A Prototype Chipset for a Large Scaleable ATM Switching Node

M. Weeks, M. B. Maaz, H. Krishnamurthy, P. Shipley, M. Bayoumi

The Center for Advanced Computer Studies,
University of Southwestern Louisiana,
Lafayette, Louisiana 70504
e-mail: mcw4900@usl.edu

Abstract-- This paper presents a chipset for a 16x16 switching node for the distributing banyan network. This chipset enables the use of a larger and much more efficient switching node than was previously available. Very high performance is required of the chips and thus a number of special circuits have been designed to achieve this performance. The chipset resulting from this design consumes low power. The chips have been designed in 1.0 micron CMOS using a mixture of static and dynamic logic. To achieve the speed needed for a larger node, a register file has been employed to store the packet headers on the control chip. It has an area of 3,150x3,750 micron, and uses 130,000 transistors. The SRAM blocks on the switch chip, which store a bit-slice of the packets, uses 228,600 transistors.

1. Introduction

The Asynchronous Transfer Mode (ATM) has emerged as a leading technology for high speed packet switching, especially in implementing Broadband Integrated Digital Services Networks (BISDN) and other high speed communication networks. ATM's high operating speed demands efficient hardware design and implementation of the switch. The most relevant and important design considerations are scaleability, cost effectiveness and efficient performance while not losing packets or violating cell sequence. There have been several ATM switch architectures proposed and designed. This paper presents the architecture and implementation of the chipset for an Asynchronous Transfer Mode (ATM) network node. The chipset consists of a control chip, and 4 switch chips, which form a scaleable 16x16 switching element that runs at 155 MHz. The architecture uses shared multibuffering, since it is less bandwidth limited than a shared buffering switch. The control logic increases in complexity, but the switch is more scaleable. Scaleability means that the network architecture can be expanded into a larger size easily and in a cost-effective way.

The chipset is designed to be part of a distributed banyan network. These networks allow the efficient construction of very large ATM switches. The network architectures all consist of a base fabric of switching nodes, interconnected using a distributed banyan packet switch [1]. Identical switching nodes can be used in both distributing and routing layers by adding a modest amount of extra control logic and a pin to indicate which type of layer it is in. The chipset of this paper efficiently implements a larger size switching node than was previously available [2]. The larger switching node achieves higher efficiency, better performance, and lower cost than previous designs.

Figure 1 shows a distributing banyan network made up of 4x4 switching elements. The new switching nodes would be interconnected in the same manner as the nodes shown in figure 1, but with 16 inputs and 16 outputs per switching element. The new switching element is made of 1 control chip and 4 switch chips. The network of figure 1 was chosen to demonstrate the distributing banyan network since it clearly conveys the structure. A 256x256 network using the chipset presented here, would be constructed in a similar fashion, but very much larger.

The following section discusses the architecture of the switching node. The third section examines the architecture of the control chip in more detail, followed by a section on the switch chip architecture. The control chip implementation is discussed in the fifth section, and the sixth section describes the implementation of the switch chip. The final section summarizes the chipset.

2. ARCHITECTURE OF THE SWITCHING NODE

A 16x16 switching node is made by connecting a control chip to four switch chips. Each node can store 256 packets. The control chips send commands to the switch chips, which store the packets in order of the commands. The switch chips store the packets, and send them to the next node on the way to their destination. The architecture is called self-routing since the packets are routed as they move through the network. This contrasts switches that use single stage shared memory since they require global control, where the switch state must be known by a central controller. A central controller must process all of the cells in the network, resulting in unacceptable delay if the switch size is expanded. Thus, the self-routing architecture allows for a more scaleable chipset.

Figure 2 shows the control chip and 4 switch chips forming a 16x16 switching element. The switch chips receive a bit-slice of each packet. For example, the first switch chip gets the first 1/4 of each packet. The control chip gets the packet headers from the switch chips, as well as input flow control. This information allows the control chip to reconstruct the sequencing of the packets and to adjust the internal flow of packets according to back-pressure. The flow control signals how much unused storage capacity the node has, allowing packets to be redirected when a node becomes full. The control chip outputs flow control data, addresses, and destination routing information to the switch chips. The control chip determines where incoming packets should be stored, as well as when and where each of the packets stored in the switch chips should be sent out.

3. CONTROL CHIP ARCHITECTURE

The control chip has four components, Fig. 3. First is the flow control unit, which communicates with the control chips in the next layer. It indicates the capacity of this switch level, from empty to full. The second component, the available register index, keeps track of which of the 16 cells are valid. The third part, the register file, makes up the bulk of the control chip (see figure 4). It holds the 256 21-bit headers, and contains over 130,000 transistors. The fourth component of the control chip, the comparator, looks for the oldest packet.

The register file stores the header data for each of the packets stored in the switch chips. The main building block of the register file is the register unit, figure 5. Sixteen register units wired together form a register module. The register file has 16 register modules, each wired to a different data bus. The register modules are selected by special decoders which enable it when the 4 bit address matches the register module's address. Note that the comparator of the register unit is a small circuit which compares 2 5-bit routing numbers. This should not be confused with the timestamp comparator discussed below.

The control chip decides which packet is the oldest via a timestamp comparator, figure 6. The 256 timestamps are fed into the comparator serially, with the most significant bit fed in first. Only one bit from each timestamp is used at a time. The 256 timestamp bits are ANDed with the previous bits (originally all 1's) in parallel, and the second register stores the results for next time. This process continues until only 1 of the 256 bits is a logical 1. Thus, the timestamp with the longest run of 1's in the most significant bits will be declared the oldest. The zero detector determines when a tie exists between two or more timestamps. It prevents the other timestamps from having a chance to compete again for the oldest timestamp. When only one timestamp has a corresponding logical 1 stored in the second register, an encoder translates the position of this "winning" bit to an 8 bit packet address. The control chip sends the address to the switch chips, and the switch chips send out the packet corresponding to this address. Eight bits of destination data are also sent from the control chip to the switch chips. The switch chips route the packet to the next node of the switch.

4. SWITCH CHIP ARCHITECTURE

The switch chip has a 4-bit wide 16x16 input crossbar, 16 blocks of static RAM (SRAM), each storing 1792 bits, and a 4-bit wide 16x16 output crossbar, Fig. 7. Each crossbar and the registers use small units to buffer control information. The packets are bit-sliced, so each switch chip stores one fourth of each packet. For the 16x16 switch, the input ports to each switch chip are 4 bits wide per channel. The input crossbar routes the packets to the SRAM block according to the high order (4 bit) address from the control chip. The low order (4 bit) address, also from the control chip, specifies which of the 16 sub-blocks should store the

bit-sliced packets. Each SRAM block of figure 8 is made up of 16 28x4 blocks. Since the switch chips are bit-sliced, they only need to operate at one-fourth the speed of the control chip, or about 40 MHz.

5. CONTROL CHIP IMPLEMENTATION

Due to the high performance requirements of the control chip, a number of specialized circuits are required for its implementation. These circuits include many components working at 155 MHz, which is four times the speed of the rest of the chipset. Pipelining is used to help achieve these speeds. Special mixed-signal devices, such as the zero-detector, are used in critical places to increase speed and save area. The combination of these features allows the chip to meet its performance goals.

The register unit has 2 sets of D-latches, one for the routing data, and one for the timestamp data. The 5 bits of routing data and 16 bits of timestamp data are referred to as the header data. The routing data is used to determine when to present the timestamp data to the comparator, forming a content addressable memory. At the beginning of a new cell cycle, the register will load the routing information in one clock cycle, and load the timestamp data in the following clock cycle. For each of the 16 buses going to the register modules, a routing filter sends 5 bits of routing information first, followed by 16 bits of timestamp data. In the routing layers, if the routing information stored by the register unit matches the 5 bits of routing information sent along the routing bus, then the register unit sends the 16 timestamp data bits out in high speed serial. By contrast, the chips in distribution layers always send the timestamp data, regardless of the routing data. Four select lines strobe the multiplexor to send all 16 timestamp bits through the output, one at a time. Register units are used instead of SRAM due to speed concerns. The register units provide higher bandwidth, and are content addressable. The register file, which holds 256 register units, takes up most of the space on the control chip, using an area of 6,300x7,500 lambda. The register modules, stacked on top of each other, created a structure that was too high to fit on the silicon die. The designers used an alternate layout, where 8 register modules are stacked on top of each other, in two columns. The outputs from the modules in the left column route their outputs under the modules on the right. Therefore, all of the register file outputs are on the right-hand side, to interface with the comparator.

The comparator is used to determine the oldest timestamp value. Its zero detector determines when all inputs are 0. When this occurs, all of the inputs are replaced with the outputs of the register in the comparator. This allows the comparison to be skipped for that bit position. In essence, the upper bits in the older timestamps are equal. Figure 9 shows a sub-circuit used both in the zero detector and the encoder. This circuit has two outputs, one corresponding to at least one input being high, and the other to more than one input being high. From these two values three states can be distinguished: no inputs high, exactly one input high, and two or more inputs high. This mixed signal circuit is used because it is both faster and much smaller than a conventional combination circuit.

6. SWITCH CHIP IMPLEMENTATION

In the crossbar, any of the 16 input channels can be switched to any of the 16 outputs. The channel outputs are found by ORing the outputs from each crossbar module together. A 4-to-16 decoder generates the 16 control lines needed for 16 crossbar modules, as well as the complement of the control lines. Due to the decoder, only one crossbar module is activated at a time. The 16x16 crossbar needs 80 control lines, since the 16 decoders need 4 address lines and 1 enable line each. A shadow register holds these control signals until they can be sent in parallel. The crossbar circuit uses about 5000 transistors.

The switch chip contains 16 SRAM blocks. Each block breaks down into 16 sub-blocks, capable of storing 28x4 bits, Fig. 10. The standard SRAM design of 6 transistors was used to make a cell. First, the 28x4 sub-block was designed, then two sub-blocks were connected by sharing a chain of D flip-flops to minimize the area needed. The chain of D flip flops allow 4 bits to be accessed by a sub-block at once. Each D flip flop enables one of the 28 SRAM words of 4 bits. This approach saves area over the original design of addressing each 4 bit word independently, which would have required 5 more address lines and larger decoders. The decoders used were made without the second metal layer, to route the SRAM data lines over them. A block of SRAM uses 14,300 transistors for the data storage, timing circuits, and interfacing. Therefore,

to have 16 SRAM blocks on the switch chip, a total of 228,600 transistors are used.

To control the output crossbar, the switch chip uses a small control circuit, Fig. 11. It receives a 5 bit address from the control chip, which it compares to the 16 blocks in the circuit's first level. When a match between addresses exists, then two packets are destined for the same output port. This is known as the hot-spot problem. To address this problem, the small control circuit sends the address causing the problem to the second level of 16 blocks. The design uses three levels of address blocks to obtain acceptable throughput, even with the hot-spot problem.

7. CONCLUSIONS

This paper describes the design of a chipset used in a distributed banyan ATM switch. A switching element consists of one control chip, and four identical switch chips. Simulations have confirmed that the new switching node enabled by the presented control chip has over twice the storage efficiency of previous nodes for the distributing banyan network. This chipset can be used as a 16x16 switching element. These switching elements allow the network designer to make a 256x256 distributed banyan network. The design uses distributed control which makes the switch scaleable. Therefore, the chipset presented here makes a large, scaleable ATM switching node.

ACKNOWLEDGMENTS

The authors would like to thank Sandeep Chaparala and Kambiz Zamani for their help in implementing this chipset.

The authors also acknowledge the support by DOE Grant # DE-AC05-840R21400.

REFERENCES

[1] Shipley, Paul, "VLSI Architectures for a New Class of ATM Switches", *Ph.D. Dissertation*, 1996.

[2] Shipley, Paul; Sayed, Sherif; Bayoumi, Magdy; "A High Speed VLSI Architecture for Scaleable ATM Switches", *Proceedings of the Sixth Great Lakes Symposium on VLSI*, pp. 72-76, March 1996.

[3] Weste, Neil; Eshraghian, Kamran; Principles of CMOS VLSI Design: A Systems Perspective, Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[4] Mukherjee, Amar, Introduction to nMOS and CMOS VLSI: Systems Design, Prentice Hall, Englewood Cliffs, New Jersey, 1986.

[5] Woodruff, G. M., Rogers, G. H., and Richards, P. S., "A Congestion Control Framework for High-Speed Integrated Packetized Transport", in *Proc. IEEE GLOBECOM* (Hollywood, FL), pp. 203-207, Nov. 1988.

[6] Giacopelli, J. N., et al., "Sunshine: A High-performance Self-Routing Broadband Packet Switch Architecture", *IEEE J. Sel. Areas in Commun.*, vol. 9, no. 8, pp. 1289-98, Oct. 1991.

[7] Kondoh, Harufusa et al., "A 622 Mb/s 8x8 Switch Chip Set with Shared Multibuffer Achitecture", *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 7, pp 808-814, July 1993.

[8] Mirfakhraei, Nader, "Design of a CMOS Buffered Switch for a Gigabit ATM Switching Network", *IEEE Journal of Solid-State Circuits*, vol. 30, no. 1, pp. 11-18, January 1995.

[9] Denzel, W. E., Engbersen, A. P. J., Iliadis, I., "A Flexible Shared-Buffer for ATM at Gb/s Rates", *Computer Networks and ISDN Systems*, vol. 27, pp. 611-624, 1995.

[10] Shipley, Paul; Weeks, Michael; Bayoumi, Magdy; "A Scaleable ATM Architecture for Bursty Traffic", *Proceedings of the Fifth International Conference on Computer Communications and Networks*, pp. 188-191, October, 1996.

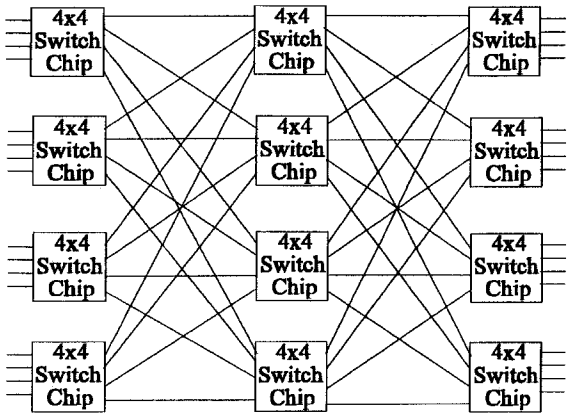

Figure 1. A 16x16 Distributed Banyan Network Composed of 4x4 Switching Elements

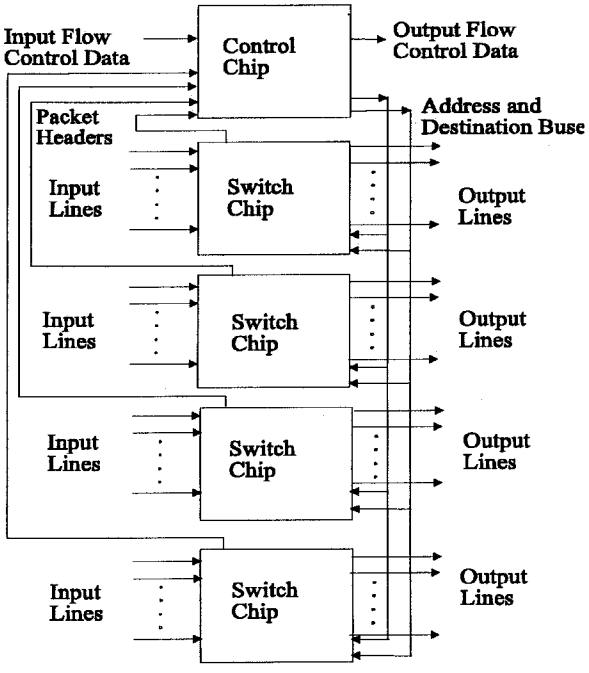

Figure 2. Diagram of the 16x16 Switching Node

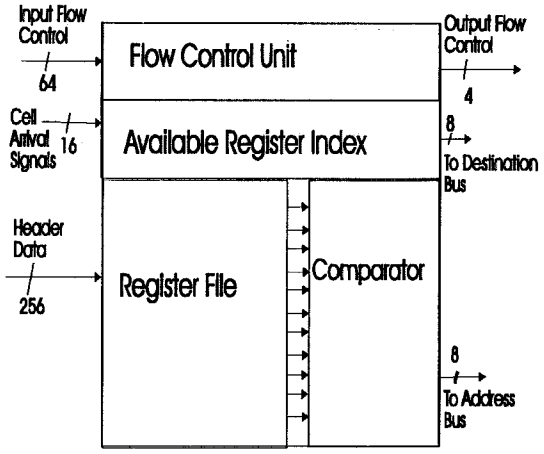

Figure 3. Structure of the Control Chip

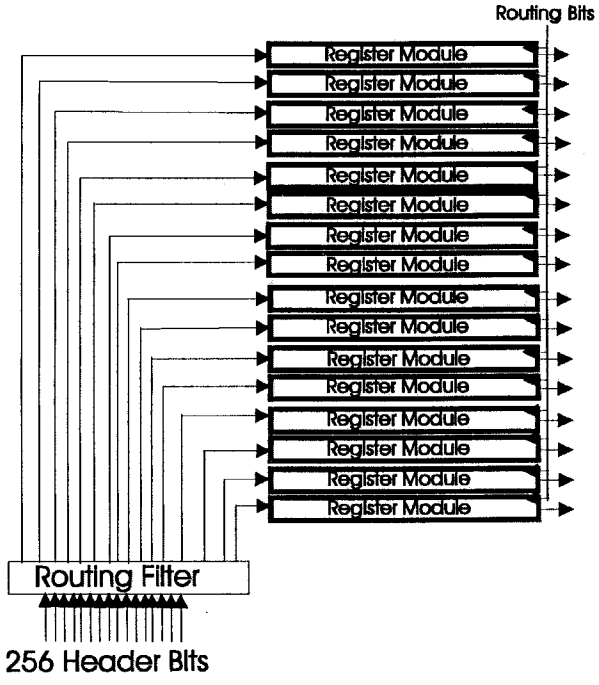

Figure 4. The Register File

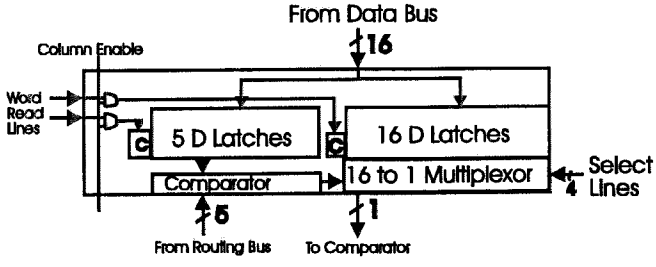

Figure 5. A Register Unit

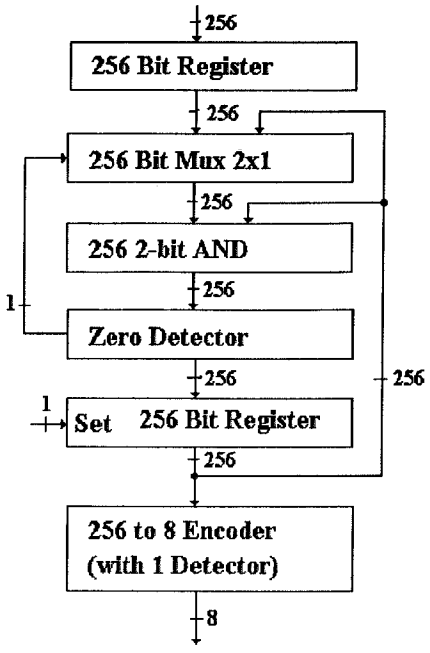

Figure 6. Structure of the Comparator

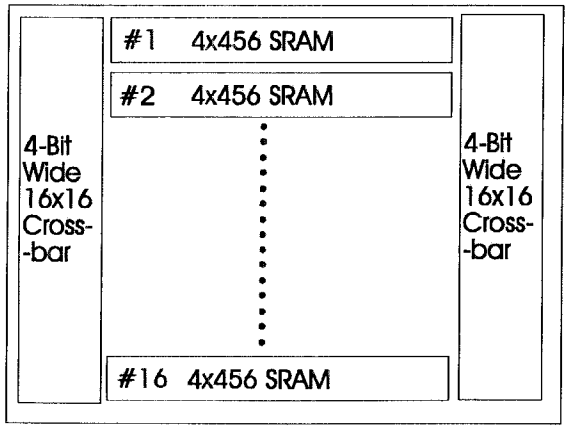

Figure 7. Structure of a Switch Chip

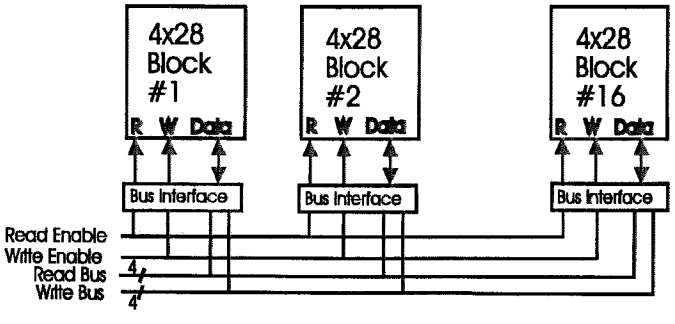

Figure 8. SRAM Memory Module

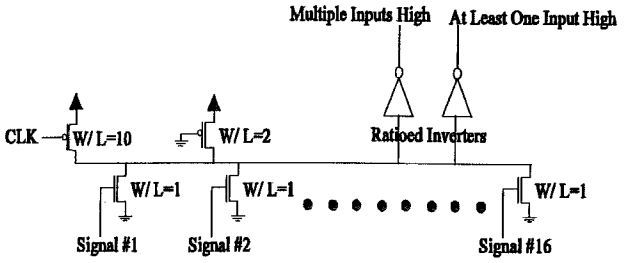

Figure 9. Circuit Determining 0,1, or More Inputs High

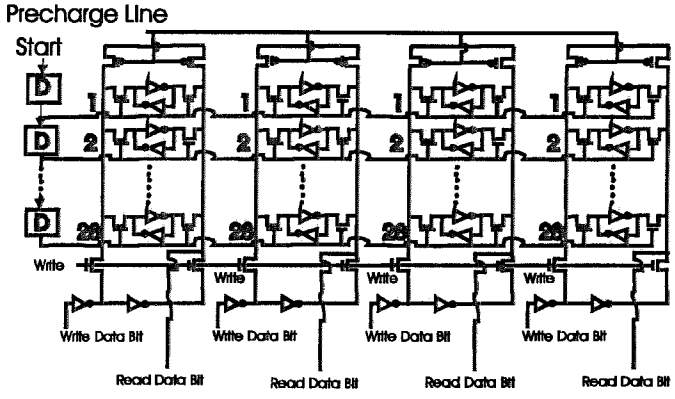

Figure 10. Diagram of 4x28 SRAM Block

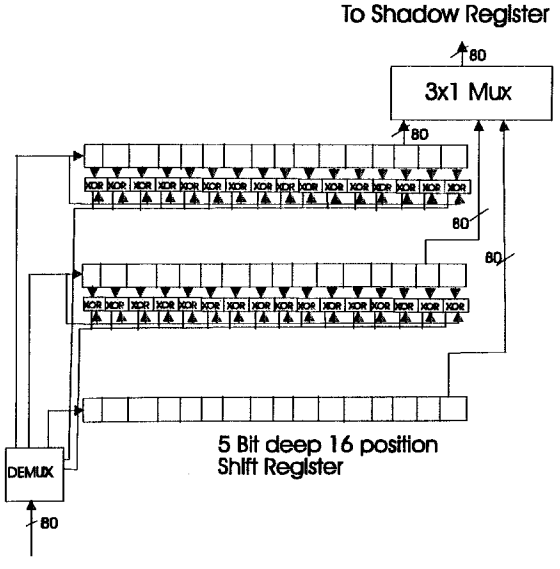

Figure 11. Output-Side Crossbar Control for the Switch Chip