J.King Saud Univ., Vol.9, *Computer & Information Sci.*, pp. 125-148 (A.H 1417/1997)

Architecture Design of an ATM Switch Based on a High Speed Network

Sami S. Al-Wakeel
Department of Computer Engineering, College of Computer and Information Sciences
King Saud university, Riyadh, Saudi Arabia

(Received, 19 February 1995; accepted for publication, 03 July 1995)

Abstract. One of the key issues that must be fulfilled to realize BISDN, is to develop high speed and high capacity ATM packet switches. For this, many technical problems must be investigated, such as design of switch architecture, development of switch protocols, and evaluation of the switch performance.

In this paper we will propose an ATM switch with a shared medium architecture. The medium is a high speed fiber optic network with reservation based access protocol. The switch uses a novel interconnection topology between the switch units to reduce the fiber optics network aggregate data rate for a high dimensionality switch. The resultant switch dimensionality estimate shows that with a switch unit network bus rate of 4 Gb/s , a fully connected broadband switch with 16384 I/O STS-3C lines (155.52 Mb/s port bandwidth) can be realized. Moreover , the architecture of the switch allows a modular growth , meets the need for heterogeneous and dynamically changing mix of traffic, and provides multi-point connection capability.

Key words: ATM switches; Fast packet switching; Shared medium switches.

Introduction

In recent years, there is a growing need for a wide variety of communication services.

Asynchronous Transfer Mode (ATM) is expected to be the ultimate transport mechanism to support diverse applications ranging from low bit rate terminals communication and audio to graphics and broadcasting of high resolution videos [1] .

To match the transmission speed of the ATM network links, it is essential to develop fast packet switches suitable for providing ATM broadband communication services. Anther important requirement for ATM switching is the switch growability. The switch architecture should allow the switch to be modular and to permit the switch size growth without performance degradation [2]. To meet these requirements, in contrast to traditional packet switching approach, ATM switches use hardware switching fabric instead of computer controlled communication processes, and implement switching protocol in hardware rather than software [3].

Various architectures for ATM switches have been proposed and analyzed in the recent years. A summary and comparison of growable ATM switching fabric architectures including Knockout, Batcher-Banyan, folded and two sided three stage fabrics, is presented in [2]. Reference [3] assesses architectural characteristics necessary for ATM switching and profiles several commercially available systems. Survey of the performance of nonblocking switches with FIFO input buffers is presented in [4,5]. Reference [6] presents an overview of ATM switching and identifies six classes of ATM switching techniques in their survey based on the internal structure of the switching fabric. Reference [7] assesses ATM switches architectural characteristics and identifies three categories of ATM switching internal structure. These categories are Shared-memory, Shared-medium, and Space-division.

In shared-medium type switches, a common high-speed medium (e.g a bus or a ring) is time-multiplexed among several input/output connections. The traditional implementation of the shared bus switch requires a bus bandwidth equal to N times the rate of a single input lines (N is number of bus line connections). This approach, however, has several shortcomings. First, the switch size is directly proportional to the medium bandwidth , and therefore ,it may not fulfil the requirement of an ATM exchange with at least 1000 ports for subscriber lines or trunk connections. Second , the switch bus capacity may fall short from several Giga bps throughput that future broadband communication services require.

To increase the capacity of the shared-medium switch, a high speed network (HSN) with high throughput is used where the traditional physical layer (e.g. coaxial cables) is replaced with optical fiber. Examples for this type of switches are given in [8-11]. Nevertheless, a major problem associated with optical bus switches is that the

electronic of bus interface circuitry must operate at the bus aggregate rate, thus limiting the overall switch capacity.

In this paper we will propose a new architecture design of an ATM switch that implements a shared-medium. The medium is a high speed fiber optic network with a reservation access protocol [12]. This network will act as a base of an ATM fast packet switch. The switch architecture has three main significant principles : a high speed reservation oriented switch protocol, a novel interconnection topology between multiple switch networks which allows implementation of a high dimensionality switch without magnifying the aggregate data rate required, and the flexibility and simplicity of the switch routing scheme.

Switch Architecture

Figure 1 shows the overall switch architecture block diagram. The switch system constitutes of basic switch units (BSU), master switch units (MSU) and a single super switch unit (SSU). The switch units are connected as a three-level structure. A single BSU represents the first level and connects a number of switch input/output lines through the line interface modules (LIM). In the second level, several BSU's are connected to each other in a single MSU through basic interface modules (BIM). The switch MSU's are connected within a single SSU via the master interface modules (MIM) in the third level. For routing purpose, each LIM is labeled by a unique code consisting of three fields that identifies its output line, BSU and MSU respectively. The BIM has a label of two fields identifying the destined BSU and MSU where the cells read by the module are to be delivered. Similarly, the MIM label has a single field label that identifies the destined MSU.

Figure 2 shows the BSU architecture. A single BSU is the minimum configuration of the switch. It can be operated as a stand alone switch or acts as a first stage of a large size ATM switch.

Each BSU accommodates a fiber optics bus, a bus controller module (BCM) , basic and master bus interface modules, and terminates the switch input/output transmission lines. The fiber optics network job is the ATM cells multiplexing and to provide the transfer media between switch input and output modules and/or units. The network consists of a single folded unidirectional broadcasting bus. Signals transmitted by the network modules (e.g LIM) use the inbound channel (first portion of the bus) to send cells to other modules on the outbound channel (second portion of the bus). R, S

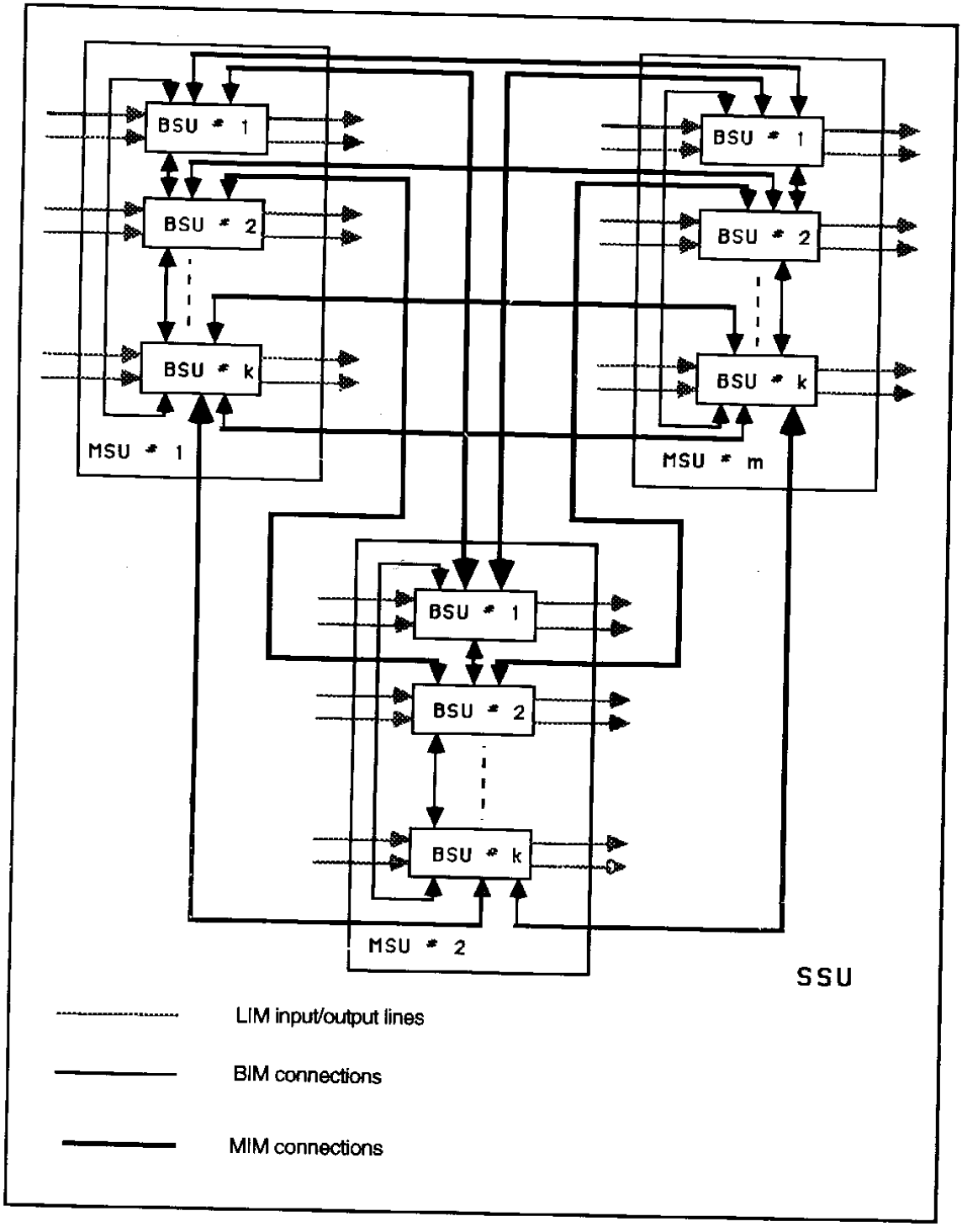

Fig. 1. Switch block diagram.

BCM Δ Bus Controller Module

LIM Δ Line Interface Module

BIM Δ Basic Interface Module

MIM Δ Master Interface Module

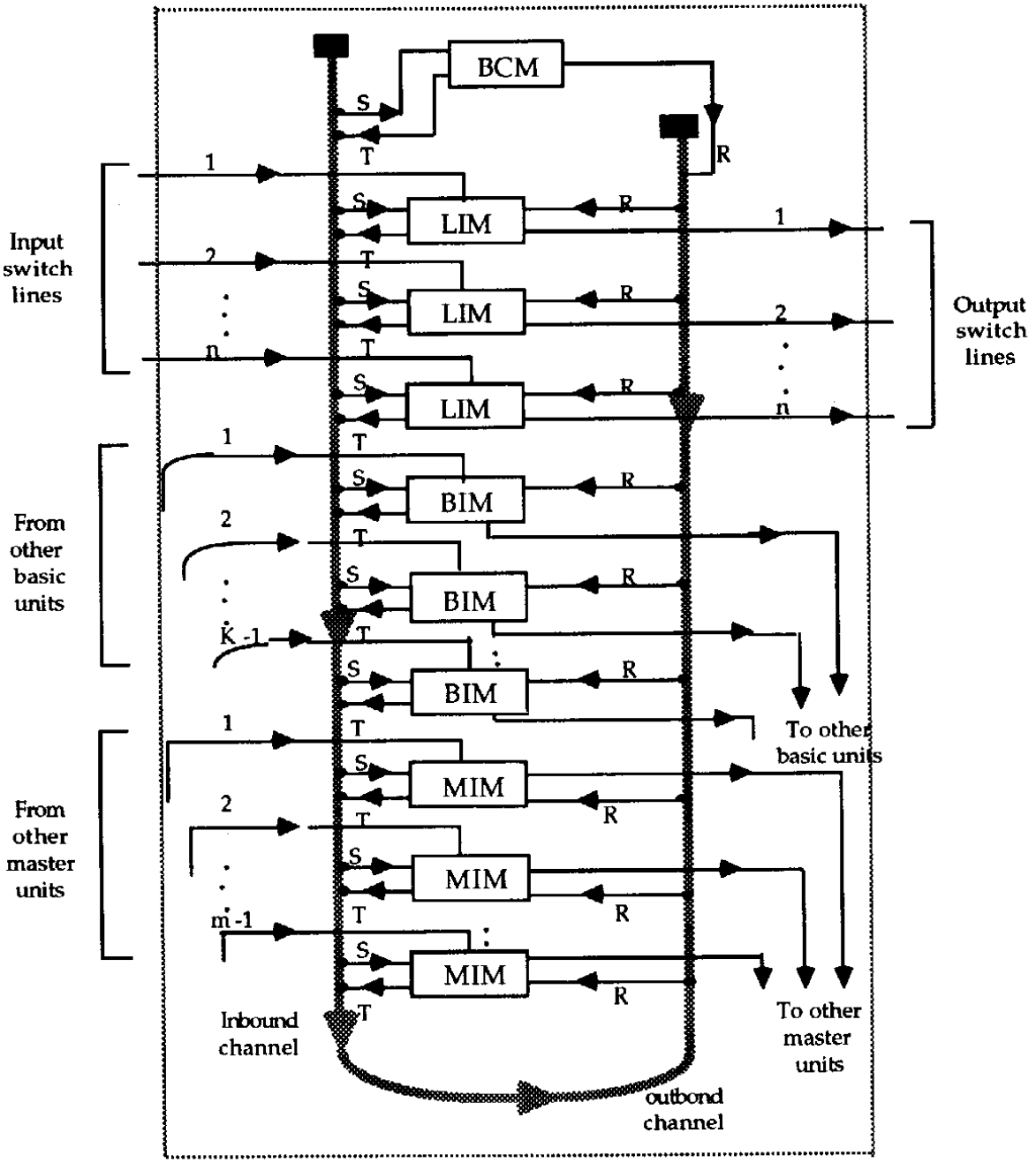

Fig. 2. Basic switching unit architecture.

and T are modules taps used to receive the ATM cells, sense carrier and to transmit cells, respectively.

The fiber optics network architecture and design are repeated at all levels of the switch hierarchy, and the same network MAC protocol is used in MSU and SSU of the switch. Details of the network architecture and protocol are given in reference [12].

The BCM is the first upstream node module connected to the fiber optics bus (other nodes are the LIM's and the BIM's and MIM's switch modules). It controls the bus operation by periodically transmitting, at its T tap, the bus timing pulses and synchronization patterns that signify the start of a slotted frame and mark the beginning of each frame slot, respectively. Besides, the BCM determines the number of slots to be generated in the next frame according to MAC procedure to be described in the next section. Other tasks performed by the BCM are to handle control and management cells transmitted over the BSU bus, and to perform maintenance jobs and fault recovery control.

The BIM's and MIM's bus interface modules are connected to both sides of the network bus and allows cells transfer between BSU's and the MSU's respectively. The interconnections between the BIM's and MIM's in both of the MSU's and SSU levels are based on CCITT Recommendation I.432.

The line termination of the BSU is done by line interface modules (LIM). These modules contain receive/send FIFO buffers, ATM cells header translation devices (HTD), fiber optics network bus interface (NBI) and carries out data link layer and physical layer control procedures. The HTD fetches the received ATM cell header which indicates the virtual channel identifier (VCI) of the call cells, and determines the outgoing route for the call cells. This is carried out by replacement of the original VCI by new VCI value and a routing header that is prebend to the cell and is used for internal switch routing only. The header specifies the cell output BSU and MSU and the output switch line where the cell is to be delivered. The NBI transfers the incoming and outgoing cells through the fiber optics bus and performs the fiber optics network medium access (MAC) protocol functions.

The LIM terminations of input and output switch lines are based on CCITT Recommendations I.413 and I.432 which adopt a cell-based and SHD (synchronous digital hierarchy) based, physical layer for transmission of ATM cells. However, the LIM may include a service interface adapter (SIA). This device is placed in front of each LIM to implements various services classes and protocols of ATM adaptation layer

(AAL). These classes are defined by CCITT Rec. I.362 and I.363. for support of information-transfer protocols not based on ATM (e.g LAPD). Thus, the SIA device allows the switch input lines to interface to existing broadband terminals, LAN's and MAN's. etc, to perform various protocols conversion and to support various new services such as SMDS (switched multigegabit services).

The switch MSU and SSU form the second and third levels of the switch hierarchy as shown in Fig. 1. Each of these units contains BSU's with fiber optics networks whose architecture is identical. The SSU implements a fully connected switch, and has three types of interface modules : the LIM which connects the lower level of the BSU's I/O lines , the BIM's and MIM which connect it to the second and third level of the SSU respectively. The MSU, however, only requires two types of interface modules namely the LIM's and BIM's to operate as a medium size switch.

Accordingly, the switch architecture distributes various switch data transfer functions, signalling and control functions over the switch modules. Table 1 shows switch functions categorization and distribution among various switch components.

Table 1. Distribution of switch functions

Layer	Function	BSU	LIM	BIM	MIM
Physical	I/O Line interface		o		
Data link	Link control	o	o		
=	Medium access	o	o	o	o
ATM	Cells multiplexing	o			
=	Switching		o	o	o
=	Cells discarding		o	o	o
=	Congestion control		o	o	o
AAL	Protocols conversion		o		
Network	Routing		o	o	o
=	Call control		o		
Higher layers	Management, control fault recovery and maintenance	o/ BCM			

Switch Medium Access Protocol

The fiber optics bus represents the core and the shared medium of the switch BSU unit. It uses a novel reservation-based access protocol. Due to the reservation approach, the performance of the network bus in not limited by operating data rate, has a unity utilization and allows bounded network medium access delay which is highly desirable feature for real time applications such as voice communications.

The network protocol imposes a slotted frame configuration. The slotted frame (generated by the BCM) is of variable size and the MAC protocol allows at the most two partial frames to exist simultaneously on the network bus. At normal load, the frame length will not exceed a given maximum length chosen such that it satisfies the maximum tolerable delay for real time or data communication through the switch.

The first slot in the frame is used as a reservation slot. The other slots are considered communication slots. Each communication slot transfers a single ATM cell. Reservation protocol operates as follows [12]:

When a cell arrives at an input line, or to be routed between BSU's, the cell is stored in the corresponding module input buffer. The line interface module prepares a reservation request to transmit, and waits until it senses the start of the reservation slot. Then, it checks for a group of eight zero consecutive bits assigned to the module within the reservation slot. If detected, the station immediately transmits its reservation request by setting some or all of these bits according to the number of slots needed in the next frame. Each module may reserve a variable number X of slots in the next frame according to the number of cells accumulated in its input buffer.

After successfully transmitting the reservation request byte, the input module, upon sensing the end of the reservation slot in a frame, waits for its previously reserved communication slot(s) in the frame. Upon arrival of the reserved communication slot(s), the module immediately transmits all its buffered cell(s). The frame transmitted cells can then be read by all other modules, at the R taps, on the network outbound channel.

For each frame, the BCM determines the next frame length (in slots) by reading, through its R tap, the status of the reservation requests made by each module in the current frame. The BCM determines the number of slots to be generated in the next frame according to the sum of reservations made by all active modules. Therefore, the frame length will vary from one frame to another depending on the reservation activity

of modules connected to the bus. More details on the network reservation protocol and architecture is presented in [12].

Switch Partitioning, and Bus Rate Estimation

The switch I/O N ports are partitioned in the first switch level into a number of BSU units, each having nxn input/output ports as shown in Fig. 3. Groups of BSU's constitute the MSU with $nkxnk$ I/O ports at the second level. With nxn input/output BSU lines, k BSU's per MSU and m MSU's, the full switch size is $N=n.k.m$. Within a single MSU each BSU is connected to every other BSU fabricated on the same MSU unit. Thus, each BSU needs (k-1) BIM's as shown in Fig. 2. To connect several MSU's together, BSU(i) of MSU (j) is connected to the corresponding BSU(i) of MSU(l) where $1<=i<=k$, and $1<= j ,l <= m$. Therefore, for SSU connection, each BSU needs (m-1) MIM interface modules.

Based on this proposed topology, it is clear that the switch has a regular and uniform structure. Thus, it is possible to have high integration density for VLSI implementation, and relaxed synchronization for clock and information signal. Besides, the number of BSU's and MSU's scales linearly with the number of switch I/O lines. The aggregate rate of the BSU bus is, however, almost constant for a large switch. This can be shown as follow :
let a be the input line access rate. For nxn input/output BSU lines, k BSU's per MSU and m MSU's per SSU, and under uniform traffic, the aggregate traffic on BSU bus R is the sum of BSU input lines traffic, transit traffic from other MSU's and external traffic destined on the BSU. The destined traffic constitutes traffic from other BSU's on the same MSU, and traffic from other external MSU's. So,

$$R = n.a + [(m-1).na.(k-1)/k.m] + [n.a.(k-1).1/(k.m) + (m-1).k.n.a.1/(k.m)]$$

or

$$R = 3\,n.a - n.a.(k+m)/(k.m)$$

Projections of Switch Size, Cost and Traffic Capacity

The switch topology and architecture make the development of a very large ATM switch realizable with current technology of electronic signal processing equipments which can handle high speed bit rate not exceeding 10 Gb/s. Besides, the

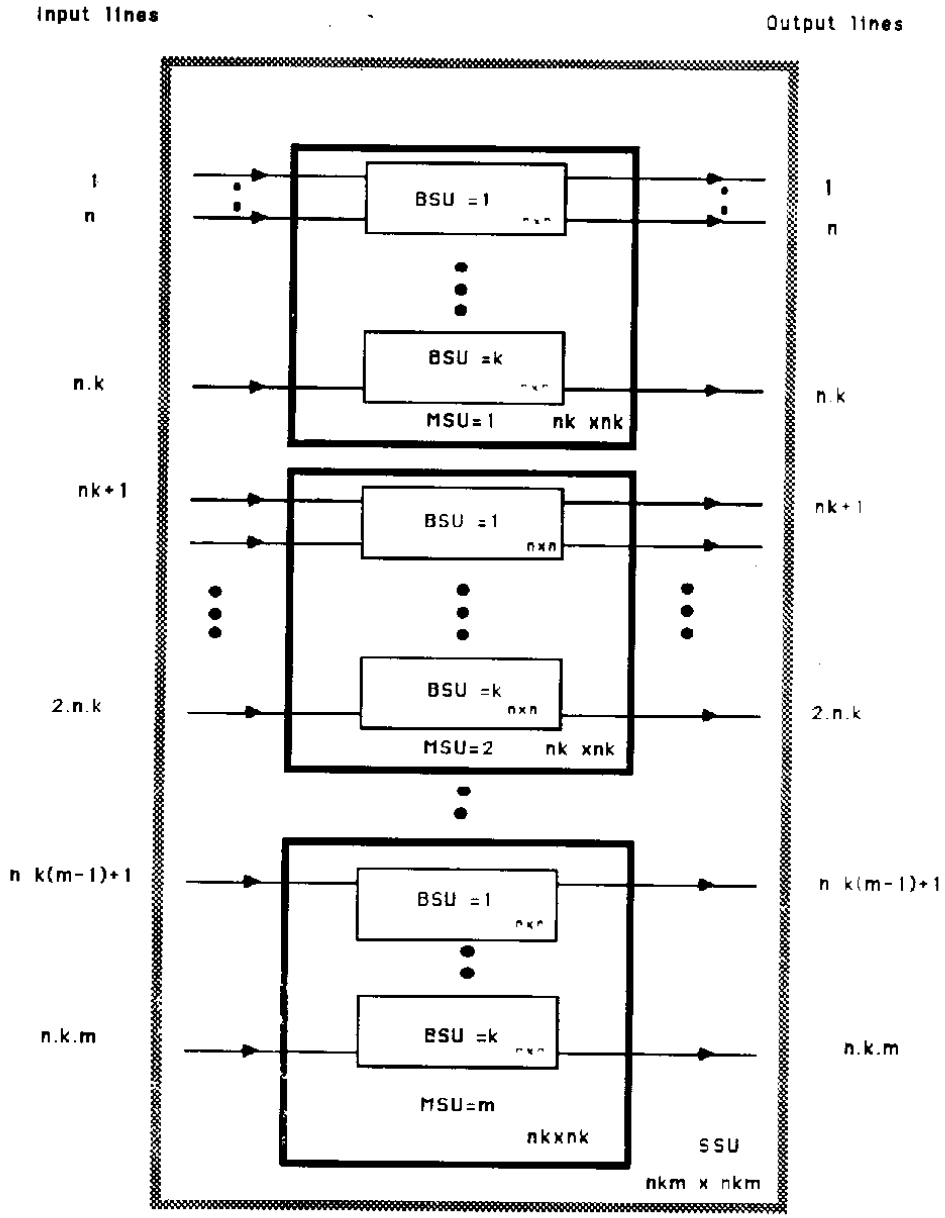

Fig. 3. Overall switch partitioning.

switch design has a clear advantage being modular in structure and can be configured easily to any size. The ATM switch size is determined by the numbers of lines terminated on the BSU, the numbers of BSU's and MSU's. The switch may be implemented as a single BSU only to serve a small size business community, or designed to construct systems for a large packet switching exchange, capable of

providing service to hundreds of lines. An example of typical values for the proposed ATM switch design parameters is: BSU with 8x8 I/O lines at 51.84 Mb/s (SONET standardized STS-1 digital data rate), 1.2 Gb/s fiber optics bus, 8 BSU's and 8 MSU's. With these values, a switch with two hierarchy levels (i.e.only BSU's and a single MSU) has a 64x64 size and a throughput about 3.2 Gb/sec. A switch with 8 MSU's connected, while having a BSU bus rate not exceeding 1.2 Gb/s, can reach a size of 512x512 and throughput about 27 Gb/sec.

Using lower data rate for I/O lines (e.g 6.312 Mb/sec of T-2 carrier) and 512 Mb/s bus data rate, each BSU can interface 32 input transmission lines, and a switch with a single MSU of 16 BSU's can easily reach the size of 512x512.

To design a switch that meets the ATM standard's requirements, the challenge is however to realize switches scalable to serve several thousands of lines at the access rate of SONET OC-3 (155.52 Mb/s) or OC-12 (622.08 Mb/s) and above. Therefore, the switch has to be capable of handling a total bandwidth of at least 2 Tb/s. For possible implementation of such switch with N=16384 OC-3 lines, there will be 32 MSU's, each interconnects 64 BSU's with 8x8 I/O OC-3 lines respectively. The BSU bus rate is in the order of 4 Gb/s. This rate is easily met by the capability of present-day signal processing electronic devices.

The hardware cost of our proposed growable switch is expressed mainly by the cost of basic switching elements (i.e BSU's), cost of BSU's interface modules and the cost of modules interconnection links. Table 2 shows the projection of the switch size, and cost of various switch components for two implementation examples. As shown in the table, the total number of switching elements is k.m, while the interface modules cost CN in a switch of size NxN lines, is given by:

$$CN = m.k.[n + (m-1) + (k-1)]$$

And, the total number of interconnection links (LN) in a switch of size NxN lines , is given by:

$$LN = k.m.[(k-1) + (m-1)]$$

To illustrate the cost saving of our switch, the number of 8x8 switching elements in our proposed switch architecture and two other switch fabrics, that resembles the well known Clos networks (obtained from [2]), are compared as shown in Fig. 4. The

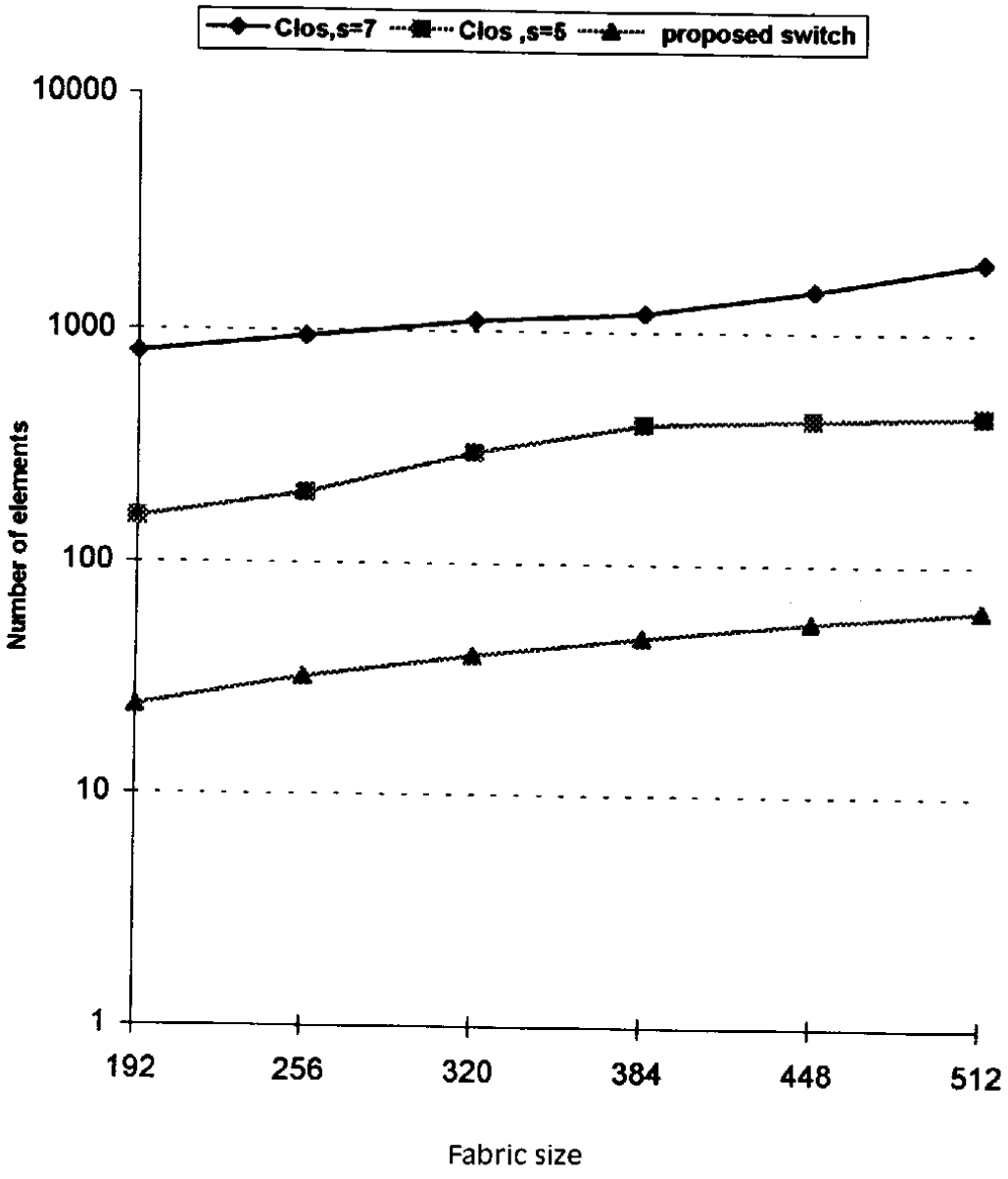

Fig. 4. Number of 8x8 switching elements in fabric of various sizes.

switching element in our architecture is a BSU with eight I/O lines. The first Clos switch has seven stages and sixteen elements in the first stage, while the second has five stages with eight elements in the first stage. It is easily to see, the number of elements in conventional Clos-based switches exceeds the elements cost of our proposed switch.

Table 2. Switch design examples

Switch parameter	Parameter variable	Example 1	Example 2
Input line rate (Mbps)	*a*	155.52	51.84
No. of LIM's per BSU	*n*	16	8
No. of BSU's per MSU	*k*	32	64
No. of MSU's per SSU	*m*	32	32
No. of BIM's per BSU	*k-1*	31.00	63
No. of MIM's per BSU	*m-1*	31.00	31
Total no. of interface modules per BSU	*n+(m-1)+(k-1)*	78.00	102
BSU bus rate (Mbps)	*3.a.n- n.a(m+k)/m.k*	7,309.44	1,224.72
Total no. of interface	*k.[n+(m-1)+(k-1)]*	2,496.00	6,528
Total no. of interface modules per SSU	*m.k.[n+(m-1)+(k-1)]*	79,872.00	208,896
No. of LIM's per SSU (switch size)	*m.k.n*	16,384.00	16,384
No. of BIM's per SSU	*m.k.(k-1)*	31,744.00	129,024
No. of MIM's per SSU	*m.k.(m-1)*	31,744.00	63,488
No. of links between BSU's of a single MSU	*k.(k-1)*	992.00	4,032
No. of links between MSU's per SSU	*k.m.(m-1)*	31,744.00	63,488
Total number of switch links	*km(k-1)+km(m-1)= km[(k-1)+(m-1)]*	63,488.00	192,512

Besides, the proposed switch has at most three stages which results in smaller cell delay, and does not require complex routing procedures.

Routing Considerations

The overall routing addressing format is attached as a routing header to the beginning of each 53 bytes cell by HTD of the module during the VCI translation process. To route cells to one of LIM's of the switch, a simple self-routing scheme

which lies between source routing [13], and routing by a unique identifier (UI), is implemented. Similar to UI routing, each output line is identified by a unique code, but on the other hand, like linear source-routing (LSR), all routing information in the cell header is generated by the source LIM. Thus, the switch routing scheme differs from UI routing technique in that no routing table lookup by other transit modules is needed beyond the source module. This results minimum delay at transit modules , eliminates the need to maintain routing tables by each transit module, and allows generalization for multi-cast routing.

There are three types of routing formats: unicast, broadcast and multi-cast cells header formats. For a LIM to LIM connection, a unicast cell is generated with three destination address fields, each is represented by 8-bit word. To route cells to one of the LIM's modules of the BSU, the first field bits is used. For routing to an external BSU of the same MSU (i.e. second level routing), the second field address bits are used. These bits identify the particular BIM used for cell transport between the BSU's. The third field consists of additional address word , needed for the third level routing, to select the destination MIM and consequently the destination MSU. The two most significant bits of each address field are used for format types identification. (01) for unicast cells, (10) for broadcast cell, (11) for multicast cells and (00) to indicate the end of address fields in a multicast cell. The remaining six bits of each field, denote up to 64 different labels for switch modules (LIM's, BIM's and MIM's).

The routing is performed by the "broadcast and select" approach on the BSU high speed bus. When an output module (LIM, BIM or MIM) receives the address header of the cell, it reads the cell header from the switch bus and processes it according to the following :

If the module is a LIM module, it compares its unique code (which identifies the LIM output line, BSU and MSU) with that of the header fields. If the module sees a match of all fields, the cell is then transported to the destined output line port. The module also strips off the header address in front of the cell.

If the module is a BIM, it compares the MSU and BSU fields to that of its label. In case of a match, the header address and the cell body are then transmitted without modification over the destined BSU.

Finally, if the module is a MIM, it compares the MSU field to that of its label. In case of a match, the header address and the cell body are then transmitted

without modification over the external destined MSU.
Accordingly, the routing algorithm enables every cell to find its way from source to destination in a maximum of three hops.

For instance a cell originating from LIM (4) of BSU(5) in MSU(1) (i.e BSU(1,5)) and destined to LIM(6) of BSU(3) in MSU (2) (i.e.BSU(2,3)) can be routed through the following path : BSU(1,5) --> BSU(2,5) --> BSU(2,3), via MIM (2) connected to BSU(1,5), BIM(3) connected to BSU(2,5) and to output port via LIM(6) of BSU(2,3).

The broadcast cell format is similar to unicast ones (i.e header consisting of three fields) except for the two format bits which is set to 10 for the address field of the broadcasting modules. For instance, if the LIM field, has format bits set to 10, then all LIM's of that BSU are considered cell destinations. If these format bits were set in the BSU address field, then all BIM's connected to source BSU are destined modules.

In contrast to unicast and broadcast cells, which utilizes one frame slot for transport of a single ATM cell and its header, the format of a multicast cell uses several slots to transmit the multicast cell. One slot carries the information cell and is preceded by the other slot(s) containing the routing header. The header consists of a multi-cast chain of address labels of the multicast lines. The chain consists of several destination labels(of three fields each) whose code follows one another. Each field bits are set to (11) while the last field of the header has the format bits set to (00) indicating end of routing header. Using three words per label address, each header slot may contain up to sixteen multicast destination output labels.

To apply source -routing technique for multicast cells , the source module computes the multicast chain from the switch topology database, and code the multicast address chain to form the cell header [13].

For all types of cells , in addition to routing fields that identifies the destination, the cell header contains a fourth field of two bits used to represent the cell's routing priority. The priority information is used to meet the cell loss rate requirement and switching delay requirement of different service classes.

Multicasting, Services Integration and Congestion Control Capabilities

A multi-cast function is needed for various distribution or conference services. According to the reservation protocol, the frame transmitted cells can be read by all BSU

modules, and there is no limit on the number of LIM's that can receive a given cell. Thus, the switch easily supports a flexible multipoint connection capability. In our switch, this multicasting feature is supported without the need for cell replication (creation of a cell to each LIM) and consequently without increasing the switch load. An input LIM just sends its cells to the fiber bus. Each output module, upon recognizing a broadcast or a multicast header, can get a copy of all multicast cells on the bus and so explicit duplication of copies is not needed. Accordingly, the switch is very suitable for ATM broadcast and voice/video conferencing applications.

The flexibility to allocate access channels rate dynamically on a per-call basis and to meet QOS for various types of services are desired features of ATM switches. Variable bit rate services in BISDN require terminals and application sources to transmit at arbitrary rate. These features can be easily realized in our switch by allowing switch input lines speed to be selectable, and by cell input buffering at the LIM followed by proper selection of X parameter for the call cells. Through setting the parameter X individually by each LIM module, a dynamic switch bandwidth allocation can be obtained and the switch architecture meets the need for heterogeneous and dynamically changing mix of traffic. Thus, the switch can support various types of input BISDN services, ranging from video phone to high definition television.

Congestion control is an important issue that need to be investigated to design a practical ATM packet switching system. Congestion usually results due to fluctuations of bursty source that exceed the switch capacity and causes buffer overflow and cell loss. Another blockage class in ATM switching is head of line blocking(HOL), where cells in buffers following a delayed cell are subject to starvation. HOL blocking occurs in pure input queuing space switches when the switch internal speed is the same as the input and output lines.[5] The design of our switch architecture and the system reservation protocol guarantees that at normal load, and by proper selection of the switch parameters, no cell congestion, or head of line blockage may occur. Congestion is avoided at the switch input buffers by having a non-blocking high speed bus architecture within each switch BSU unit. The switch architecture runs much faster than the input and output lines. Thus, it can transfer all input cells at a particular input time slot before the arrival of the next input slot. Within the switch, the switch reservation access scheme ensures that each BIM or MIM module transmits a reservation request for a variable number of slot, in every frame. Therefore, an overloaded BIM or MIM module can use the available un-used bus bandwidth share of the under-loaded modules. Furthermore, the switch routing addressing formats make it always possible, in case of internal switch congestion , to establish simultaneous, independent internal paths between any arbitrary pairs of input and output lines.

However, as network bus frame length varies from one frame to another, it is possible that, for a relatively low speed BSU bus, when many modules are overloaded, the bus frame approaches its maximum allowable length. The maximum frame length is a design parameter, chosen to satisfy the maximum tolerable delay for real time traffic or data communication. In such situation, no input modules will then do any more reservation beyond the maximum frame length and the module cells will be delayed for several frames, and eventually a series buffer overflow may happen. Congestion due to bus frame at its maximum length is easily controlled and may be completely prevented by appropriate design and selection of switch network bus parameters such as bus data rate and/or the number of BSU I/O lines. The amount of buffering required in each input/output port depends on the model of cell arrival and cell loss requirement. Modeling the queuing process of the input LIM's buffer as M/D/1, it can be shown that with a buffer size of 5 cells , the probability of losing a cell within a bus interface is less than 10^{-6} for an 84% load. [10]

HOL blockage is prevented simply because an output module can receive cells routed to it by several input modules as a result of the high speed of the switch fabric, and because of the broadcasting nature of the switch units. In contrast to space switches, where cell switching is done over parallel paths at the same time, our switch is basically a time switch, where cells are transmitted sequentially over the switch buses. Due to the reservation scheme and the unidirectional nature of the switch fiber optics bus, cells transmission by different inputs modules - and addressed to the same output - will not contend for the output. Rather, these cells will be broadcasted in the cascaded communication slots of the bus frame and reach the destination in timely order. Thus, no cell will be blocked due to contention for the network bus or external conflict at the output module.

Maintainability and Fault Tolerance

ATM switching necessitates a fault tolerant design. The System must be structured such that it provides a reliable and continuous service without interruption , even with some defective elements. An isolated failure in one of the LIM's will affect one port only. The impact of a BIM and a MIM failure is limited to virtual connections broadcasted on the source BSU and switched across the BIM/MIM destined BSU and MSU respectively. With a fiber network bus failure, service will be interrupted for all I/O lines and modules connected to the bus. However, the possibility of a passive component such as the fiber bus failing, is remote. In our switch, since all interface modules are identical, reliability requirement can be met by providing additional spare interface modules connected to each BSU network. These modules can take over if any

LIM, BIM or MIM failed. To further increase system reliability, redundancy is provided. A duplicate of each BIM and MIM modules can be built to serve as a standby in the event of failure of these modules. The duplication principle may also be combined with routing broadcasting function to provide an alternate switching path in case of failure of BIM or a MIM. To achieve fault tolerance against BCM failure, every LIM module can be a potential BCM whenever it detects no timing pulses from an upstream module for a time out period or by introducing a redundancy for the BCM.

Switch Performance Model and Analysis

To model our architecture of a (nkmxnkm) switch, a multidimensional complex model is needed that does not lend itself easily to exact analysis. However, an approximate model can be developed based on the well known queuing techniques for ATM switching systems where the switch system is modeled as queues in tandem [14-16].

Modeling and evaluation of the proposed switch performance using such techniques and comparison with performance measures obtained by simulation is a subject of further study. Nevertheless, we will present here a selected performance measures of the switch, obtained by the simple model of GI/D/1 queuing system. The assumption for this model is that no cell queuing occurs at the switch input due to the high speed of BSU fiber bus, compared to the input switch line speed. The service time of this model corresponds to the constant output service times of the output LIM buffer. Note that the cells arrival process to switch output buffer cannot be assumed Poisson, since the cell switching time has an arbitrary distribution. For a uniform traffic destinations among switch lines, the rate at which cells reach the output buffer is virtually equal to the rate of incoming cells λ on the switch input line (i.e input offered load). Using approximate results derived by Kobayashi in references [17,18] for GI/GI/1 queue, we have for our GI/D/1:

W_o=Waiting time at output LIM buffer

$= \sigma/[\mu.(1- \sigma)]$

where $\sigma = \exp\{ -2(1-\rho)/(\rho (co^2 + cs^2/\rho))\}$

μ = deterministic output line service rate (slot/unit time)

ρ = output traffic intensity = λ / μ

cs^2 = squared coefficient of variation for output buffer service time. = 0 for deterministic service time.

co^2 = squared coefficient of variation for cell arrival process to the output LIM buffer.
= squared coefficient of variation of switch cells inter-departure time

This parameter is a function of the switch size, cell arrival distribution to switch input line and cells output destinations distribution. For a large size switch, with uniform destination distribution, almost all arriving cells will need three hops to traverse the switch units to the output buffer. Thus, it suffices to assume a deterministic cell switching time. In this case, co^2 may be approximated by using [19, eq.(38)].

$$co^2 = (1 - \rho i^2)\, ca^2$$

where ca^2 squared coefficient of variation for cell inter-arrival time at switch input LIM's. For Poisson arrival $ca^2=1$
and ρi = input line traffic intensity = $\lambda \cdot \tau$

τ is the average time an arriving cell takes to traverse all switch units.
Alternatively, if we assume the switch service time has an Erlang distribution of m stages, then using [20, eq. (2) & (11)] we have :

$$co^2 = (\tau^2/m + 1/\lambda^2)/(\tau + 1/\lambda)^2$$

Using the above we easily derive : -

S_O = total system time delay = $W_O + 1/\mu$

N_O = Average queue size of output LIM buffer (cell) = $\lambda \cdot S_O$

Maximum Switch throughput (G_{max}) may also be derived as follow [14]:

$$G_{max} \cdot \mu = 1 \text{ or}$$

$$G_{max} = 1/\mu$$

The above equations are used to plot the switch system delay , and mean output queue sizes as a function of switch input load and mean switching time (assuming both deterministic and Erlang distributions) as shown in Figs. 5 and 6. As shown in Fig. 5, for light and moderate switch loads, there is a good agreement between our approximate model results and simulation, as the difference in switch system delay is less than one time slot.

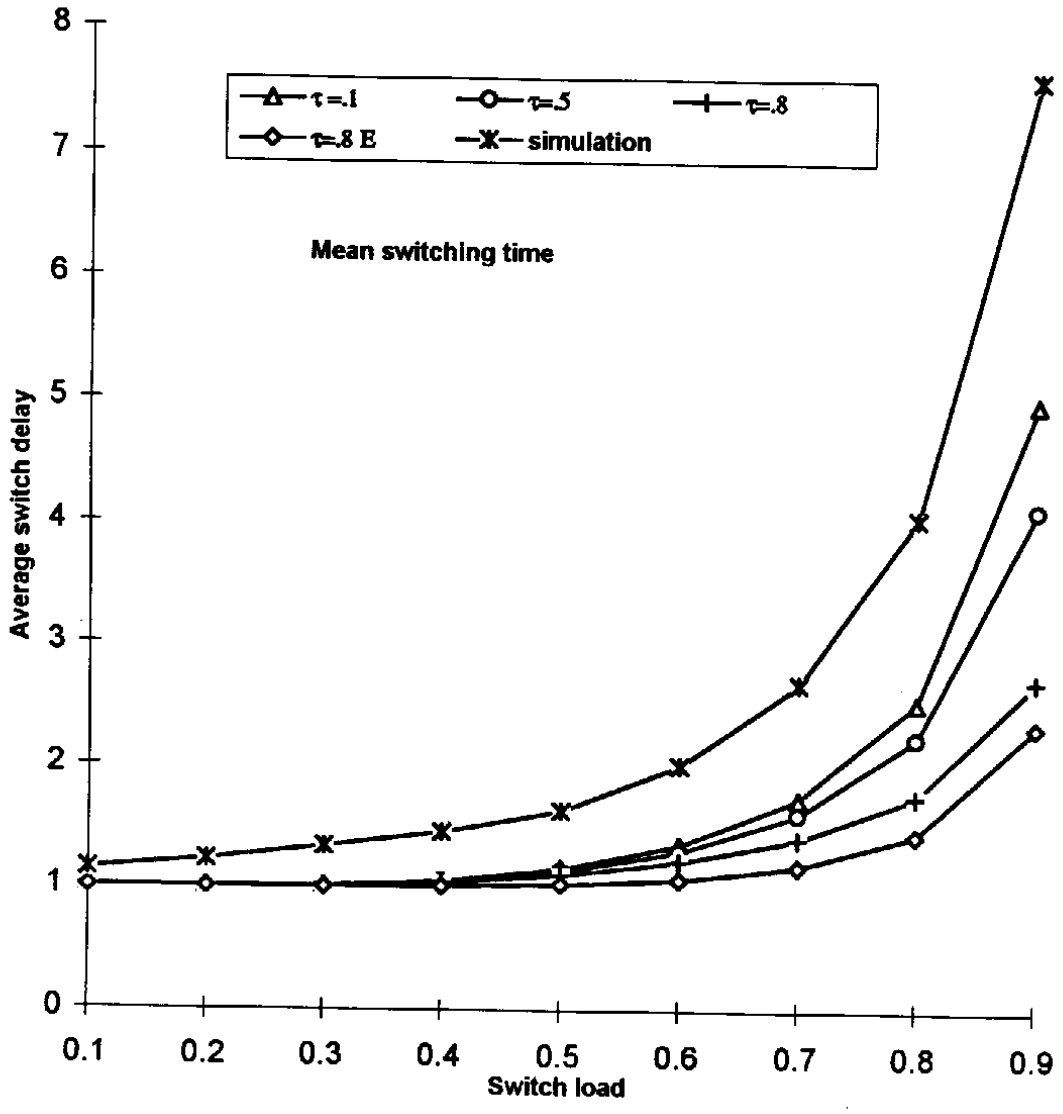

Fig. 5. Average switch delay as a function of switch load.

Conclusion

In this paper we proposed an architecture design for a large scale ATM switch , that exploits a high speed optical network as a shared medium for fast switching of ATM cells. The network implements a reservation based access protocol. The switch I/O lines are partitioned by a novel topology into multiple groups over a three level structure. Several major functions of the ATM switch were investigated. These include : cells routing, congestion control, network MAC protocol, and switch multicasting capabilities. Specific values for the switch cost in terms of number of switch units,

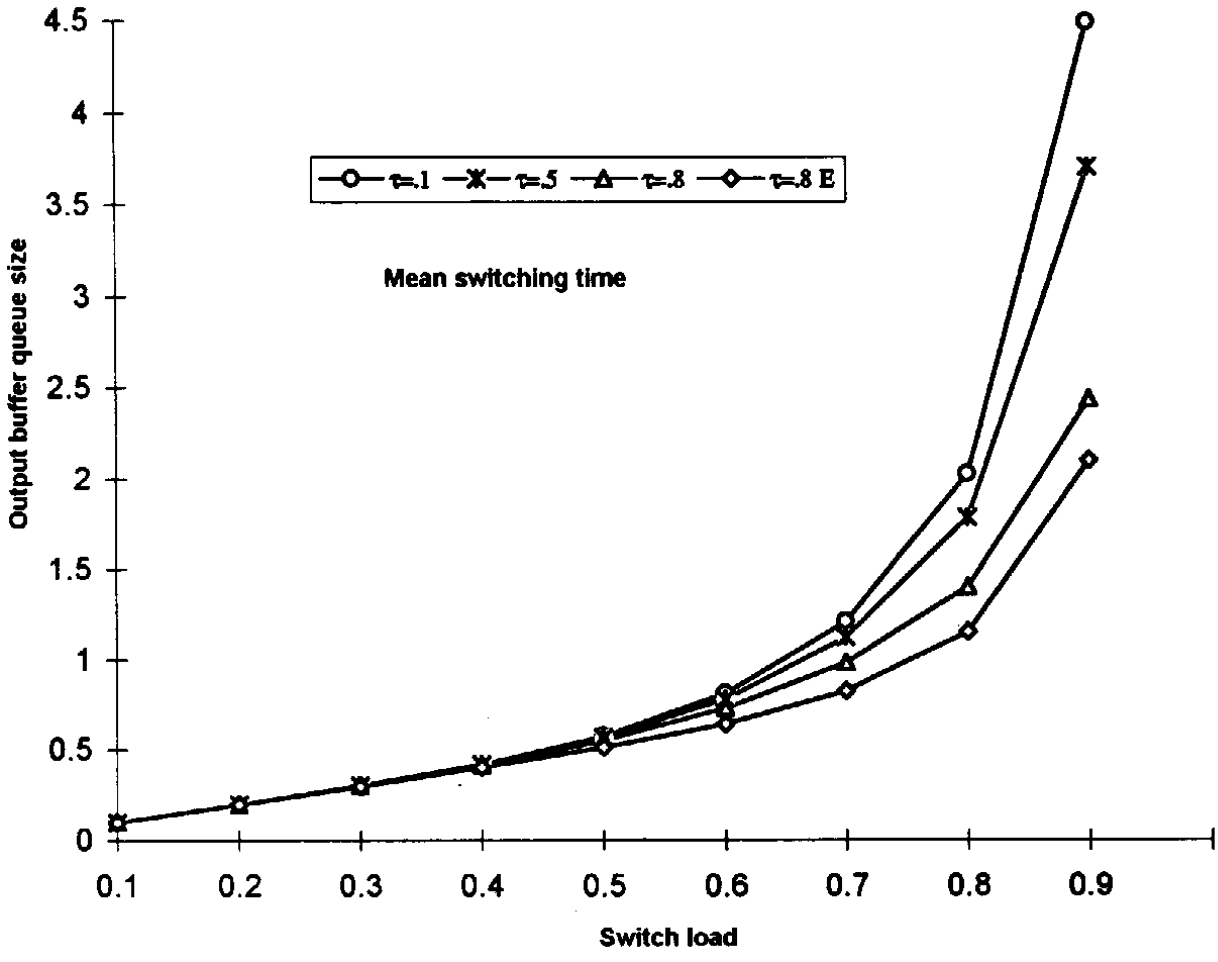

Fig. 6. Average output LIM buffer queue size as a funciton of switch load.

modules and interconnection links needed to build a large ATM switch are presented. These numbers are based on realistic projections of the technologies involved. The switch architecture and topology allow the switch to expand to a very large size without exceeding the limits of electronic signaling processing speed on the switch buses. An approximate performance model and simulation were used to derive various switch performance measures including buffer queue size, mean delay and cells loss probability.

We conclude that by using a novel system topology and a high speed networks with reservation based protocol network, a new ATM switch can be built that has the advantages of:

- A modular and non-blocking structure .
- Possible achievement of high throughput and capacity required for largescale ATM network switching (e.g 2.5 Tb/s and 16384 I/O lines).
- System configuration that supports various communication services and meets the need for heterogeneous and dynamically changing mix of traffic.

Future work is the presentation of VLSI technology for switch implementation , analytic modeling and extensive evaluation study of the proposed switch performance with various switch design options, input traffic process characteristic and traffic destination distributions.

References

[1] Minzer, S. "Broadband ISDN and Asynchronous Transfer Mode(ATM)." *IEEE Comm. Magazine* (September. 1989), 17-57.

[2] Andrezej Jajszczyk and Wojciech Kabacinski. "A Growable ATM Switching Architecture." *IEEE Trans. on Comm.*, 43, No.4 (April 1995), 1155-1162.

[3] Rooholamini, Reza and Cherkassky, Valdimir. " Finding the Right ATM Switch for the Market." *IEEE Computer Magazine* , 27, No.4 (April 1994).

[4] Yuji Oie; Tatsuya Suda; Masayuki Muratata, and Hideo Miyahara, "Survey of the Performance of Nonblocking Switches with FIFO Input Buffers." *IEEE ICC'90, Conference Proc.* (1990), 737-741.

[5] Achille Pattavina. "Nonblocking Architectures for ATM Switching." *IEEE Communication Magazine*, (Feb. 1993), 38-48.

[6] Ahmadi, Hamid and Denzel, Wolfgang. " A Survey of Modern High-performance Switching Techniques." *IEEE J. on Selected Areas in Comm.*, 7, No. 7 (Sept. 1989), 1091-1103.

[7] Tobagi, Fouad A." Fast Packet Switching Architecture for Broadband Integrated Services Digital Networks". *Proceedings of the IEEE*, 78, No. 1 (Jan. 1990), 133-166.

[8] Suzuki, Hiroshi *et al.* " Very High-Speed and High-Capacity Packet Switching for Broadband ISDN." *IEEE J. on Selected Area in Comm.*, 6, No.9 (Dec. 1988), 1556-1564.

[9] Philippe, A.P. and Paul, R.P. "High-Dimensionally Shared-Medium Photonic Switch." *IEEE Trans. on Comm.*, 41, No.1 (Jan. 1993), 224-236.

[10] Yu-Shuan Yeh; Michael G. Hluchyj, and Anthony S. Acampora. "The Knockout Switch: A Simple, Modular Architecture for High Performance Packet Switching." *IEEE J. on Selected Area in Comm.*, SAC-5, No.8 (October 1987), 1274-1283.

[11] Jonathan Chao, H. " A Recursive Modular Terabit/Second ATM Switch." *IEEE J. on*

Selected Area in Comm., 9, No.8 (October 1991), 1161-1172.

[12] Al-Wakeel, S.S. and Ilyas, M. " R-NET A High Speed Fiber Optics Network with Reservation Access Protocol ." *Inter. J. of Digital and Analog Comm. Systems*, 5, (1992),1-13.

[13] Yum, Tak-Shing Peter, and Chen, Mon-Song. "Multicast Source Routing in Packet-Switched Networks." *IEEE Trans. on Comm.*, 42, No.2/3/4 (Feb/March/April, 1994), 1212-1215.

[14] Chen, Jeane S.C. and Stern, Thomas E. " Throughput Analysis , Optimal Buffer Allocation , and Traffic Imbalance Study of a Generic Nonblocking Packet Switch." *IEEE Trans. on Comm.*, 9, No. 3 (April 1991), 439-449.

[15] Delre, Enrico and Fanacci, Romano. "Performance Evaluation of Input and Output Queuing techniques in ATM Switching Systems. " *IEEE Trans. on Comm.*, 41, No. 10 (October 1993), 1565-1574.

[16] Pattavina, Achille and Bruzzi, Giacomo. " Analysis of Input and Output Queuing for Nonblocking ATM Switches. " *IEEE/ACM Trans. on Networking* , 1, No. 3 (June 1993), 314-328.

[17] Myskja, Arne. " On Approximation for GI/GI/1 Queue" , *Computer Networks and Isdn Systems*, 20 (December 1990), 285-295.

[18] Kobayashi, H. "Application of Diffusion Model Approximation to Queuing Networks, I: Equilibrium Queue Distributions." *Journal ACM*, 21, No. 2 (1974).

[19] W. Whitt, "The Queuing Network Analyzer ". *Bell System Technical Journal*, 62 , No. 9 (November 1983), 2779-2815.

[20] C. Pierre-Jacques and G. Scheys. "Minimization of the Total Loss Rate for Two Finite Queues in Series." *IEEE Trans. on Comm*, 39, No. 11(November 1991), 1651-1661.

التصميم المعاري لمقسم ATM تعتمد على شبكة فائقة السرعة

د. سامي صالح الوكيل

قسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود، الرياض

ملخص البحث . يعد تطوير مقاسم ATM ظرفية ذات سرعة فائقة وسعة عالية من الضروريات لتحقيق شبكة الخدمة المتكاملة الرقمية ذات النطاق الواسع BISDN. ولإنجاز ذلك يلزم البحث في العديد من الجوانب الفنية الخاصة بالتطوير للمقسم نحو تصميم مداولات المقسم وتطوير عمارته وتحليل أدائه.

وفي هذا البحث نقترح تصميماً لعمارة مقسم ATM ذي معبر مشترك. ويتكون المعبر من شبكة فائقة السرعة من الألياف البصرية تعتمد على نظام الحجز في مداولة الاتصال بها.

ويقوم تصميم المقسم على هيكلية ارتباط مبتكرة بين وحدات المقسم بهدف تكوين مقسم فائق السعة دون حصول زيادة بالغة في معدل التراسل للبيانات الإجمالي على شبكة المقسم. وبناءً على ذلك جرى التوصل إلى مقسم بسعة ١٦٣٨٤ خط إدخال وإخراج من نوع STS-32 (ذو سرعة تبلغ ١٥٥ر٥٢ ميجا بت/ ثانية) باستخدام شبكة ألياف بصرية بمعدل تراسل لايتجاوز ٤ جيجابت/ ثانية. بالإضافة إلى ذلك يهيىء التصميم المعماري للمقسم إمكانية النمو التدريجي في سعة المقسم، كما يحقق متطلبات التراسل لأنواع مختلفة متغيرة من الحركة، كما يتيح خدمة التراسل ذات الاتصال المتعددة.