An LSI implementation of the Simple Serial Synchronized Multistage Interconnection Network

Takayuki Kamei Masashi Sasahara Hideharu Amano

Keio University
Dept. of Computer Science
3-14-1, Hiyoshi Yokohama, Japan
Tel: +81-45-560-1063
Fax: +81-45-560-1064
e-mail: {kamei,sasahara,hunga}@aa.cs.keio.ac.jp

I. Motivation and outline od the design

A high speed switch is a critical component of multiprocessors. Multistage Interconnection Network (MIN) has been utilized as a switch for connection processors and memory modules in multiprocessors. Unlike the crossbar, it consists of small switching elements, and provides a high bandwidth with relatively small hardware.

Most of traditional MINs are blocking networks and packets are transferred in the store-and-forward manner between switching elements with bit-parallel(8-64bits) lines. Since the width of communication paths and transferrd mannar cause pin-limitation problem and complicated structure, the high density implementation and high speed clock is not utilized.

In order to solve these problems, we implemented the SSS-PBSF chip. This switch uses the PBSF connection structure which can obtain a higher bandwidth than that of crossbar with connecting banyan networks in 3 dimensional direction. Simple Serial Synchronized (SSS) style control mechanism is adopted both for high speed operation and solving the pin-limitation problem.

II. The SSS style control mechanism

All packets are inserted into SSS-MIN serially (in a few bits parallel) synchronized with a common frame clock. Since each switching element stores only one bit (or a few bits) of the packet, the SSS-MIN behaves like a set of shift registers with the switching capability. After the delay for passing through all stages, head of packets come out from the outlets of the MIN.

When a conflict occurs, one of the conflicting packets set the conflict bit, and must be routed to the incorrect direction, since the SSS-MIN provides no packet buffers in each switching element. The conflicting packet is treated as a dead packet, and never interferes the other packets.

When the top of the address packet reaches at the outlet of the MIN, it can be judged whether the packet is correctly routed or not (conflicted), An acknowledge signal is returned to the processor which issues the access request. If the acknowledge signal is not returned, the packet is inserted again from the input buffer synchronized with the next frame signal.

Since each switching element and control logic is quite simple and utilized a high density implementation, a whole network can be implemented in a single chip. Thus, these operations described above can be act with a high frequency clock, and the performance of a whole network is more than most of MINs.

Using these features, almost the same performance can be achieved with only 1/8 hardware or chip numbers compared with conventional MIN.

III. Piled Banyan Switching Fabrics

The PBSF (Figure 1) [1] consists of banyan networks which are connected in the three dimensional direction. A switching element except in the highest and lowest layers provides four inputs/outputs(two for horizontal, and two for vertical direction.).

Packets are inserted into the highest layer banyan network, and transferred to the horizontal direction. When two packets collide with each other, a packet is fed to the corresponding switching element in the next lower layer banyan network with a clock delay. The vertically transferred packet may collide with both packets which are transferred to the horizontal direction. In this case, one of horizontally transferred packets is selected and sent the next lower layer network.

When three (one from the vertical, and two from the horizontal direction) packets for the same direction conflict in a switching element, a packet is routed to the correct direction, a packet is routed vertically, but the other packet cannot be routed to any direction. Such a packet is routed to the incorrect direction, and treated as a "dead-packet". On the lowest layer, such a packet is discarded.

The total throughput of the PBSF is higher than the blocking networks used in conventional MINs, because a conflicted packet falls lower layers and continues routing.

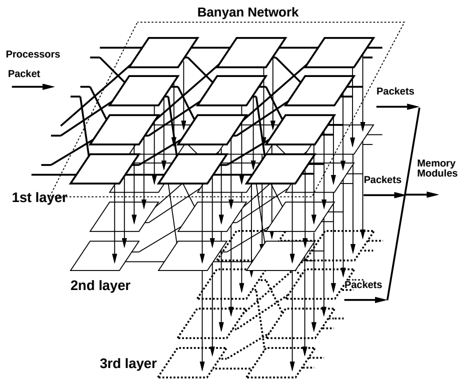

Fig. 1. Piled Banyan Switching Fabrics

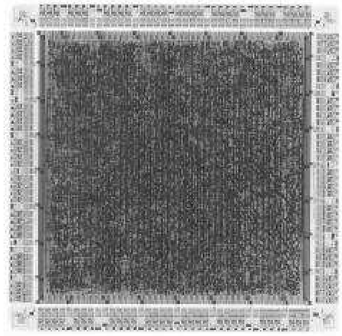

Fig. 2. A microscopic photograph

IV. Chip Design and Implementation

A. Chip Design

The SSS-PBSF chip was developed by the pilot program of the chip design and fabrication service organization for LSI design education in Japan. In the pilot program, the type of chip is limited to be the $0.5\mu m$ CMOS SOG with 16k gates at maximum. Since the PGA package with 176 pins (116 pins for signals) can be used, the size of the network is set to be 16×16 which is the maximum one with this number of pins. The structure and function of the PBSF implemented in the chip is optimized so as to maximize the performance under the pin and gate limitation according to results of the estimation with the computer simulation.

As a result, we decide that the structure of PBSF chip has two layers and multiplexed output, and that the function is 2bit packet priority control, and message combining on the first layer.

B. Implementation

The implemented PBSF consists of the input interface, PBSF core, and the output interface including the multiplexor. The implemented circuit is almost the same from the basic implementation of the SSS-MIN, but a little different.

Input interface The input interface, similar to the basic of SSS-MIN, holds a small buffer to store the header of the address packet while the trace is established. The different point to the simple SSS-MIN is added 2-bits extra wires, in order to reduce the latency for setting up the trace with priority control. The control logic for routing timing is included in the input interface.

PBSF core consists of the address switch with control block, acknowledge switch, and data switch.

in order to reduce the latency with priority control, 2 bits extra wires are provided. Thus, the number of wires for the address switch is expanded to be 4. As the address switch is forward direction, there is paths to the second layer for bypassing the conflicting packet.

Output interface Output interface selects one of packets from two layers with the multiplexor. If two address packets are entirely the same, these packets are combined in the multiplexor.

TABLE I
THE SPECIFICATION OF THE PBSF CHIP

Clock	90MHz
Bit width	2bit(address) 1bit(data)
Size	16-in/out
Max bandwidth	$270 Mbits/sec \times 16$
The number of cell	17356
Signal pin	116pin
Die size	7.34mm × 7.34mm
Power consumption	20W(with 90MHz clock)
Technology	$0.5\mu m$ CMOS sea-of-gates

C. Specification

The chip is designed using the RTL(Register Transfer Level) description of Verilog-XL. The design was done only with three months by two students. Figure 2 shows a microscopic photograph of the chip. Since the logic synthesis, placement and routing is done automatically (also the regulation of clock skew), switching elements can not be observed in the photograph.

Table I shows the specification of the implemented PBSF chip. Although the bit width is small because of the pin limitation, 540Mbyte/sec total throughput can be attained with 90MHz clock.

Acknowlegement

The authors would appreciate with the staff of the pilot program and NTT Electronics Technology Co. Ltd. for their sincere advice for the chip design.

References

[1] T. Hanawa, H.Amano, Y.Fujikawa, "Multistage Interconnection Networks with multiple outlets," *Proc. of the International Conference on Parallel Processing*, pp.I1-8, Aug. 1994

[2] T. Kamei, M. Sasahara, H. Amano, "An LSI implementation of the simple sereal synchronizes multistage interconnection network," *Proc. of SASIMI'95*, pp.199-206, Aug. 1995