Chapter Seven
Summary

Significant improvements in the performance of communication switches can be achieved by using multi-queue input buffers instead of FIFO buffers. A crossbar is often used to connect multi-queue input buffers to the output ports of the switch and an arbiter is needed to resolve conflicting requests for crossbar resources from the various queues. If all the queues of each input buffer are connected to a single crossbar input, the arbitration task is *symmetric* with respect to inputs and outputs. Since the result of the arbitration for each crosspoint depends on or influences the arbitration of all the crosspoints in the corresponding row and column, the arbiter is relatively complex and cannot be partitioned into independent row or column arbiters.

We have evaluated different symmetric crossbar arbitration schemes using static probabilistic analysis, single switch event-driven simulations, and simulations of a buffered Omega network. We have shown that in synchronous networks, a good symmetric crossbar arbitration scheme is essential to realizing the potential for performance improvement from multi-queue buffers. In particular, a poor scheme may result in lower performance than with FIFO buffers while a good scheme can increase the maximum throughput of the network by more than 40%. As the buffer size and switch size increase, the benefits of a good arbitration scheme increases.

Two arbitration schemes, which are based on the propagation of an arbitration wave across an array of arbitration cells, were shown to achieve nearly the same performance as the complex theoretical schemes. The proposed wave front and wrapped wave front arbiters are amenable to simple regular implementation in VLSI. Circuit simulation of our CMOS implementation of the 4×4 wave front arbiter, using 2 μ technology, indicates that, for a 4×4 crossbar, this arbiter can produce a valid and efficient crossbar configuration in 15.5 ns.

For asynchronous networks, the choice among the various symmetric crossbar arbiters has less impact on performance than for synchronous networks. The impact of the arbitration scheme on the performance of an asynchronous network depends on the packet size, switch size, and input buffer size. If the packet size is large and the switch size and input buffer size are small, the quality of arbitration of a poor arbiter is approximately the same as that of a good arbiter. As the packet size decreases and the switch size and input buffer size increase, the difference in the quality of arbitration between a poor arbiter and a good arbiter increases. However, even if the packet size is small and the switch size and buffer size are large, the difference in the quality of arbitration between a good arbiter and a poor arbiter is less significant than in synchronous networks. Since the differences among the various arbiters in terms of quality of arbitration is small, in many situations circuit speed and circuit size should be the deciding factors when choosing an arbiter.

With the advances in VLSI fabrication and packaging technology, larger crossbars are becoming practical. However, the time required to arbitrate conflicting requests to crossbar switches increases with the size of the crossbar. We have shown that by decomposing the arbitration process, the crossbar bandwidth can be increased while the average latency is reduced. The additional implementation cost compared to a nondecomposed scheme is minimal. This design lends itself to efficient modular implementation of large crossbars using small crossbar building blocks.

Switches based on DAMQ buffers manage, at each input port, a separate queue for each output port. The circuit complexity for managing multiple queues in an input buffer increases as the number of queues is increased. Furthermore, with the original DAMQ design, the number of *request* and *grant* lines between each buffer and the arbiter increases linearly with the size of the switch. Based on these considerations, for a large crossbar, it is undesirable to have at each input port a separate queue for each output port. We have shown that this problem can be alleviated by reducing the number of queues per input buffer. For large switches with limited input buffer size, even if the number of queues is reduced significantly, there is only a small reduction in performance compared to switches that use full DAMQ buffers.

While high bandwidth and low latency are critical requirements from interconnection networks, fairness and a guarantee of timely delivery of *every* packet must also be provided. In an asynchronous network where the communication switches use multi-queue input buffers, eventual packet delivery cannot be guaranteed unless special starvation prevention mechanisms are employed. We have presented a simple starvation prevention mechanism for the wave front arbiter used in such communication switches. We have laid out in custom VLSI wave front arbiters with and without the starvation prevention mechanism. The overhead of the mechanism was found to be 11% in terms of layout area and, in the worst case, 15% in terms of arbiter circuit performance.

The mechanism presented does guarantee eventual packet delivery. Furthermore, simulation studies have shown that, under certain nonuniform traffic patterns, the mechanism can improve performance for packets in what would otherwise be a particularly slow queue at a cost of a relatively small decrease in performance for packets in some other queues.

We have discussed the arbiter designs which can support different types of network traffic. One of the issues is the arbiter support for adaptive routing. Adaptive routing can be used to improve network performance and provide fault tolerance in multiprocessor interconnection networks. With adaptive routing, the interconnection network achieves better link utilization and fault tolerance by routing packets around congested or faulty areas. We have shown that the proposed symmetric crossbar arbiters can be used to support adaptive routing in communication switches efficiently. We have also discussed the issue of arbiter support in the communication switch for high-priority traffic in interconnection networks. A crossbar arbiter based on DAMQ buffers with dedicated high-priority queues is described. We showed that such an arbiter can be efficiently implemented by modifying the wave front arbiter.

Bibliography

[Agar91] A. Agarwal, ‘‘Limits on Interconnection Network Performance,’’ *IEEE Transactions on Parallel and Distributed Systems* **2**(4)(October 1991).

[Bert87] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall (1987).

[Bhuy87] L. N. Bhuyan, ‘‘Analysis of Interconnection Networks with Different Arbiter Designs,’’ *Journal of Parallel and Distributed Computing* **4**(4) pp. 384-403 (August 1987).

[Chen90] M. S. Chen and K. G. Shin, ‘‘Adaptive Fault-Tolerant Routing in Hypercube Multicomputers,’’ *IEEE Transactions on Computers* **39**(12) pp. 1406-1416 (December 1990).

[Chie86] A. A. Chien, ‘‘Congestion Control in Routing Networks,’’ *Master Thesis*, Massachusetts Institute of Technology, (October 1986).

[Chow87] E. Chow, H. Madan, and J. Peterson, ‘‘A Real-Time Adaptive Message Routing Network for the Hypercube Computer,’’ *8th Real-Time Systems Symposium*, pp. 88-96 (1987).

[Crow85] W. Crowther, J. Goodhue, R. Gurwitz, R. Rettberg, and R. Thomas, ‘‘The Butterfly Parallel Processor,’’ *IEEE Computer Architecture Newsletter*, pp. 18-45 (September/December 1985).

[Dall86] W. J. Dally and C. L. Seitz, ‘‘The Torus Routing Chip,’’ *Distributed Computing* **1**(4) pp. 187-196 (October 1986).

[Dall88] W. J. Dally, ‘‘Fine-Grain Message Passing Concurrent

Computers,'' *Proceedings of the Third Conference on Hypercube Concurrent Computers* **1** pp. 2-12 (January 1988).

[Dall90] W. J. Dally, ''Virtual-Channel Flow Control,'' *17th Annual International Symposium on Computer Architecture*, (May 1990).

[Dias81a] D. M. Dias and J. R. Jump, ''Analysis and Simulation of Buffered Delta Networks,'' *IEEE Transactions on Computers* **30**(4) pp. 273-282 (April 1981).

[Dias81b] D. M. Dias and J. R. Jump, ''Packet Switching Interconnection Networks for Modular Systems,'' *Computer* **14**(12) pp. 43-53 (December 1981).

[Feng81] T. Feng, ''A Survey of Interconnection Networks,'' *Computer*, pp. 12-27 (December 1981).

[Fraz89] G. L. Frazier and Y. Tamir, ''The Design and Implementation of a Multi-Queue Buffer for VLSI Communication Switches,'' *International Conference on Computer Design*, pp. 466-471 (October 1989).

[Gall77] R. G. Gallager, ''A Minimum Delay Routing Algorithm Using Distributed Computation,'' *IEEE Transactions on Communications* **COM-25**(1) pp. 73-85 (January 1977).

[Gott83] A. Gottlieb, R. Grishman, C. Kruskal, K. McAuliffe, L. Rudolph, and M. Snir, ''The NYU Ultracomputer - Designing an MIMD Shared Memory Parallel Computer,'' *IEEE Transactions on Computers* **C-32**(2) pp. 175-189 (February 1983).

[Gust84] D. B. Gustavson, ‘‘Computer Buses - A Tutorial,’’ *IEEE Micro* **4**(4) pp. 7-22 (August 1984).

[Hals86] R. H. Halstead, T. L. Anderson, R. B. Osborne, and T. L. Sterling, ‘‘Concert: Design of a Multiprocessor Development System,’’ *13th Annual Symposium on Computer Architecture,* pp. 40-48 (June 1986).

[Hsu90] J.-M. Hsu and P. Banerjee, ‘‘Hardware Support for Message Routing in a Distributed Memory Multicomputer,’’ *International Conference on Parallel Processing*, (1990).

[Joer87] C. F. Joerg, ‘‘Design of a Packet Switched Routing Chip for the Dataflow Supercomputer,’’ *Bachelor Thesis,* Massachusetts Institute of Technology, (May 1987).

[Kerm79] P. Kermani and L. Kleinrock, ‘‘Virtual Cut Through: A New Computer Communication Switching Technique,’’ *Computer Networks* **3**(4) pp. 267-286 (September 1979).

[Kons90a] S. Konstantinidou and L. Snyder, ‘‘Chaos Router: A Practical Application of Randomization in Network Routing,’’ *Symposium of Parallel Algorithms and Architectures* , pp. 21-30 (1990).

[Kons90b] S. Konstantinidou, ‘‘Adaptive Minimal Routing in Hypercubes,’’ *6th MIT Conference on Advanced Research in VLSI,* pp. 139-153 (1990).

[Kons91] S. Konstantinidou and L. Snyder, ‘‘Chaos Router: Architecture and Performance,’’ *International Symposium on Computer Architecture*, pp. 212-221 (1991).

[Kuma84] M. Kumar and J. R. Jump, ‘‘Performance Enhancement in Buffered Delta Networks Using Crossbar Switches and Multiple Links,’’ *Journal of Parallel and Distributed Computing* **1**(1) pp. 81-103 (1984).

[Lang82a] T. Lang, M. Valero, and I. Alegre, ‘‘Bandwidth of Crossbar and Multiple-Bus Connections for Multiprocessors,’’ *IEEE Transactions on Computers*, pp. 1227-1233 (December 1982).

[Lang82b] T. Lang and M. Valero, ‘‘M-users B-servers Arbiter for Multiple Busses Multiprocessors,’’ *Microprocessing and Microprogramming*, pp. 11-18 (October 1982).

[Lawr75] D. H. Lawrie, ‘‘Access and Alignment of Data in an Array Processor,’’ *IEEE Transactions on Computers* **C-24**(12) pp. 1145-1155 (December 1975).

[McMi80] R. J. McMillen and H. J. Siegel, ‘‘The Hybrid Cube Network,’’ *Distributed Data Acquisition, Computing, and Control Symposium*, pp. 11-22 (December 1980).

[Mitr82] I. Mitrani, *Simulation Techniques for Discrete Event Systems,* Cambridge University Press (1982).

[Mudg82] T. N. Mudge and B. A. Makrucki, ‘‘Probabilistic Analysis of a Crossbar Switch,’’ *9th International Symposium on Computer Architecture*, pp. 311-319 (April 1982).

[Mudg87] T. N. Mudge, J. P. Hayes, and D. C. Winsor, ‘‘Multiple Bus

Architectures,'' *Computer* **20**(6) pp. 42-48 (June 1987).

[Ngai89] J. Y. Ngai and C. L. Seitz, ''A Framework for Adaptive Routing in Multicomputer Networks,'' *ACM Symposium on Parallel Algorithms and Architectures,* (1989).

[Pate79] J. H. Patel, ''Processor-Memory Interconnections for Multiprocessors,'' *International Symposium on Computer Architecture,* pp. 168-177 (April 1979).

[Pear75] R. C. Pearce, J. A. Field, and W. D. Little, ''Asynchronous Arbiter Module,'' *IEEE Transactions on Computers,* pp. 931-932 (September 1975).

[Pfis85a] G. F. Pfister, W. C. Brantley, D. A. George, S. L. Harvey, W. J. Kleinfelder, K. P. McAuliffe, E. A. Melton, V. A. Norton, and J. Weiss, ''The IBM Research Parallel Processor Prototype (RP3): Introduction and Architecture,'' *1985 International Conference on Parallel Processing,* pp. 764-771 (August 1985).

[Pfis85b] G. F. Pfister and V. A. Norton, '''Hot Spot' Contention and Combining in Multistage Interconnection Networks,'' *IEEE Transactions on Computers* **C-34**(10) pp. 943-948 (October 1985).

[Reed87] D. A. Reed and R. M. Fujimoto, *Multicomputer Networks: Message-Based Parallel Processing,* The MIT Press (1987).

[Stev86] K. S. Stevens, S. V. Robison, and A. L. Davis, ''The Post Office - Communication Support for Distributed Ensemble Architectures,'' *The 6th International Conference on Distributed Computing*

Systems, pp. 160-166 (May 1986).

[Swop86] S. M. Swope and R. M. Fujimoto, ‘‘Simon II Kernel Reference Manual,’’ Technical Report UUCS 86-001, University of Utah, Salt Lake City, UT (March 1986).

[Tami88a] Y. Tamir and G. L. Frazier, ‘‘High-Performance Multi-Queue Buffers for VLSI Communication Switches,’’ *15th Annual International Symposium on Computer Architecture*, pp. 343-354 (May 1988).

[Tami88b] Y. Tamir and G. L. Frazier, ‘‘Support for High-Priority Traffic in VLSI Communication Switches,’’ *9th Real-Time Systems Symposium*, pp. 191-200 (December 1988).

[Tami91] Y. Tamir and Y. F. Turner, ‘‘High-Performance Adaptive Routing in Multicomputers Using Dynamic Virtual Circuits,’’ *Distributed Memory Computing Conference* , (April 1991).

[Tami92] Y. Tamir and G. L. Frazier, ‘‘Dynamically-Allocated Multi-Queue Buffers for VLSI Communication Switches,’’ *IEEE Transactions on Computers* **41**(6) pp. 725-737 (June 1992).

[Tami93] Y. Tamir and H.-C. Chi, ‘‘Symmetric Crossbar Arbiters for VLSI Communication Switches,’’ *IEEE Transactions on Parallel and Distributed Systems* **4**(1) pp. 13-27 (January 1993).

[Welc83] P. D. Welch, ‘‘The Statistical Analysis of Simulation Results,’’ pp. 267-329 in *Computer Performance Modeling Handbook*, ed. Stephen S. Lavenberg, Academic Press (1983).

[Whit85] C. Whitby-Strevens, ‘‘The Transputer,’’ *12th Annual Symposium on Computer Architecture*, pp. 292-300 (June 1985).

[Yoon90] H. Yoon, K. Y. Lee, and M. T. Liu, ‘‘Performance Analysis of Multibuffered Packet-Switching Networks in Multiprocessor Systems,’’ *IEEE Transactions on Computers* **39**(3) pp. 319-327 (March 1990).