deterministic routing algorithm depending upon the traffic. In this algorithm, each router in the network continuously checks its local network load and chooses algorithm type based on this information. When the network is not congested, a DyAD router works in a deterministic mode.

IV. Power Gating

In a SoC, except the processing element, routers themselves consume a considerable amount of power. When there is no traffic, still 35.7% of the standby power is consumed by the routers at 75° C [3]. As the technology scales down, dynamic power will decrease but as the number of elements keep on increasing, leakage power will increase. So attempt should be made to decrease this leakage power. Previous works have been done for this purpose. Among them, DVFS method [4], [5], power gating techniques [6]-[9]. Power gating is mainly of two types: coarse-grained and fine-grained. In coarse-grained power gating, a particular IP module can be turned on or off using power switches. This method is simpler to implement. But for routers, more control is required over the router elements.

For this purpose, fine-grained power gating technique is well suited. In [8], the elements are divided into various micro power domains which share the same signal.

Therefore, unlike coarse-grained power gating, power switches are inserted in the micro power domains between VGND and GND. It provides greater flexibility to control the power. On the other hand, more power switches means more power consumption. Therefore, this trade off has to be kept in mind while implementing a fine-grained power gating design. Fig. 3 shows fine-grained power gating discussed above.

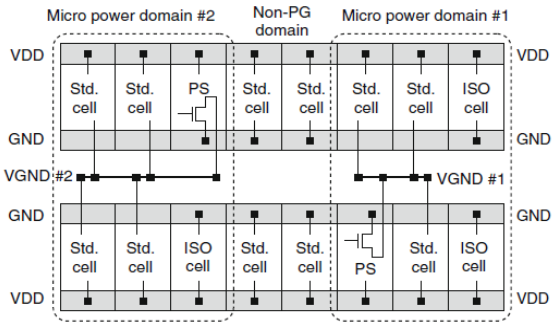

Fig. 3. Fine-grained power gating [3].

V. Wake up Latency

Another important aspect of power gating technique is the wake up latency. If the power switches takes too long to wake up, there will be pipeline stall. If the power switches wake up too early, they will add up to the dynamic power consumption. Therefore, certain mechanism should be there for maintaining an optimum wake up latency. Various approaches have been adopted like Look ahead method [10], where the output port of a flit is calculated in advance to reduce wake up latency. Another technique called Look ahead with ever-on may be used where the first hop is always on and no wake up signal is required for that. In Look ahead with active buffer method a part of the buffer is always on. Generally VC buffers use Active buffer window method. Other modules use Look ahead method.

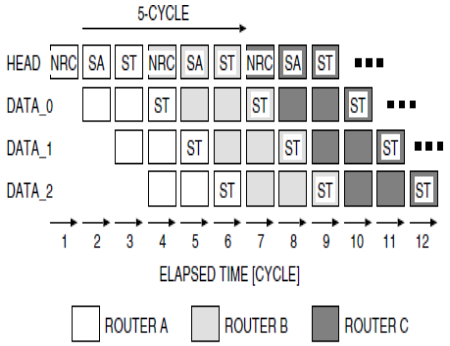

Fig. 4. Pipelined router with Look ahead routing [10].

Fig. 4 shows pipeline router with look ahead carry. Here the Routing Computation unit nth router computes the output port of the n+2nd router on the path. The header flit goes through three stages called Next Routing Computation (NRC), Switch Allocation (SA) and Switch Traversal (ST).

VI. ViChaR Architecture

The router structure discussed in section II is prone to leakage power consumption as all the VCs are not active at a time and also the VCs are underutilized. Nicopoulos et al [2] proposed a Dynamic Virtual regulator for NOCs consisting of Unified Buffer Structure and Unified Control logic. fig. 7 shows ViChaR block and Fig. 8 shows the ViChaR architecture. This architecture gives stress upon the fact that small number of deep VCs are suitable for light traffic whereas a large number of shallow VCs are best suited for

heavy traffic. Fig. 5 explains this statement. For fixed buffer allocation, this flexibility cannot be achieved. In ViChaR, each port can avail the buffers on a basis of dynamic allocation. The allocation depends upon the traffic condition. It was shown in [2] that ViChaR architecture results into 4.03% of area savings and 1.74% power overhead. Also throughput wise, it is superior to some other router architectures proposed in literature. In traditional router architecture, a packet will be blocked if the packet at the head of a VC is blocked. This is called Head-of-line blocking. Also the same FIFO cannot be used by other packets as this will lead to packet mixing. Fig. 6 shows these two concepts.

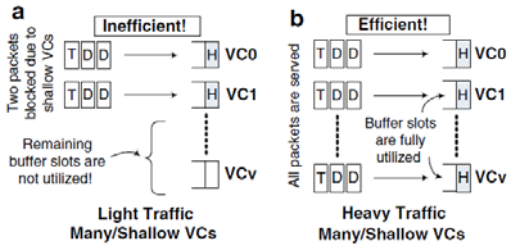

Fig. 5. (a) Light traffic many/shallow VCs. (b) Heavy traffic many/shallow VCs. [2].

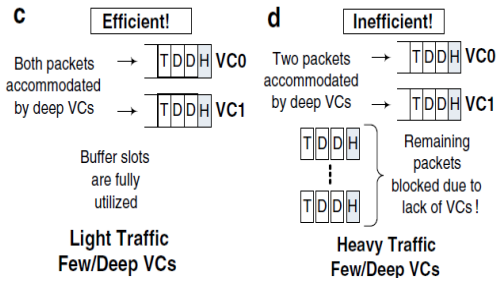

Fig. 5. (c) Light traffic few/deep VCs. (d) Heavy traffic few/deep VCs [2].

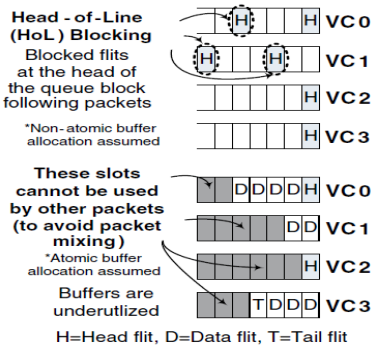

Fig. 6. Limitation of rigid structure [2].

VII. Literature Survey

A number of buffer architectures are proposed in literature. In [2] it was shown that network performance is directly related to buffer size and buffer organization. Dynamically allocated buffer structured was proposed in [11]. DAMQ with self compacting buffers [12], FCCB [13], VC allocation on traffic condition [14], Chaos router [15], BLAM routing algorithm [16], Geyser 1 [17] are also discussed in literature. Some power aware models are discussed in [18], [19], [20]. Concept of virtual channel flow control was introduced in [21]. Energy comparison of different types of routers was shown in [22].

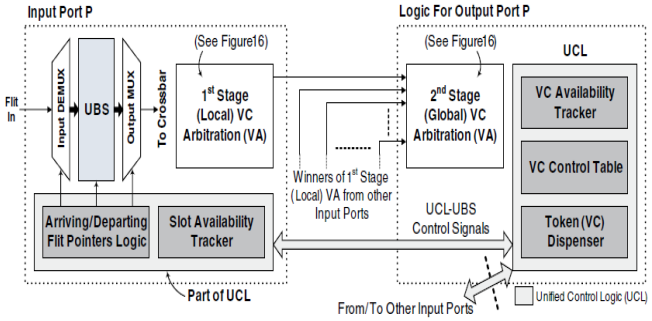

Fig. 7. ViChaR block [2].

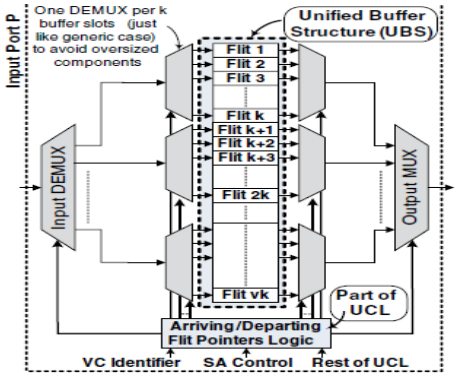

Fig. 8. ViChaR Router architecture (Courtesy Nicopoulos et al [2]).

VIII. Proposed Architecture

ViChaR architecture provides a flexible structure for NOC router. The problem with the architecture is that it is not suitable for fine grained power gating i.e. it is hard to find the components that would consume similar amount of power or that would be on or off at the same time. Although because of dynamic router architecture the VCs are properly utilized but still all the routers are not needed at all times. Therefore, leakage power is still there. A technique is introduced by which the leakage power can be reduced. Here instead of unified buffer structure, a segmented buffer structure is used.

The total number of buffers is divided into various groups i.e. in different segments. For each segment, one control logic called Segmented Control Logic is used. The number of segments depends upon the granularity of power gating. More number of segments means finer grained power gating. Fig. 9 shows a part of segmented buffer architecture.

From the figure it is seen that the control logic is same for flit 1 to flit 2k. It is worth stating here that this range may vary from user to user as discussed previously. Obviously, as number of segments increases, area overhead will also increase. For example for four segments, the area overhead will be 6.96%. Therefore, if area can be compromised, the number of segments may be increased for finer power gating.

Also, as each of the buffers, switch and MUXes can be connected to a power switch, number of power switches will also add up to the area overhead. Therefore, an optimum number of power switch insertion have to be kept in mind.

In the architecture, each virtual channel is capable for storing four 64 bit flits. Fig. 10 shows the technological schematic of one virtual channel. The simulation was done on Xilinx ISE Project Navigator 12.4.

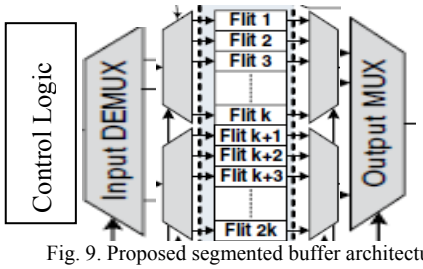

Fig. 9. Proposed segmented buffer architecture.

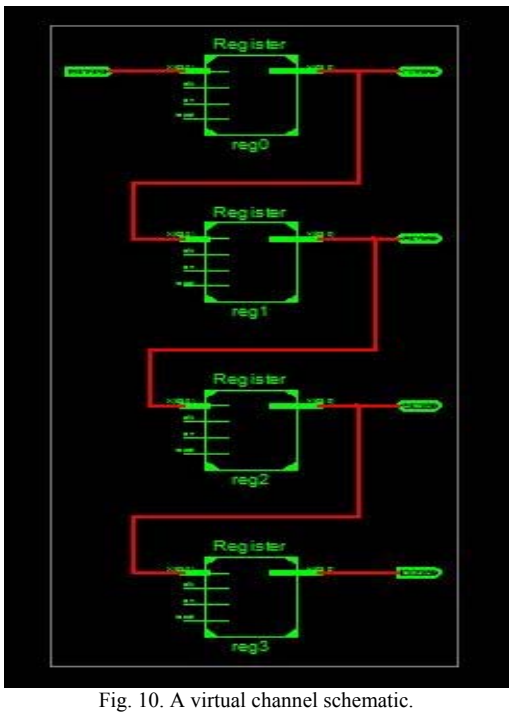

Fig. 10. A virtual channel schematic.

IX. Conclusion

A segmented buffer structure has been proposed in this paper for which finer power gating is possible though area overhead will increase with increasing numbers of segments. In this paper, a technique is proposed to increase the efficiency of the existing ViChaR architecture.

In future, this structure may be simulated and performance may be evaluated based on energy consumption.

Acknowledgment

We acknowledge the developers of NOXIM and NIRGAM [23] simulators which are based on SystemC.

References

[1] C. Nicopoulos, V. Narayanan, and C. R. Das, *Network-on-Chip Architectures: A Holistic Design Exploration*, ch. 2, Springer.

[2] C. Nicopoulos, V. Narayanan, and C. R. Das, "ViChaR: A dynamic virtual channel regulator for NOC routers," *Lecture Notes in Electrical Engineering*, vol. 45, pp. 19-40, 2010.

[3] H. Matsutani, M. Koibuchi, H. Nakamura, and H. Amano, *Run-Time Power-Gating Techniques for Low-Power On-Chip Networks*, *Low Power Networks-on-Chip*, Springer New York Dordrecht Heidelberg London, pp. 21-43.

[4] L. Shang, L. S. Peh, and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks," in *Proc. the International Symposium on High-Performance Computer Architecture (HPCA'03)*, 2003, pp. 79–90.

[5] V. Soteriou and L. S. Peh, "Exploring the design space of self-regulating power-aware on/off interconnection networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 3, pp. 393–408, 2007.

[6] D. Ikebuchi *et al.*, "Geyser-1: A MIPS R3000 CPU core with fine grain runtime power gating," in *Proc. the IEEE Asian Solid-State Circuits Conference (A-SSCC'09)*, 16-18 Nov, 2009, pp. 281-284.

[7] N. Seki *et al.*, "A fine-grain dynamic sleep control scheme in MIPS R3000," in *Proc. the International Conference on Computer Design (ICCD'08)*, 2008, pp. 612–617.

[8] K. Usami and N. Ohkubo, "A design approach for fine-grained run-time power gating using locally extracted sleep signals," in *Proc. the International Conference on Computer Design (ICCD'06)*, 2006, pp.155-161.

[9] Z. G. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban, H. Jacobson, and P. Bose, "Microarchitectural techniques for power gating of execution units," *ISLPED,* pp. 9–11, Newport Beach, California, USA, 2004.

[10] H. Matsutani, M. Koibuchi, D. H. Wang, and H. Amano, "Run-Time power gating of on-chip routers using look-ahead routing," in *Proc. the. ASPDAC 2008*, 2008.

[11] Y. Tamir and G. L. Frazier, "High-performance multiqueue buffers for VLSI communication switches," in *Proc. the 15th Annual International Symposium on Computer Architecture (ISCA)*, 1988, pp. 343–354.

[12] J. Park, B. W. O'Krafka, S. Vassiliadis, and J. D. Frias, "Design and evaluation of a DAMQ multiprocessor network with self-compacting buffers," in *Proc. IEEE Supercomputing '94, the Conferenceon High Performance Computing and Communications,* 1994, pp. 713–722.

[13] N. Ni, M. Pirvu, and L. Bhuyan, "Circular buffered switch design with wormhole routing and virtual channels," in *Proc. the International Conference on Computer Design (ICCD)*, 1998, pp. 466–473.

[14] Y. Choi and T. M. Pinkston, "Evaluation of queue designs for true fully adaptive routers," *Journal of Parallel and Distributed Computing*, vol. 64, no. 5, pp. 606–616, 2004.

[15] S. Konstantinidou and L. Snyder, "The Chaos router," *IEEE Transactions on Computers*, vol. 43, pp. 1386–1397, 1994.

[16] M. Thottethodi, A. R. Lebeck, and S. S. Mukherjee, "BLAM: A high-performance routing algorithm for virtual cut-through networks," in *Proc. the International Parallel and Distributed Processing Symposium (IPDPS)*, 2003.

[17] D. Ikebuchi *et al.,* "Geyser-1: A MIPS R3000 CPU core with fine grain runtime power gating," in *Proc. IEEE Asian Solid-State Circuits Conference,* pp. 281-284, 2009.

[18] A. Banerjee, R. Mullins, and S. Moore, "A Power and energy exploration of network-on- chip architectures," in *Proc. the International Symposium on Networks-on-Chip (NOCS'07)*, 2007, pp. 163–172.

[19] E. Beigne *et al.*, "An asynchronous power aware and adaptive NoC based circuit," *IEEE Journal of Solid-State Circuits,* vol. 44, no. 4, pp. 1167–1177, 2009.

[20] X. Chen and L. S. Peh, "Leakage power modeling and optimization in interconnection networks," in *Proc. the International Symposium on Low Power Electronics and Design (ISLPED'03)*, 2003, pp. 90-95.

[21] W. J. Dally, "Virtual-channel flow control," *IEEE Transactions on Parallel and Distributed Systems*, vol. 3, no. 2, pp. 194–205, 1992.

[22] B. Debbarma, S. N. Pradhan, and A. Jamatia, "Energy comparison of different NOC routers," in *Proc. ACEC 2013*, 2013, pp. 12–14.

[23] NIRGAM. [Online]. Available: http://www.nirgam.ecs.soton.ac.uk

Biman Debbarma is presently working as an assistant professor at Electronics & Communication Engineering Department, National Institute of Technology Agartala, India. He received his B.Tech in electronics engineering from NIT Jamshedpur and M.Tech in microelectronics and VLSI design from NIT, Agartala, and indina. He is currently pursuing PhD (VLSI) at NIT Agartala. His research interests include network on chip, image processing etc. He has academic

Sambhu Nath Pradhan received his Ph.D. from Indian Institute of Technology, Kharagpur and M.E. from Bengal Engineering and Science University in 2010 and 2004 respectively. He is currently working as an assistant professor in ECE Dept. National Institute of Technology Agartala. His research interests include low power design and testing and thermal aware design.