FACTA UNIVERSITATIS (NIŠ)
Series: Electronics and Energetics vol. 9, No. 2 (1996), 207–217

DESIGN OF A NEW ATM SWITCH BASED ON THE CENTRALIZED CONTROL

Zoran Petrović, Milenko Cvetinović and Vladimir Skulić

Abstract. This paper presents a new architecture for small to medium ATM switches at 150Mb/s based on the channel grouping principle. The channel grouping is used to enhance performances and centralized hardware implements port allocation procedure. The detailed description of the algorithm and the proposed hardware realization are given. A simple and flexible standard random access memory based hardware is suggested for the central control unit realization. In addition, it is shown that duplication further increases performances and provides graceful degradation in the case of failure.

1. Introduction

In this paper a new ATM switch is proposed intended for networks where channel grouping or link grouping routing strategy is applied. As described in [1] a switching node assigns a new call to a link group identified by a link group number. All the links belonging to a link group are connected to the same adjacent switching node so that all physical links connecting a pair of nodes may be treated as a single high bandwidth link. In that way high bit-rate services are supported, bursty traffic can be through multiplexing averaged, and the supervision and the control for a link group instead of every single channel are simplified. It should be pointed out that in this way switch throughput can be considerably increased. The proposed switch architecture is based on input/output queuing and multistage nonblocking network with selfrouting cells.

Manuscript received April 17, 1996.

The authors are with Faculty of Electrical Engineering, University of Belgrade, Bul. Revolucije 73, P.O.Box 816, 11001 Belgrade, Yugoslavia.

In paper [1] a switch architecture was proposed (Fig. 1) with two Distribution networks, Buffer subsystem, Batcher sorter and Banyan routing network. An original multilink access algorithm (MLA) was proposed to realize link group routing. This algorithm contains eight steps and requires feedback paths from output ports to input ports. In that paper performance analysis was accomplished concerning switch capacity, delay- throughput and packet loss probability. It is shown that this solution provides more efficient utilization of the transmission bandwidth between two switching nodes than unilink routing. It should be noticed that this solution is an integration of three–phase algorithm proposed in [2] and MLA.

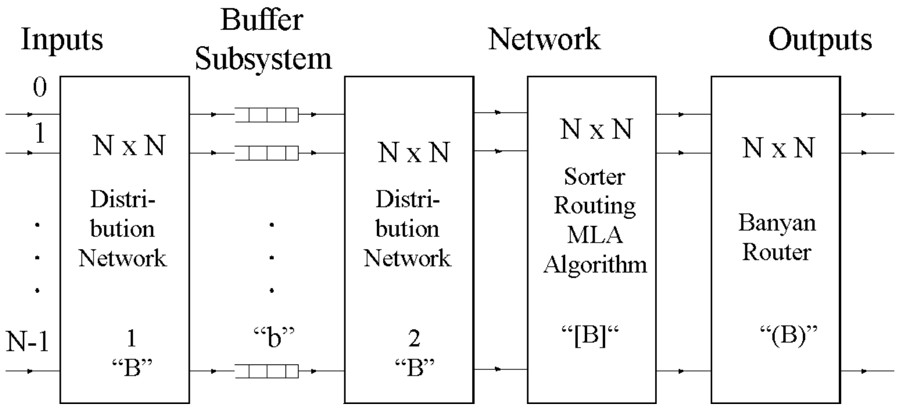

Figure 1. BbB[B](B) switch architecture.

In this paper a new ATM switch architecture is proposed with the same performance level as in [1] and with much simpler hardware. The working algorithm principle is similar to ring-reservation input queuing switch described in [3]. In ring–reservation switch (Fig. 2.), a reservation frame is serially transferred along the ring, so that each port controller PC_i ($i = 1, \ldots, N$) can reserve the output requested by its head–of–line (HOL) cell, if not already reserved by upstream PCs. Those PCs that successfully booked a switch output can transmit their HOL cell in the current time slot through the nonblocking network.

In this paper it will be shown that the proposed solution is further improved and is more efficient than the two previously mentioned and can use advantages of combined input and output queuing strategy [4].

2. Proposed architecture

2.1. Description of new architecture

The majority of the proposed solutions for ATM switches is based on decentralized switching fabric as well as control unit [5,6]. No doubt that such an approach is appropriate and necessary for high size switches based on standard solid state electronics technology. For lower and middle sizes, at 150Mb/s, a partially centralized approach offering simpler and more flexible control unit architecture is proposed in this paper. In practice, small switches may be useful in local exchanges where the link capacities between switching nodes may be different. To enhance performances, grouping channels approach like [1] and [8] is accepted.

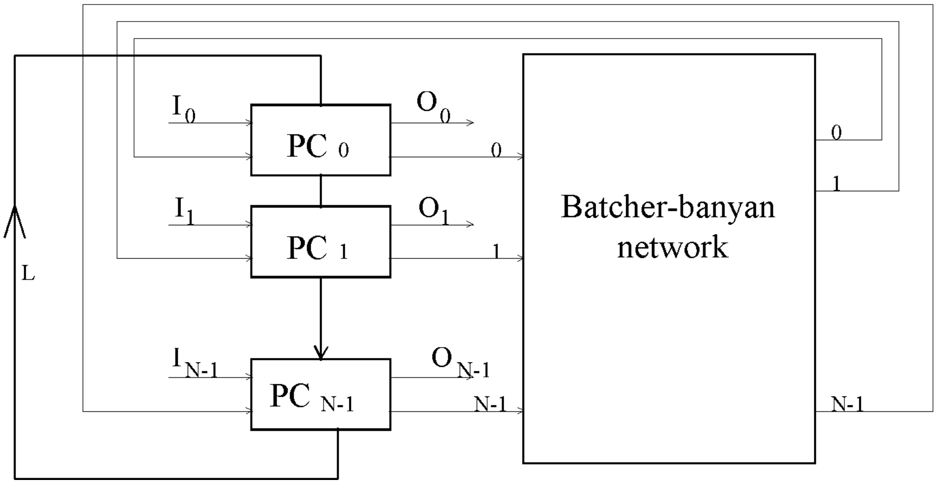

Figure 2. Ring reservation input–queuing switch.

The principal solution for a new ATM switch is presented in Fig. 3. It consists of input buffers (IQ), port controllers (PC), standard Batcher sorter as a switch fabric, and central control unit (CC). Port controllers are connected by parallel bus to CC. Port controllers provide data transfer from an external serial line to input buffer and access to cell header. With the help of central control unit, PC generates routing flag and sends data cells through switching fabric to output. Basically, port controllers provide the same functionality as in [2].

More detailed functional diagram for CC is in Fig. 4. CC contains port address register (Acnt), appropriate number of counters (Bcnt) for a number of reserved ports in a group, with busy group flip-flop (B) and necessary decoding (Dec) and sequencing logic (CCU).

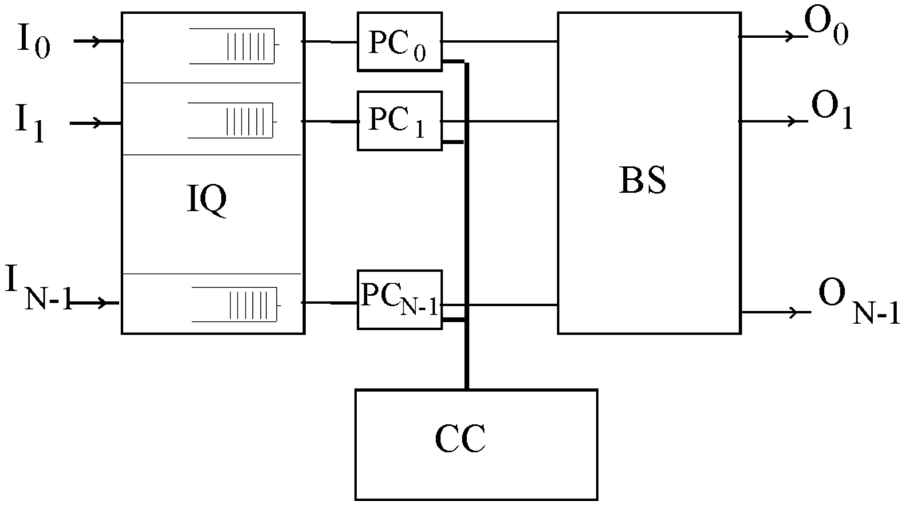

Figure 3. Proposed ATM switch structure.

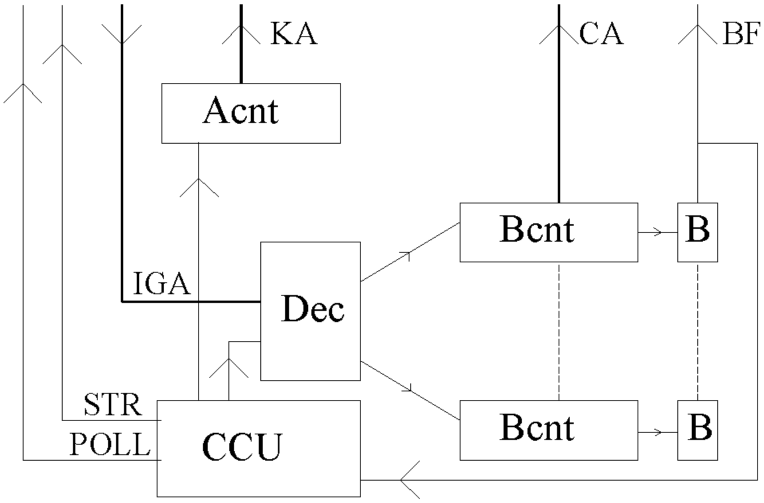

Figure 4. CC structure.

2.2. Description of the working algorithm

The working algorithm is as follows: in every time slot CCU serially selects a port controller (by activating strobe signal, STR, from CCU, and port address, KA, from Acnt). PC then sends the requested group number (IGA) according to HOL cell header. The decoded group number selects counter of already reserved ports in a group (Bcnt) and sends its content (CA) together with BF to PC. Bcnt is then incremented and BF set if the end of the group is reached. Control signals and timing diagrams during one cell time slot are shown in Fig. 5.

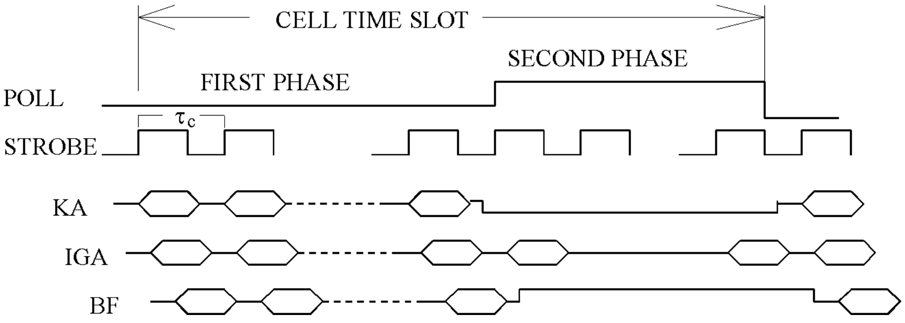

Figure 5. Controller timing diagrams.

As previously mentioned, switch outputs are grouped. Let N_i define a number of ports in a group i and N_g a number of groups. Then for switch $N \times N$, $N = \sum_{i=1}^{N_g} N_i$ is the number of ports. If n_i is the requested number of ports in a group i then if $n_i < N_i$ all requests will be fulfilled and $N_i - n_i$ ports should be idle. In the opposite case, if $n_i > N_i$ then N_i requests will be accepted and other, $n_i - N_i$, should wait for the next time slot. PCs that received permission (BF not set) can send their HOL cell through sorter. If banyan router is added after sorter the nonblocking switch is completed.

Similar to [7] we propose an additional step to eliminate the need for banyan router. This can be done by providing that idle port controllers (because they have no cells to send or requested output is not available) send test packets to unused outputs. In that way all sorter outputs will be occupied in the proper order. In other words, the sorter alone provides complete cell routing. Added test packets enable on line functional self testing.

The algorithm for the proposed additional step (second phase) is as follows: Control unit (CCU) activates POLL signal which is daisy chained through PCs so that each unreserved PC in succession receives the address of the next non reserved port. CCU provides address (Dec) for selecting Bcnt through its internal counter. When all ports in a group are reserved, BF flag is set and internal CCU group counter is incremented to point to the next group. This procedure provides that all port controllers and all output ports become active.

It can be seen that in both previous steps Batcher sorter (BS) is idle and so pipelining is easy to implement. During the transfer of a cell through BS, routing flag is prepared for the next one. In that case approximately $2.8\mu s$ (slot time at 150Mb/s) can be used for the proposed algorithm. This value defines, for the chosen integrated circuits technology, maximum switch capacity.

If, in each cell time slot, port scan counter (Acnt) starts from the next PC address we have a system with rotating port priorities. This helps against HOL blocking. Introducing additional RAM translation table in CA path much more flexible priority system (if necessary) can be implemented.

2.3. Maximum capacity of the proposed switch architecture

The maximum switch size using this architecture can be determined in this way. The first phase duration is fixed $N \cdot \tau_c$, where τ_c represents time required to service one PC request. The second phase duration is a variable number of cycles (which should be approximately of the same length) but in the worst case is $N - N_{imin}$ cycles if all PCs request the same group with the minimum number (N_{imin}) of ports in a group. With supposed $10ns$ cycle time and $2.8\mu s$ per cell, this architecture provides capacity of up to 128 ports per switch. Each cycle time (τ_c), with simple internal pipeline (Bcnt is incremented during non critical part of the cycle), should be greater than the sum:

$$t_b + t_{dPC} + t_b + t_{CC} + t_b + t_{sPC}$$

where: t_b - bus delay,

t_{dPC} - delay in PC from accepting address to sending requested group number,

t_{CC} - delay in CC from received IGA to send CA and

t_{sPC} - set up time for PC routing flag register.

This can be achieved using modern high speed CMOS technology with hardware proposed in the next section.

2.4. Suggested practical realization of CC

The suggested practical realization of CC is proposed in Fig 6. An array of counters and decoders is replaced by standard RAM memory with additional incrementing (INC) unit. Block RT is a part of RAM containing a number of ports in a group. Comparator (COMP) sets B flag (again a bit in RAM) based on up to that moment reserved number of ports in a selected group and available ports. Mx multiplexes externally (phase I) or internally (phase II) generated addresses. The required number of words in RAM is equal to maximum expected number of groups, N_{imax} in a switch. Required RAM width is $\log_2 N + \log_2 (N_{i\,\max}) + 1$.

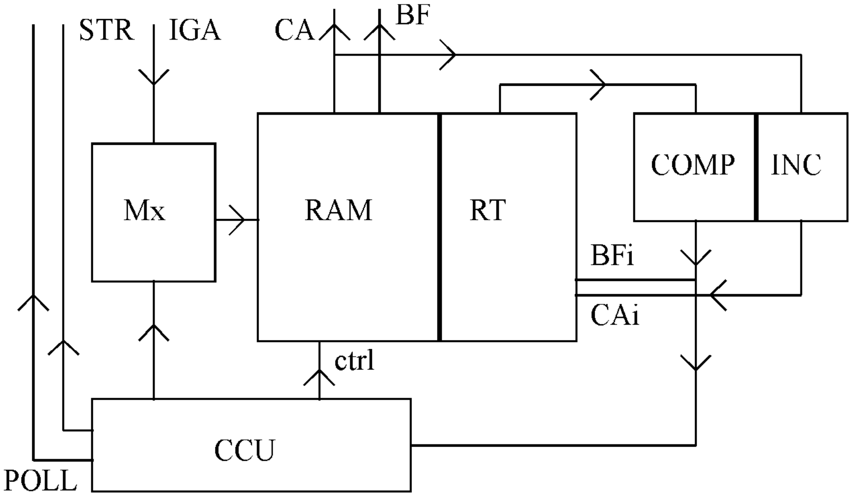

Figure 6. CC RAM based implementation.

Two disadvantages of the proposed solution should be pointed out. The first one is that the control unit is centralized and hence susceptible to catastrophic failure. The standard procedure to avoid this disadvantage is to duplicate the switching and control part of the system. Providing that additional controller takes the next cell from the input queue a sort of windowing (depth 2) is implemented. The second disadvantage is that this architecture employs input buffering scheme that is less efficient than input/output buffering, especially characterized by throughput [4]. Adding output buffer can bring additional performance enhancement and overcomes this drawback.

Duplicated switch architecture (Fig. 7.) is now with combined, input, (IQ), and output, (OQ), buffering and so provides an increased throughput for high traffic load. So, two additional advantages in improved ATM switch are: fault tolerance and HOL blocking avoidance.

3. Simulation results

In order to compare simpler single and duplicated switch architecture we developed simulation programs. Uniform traffic at switch inputs is assumed, cell arrival is the Bernoulli process with parameter λ and the requested destination has uniform distribution with probability $1/N$. We define the cell arrival rate λ as the probability that a cell will arrive on a particular input. This can be also interpreted as the offered load. Simulations are performed on an ensemble of 5,000,000 time slots. Figs. 8. and 9. present simulation results for a switch with 32 input and 8 output groups with 4 outputs per group. Fig. 8 shows cell loss probability as a function of input buffer length for different traffic loads λ (0.65, 0.70 and 0.75).

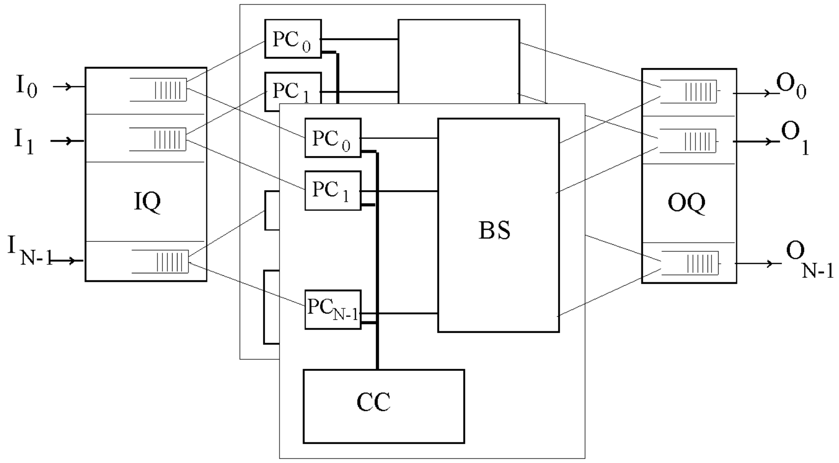

Figure 7. Improved implementation of ATM switch.

Fig. 9. shows a cell loss probability as a function of combined (summed) input and output buffer lengths. Actually, in duplicated switching fabric with added output queues, the cell loss is possible at both ends, input and output. A significant number of simulation runs shows that for uniform traffic distribution, even at high loads, input buffer length of 3 makes cell loss probability at input negligible compared to loss at the output.

The obvious conclusion from the carried out simulation for the duplicated switch is that the same cell loss probability can be achieved with smaller buffers. This effect is more visible at higher traffic loads. At 0.75 load and cell loss probability about $2.5 \cdot 10^{-7}$ single switch requires input buffer of 34, and duplicated only 9 $(3+6)$ cells.

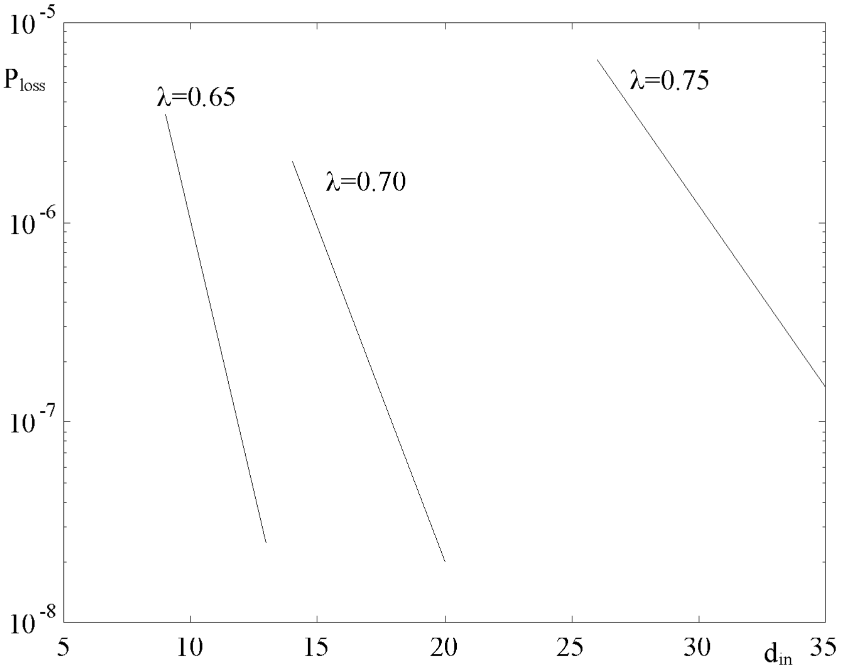

Figure 8. Cell loss probability vs. buffer length for input buffers, computed for uniform traffic at different loads, obtained by simulation of non–doubled switch.

It can be seen from Fig. 9. that with buffer size of 20 (3 for input and 17 for output buffer) per port, cell loss probability is 10^{-6} at traffic load $\lambda = 0.9$. As previously mentioned, input buffer size increase, does not considerably decrease cell loss probability. In spite of this, we suggest to extend input buffer size. Then, in case of one controller failure, with input buffer length of, for example 15, uniform traffic load $\lambda = 0.7$ can be handled with the same loss probability 10^{-6}. So the results on Fig. 8 can be interpreted as a cell loss probability versus input buffer length in the case of one switch failure.

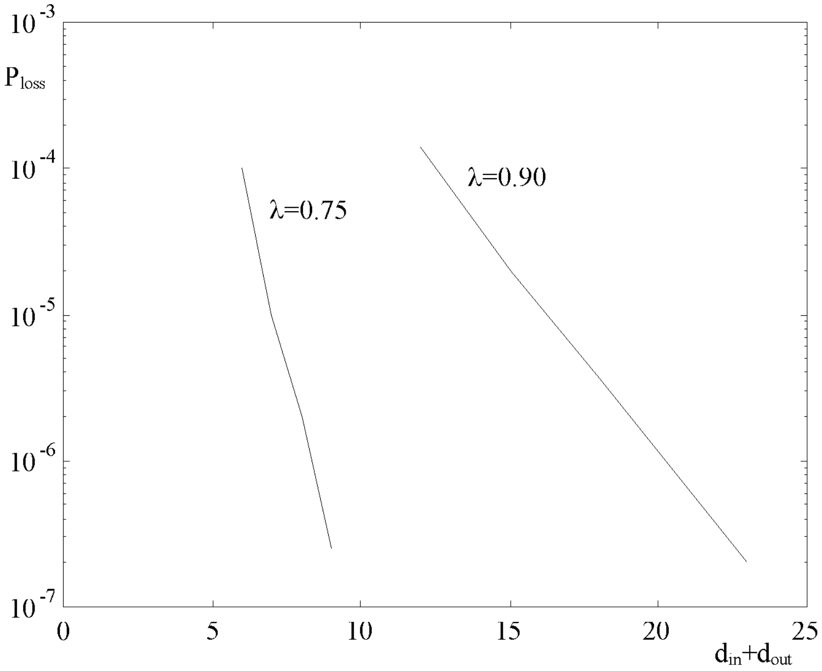

Figure 9. Cell loss probability vs. total buffer length for input and output buffers, computed for $d_{in} = 3$ and uniform traffic of two different loads, obtained by simulation. The size of switch is 32 × 32 and number of the output link groups is 8.

4. Conclusion

An architecture for small to medium ATM switches at 150Mb/s is proposed. Partially centralized control unit offers a simpler and more flexible solution when channels are grouped. The Batcher sorter is not used during routing flag determination which makes possible simple pipeline implementation. Banyan router is eliminated by a two phase algorithm similar to [7]. Rotated priority and duplication with load sharing enhance performance by lowering probability of head of line blocking and provide graceful degradation in the case of failure. Centralized architectures are prone to fatal errors and this approach combines high reliability and capacity. Low throughput, a feature inherent to input buffering, is avoided. RAM based channel grouping eases a system reconfiguration.

REFERENCES

1. P. Lau and A. Leon–Garsia: *Design and Analysis of a Multilink Access System Based on the Batcher–banyan Network Architecture.* IEEE Trans. on Com., vol. 40, No. 11, November 1992, pp. 1757–1766.
2. J. Hui and E. Arthus: *A Broadband Packet Switch for Integrated Transport.* IEEE SAC, vol. SAC-5, No.8, Oct. 1987, pp. 1264–1273.
3. B. Bingham and H. E. Bussey: *Reservation–Based Contention Resolution Mechanism for Batcher–banyan Packet Switches.* Electron. Lett. vol. 24, No.13, June 1988, pp. 722–723.
4. L. Yashe, L. Zengji: *Performance Analysis of ATM Switch Fabric with combined Input/Output Buffering.* ICCT96, Beijing, China, May 1996, pp. 508–512.
5. H. Ahmadi and W. E. Denzel: *A Survey of Modern High–performance Switching Techniques.* IEEE JSAC, vol. 7., No. 7, Sept. 1989, pp. 1091–1103.
6. E. W. Zegura: *Architecture for ATM Switching Systems.* IEEE Com. Magazine, Feb. 1993, pp. 8–37.
7. A. Pattavina: *Nonblocking Architecture for ATM Switching.* IEEE Com. Magazine, Feb. 1993, pp. 38–48.
8. A. Pattavina: *Multichannel Bandwidth Allocation in a Broadband Packet Switch.* IEEE JSAC, vol. 6, No. 9, Dec. 1988, pp. 1489–1499.