A shared buffer memory switch for an ATM exchange

Hiroshi Kuwahara, Noboru Endo, Mineo Ogino, Takahiko Kozaki

Central Research Laboratory, Hitachi Ltd.

Yoshito Sakurai, Shinobu Gohara

Totsuka Works, Hitachi Ltd.

Abstract

Of the various ATM (Asynchronous Transfer Mode) switch architectures proposed to date, the memory switch type seems to have the highest hardware-utilization efficiency. In other words it requires the least hardware to realize a certain packet throughput.

This paper proposes a shared buffer memory switch in which output buffer memories are shared by all the switch output ports and are alloted to one particular output port as the occasion demands. This switch architecture can further improve the hardware-utilization efficiency of the memory switch by increasing the buffer memory usage rate.

Discussion on switch traffic characteristics indicates that buffer sharing reduces the required memory size to less than 0.14 of that otherwise required for adequate switch size and estimates roughly the buffer size required for the switch.

The resultant LSI count, for example, is about 15 chips for the main part of a 32×32 switch (150Mbps for each port) which can be mounted on one printed board. The switch is partitioned into a buffer memory LSI and a control LSI to make them flexible to change in ATM switch specifications.

1. Introduction

The Asynchronous Transfer Mode (ATM) is considered a promising technique to transfer and switch various kinds of media, such as telephone speech, data and motion video, in broadband ISDN. It can improve the utilization efficiency of trasmission and switching equipments by statistical multiplexing fixed length packets of these media on a broadband transmission line (150Mbps or 600Mbps). The major items for developing ATM are switching architecture, protocol structure and traffic control technique.

This paper proposes a new ATM switch architecture comprising a shared buffer memory switch which seems to have the potential to provide good traffic characteristics and easy LSI implementation. In this architecture, the buffer memories for output-queueing are shared by all the switch output ports and are alloted to one particular output port as the occasion demands.

We will discuss the effects of memory quantity reduction enabled by buffer memory sharing and roughly estimate the size of buffer memory required. Implementation of the switch in LSIs is also discussed from the viewpoints of partitioning to permit easy modification corresponding to an item of ATM switch requirement change, using an updated CMOS technology. Priority control functions for service classes and the broadcast function are also discussed. It will probably be necessary to add these functions to the ATM switch without inherently changing the switch architecture.

2. Requirement guideline for ATM switch

We assume the requirement guideline of an ATM exchange office in 1990's to examine switch architectures, as shown in Table 1.

At the first stage of broardband ISDN, the line bit rate of 150Mbps will be mostly used. This bit rate can transmit NTSC/HDTV motion-picture information.

Total throughput can be regarded as a scale measuring ATM switch capacity. In the B-ISDN public network an ATM switching system may be required to accommodate almost 1,000 trunks. This means it is required to have an approximate throughput of 150Gbps.

To realize such large capacity in a switching system, a multistage switching network can be adopted. Although the method of expanding switch capacity remains for further detailed studies, a 3-stage-switching network with 32-port-unit switches must be appropriate to constitute a 1,000-port-switch network.

Table 1 Guideline requirements of an ATM exchange office

Items	Requirement
Application Field	Public Network
Line Bit-rate	155.52Mbps
Total Throughput	150Gbps
Total Line Capacity	About 1,000 lines
Unit Switch Capacity	About 32 lines
Trunk Utilization	More than 70%
Cell Loss Probability	$< 10^{-9}$
Cell Structure	Header 6B User's Data 66B Total 72B
Service Class Broadcasting	Under study in detail

CH2655-9/89/0000-0118 $1.00 © 1989 IEEE

3. ATM switch architecture

An ATM switch consists of two functional components: a buffer memory in which ATM cells are stored for queueing, and a switching element through which the cells are transferred from an input port to an output port. An input-buffer type switch, which has input-buffers for the queueing in front of the switching element can utilize only less than 60% of its full bandwidth of the switch components , because of Head of Line (HOL) blocking [3].

An output-buffer-type switch which has output-buffers at the rear of the switching element can achieve a higher utilization efficiency because they have no congestion factor such as HOL blocking. A memory switch in which the sequence changes between memory-write and memory-read carry out buffering for both the queueing and switching functions is also classified into this category with regard to HOL blocking.

To minimize the quantity of hardware, the memory switch is made smaller than the output-buffer-type switch. This is because the handling of multiplexed cells from and to all the input and output ports efficiently utilizes a switch's switching circuit part and control part. The high-speed cell-handling requirement of the memory switch, which is imposed in return for multiplexing, will be overcome by using a submicron CMOS technology and a bit-parallel logic circuit. Thus, the memory switch seems to have the greatest potential of the various ATM switch architectures proposed to date[1],[2],[3],[5],[6].

3.1 A shared buffer memory switch architecture

The shared buffer memory switch proposed in this paper is depicted in Fig.1. The operation of this switch is as follows.

Input cells are converted from serial-data format to parallel-data format in S/P. Further, their header parts are sent to HD CNV to obtain the output-port addresses to which they should be sent , and to obtain new headers to be attached to the cells when they are sent from a switch. Subsequently, cells from all input ports are multiplexed in MUX and written one-by-one into CELL BF PART of BFM. The write address used is obtained from the WA corresponding to the destination output port of each cell and which is designated by the output of the RT DEC. The RT DEC decodes the output port address for each cell received from MUX. A BFM idle address is simultaneously read from IA BF and written into ADR CHAIN POINTER in the address into which the input cell is written, and is also stored in the WA ,overwriting the former WA content. This operation constructs an address chain within ADR CHAIN POINTER, corresponding to each output port and which performs the role of an output queue.

The address chain of output queue is shown in Fig.2. WA0 and RA0 correspond to a certain output port, indicating the end and beginning of the chain, respectively, and WA1 and RA1 correspond to another output-port. The end of the chain indicates the address into which the next input cell to the output queue will be written.

When a cell is output from the switch, OUT CNT sequentially indicates all the switch output ports for the demultiplexing of the cells read out from the BFM. The output cell is read out from the BFM with the address obtained from the RA indicated by the OUT CNT through OUT DEC and is sent out through DMUX and P/S. The content of the RA is also stored in the IA BF, because this address becomes idle and is replaced by the next chain address read from the same address as the output cell. The IA BF has a FIFO(First-in First-out) function and it stores idle addresses of the shared buffer memory.

This mechanism permits each BFM address to be alloted temporarily to any output port as the occasion demands, not permanently to one particular output port. The output queues are also stored in the same BFM as the ATM cells. This dynamic allotment explains why we call this switch architecture a shared buffer memory switch, which reduces the required buffer memory quantity as discussed in section 4.

3.2 Priority control for various service classes

The various media handled by an ATM switch have different requirements with regard to switching delay and cell loss probability. To meet these requirements, it is preferable to handle a cell differently during the user information transfer phase, according to the service class to

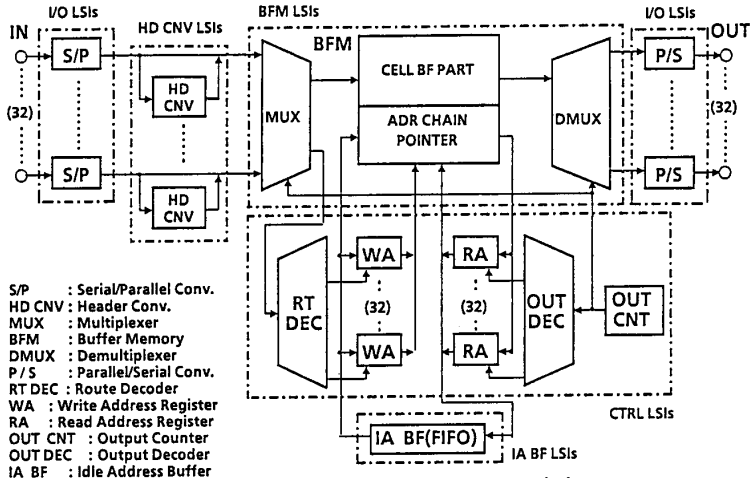


Fig.1 Shared Buffer Memory Switch

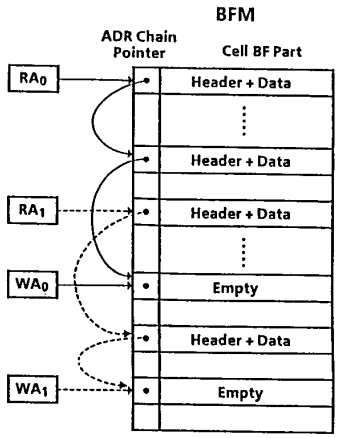


Fig.2 Output Queue Chain in BFM

which the cell belongs.

The shared buffer memory switch can be easily modified to handle plural service classes by providing a pair of WA and RA registers to correspond with each combination of service class and output port, respectively. An increase in shared buffer memory capacity caused by the addition of a service class function will be minimal because of the buffer sharing effect among the output-ports and service classes.

If it is necessary to restrict the length of the output queue for a certain service class, an up/down counter can be added to the pair of WA and RA registers for that service class. It will be incremented a cell input to the queue and decremented by a cell output from the queue. When the number indicated by the counter exceeds an upper limit , a new input cell is inhibited from being chained to the output queue.

3.3 Broadcast and service classes

A broadcast function, used for various distribution services or conference services, can be implemented with the addition of a broadcast circuit, as shown in Fig.3, to a shared buffer memory switch.

Broadcast cells form a broadcast queue in BFM irrespective of their destination output port. This is regarded as the same as output queues and has a set comprising a WA, an RA, an RT DEC output and an OUT DEC output. The broadcast function is explained using Fig.3, as follows.

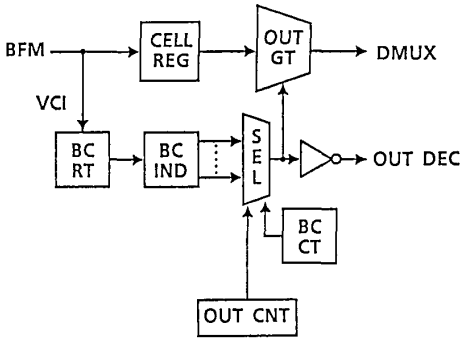
When a broadcast cell is sent out from the switch, it is read out from the broadcast queue chain in the BFM as described in section 3.1 and is stored in CELL REG. The cell header is sent to BC RT to be converted into cell routing information, i.e. the output port number to which the cell is broadcast. BC RT is a memory table, for example, whose addresses are designated by VCI and whose read data is a bit-pattern of each bit corresponding to the switch's output port . The routing information of the bit pattern is stored in BC IND.

The high- or low- logic signal level of the each bit indicates whether or not the cell in CELL REG is broadcast to the corresponding output port. The routing information corresponding to a call connection is written in BC RT by a call processor at call-set-up time. The output counter OUT CNT designates each bit in BC IND by a cyclic one-by-one round trip of all BC IND bits . When the SEL output is high, it opens OUT GT and inhibits the output of OUT DEC so as to skip the BFM read -out corresponding to the output-port number indicated by OUT CNT, and when it is low it closes OUT GT and enables the BFM read-out. When OUT CNT passes through all the output-ports, broadcasting of the cell in CELL REG is completed and the next broadcast cell is read out from the BFM.

BC CT designates whether or not broadcasting has been executed in each cycle of the round trip of the OUT CNT. The frequency of the broadcast execution cycle is determined by the call set up program when the program assigns a bandwidth to each broadcast virtual channel identifier.

In this broadcast method , because the broadcast cell is copied just when the cell is sent out, the copied cells do not occupy BFM and do not disturb the non-broadcasting cell input into BFM.

In this broadcast architecture, the throughput of the broadcast connection is limited to 1/16 of non-broadcast connection in the worst case, because just one cell can be broadcast during each cycle of the round trip of the OUT CNT. If the impovement in broadcast connection throughput is required by a broadcast application, this is made possible by modefying this broadcast circuit.



BC IND : Broadcast Output-port Indicator
CELL REG : Broadcast Cell Register
OUT GT : Output Gate
BC CT : Broadcast Cycle Indication Table
BC RT : Broadcast Routing Table

Fig.3 Broadcast Circuit

4. Buffer size estimation

In this section, we estimate a buffer size satisfied by a given cell loss probability and the reduction effect of total memory quantity by buffer sharing in comparison with a separated buffer-type switch in which a fixed-length buffer alloted to each output port. It is assumed that the input cells arrive at the switch randomly, and the cells are uniformly distributed across the output ports. In evaluating a separated buffer-type switch, we use the M/D/1/K model as a buffer queueing model at each output port.

In the shared buffer-type switch, the cell loss probability R is determined by the traditional approach of truncating the tail of the distribution obtained by assuming infinite buffer size at each output port [8]. Truncation at the far end of the distribution tail should provide a reasonably close approximation. The number of cells in the shared buffer Y is the sum of cells X in each queue at an output port, so the distribution of cells in the shared buffer is the convolution of the queue distribution of each output-port.

The Chernoff bound is available for bounding the tail of the sum of a large number of independent random variables, [7]. The Chernoff bound for the tail of a density function is denoted as

$$P[Y \geqq n\gamma_X^{(1)}(v)] \leqq \exp\left(n[\gamma_X(v) - v\gamma_X^{(1)}(v)]\right), \qquad (v \geqq 0)$$

$$\left(\ \gamma_X(v) \ = \ \log M_X(v)\ \right)$$

here, $M_X(v)$ is the moment-generating function of X and n is the switch size.

In the M/D/1 model,

$$M_X(v) = [(1-\rho)(1-\exp(v)]/[1-\exp(v)\cdot\exp((1-\exp(v))\cdot\rho)]. \qquad (v \geqq 0)$$

$$\therefore\ \gamma_X(v) = \log(1-\rho) + \log(1-\exp(v)) - \log[1-\exp(v)\cdot\exp((1-\exp(v))\cdot\rho)]$$

Using this bound, we estimate the buffer size which satisfies a given cell loss probability R. The relationship between the buffer size B and the cell loss probability R is shown in Fig.4. The shared buffer type requires fewer buffers than the separate buffer type. Figure 5 shows the buffer reduction ratio, which is the ratio between buffer size of the shared buffer type and that of the separate buffer type for the same cell loss probability, as a function of switch size n. The shared buffer type requires 0.14 of the buffer size of the separate buffer type under condition of cell loss probability $R=10^{-9}$, link utilization $\rho=0.8$, and switch size $n=32$.

We make a rough estimate of the buffer size required under bursty traffic conditions for LSI implementation of the proposed shared buffer memory switch. It is clear that the variance of the offered load between the output ports of the switch under bursty traffic conditions is larger than that under random traffic conditions. Therefore, the buffer reduction ratio with buffer sharing under bursty traffic conditions is smaller than that under random traffic conditions.

We estimate the buffer size of the shared-buffer-type switch under bursty traffic conditions using the buffer size of the separate-buffer-type switch under bursty traffic conditions and the buffer reduction ratio under random traffic conditions. With bursty input traffic, it is assumed that the bursts arriving during a unit service interval Poisson-distributed, and that the burst lengths are geometrically distributed with a mean $l=10$ cells. It is also assumed that the switch size $n=32$, link utilization $\rho=0.8$ and cell loss probability satisfies $R=10^{-9}$.

With calculation using $M^{[X]}/D/1/K$ model[9], 898 buffer cells per output port are needed in the separate-buffer type under the bursty traffic conditions mentioned above. The buffer reduction ratio under random traffic conditions is 0.14 as stated above. In the shared buffer type it is estimated that 125 cells of buffer are needed per output port to satisfy the cell loss probability $R=10^{-9}$ under bursty traffic conditions.

Thus, we set a buffer size for a 32×32 switch at 4k cells (128 cells / output port), to estimate the hardware quantity required for the LSIs.

5. LSI implementation

The proposed switch is partitioned into five kinds of LSIs, as depicted by the broken lines in Fig.1. The LSI arrangement structure is mounted on two kinds of printed boards with the chip count for each kind of LSI, as shown in Fig.6. The hardware quantity and circuit operation speed required for the Switch-main-part LSIs which is mounted on SW board are estimated as shown in Table 2.

The condition for this estimation are

(1)	Switch size:	32 input-ports by 32 output-ports (150Mbps / port),
(2)	Buffer quantity:	4k cells (128 cells / port),
(3)	Service class:	3 classes,
(4)	Broadcast:	Not implemented,
(5)	Cell size:	72 Octets,
and (6)	LSI technology:	Submicron CMOS process

In Fig.6, the BFM LSIs are arranged like a bit-slice i.e. each LSI corresponds to a bit of an 8 bit parallel cell format . This parallel format is converted to and from serial format by an I/O LSI. Each CTRL LSI corresponds to each priority service class, respectively. This partition is able to minimize the LSI modification caused by the changes in ATM switch specifications.

Only the BFM LSI has to be modified in response to the specification changes concerning the cell transfer part of the switch, such as ATM cell size or buffer capacity.On the other hand, only the CTRL LSIs provide switching control functions such as service class control, buffer overflow protection and broadcast function. Under certain conditions several kinds of CTRL LSIs have to be designed to correspond to particular control functions. For example, a broadcast function requires one more CTRL LSI to be added to the arrangement of Fig.6.

It seems to be feasible for a 32×32 switch to be mounted on one printed board (SW board, $300mm\times330mm$) considering the total chip count of about 15 LSIs for the main part of the switch (excluding the I/O LSI because it may be included in an input/output interface circuit of an exchange) as shown in Fig.6.

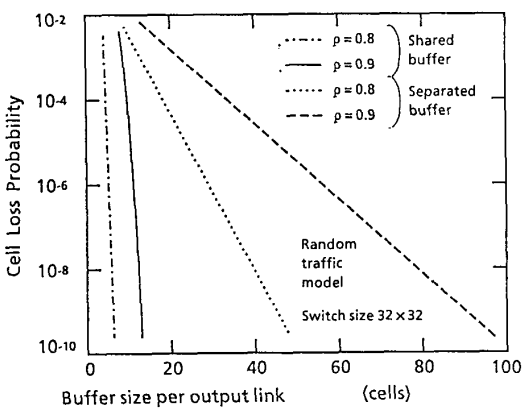


Fig.4 Cell Loss Probability on the memory switch

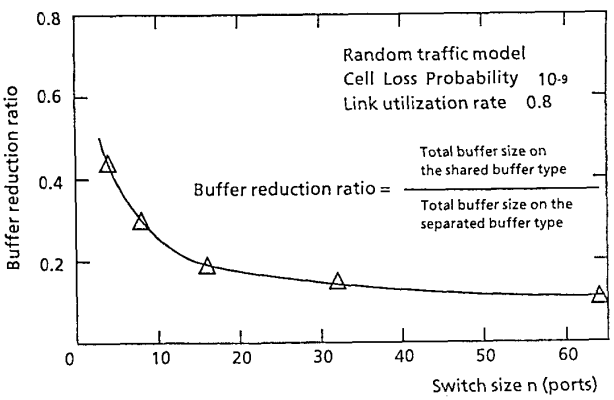


Fig.5 Buffer reduction by the buffer sharing

6. Summary

Compared to the ATM switch architectures proposed to date, the memory- switch type seems to have the highest hardware-utilization efficiency, or in other words it requires the least amount of hardware to realize a certain packet throughput.

This paper proposed a shared buffer memory switch in which buffer memories are shared among all the switch output-ports and are alloted to one particular output port as the occasion demands. This switch architecture can further impove the hardware-utilization efficiency of the memory switch by increasing the buffer memory utilization.

The discussion on switch traffic characteristics shows that buffer sharing reduces the required memory quantity to less than 0.14 of that otherwise required for adequate switch size and roughly estimates the buffer size required for the switch.

The resultant LSI count, for example, is about 15 chips for the main part of a 32×32 switch (150Mbps for each port) which can be mounted on one printed board. The switch is partitioned into a buffer memory LSI and a control LSI to make them flexible for changes in ATM switch specifications accompanying by some commercial available standard LSIs .

Table.2 LSI Specifications for a 32 × 32 sw

LSI		Gates		Memory		Chip	Signal	Process
		Count	Delay	Count	Cycle time	Count	Pins	
BFM	Custom made	55K	<0.8ns	360Kbits (4kW ×90bits)	<25ns	8	190	CMOS
CTRL	Gate Array	60K		-	-	3	250	
IA BF	Commercial Available FIFO	-	-	16Kbits (4kW ×4bits)	<50ns	2~4	28	
Total count						13~15		

The experimental logic circuit of the proposed switch architecture was fabricated using standard TTL ICs and is being evaluated from the viewpoints of traffic characteristics corresponding to various traffic models of ATM cell input.

Acknowledgment

The authors wish to express their gratitude to Messrs. Eiichi Amada and Hirotoshi Shirasu of the Central Research Laboratory, Hitachi Ltd., and to Messrs. Ken-ichi Ohtsuki and Takao Kato of the Totsuka Works, Hitachi Ltd., for their valuable contributions in discussion and guidance on ATM switch architecture, and to Mr. Ken-ichi Asano of the Central Research Laboratory, Hitachi Ltd., for his logic circuit design and estimation of the hardware quantity.

References

[1] Ohtsuki, K., et al. "A New Switching System Architecture for the 1990s" Proc. Forum, 1987, Vol.2, pp.15-19.

[2] Gohara, S., et al. "A New Distributed Switching System Architecture for Media Integration" Proc. ICC, 1987, pp.373-377.

[3] Hui, J., " A Broadband Packet Switch for Multi-rate Service " Proc. ICC, 1987, pp. 782-788

[4] Karol, M. J., et al. " Input versus Output Queueing on a Space-Division Packet Switch " IEEE Trans. Commun. vol. COM-35, pp. 1347-1356, Dec. 1987

[5] Eng, K. Y., et al. " A Knockout Switch for Variable Length Packets " Proc. ICC, 1987, pp. 794-799

[6] Thomas, A., et al. " Asynchronous Time-division Techniques: An Experiment Packet Network Integrating Videocommunication" Proc. ISS, 1984

[7] Kleinrock, L., "Queueing Systems," Vol. 1 : Theory, John Wiley and Sons, Inc.

[8] Eckberg, A.E. and Hou, T.-C., " Effects of Output Buffer Sharing on Buffer Requirements in an ATDM Packet Switch, " Proc. Infocom.,1988 pp.459-466.

[9] Chu, W.W., " Buffer Behavior for Batch Poisson Arrivals and Single Constant Output, " IEEE Trans. Commun. vol.COM-18,no.5, pp.618-618, Oct 1970.

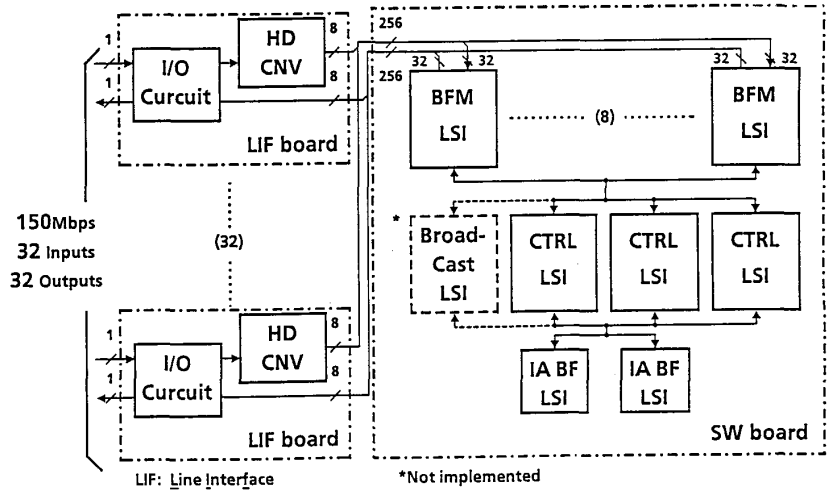


Fig.6 LSI arrangement structure