A 1.2 μm CMOS DEADLOCK-FREE ROUTER FOR HYPERCYCLE NETWORKS

R. Sivakumar & N. J. Dimopoulos

Dept of Electrical & Computer Engineering
University of Victoria
B.C, CANADA - V8W 3P6

Abstract

In this work, we consider deadlock-free routing in Hypercycles which are class of multidimensional graphs used for modeling interconnection networks. Hypercycles offer simple routing, incremental expansion, variable diameter, enhanced fault-tolerance and hence can be tailored specifically to the topology of a particular application. This paper deals with the hardware implementation of a 1.2μm CMOS chip that implements deadlock preventing routing. The chip consists of about 20,000 transistors and has an expected throughput of 20 Million decisions per second.

1. Introduction

Hypercycles can be considered as products of circulant graphs and a detailed description of the terminologies can be found in [1]. Hypercycles range in complexity from simple rings to fully connected graphs and are generalizations of several popular interconnection networks such as rings, toruses [2], binary n-cubes, k-ary *n*-cubes [3] and generalized hypercubes [4]. Many properties and algorithms used, for instance in routing and processor allocation can be extended to the entire class of hypercycles making it possible to choose a topology that best suits the system requirements of a specific class of applications.

1.1 Deadlock-Free Routing

A number of different routing policies have been introduced for a variety of interconnection networks. Deadlock free routing algorithms are robust and ideal for message passing concurrent machines such as the NCUBE, Cosmic cube [5], MARK III Hypercubes [6] and Intel iPSc [7]. In [8], Dally gives a comprehensive discussion on deadlock free routing in multiprocessor networks in which the necessary and sufficient conditions for a routing algorithm to be deadlock free are explained by means of a channel dependency graph

It has been shown in [8,9] that the *e-cube* or prioritized dimension routing for hypercubes is deadlock free since the resources are ordered and there are no cycles. In ecube, a lower order link cannot be issued unless a higher order resource is obtained. If messages are routed in order of decreasing dimension and hence decreasing channel subscript, no cycles are formed in the dependency graph and the e-cube algorithm is deadlock free.

Similar approaches have been devised for k-ary n-cubes in which virtual channels [3] are introduced to break circular dependencies preventing deadlocks. Backtracking strategies [1,10] avoid deadlocks by retry when a path is blocked. Despite their high throughput for moderate loads, backtracking methods suffer from excessive thrashing at heavy loads.

1.2 Prioritized dimension Routing

In [11], a prioritized dimension routing algorithm based on ecube circuit switching technique is discussed for a certain class of hypercycles and it is proved that it is deadlock preventing. A category of one dimensional hypercycles with diameter 1 (fully connected) and diameter 2 has been proved to fall under e-cube routing in [11].

Algorithm
The ecube routing proceeds as follows: Routing from a source to destination node is accomplished in order of decreasing dimension. Given an *n*-dimensional hypercycle, the following priority levels are assigned by definition. Dimension *n* has the highest priority while dimension 1 has the least priority. When a destination address is presented, the algorithm checks whether the destination has been reached in dimension *i* (*i*=*n* ... 1) by comparing the source(-current) and destination address bits. If it has not, then the next shortest path address in dimension *i* is generated. The computed address is then validated based on the available ports and this constitutes the next port address to forward the message; otherwise a break signal is generated. On the other hand, if the destination has indeed been reached in the i^{th} dimension, routing proceeds in a similar fashion in the

0-7803-0971-5/93/$3.00 © 1993 IEEE

remaining j dimensions (of lower priority) for $j=i-1$ ... 1 till the source and destination address bits are equalized.

2. Architecture

Designing a simple and testable ASIC is of paramount importance from the stand point of performance and reliability. The proposed deadlock-free router chip has been designed based on the following assumptions.

1. The maximum dimensionality of the chip is 4.
2. The maximum number of available ports is 16.
3. Each dimension can have a maximum population of 15 nodes.
4. The maximum connectivity in each dimension is 5.
5. The degree of the network cannot exceed 16.

Fig. 1 shows the major functional blocks of the router which includes the following:

- Input Registers (Memory bank)
- Deadlock-Free Router combinational Logic block
- Output registers
- Control block

The Memory bank consists of four 19-bit registers and one 16-bit register while the output register is 8-bit in length. The control block provides the necessary timing signals for latching the data into the input/output registers and operates on a clock with a period of 8ns. The control unit is a simple 2 bit counter that sequences through 4 possible states generating the appropriate timing signals for operating the chip.

The main arithmetic units in the deadlock-free router combinational logic illustrated in Fig. 2. comprise of:

1. Four modules of Ecube decoder, one for each of the 4 dimensions.
2. Four modules of Next Port Generators (NPG).
3. Port Selector and Validator.

The ecube-decoder determines whether a given configuration of a 4-D hypercycle can perform ecube routing. The NPG computes the shortest path (*greedy routing*) [1] destination address based on the network configuration and current address. This is incorporated in the combinational block along with a module for determining the *odd-even preference routing*[11]. The Port selector and Validator modules then determine a forwarding port based on the available ports and the prioritized dimension algorithm discussed previously.

2.1 Data Path

Each of the 19-bit input registers of the router contain the following data in dimension i, i=4 ... 1.

1. Connectivity vector
2. Source or current address
3. Destination address
4. Population
5. Offset of previous dimensions

The connectivity vector, current address, destination address, population and offset are required by each NPG in the i^{th} dimension. Each of the above inputs in dimension i is 4-bit in length except the connectivity vector which is a 3-bit number. The 16-bit input register holds the available port address with each bit corresponding to a port. The 8-bit output register latches the computed output port (5-bit result) along with the following signals

1. *Ready* - this signal informs the availability of the result as soon as it is computed.
2. *Ecube* - if this line is asserted, it implies that the network conforms to the deadlock-free routing scheme.
3. *Destination Reached* - if the source and destination address bits match.
4. *Wait* - this signal signifies a break condition during routing due to the nonflammability of the computed forwarding port.

3. Operating Modes

Each decision cycle of the Hypercycle router is composed of two phases, namely

1. *Load phase*
2. *Execute phase*

In the *Load phase*, the router is initialized with data needed for its operation during system configuration. In this phase, the controller loads the router with information such as the connectivity, current address, destination address, population, offset and available ports. Note that the above data is needed only once during configuration.

In the *Execute phase*, the data loaded into the router is processed and the resulting next port address (if an available one is found) together with the destination reached or possible break is generated. The router operates on a synchronous mode with the clock, providing a result every 4 clock cycles after the data has been loaded into the memory bank. It takes one clock cycle to load each register. The total execution time for each decision is therefore dictated by the time taken for the load cycle and the execute cycle. Note that the load cycle time is maximum during system power up when the six registers have to be initialized with the relevant parameters. During normal operation, it suffices to load the destination address and the available ports since these change dynamically. The load phase therefore takes 2 clock cycles to be initialized.

The *throughput* (TP) of the router is defined as the number of decision cycles executed per second (DPS)

$$TP = \frac{1}{T_l + T_e} = \frac{1}{2T + 4T} = \frac{F}{6} \qquad \text{(Eq 1)}$$

where T_l and T_e are the load and execute phase delays. Thus, the effective throughput of the router is one sixth of the maximum clock frequency.

4. Hardware Implementation

The micrograph of the deadlock-free router fabricated in 1.2 μm NTE CMOS [12] technology in given in Fig.3. The CAD tools used in the design process comprise of Cadence, SILOS, VERILOG and Logic-III (UVIC).The designed router contains 3828 logic gates using 19,906 transistors. The area of the chip is 4880 x 4080 sq.μm. The chip has 51 pins housed in a 68-pin PGA. The chip is synchronous and is designed to operate at a maximum frequency of 125 Mhz. Substituting in Eq. (1), it can be seen that the throughput of the router is in excess of 20 Million decisions per second. It is fully programmable and checks whether a given configuration is admissable for ecube routing. The internal registers are connected as a *scan chain* and the router is fully testable with a fault coverage of 97 % for single stuck-at faults. In this paper, we have demonstrated a practical and viable silicon implementation of a deadlock-free router for hypercycle networks. The characterization of the manufactured device along with system integration will be taken up in the future.

Acknowledgement

The authors would like to express their gratitude to the Canadian Microelectronics Corporation for its assistance in fabricating the router chip and also to NSERC and IRIS for funding this research. They also thank Mr. Rod Byrne, Mr. W.A Keddy, Mr. K. Jones, Mr. R. Kelly and Mr. W. Kastelic for their help in the VLSI laboratory.

5. References

[1] N.J. Dimopoulos, D. Radvan, K.F. Li "Performance Evaluation of the Backtrack to the Origin and Retry Routing for Hypercycle based Interconnection Networks" *Proceedings of the Tenth International Conference on Distributed Systems*, Paris, France, pp. 278-284 (June 1990).

[2] K. Hwang and A. Briggs, *Computer Architecture and Parallel Processing*, McGraw Hill series, 1987

[3] W. J. Dally "Performance Analysis of k-ary n-cube Interconnection Networks" *IEEE Trans. Computer.*, Vol. 39, no. 6, pp. 775-784, June 1990.

[4] L.N. Bhuyan, and D. P. Agrawal, "Design and Performance of Generalized Interconnection Networks" *IEEE Trans. Computer.* Vol. C-32, No. 12, pp. 1081-1090, Dec. 1983.

[5] C.L. Seitz. "The cosmic cube" *CACM* vol. 28, pp.22-33, Jan. 1985

[6] R.D. Rasmussen, N. J. Dimopoulos, G. S. Bolotin, B. F. Lewis, and R. M. Manning "MAX: Advanced General Purpose Real-Time Multicomputer for Space Applications" *Proceedings of the IEEE Real Time Systems Symposium* pp. 70-78, San Jose, CA.,(Dec. 1987).

[7] iPSC User's Guide, No. 17455-3, Intel Corp., Portland, Ore., 1985.

[8] W.J. Dally, and C. L. Seitz "Deadlock-Free Message Routing in Multiprocessor Interconnection Networks" *IEEE Trans. Computer.* Vol. C-36, No. 5, pp. 547-553, May 1987.

[9] W. J. Dally, J. A. Stuart Fiske, J. S. Keen, R. A. Lethin, M. D. Noakes, P. R. Nuth, R. E. Davison, G. Fyler "The Message-Driven Processor: A Multicomputer Processing Node with Efficient Mechanisms" *IEEE Micro* pp. 23-40, April 1992.

[10] R. Sivakumar, N. J. Dimopoulos, V. Dimakopoulos, M. Chowdhury, D. Radvan "Implementation of the Routing Engine for Hypercycle Based Interconnection Networks" *Proceedings of the 1991 Canadian Conference on Very Large Scale Integration* pp. 6.4.1-6.4.7, Kingston, ON. (Aug. 1991).

[11] N. J. Dimopoulos, M. Chowdhury, V. Dimakopoulos, and R. Sivakumar, "Routing and Broadcasting in Hypercycles, Deadlock Free and Backtracking Strategies", Proceedings of PARLE 92, Paris, July 1992.

[12] *Layout specifications for the 1.2 micron Standard Cell library*, VLSI Implementation Centre report, CMC, April 1990.

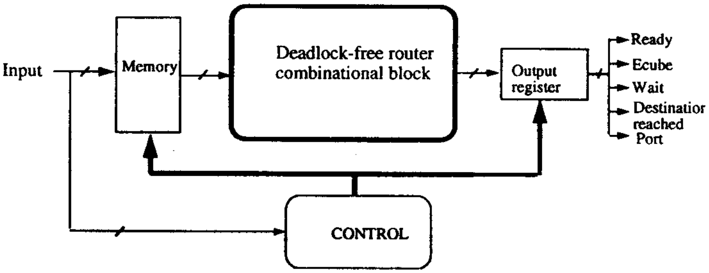

Fig. 1. Top view of the deadlock-free router chip showing the major components.

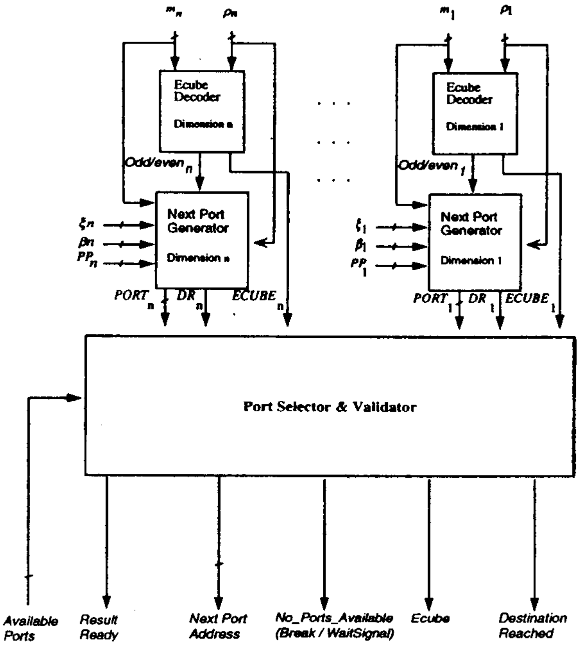

Fig. 2. Block Diagram of 4-D Deadlock free router for Hypercycles.

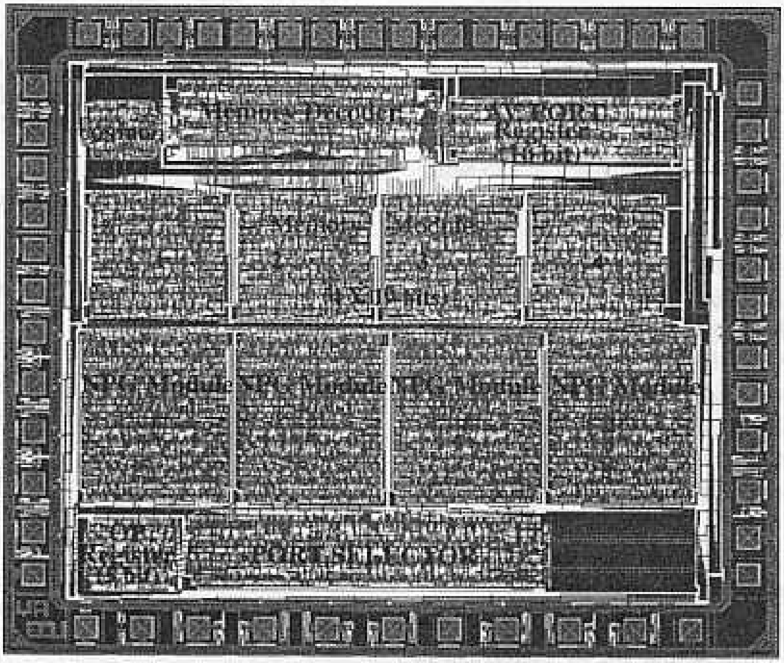

Fig. 3. Layout of the 4-D deadlock-free router in 1.2 µm NTE CMOS technology