A 50,000 Transistor Packet-Switching Chip for the StarBurst ATM Switch

Paul Chow[†], David Karchmer, Paul Chow[‡], Ron White, Tony Ngai, Paul Hodgins, David Yeh, Jeewika Ranaweera, Indra Widjaja[††], and Al Leon-Garcia

Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario, Canada M5S 1A4

Abstract

This paper describes the design and implementation of a 16x16 packet-switching chip that is used as the primary building block in a flexible, output-buffered, packet-switching architecture, called Starburst. This device contains 50,000 transistors implemented in a 1.2 µm CMOS process. Even with a very conservative two-phase clocking methodology and static logic, it can be clocked at 50 MHz. Sixteen chips have been used to construct a prototype 16x16 Starburst ATM switch with a maximum throughput of 2.5 Gbits/sec.

1. Introduction

This paper describes a chip that can be used as the primary building block in a 16-input by 16-output packet switch based on a flexible, non-blocking, output-buffered packet-switching architecture, called Starburst (1)(2). The Starburst architecture features a novel buffer management scheme that combines dedicated and shared buffering, which provides lock-out protection and buffer reduction, respectively. An analytic assessment of the performance of this switch architecture has been conducted. The results show that the Starburst switch is better at handling bursty traffic than output-buffered switches with complete partitioning because of its buffer sharing capability. Under non-uniform traffic conditions, the Starburst switch outperforms switches with complete buffer sharing. Other features include support for multiple priority service and preservation of packet sequence. Furthermore, the modular design of the Starburst architecture allows the performance of the switch to be improved by the addition of more chips. The chip is a synchronous, self-routing network based on the Batcher/Banyan shuffle/exchange network which is well documented in the literature (3)(4)(5). This chip was implemented using a pipelined approach, but the conservative use of static logic and a robust two-phase clocking strategy resulted in a maximum operating frequency of 50 MHz. Sixteen of these chips were used to construct a prototype 16x16 Starburst ATM switch with a maximum throughput of 2.5 Gbits/sec.

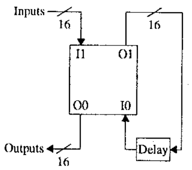

Fig. 1. Single chip Starburst switch.

2. System Architecture

The chip consists of a series of networks whose overall function is to route the input streams of packets to their desired destinations. Packets that are not delivered because of a contention at its output are buffered and delivered in subsequent packet times. Fig. 1 shows the stand-alone chip connected to form a 16x16 packet switch with a single level of buffering. The delay element is required to hold a single packet, and can be implemented by a shift register or a memory.

One key benefit of the Starburst architecture is its modular design. By cascading identical chips, as shown in Fig. 2, the performance of the switch can be incrementally improved. Each additional chip will provide an extra level of buffering that reduces the probability of dropped packets. The first N-1 chips in the chain implement dedicated output buffering while the last chip implements a buffer that is shared by all the outputs.

To increase the bandwidth, the chips can be combined in parallel. A serial input packet is first converted into N parallel bits before passing through N parallel switches. The parallel output from the switches is then serialized to again form a single bit stream for each port. Using this approach, the amount of hardware is increased by a factor of N, but the

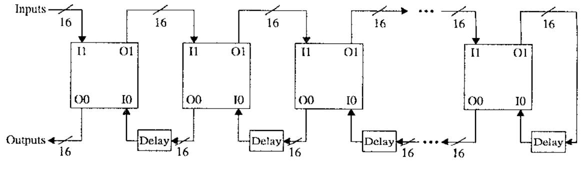

Fig. 2. Multi-chip Starburst switch.

[†] Student
[‡] Professor
[††] Now at University of Arizona

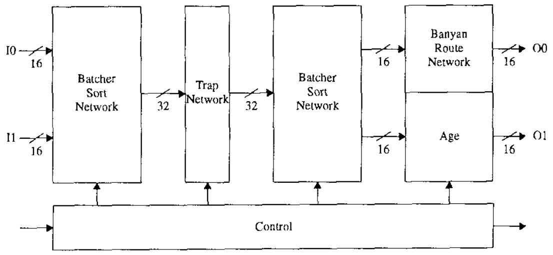

Fig. 3. Block diagram of chip.

operating frequency of the switch fabric is reduced by the same factor.

3. Chip Architecture

Fig. 3 shows the block diagram of the chip architecture. It is composed of two Batcher sorting networks, a Trap network, a Banyan routing network, and an Aging network. The first Batcher network sorts incoming packets in ascending order based on the packet header shown in Fig. 4. The packet header consists of a split bit, an activity bit, an address field, a priority field, and an age field. The split bit is used to mark packets with identical destination addresses. The activity bit indicates whether a time slot contains a packet (activity bit is '0') or is empty (activity bit is '1'). The 4-bit address field specifies the address of the output port. The priority field is used to divide the packets into four classes; class 0 has the highest priority. The age field is used to maintain packet sequencing for virtual circuits. Packets entering the chip for the first time will have their age field set to the maximum value. Each time a packet is recirculated through the switch, its age field will be decremented by one. Thus, older packets that have remained in the switch longer will have lower age values than new packets entering the switch. The payload can accommodate packets of any predetermined length, thus it can be used to transport the 53-byte cells conforming to the ATM standard.

After the Batcher network sorts the incoming packets, the packets with the same destination addresses will end up next to each other. The Trap network then compares adjacent packets leaving packets with unique destinations unaltered. When two or more packets are destined for the same output port, the first packet is unchanged, but subsequent packets with the same destination are marked by setting the split bit in the header. The packets leaving the Trap network are re-sorted by the second Batcher network, once again in ascending order. In this way, the marked packets are sorted toward the bottom of the network. The top sixteen packets are then sent to the Banyan network to be routed to the appropriate destination output ports, and the bottom sixteen are sent to the aging network where the age field in the header is decremented, and then the packets are recirculated through the switch fabric.

Split	Activity	Destination Address	Priority	Age	Payload
S	A	d3 d2 d1 d0	p1 p0	a1 a0	

Fig. 4. Packet format.

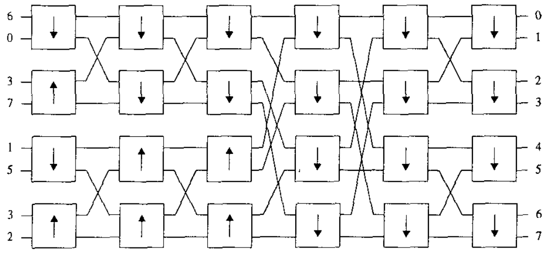

Fig. 5. Example 8x8 Batcher sort network.

The implementations of the above mentioned networks are described in detail in the following section.

4. Networks

4.1. The Batcher Sort Network

Packets coming into the chip are first sorted in ascending order by Batcher's bitonic sorter (4). Fig. 5 shows an example of an 8x8 sorter composed of a rectangular array of 2x2 sorting elements. The 2x2 elements compare the two numbers at their inputs and the larger number is routed to the output indicated by the arrow. The chip contains a 32x32 sorter that requires 15 columns of 16 elements each. Fig. 6 shows a circuit diagram of a 2x2 Batcher element that sorts the largest number to the upper output.

Before a new packet arrives to the Batcher element, the Sort Element Reset (SER) line is pulsed. Next, a bit-serial comparison (i.e. an XOR) is performed on each incoming pair of bits. As long as both inputs have the same value the

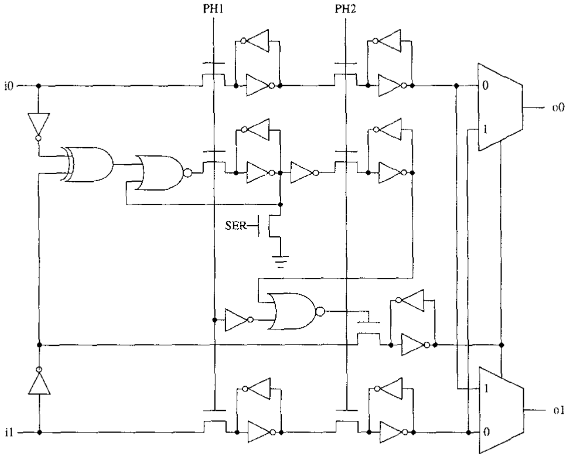

Fig. 6. Circuit diagram of Batcher element

switching circuit does not change state. When a difference at the inputs is detected by the XOR gate, the comparison circuit is disabled so that any subsequent differences are ignored. The difference signal also latches the control signal that switches the multiplexers; in this case, the larger address is switched to the upper output. The 2x2 Batcher element has 70 transistors, with a size of 106 x 240 μm^2.

4.2. The Trap Network

A 32x32 Trap network is used to separate input packets with the same output destination. It consists of a single column of sixteen 2x2 elements. Since the input packets have already been sorted, the Trap function is accomplished by comparing all the destination addresses of adjacent rows (i.e. adjacent input packets). If the two destination addresses are not equal, the packet header is left unchanged. However, if they are equal, then the lower packet is marked by setting the split bit in the header. A second sorting network is used to re-sort the packets after the packets exit the Trap network.

As seen in Fig. 7, the Trap element stores the address bits and compares them with the address bits coming from the element immediately above it. When the latches shown in Fig. 7 contain the bit sequence d0, d1, d2, d3, A, S from the header, the SEL signal is asserted and the split bit S is exchanged with the result of the comparison on the next clock pulse ('1' if the addresses in adjacent packets are equal and '0' if they are not).

The Trap network is followed by a second Batcher sorting network that again sorts the packets in ascending order. In Fig. 7, note that only active packets (A=0) are marked so that they get sorted toward the lower half of the chip for recirculation, and the unmarked inactive packets (A=1) that get sorted to the upper half will be discarded. After the packets have been re-sorted, but prior to entering the Banyan routing network, the packets with split bits equal to one are replaced by empty packets by setting A=1. Then the split bit in all the packets is reset.

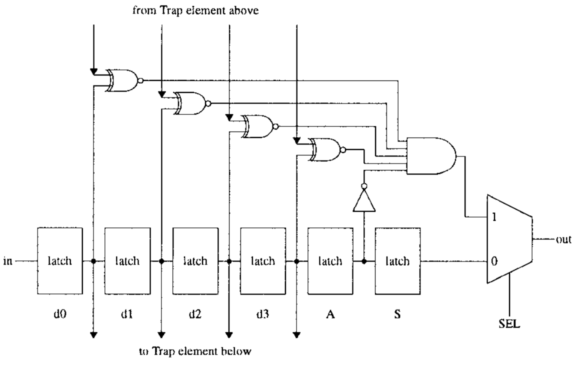

Fig. 7. Circuit diagram of Trap element.

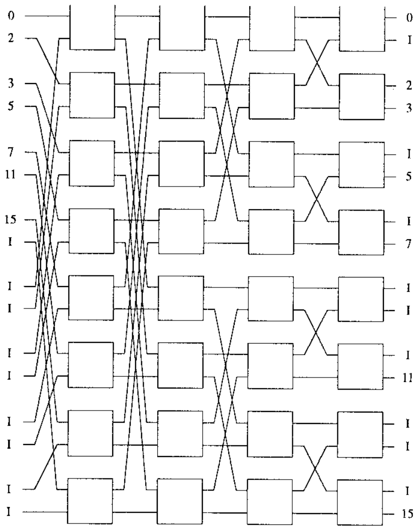

Fig. 8. Example 16x16 Banyan routing network.

4.3. The Banyan Route Network

The 16x16 self-routing, non-blocking Banyan network shown in Fig. 8, is composed of smaller 2x2 processing elements organized in an array of eight rows by four columns. The Banyan network routes the sorted packets from the second Batcher network to their desired output by examining the destination address bits of each packet (packets designated I are inactive). Each column of the Banyan network makes a routing decision based on one bit of the address field in the header. The first column routes on the most significant bit of the address, the second column routes on the next bit, and so on. For each Banyan element to remain identical so that they can be easily replicated, the address bit that determines the routing decision must be rotated so that it immediately follows the activity bit. When the packet emerges at the output of the network after four rotations, the address field is restored to its original form. Fig. 9 shows a circuit diagram of a 2x2 Banyan element. Its state is determined by the activity bit and the destination

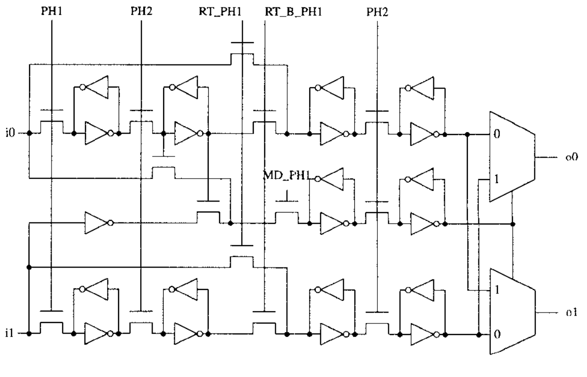

Fig. 9. Circuit diagram of Banyan element.

address bit that follows. After the activity bit has been shifted into the first latch, the MD_PH1 line is asserted. If the packet at the upper input is active (activity bit is '0'), then the routing decision is based on the address bit of the packet at the upper input; otherwise it is determined by the address bit of the packet at the lower input. Furthermore, a '0' address bit in the selected packet will set the element to a state that directs the packet to the upper output, and a '1' address bit in the selected packet will set the element to a state that directs the packet to the lower output. The address rotation is performed by the multiplexing latches that are controlled by the RT_PH1 and RT_B_PH1 signals. Each 2x2 Banyan element contains 74 transistors, and occupies 106 x 240 μm^2.

5. Results

A full-custom packet switching chip for the Starburst switch has been fabricated in a 1.2 µm double-metal CMOS processing technology. Table 1 summarizes the technical data of the chip and Fig. 10 shows a microphotograph of the chip.

Table 1. Technical data. of chip

Technology	1.2 µm CMOS
Die Area	5640 µm x 7840 µm
Core Area	4168 µm x 5821 µm
I/Os	114
Maximum operating frequency	50 MHz
Transistors	50,000

This chip was designed during a one-semester course in VLSI systems design by seven inexperienced circuit designers. The successful result speaks to the simplicity of its architecture.

A prototype 16x16 Starburst switch with a maximum throughput of 2.5 Gbits/sec has been constructed using sixteen chips. This prototype switch contains two levels of buffering and the operating frequency of the chips is 25 MHz. Since the switch is intended to operate at the STS-3 rate of 155 Mbits/sec, eight chips connected in parallel are required to satisfy the bandwidth requirements. The serial-to-parallel conversion at the input to the switch and the parallel-to-serial conversion at the output of the switch are performed by peripheral circuitry. The resulting Starburst ATM switch, shown in Fig. 11, consists of 16 chips (eight in parallel x 2 levels of buffering) on four printed circuit boards connected to a VME backplane. For clarity, the figure shows a single serial input that is converted into a 8-bit parallel quantity.

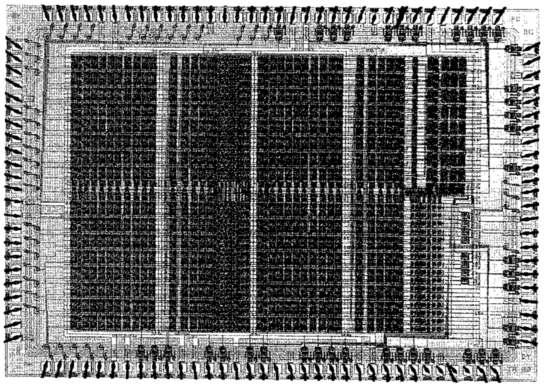

Fig. 10. Chip microphotograph.

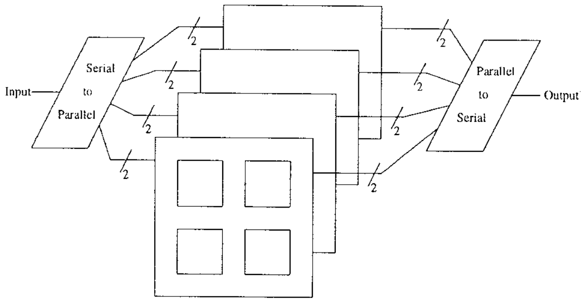

Fig. 11. Prototype 16x16 Starburst ATM switch.

6. Conclusion

In this paper, we have described the design of a full-custom integrated circuit that is used to construct a 16x16 Starburst ATM switch. With more aggressive circuit design and technologies, we expect that a single-chip 64x64 Starburst switch clocked at 500 MHz is attainable.

Acknowledgments

The authors would like to thank NSERC, Micronet, and ITRC for their financial support, and CMC for fabricating the chip. The authors also thank Jaro Pristupa for his technical support.

References

(1) I. Widjaja and A. Leon-Garcia, "StarBurst: A flexible packet switch with dedicated and shared output buffering", *International Switching Symposium* (ISS'92), October 1992.

(2) I. Widjaja and A. Leon-Garcia, U.S. Patent Number: 5274642, Issued: Dec. 28, 1993.

(3) A. Huang and S.Knauer, "Starlite: A wideband digital switch," in *Proc. GLOBECOM'84*, Atlanta, GA, pp. 121-125, Dec. 1984.

(4) K. Batcher, "Sorting networks and their applications," in *Proc. AFIPS*, pp. 307-314, 1968.

(5) William S. Marcus, "A CMOS Batcher and Banyan chip set for B-ISDN packet switching," *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 6, pages 1426-1432, December 1990.