Low-power low-area network-on-chip architecture using adaptive electronic link buffers

A. Sarathy, A.K. Kodi and A. Louri

In the deep sub-micron regime, the performance of network-on-chip (NoC) architectures is bound by the limited power and area budget. Proposed is a low-power low-area NoC architecture using a novel power-efficient control circuit that enables repeaters along the inter-router links to function as adaptive link buffers, thereby reducing the number of buffers required in the router. Simulation results in the 90 nm technology show power savings of nearly 45% and area savings of 50% for the proposed technique.

Introduction: The increasing wire delay constraints in deep sub-micron very large scale integrated (VLSI) circuit design has driven the development of modular and scalable network-on-chip (NoC) architectures [1]. In packet-switched NoCs, every processing element is connected to a NoC component (router) as shown in Fig. 1, and data transmission occurs along the inter-router links. Current NoCs adopt wormhole switching with virtual channels (VCs) per input port and router buffers allocated to each VC. Research into optimising NoC architectures has shown that the router buffers account for an increasing fraction of the total power, thereby necessitating power-efficient buffer design. In this Letter, we propose a low-power low-area NoC architecture with adaptive link buffers for data transmission as well as data storage when required, thereby reducing the number of router buffers. A novel control circuit enables the repeaters along the inter-router links to function as adaptive link buffers during congestion. This power-efficient control circuit operates accurately at variable clock frequencies. Implementation of the proposed architecture in the 90 nm technology shows that reducing the router buffer size by half and using the adaptive link buffers saves nearly 45% in power and 50% in area without significant degradation in network performance.

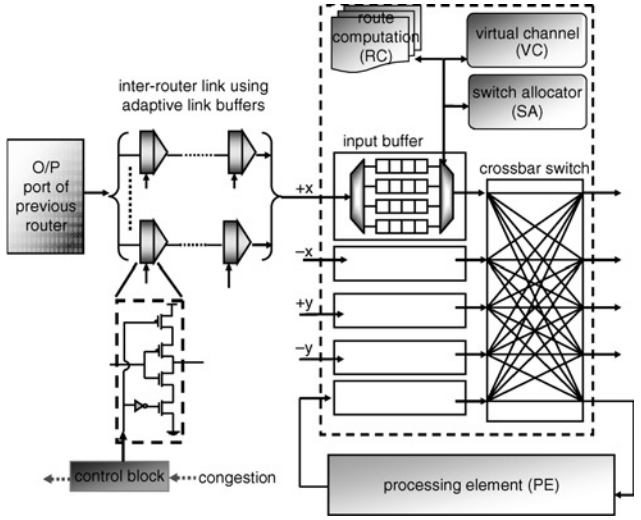

Fig. 1 *NoC router architecture, with proposed adaptive link buffers along the inter-router link*

Adaptive link buffer implementation: Current high-speed VLSI designs require repeater insertion along the wires to meet stringent timing requirements. In Fig. 1 the conventional repeaters along an inter-router link have been replaced with three-state repeaters capable of sampling and holding the data [2]. We propose a novel control circuit enabling the three-state repeaters to function as adaptive link buffers during congestion in the network. Fig. 2 shows the proposed control circuit, where the pass transistors are controlled by clocks – CLK1 and CLK2 having half the frequency of the global clock – CLK (the congestion control signal which controls the repeaters is synchronous with respect to CLK). The capacitor is charged and discharged through the pass transistors, delaying the congestion signal by one clock (CLK) cycle. In the next clock cycle, the corresponding repeater stage is tri-stated to hold the data in position and the congestion signal travels to the next control block. Fig. 3 shows waveforms from a Verilog simulation of the adaptive link buffers using Synopsys VCS simulator, for four link buffer stages. When the congestion input to a stage is 'high', the output is tri-stated (as seen on data output from stage 1 at 10 ns). Data held is released once the congestion input goes 'low' (as seen on data output from stage 1 at 22 ns).

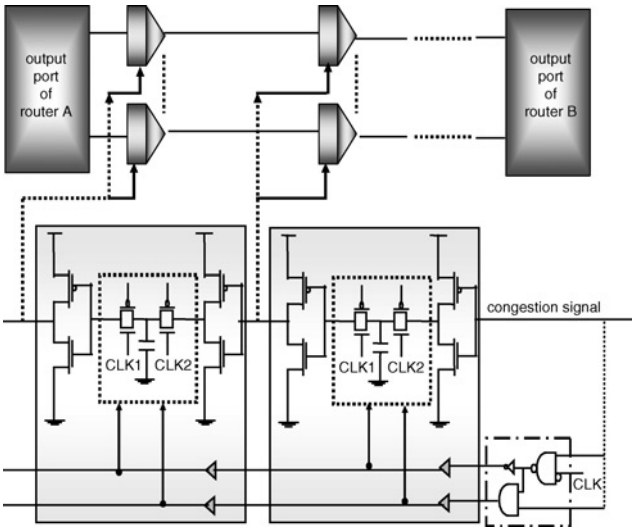

Fig. 2 *Proposed control blocks interfaced to adaptive link buffer stages*

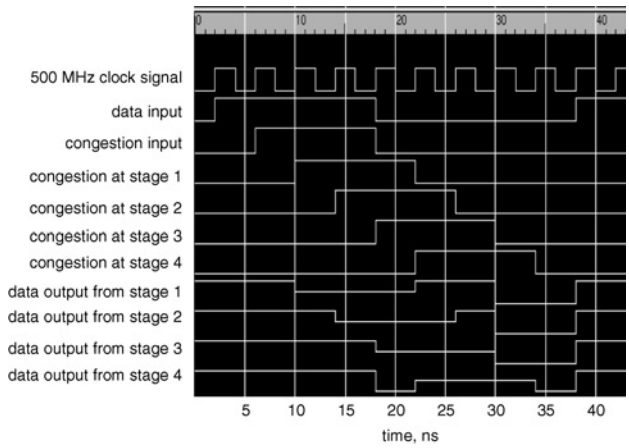

Fig. 3 *Simulation of data being held and released by adaptive link buffers*

The proposed control circuit behaves as a delay module as well as a repeater for the congestion signal. Unlike conventional repeaters, this power-efficient control circuit can be turned 'OFF' in the absence of congestion. Moreover, a single control block can drive the control input of all the adaptive link buffers in a given stage. In addition, the simple switched capacitor design operates accurately at variable frequencies and retains the signal stability even at high clock speeds.

Evaluation: We evaluated the proposed architecture in terms of power, area, and overall network performance. We considered an 8 × 8 mesh network with 128-bit flits (the basic flow control units – a packet consists of several flits). The inter-router links are 2 mm long and have eight optimally spaced repeaters along the wires. 90 nm technology parameters were considered at an operating frequency of 500 MHz and a supply voltage of 1 V, for the repeater-inserted inter-router links [3] and the SRAM-based input buffers [4]. The test cases are represented as vn_V-rn_R-cn_C, where n_V is the number of VCs per input port, n_R is the number of router buffers per VC and n_C is the number of adaptive link buffers. For example, the baseline case is denoted as v4-r4-c0, implying four VCs/input ports, four router buffers/VC and 0 adaptive link buffers.

The power consumed per flit traversal in each case is back-annotated into a cycle-accurate on-chip network simulator under uniform random traffic. Fig. 4 shows the power dissipated by the router buffers at a network load of 0.5. By reducing the router buffer size in half (v4-r2-c8) compared to the baseline, 45% savings in power and 50%

savings in buffer area are achieved. Fig. 5 shows that the throughput of the network drops by only about 3% for the v4-r2-c8 case. Therefore, the proposed low-power low-area NoC architecture using adaptive electronic link buffers saves significant router power and chip area without degrading network performance.

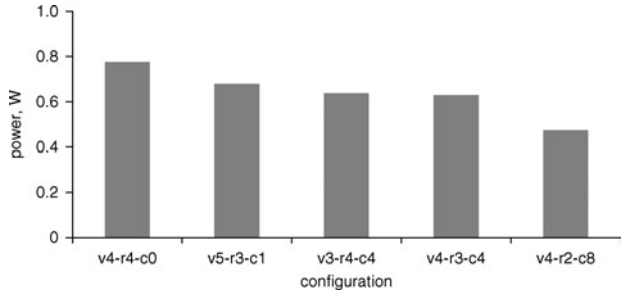

Fig. 4 *Power dissipated in router buffers of 8 × 8 mesh network (uniform random traffic) at network load of 0.5*

vn_V-rn_R-cn_C; nV: number of VCs/input port; nR: number of router buffers/VC; nC: number of adaptive link buffers

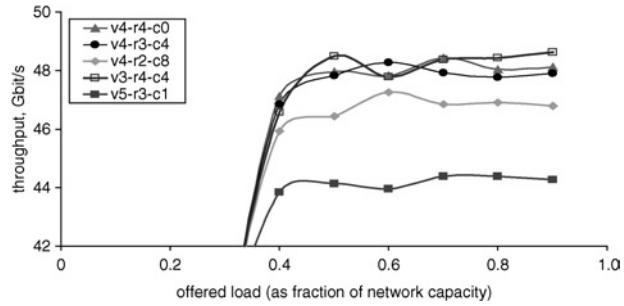

Fig. 5 *Throughput for 8 × 8 mesh network (uniform random traffic)*

Acknowledgments: This research was partially supported by NSF grant ECCS-0725765. We thank D. (Brian) Ma and M. Song for assistance with the switched capacitor control block.

28 January 2008
Electronics Letters online no: 20080239
doi: 10.1049/el:20080239

A. Sarathy and A. Louri (*Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona, USA*)

E-mail: sarathya@ece.arizona.edu

A.K. Kodi (*Department of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, USA*)

References

1 Benini, L., and Micheli, G.D.: 'Networks on chips: a new SoC paradigm', *IEEE Comput.*, 2002, **35**, pp. 70–78
2 Mizuno, M., Dally, W.J., and Onishi, H.: 'Elastic interconnects: repeater-inserted long wiring capable of compressing and decompressing data'. Proc. IEEE Int. Solid-State Circuits Conf., San Fransisco, CA, USA, 2001, pp. 346–347
3 Ho, R., Mai, K.W., and Horowitz, M.A.: 'The future of wires', *Proc. IEEE*, 2001, **89**, pp. 490–504
4 Wang, H.S., Zhu, X., Peh, L.S., and Malik, S.: 'Orion: a power-performance simulator for interconnection networks'. Proc. 35th Annual ACM/IEEE Int. Symp. on Microarchitecture (MICRO), Istanbul, Turkey, 2002, pp. 294–305