ATM Technology for Corporate Networks*

Peter Newman[†]

N.E.T. Adaptive Division
800 Saginaw Drive
Redwood City, CA 94063

The bandwidth requirements of data traffic within commercial organizations have been increasing steadily for some time both in the local area and within private wide area networks. Workstations are beginning to introduce multimedia applications to the desktop that include components of voice, video, image, and data. Such developments will require networks of much greater bandwidth than at present with the capability of handling multiservice traffic.

The Asynchronous Transfer Mode (ATM) is being developed as a high- speed networking technique for the public network capable of supporting many classes of traffic. ATM is also well suited for applications within corporate networking: as a private wide area network; as the campus backbone network; and as a high-speed local area network. This article discusses ATM technology in relation to the requirements of corporate networking. An introduction to ATM switch architecture is presented. The various approaches to ATM switch design that have appeared in the literature are reviewed. Finally, a discussion is presented of some of the current issues facing the development of ATM networks. It is argued that much simpler solutions to many of these issues may be adopted in the context of the corporate network than is permissible for the public broadband network.

Corporate Networking

We are all aware of the continual expansion in the capabilities of the desktop workstation. The steady growth in the processing speed of reduced instruction set processors seems likely to continue [26]. The client/server model of networked operation continues to gain favor in the commercial environment. Even the humble personal computer no longer stands alone but is increasingly being networked. Already these developments are causing the network bandwidth requirements at the desktop to exceed the shared 10 Mbits/s of most current local area networks. Furthermore, video and multimedia applications are beginning to move from a state of speculation and research, to active commercial consideration. Several workstation manufactures are currently integrating video capabilities into their desktop workstations. Such products are already within striking distance of an acceptable cost. These developments will demand networks with a capacity of several orders of magnitude beyond the current shared 10 Mbits/s. In addition, multimedia applications require the ability to handle multi-service traffic in a single integrated network. Commercial demand for local area bandwidth measured in gigabits/s is fast becoming a reality.

In the last few years the physical topology of local area networks has migrated from the ring and the multidrop bus towards a star configuration (the hub) even though the technology remains shared medium. Even at 10 Mbits/s this topology is easier to manage and offers higher reliability. As the

bandwidth requirement of the local area approaches the gigabit/s range, switched star topologies are the most likely to be favored in the commercial environment. The majority of desktop applications are unlikely to require individual access much in excess of one or two hundred Mbits/s. Shared medium access in the gigabit/s range is likely to remain significantly more expensive than access at around 100 Mbits/s for some time. Thus for commercial applications the switched star approach with 150 Mbits/s to the desktop and an aggregate capacity in the gigabit/s range will undoubtedly prove more cost effective than a gigabit/s shared medium design. This implies an approach based upon ATM technology.

ATM is a high-speed packet switching technique using short fixed length packets called cells. Fixed length cells simplify the design of an ATM switch at the high switching speeds involved. The selection of a short fixed length cell reduces the delay and most significantly the jitter (variance of delay) for delay sensitive services such as voice and video. ATM thus presents a single integrated switching mechanism capable of supporting a wide range of traffic types such as voice, video, image and various classes of data traffic. ATM has been selected as the multiplexing and switching technique for use in the public Broadband Integrated Services Digital Network (B-ISDN) and is receiving much standardization activity. A local area network based upon ATM will therefore offer: high capacity networking; the ability to handle many classes of traffic; and standardized access to broadband public network services when B-ISDN becomes available. ATM is a switching technique developed for the wide area so a local area ATM network will also offer seamless access to private wide area networking as well as to the public broadband network. Thus high capacity multi-service local networks are more likely to be based upon ATM than upon other proprietary high-speed packet switching schemes.

In the literature ATM technology is almost always discussed in the context of its application to broadband ISDN. The application of ATM technology to corporate networking (i.e. high capacity lo-

cal area networks, campus area networks, backbone networks, private broadband networks) results in a rather different set of requirements than those of B-ISDN. The most obvious difference in the design of an ATM switch for commercial applications is that a smaller aggregate capacity is required than that envisaged for the public B-ISDN. The initial commercial market is likely to center on the high capacity interconnection of existing data networking applications. So initially an aggregate switch capacity in the region of 1 or 2 Gbits/s will probably be sufficient. As high-speed workstation access and multimedia applications gain acceptance in the workplace, the switch design must be capable of significant growth. With growth to switch capacities of 10 Gbits/s and beyond, reliability is likely to become an issue so redundancy will be required of the switch design. Cost per access port will be a sensitive parameter for direct ATM access by workstations so the ability to offer various degrees of concentration on access to the switch will be important. High-speed access ports at 600 Mbits/s may also be required to support applications such as high performance servers.

ATM Switch Architecture

ATM is connection oriented. All cells belong to a pre-established virtual connection. All traffic is segmented into cells for transmission across an ATM network. The sequence integrity of all cells on a virtual connection is preserved across each ATM switch to simplify reconstruction of the original traffic at the destination. The ATM standard for broadband ISDN defines a cell of length 53 bytes with a header of 5 bytes and a payload of 48 bytes. The header of each cell contains a virtual channel identifier (VCI) to identify the virtual connection to which the cell belongs. To avoid problems of unique global allocation these identifiers have only local significance. In general, the VCI is local to each switch port. As each cell traverses a switch, the VCI is translated to the value assigned for the next link in the virtual connection.

An ATM switch will handle a minimum of several hundreds of thousands of cells per second at every
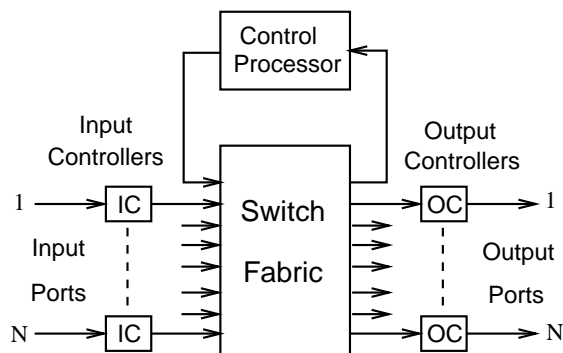
Figure 1: General structure of an ATM switch.

switch port. Each switch port will typically support a throughput of at least 50 Mbits/s, while 150 Mbits/s and 600 Mbits/s are proposed as standard port speeds and transmission rates for broadband ISDN. Proposed switch sizes range from a few ports to some thousands of ports with anything beyond a hundred ports or so currently considered to be a large switch.

The general structure of an ATM switch is illustrated in fig. 1. All of the per cell processing functions are performed in hardware by the input controllers (ICs), the switch fabric, and the output controllers (OCs). The control processor is required only for higher level functions such as connection establishment and release, bandwidth allocation, maintenance, and management. All input controllers are generally synchronized so that all cells arrive at the switch fabric with their headers aligned. This simplifies the design of the switch fabric and permits cells to be accepted according to their priority. The switch fabric operates synchronously and typically, during each timeslot, one cell may be transmitted across the switch fabric from each input controller. The control processor may communicate with the input and output controllers either by a direct communication path or via cells across the switch fabric. External interfaces to the switch are generally bi-directional and are formed by grouping an input and an output controller together to form a port interface.

All cells have their VCI translated in the input controllers before being submitted to the switch fabric. This operation is performed by table lookup on the incoming VCI in a connection table. The connection table may also contain a routing field to specify the output port of the switch through which the virtual connection is routed. Other information may be included in the table on a per connection basis such as the priority, class of service, traffic type of the connection, and so on. Some switch designs will add this information to the front of each cell to assist the switch fabric in making its routing and buffering decisions. This additional information is removed by the output controllers. Some designs of switch support multicast operation by replicating multiple copies of incoming multicast cells within the switch fabric and routing each of the copies to its required output port. In these switch designs VCI translation will also be required in the output controllers to enable each of the copies of a multicast cell to exit the switch with any required value of VCI.

In an ATM switch, cell arrivals are not scheduled. At the switch fabric, a number of cells from different input ports may simultaneously request the same output port. This event is referred to as output contention (or conflict) and is present in every design of ATM switch. A single output port can only transmit one cell at a time. Thus one cell must be accepted for transmission and any others simultaneously requesting that port must either be buffered or discarded. The location of the buffers has a major effect on the overall performance of
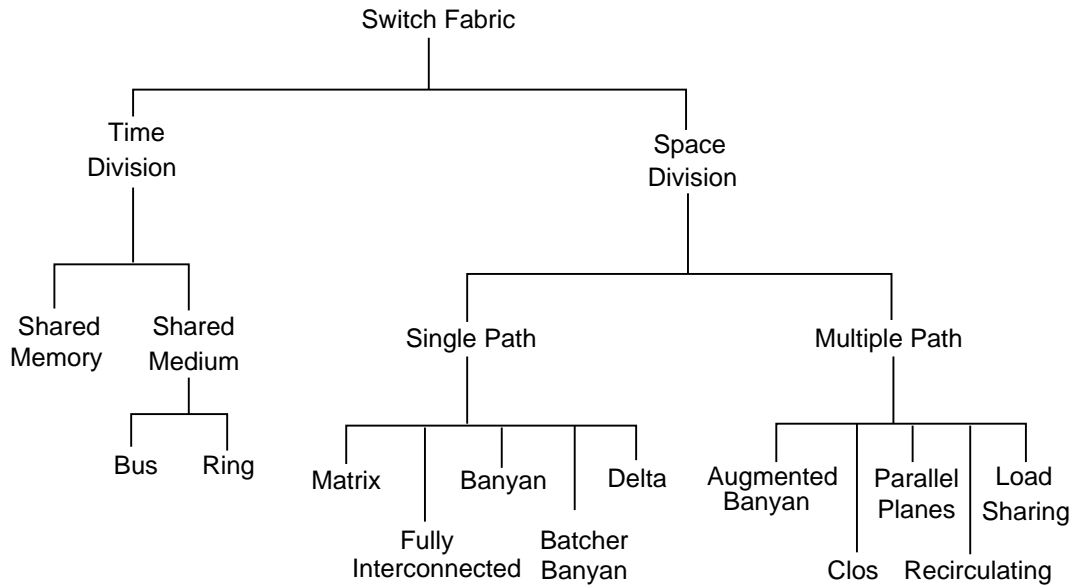
Figure 2: Classification of switch fabrics.

the switch and also affects the complexity of the switch fabric. If buffers are not located at every point where contention occurs, a contention resolution technique is required to detect contention and direct cells that cannot immediately be handled into the buffers. The topology of the switch fabric, the location of the cell buffers, and the contention resolution mechanism are the most significant aspects of an ATM switch design. Most ATM switch designs may be characterized by considering the approach taken in each of these three dimensions. The solutions proposed in the literature for each of these three design dimensions will be discussed and a simple classification presented for each one.

The Switch Fabric

The structure of the switch fabric is the most significant component of the hardware design of an ATM switch. It affects the cost, performance, capacity, growth capability, and complexity of the switch design. Many structures of switch fabric have been suggested in the literature. Fig. 2 offers a simple classification of switch fabric design that includes most of the proposed approaches.

Most switch fabrics are constructed from a num-ber of fundamental switching components called switching elements. The switch fabric of a large switch will generally be constructed from a number of interconnected switch modules. In many designs the switch module is itself a small switch constructed from a number of switching elements. A switching element is frequently implemented within an integrated circuit and a switch module on one or more circuit cards.

Time Division

In a switch fabric based upon time division all cells flow across a single communication highway shared in common by all input and output ports. This communication highway may be either a shared medium such as a ring or a bus, fig. 3(a), or a shared memory, fig. 3(b). The throughput of this single shared highway defines the capacity of the entire switch fabric and thus fixes an upper limit on the capacity for a particular implementation beyond which it cannot grow. Since every cell flows across a single shared communication highway, this class of switch fabric may easily support multicast operation. Many switching element designs have been proposed that use time division internally. A
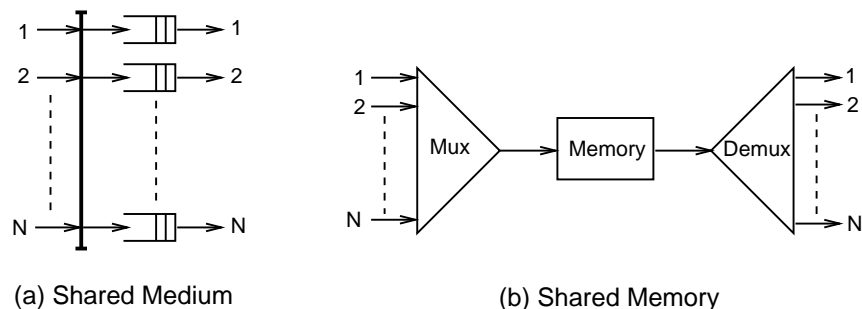
(a) Shared Medium   (b) Shared Memory

Figure 3: Time division ATM switch fabrics.

fixed round-robin scheme is generally used to distribute the bandwidth of the shared medium among the ports of the switching element.

For the construction of small switches with a capacity of up to a few gigabits/s, time division can be a very flexible technique. Access to the bandwidth of the shared medium may be arbitrated dynamically among the switch ports on a demand basis. This supports the efficient multiplexing of interfaces with widely differing access rates. Traffic concentration may also be offered to reduce the cost per port.

Shared memory designs have been reported operating with a total capacity of up to about 5 Gbits/s. PRELUDE from the French CNET [17] was one of the first ATM research projects. A number of manufacturers have developed a shared memory switching element as a single chip or a small chip set: Hitachi [44]; Toshiba [69]; Alcatel [4]; and the European RACE collaborative research program [23]. These devices are intended to serve in switch modules within a larger switch design but they could certainly be applied to a switch designed for the corporate networks market.

Shared medium designs with a total capacity of up to about 10 Gbits/s based upon a bus or ring have also been implemented. Several manufacturers have developed a shared medium switching element as a single chip or a small chip set: [14, 5] from Alcatel; the ATOM switch [37] from NEC; and the Atmospheric ring switch [27] from the RACE research program. The ATOM Switch is a large switch design using shared medium switch modules. Both the ATOM Switch and the Alcatel designs could be applied to a switch designed for the corporate networks market.

Space Division

Whereas in time division a single communication highway is shared by all input and output ports, in space division, a plurality of paths is provided between the input and output ports. These paths operate concurrently so that many cells may be transmitted across the switch fabric at the same time. The total capacity of the switch fabric is thus the product of the bandwidth of each path and the number of paths that on average can transmit a cell concurrently. The upper limit on the total capacity of the switch fabric is therefore theoretically unlimited. In practice, however, it is restricted by physical implementation constraints such as device pinout, connector restrictions, and synchronization considerations, which together limit the size of the switch fabric.

With a plurality of paths in the switch fabric a routing function is now required to select a path to the appropriate output port for each cell. Either a self-routing or a label routing technique may be used. In the self-routing approach the switch fabric is constructed from a self-routing interconnection network. Each input controller on the switch prefixes a routing tag to every incoming cell using the same table look-up mechanism it uses for VCI translation. This routing tag specifies the number

of the output port to which the cell must be delivered. The properties of a self-routing interconnection network permit each switching element in the switch fabric to make a very fast routing decision simply by inspecting the routing tag. Each cell will arrive at the required destination regardless of the switch port at which it enters. The majority of switch designs based upon a space division switch fabric employ a self-routing scheme.

In label routing, the VCI (label) is used to index routing tables within the switching elements of the switch fabric. This routing scheme does not rely on the regular properties of the interconnection network, so any arbitrary interconnection network of switching elements may be employed. The label routing approach permits a simple implementation of multicast operation but requires a large number of routing and translation tables to be maintained within the switch fabric. It is possible to combine self-routing for point-to-point traffic with label routing for multicast traffic in some switch designs.

With the introduction of multiple cells flowing through the switch fabric concurrently we introduce the possibility of conflict between cells requesting the same path or the same output port. This conflict is resolved by a contention resolution scheme or by introducing cell buffers at every point of conflict. If contention for a link occurs inside the switch fabric it is referred to as blocking. Contention for the output ports of the switch fabric may occur even though the switch fabric itself is non-blocking.

Interconnection networks for a space division switch fabric may be divided into two basic classes: single path networks; and multiple path networks. A single path network has a unique path through the interconnection network between any given input and output pair. A multiple path network has a number of different paths available between any input and output.

Single Path Networks

The single path self-routing interconnection networks most often proposed for use in ATM switch

designs are illustrated in fig. 4.

Crossbar: The term 'crossbar' derives from a particular design of single path non-blocking switch fabric developed for analog telephony. It used the topology of fig. 4(c) in which each active element, or crosspoint, was a single electrical contact. These days the term is often used to describe any single path non-blocking network that has a complexity that grows as a function of N^2 (where N is the number of input and output ports) [9]. In this sense the topologies of fig. 4(a), (b) and (c) are all crossbar designs, differing only in the use of a bus or disjoint paths on the input and output ports. The fully interconnected networks of fig. 4(a) and (b) are also known as networks with N^2 disjoint paths [72].

Crossbar designs have a complexity in paths or crosspoints that grows as a function of N^2. Thus they do not scale well to large sizes. They are, however, very useful for the construction of non-blocking self-routing switching elements and switches of modest size. A switching element with the fully interconnected structure of fig. 4(a) is described in [1]. By interleaving the distribution and concentration stages of fig. 4(a) the Christmas Tree Switch uses less than N^2 paths [77]. The Knockout Switch [79] and [54, 12] are examples of the structure of fig. 4(b). A matrix structure with cell buffers in each of the crosspoints is used in the switch designs from Fujitsu [57, 40] and in [41] from the RACE program. The Fujitsu switch has been designed as a large public switch but this structure could certainly be applied to a smaller switch design.

Banyan: The banyan network, as originally defined [30], covered a large class of interconnection networks that had only a single path between any input and any output. These days the term 'banyan' is applied to a family of self-routing networks constructed from 2×2 switching elements with a single path between any input output pair, fig. 4(f). A number of specific topologies belong to the banyan family but all offer an equivalent performance for random traffic. The banyan network has a complexity of paths and switching elements of order $N \log N$. This makes it much more suitable
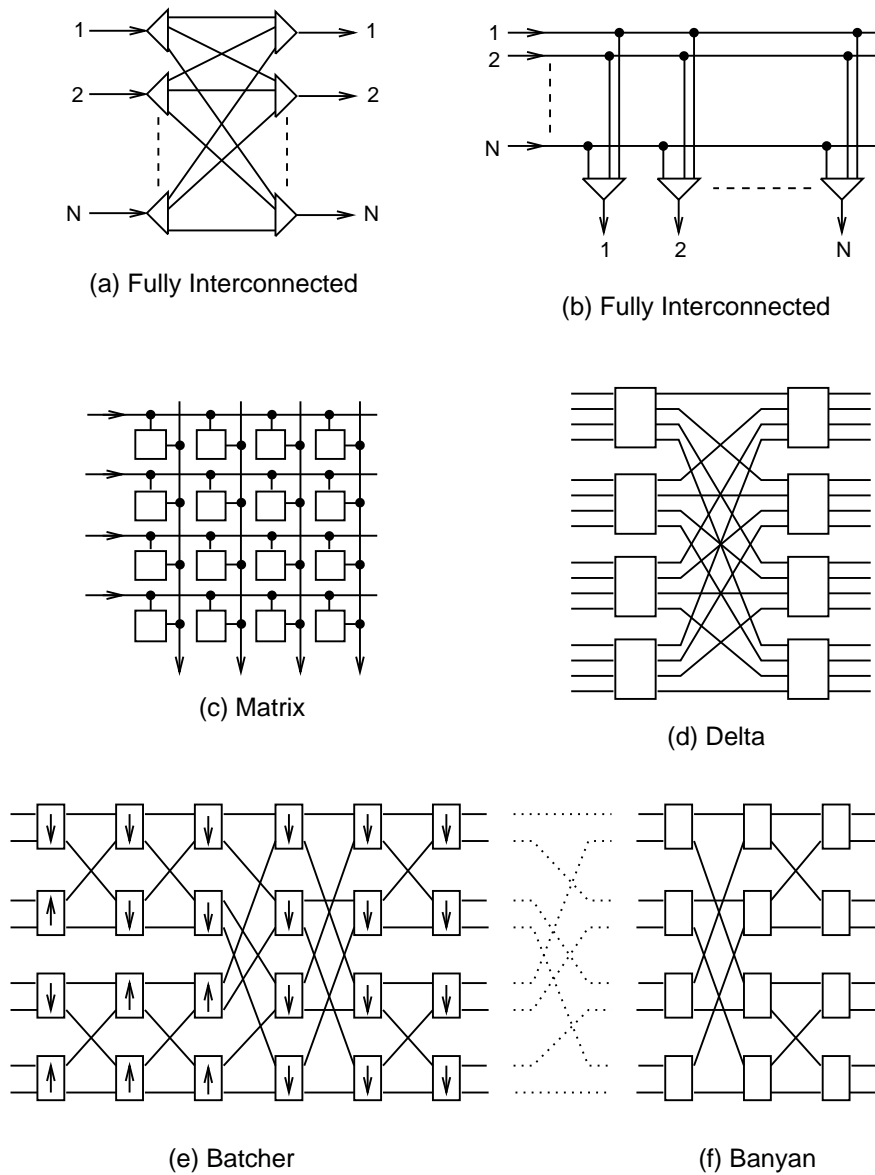
(a) Fully Interconnected

(b) Fully Interconnected

(c) Matrix

(d) Delta

(e) Batcher

(f) Banyan

Figure 4: Single path interconnection networks.

than the crossbar structures (of order N^2) for the construction of large switch fabrics.

Unfortunately, the banyan is a blocking network and its performance degrades rapidly as the size of the network increases. Also, the degree of blocking is related to the specific combination of destination requests present in the incident traffic. Thus some instantaneous patterns of incident traffic will give rise to a much poorer performance than oth-

ers. Switches that use a banyan interconnection network include Turner's switch [74] and [62, 70]. The performance may be improved if switching elements larger than 2×2 are employed. This leads to the class of delta networks.

Delta: Delta networks, fig. 4(d), are self-routing multistage interconnection networks with a single path between any input and output [60, 20]. While the performance of a delta network can be signifi-

cantly better than that of the banyan network [55] it is still a blocking network, its performance degrades as it increases in size, and it is also sensitive to the incident traffic pattern. This degradation in the performance of delta and banyan networks may be reduced by increasing the speed of the internal links within the network with respect to that of the input and output ports [58] or by introducing cell buffers into the switching elements [10, 19]. The sensitivity to the traffic pattern may be removed by randomizing the ports at which the traffic enters the network. Switches that use a delta interconnection network are reported in [43, 55, 25, 71].

Batcher-banyan: A banyan network will offer non-blocking performance if the incoming cells are sorted into order based upon their output port requests; all active cells are grouped together on a set of adjacent input ports; and no output port is requested by more than a single cell at any one time. The Batcher network, fig. 4(e), can sort an arbitrary set of cells into order based upon their routing tags and group the active cells together [6]. Thus the combination of a Batcher and a banyan network will offer non-blocking performance if some means is provided to prevent multiple cells requesting the same destination at the same time. (While a number of specific topologies belong to the general family of banyan networks, to achieve non-blocking operation, the specific topology shown in fig. 4(f) is required, connected to the Batcher network with the perfect shuffle connection pattern as indicated.)

While the individual 2×2 sorting elements of the Batcher network are very simple to implement, a large Batcher-banyan network is not easy to partition into integrated circuits and maintaining synchronization across the whole structure becomes increasingly difficult with size. Also, the growth of a Batcher network is of order $N(\log N)^2$, so many more switching stages are required in the Batcher network than in the banyan network. The Batcher-banyan network is useful for the construction of non-blocking switch fabrics of much greater size than may be achieved with a crossbar design.

The Batcher-banyan interconnection network was first proposed in the Starlite switch design from ATT [34]. The construction of a 256×256 Batcher-banyan network (requiring 36 switching stages in the Batcher network and 8 stages in the banyan network) is reported by Bellcore in [32]. It is being implemented in a set of five chips and forms the switch fabric of the Sunshine switch [29]. A Batcher-banyan switch fabric is also proposed in: [35, 50, 48] from Bellcore; the NEMAWASHI switch [3] from OKI; and also [61]. Most of the interest in the Batcher-banyan switch fabric has come from the research community rather than from equipment manufacturers. It seems unnecessarily complex for the scale of switch required in corporate network applications.

Multiple Path Networks

Multiple path networks are used to improve the performance of a single path network or to construct large switches from switch modules. Since multiple paths are available between every input output pair an algorithm is required to select one of the paths.

Augmented Banyan: Multiple paths may be introduced into a banyan or delta network by adding extra stages of switching elements, fig. 5(a). The number of paths between each input and output port is doubled for each extra stage added to the banyan network. If $(\log_2 N - 1)$ stages are added to a banyan network the Beneš network results, fig. 5(b). The extra stages can be used to distribute the traffic evenly across the banyan (or delta) network, to remove the sensitivity of the network to the incident traffic pattern, and to improve the performance of the network. Switch designs that propose this approach include: Turner's switch [74] and also [2, 55] and the MARS switch reviewed in [51]. Turner's switch has been implemented and could be applied to the corporate networking environment. Extra switching elements and interconnection links may also be added between the stages of a banyan or delta network to provide redundant paths for fault tolerance and to improve the performance [36, 78].

An alternative to using the additional stages to distribute the traffic across the input ports of the banyan routing stage is to use all the stages as rout-

(a) Augmented Banyan

(b) Benes

n x m   r x r   m x n

(c) Two-sided Clos

r x r

(d) Folded Clos

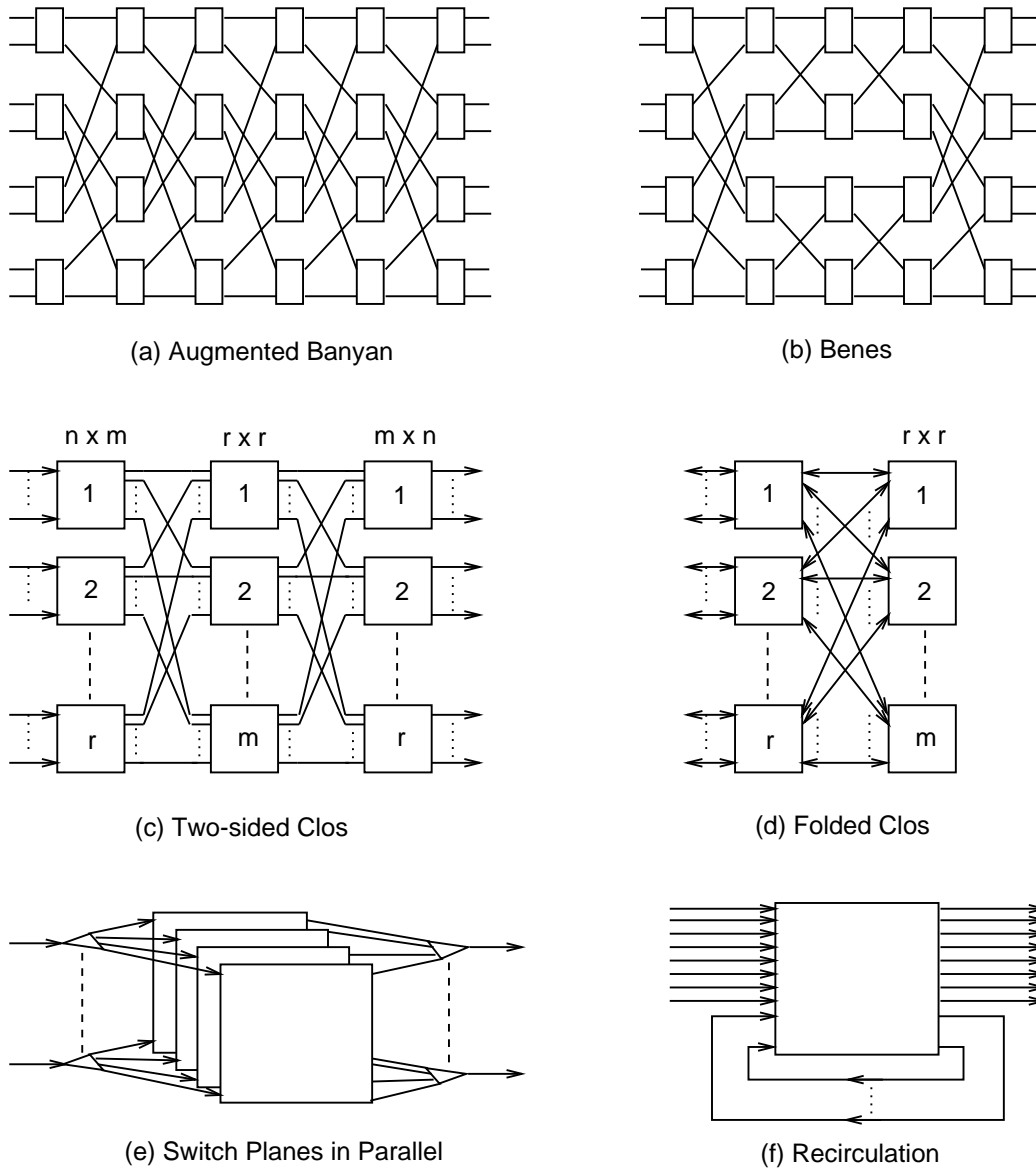(e) Switch Planes in Parallel

(f) Recirculation

Figure 5: Multiple path networks.

ing stages. Cells are removed from the network at the earliest possible switch stage. Conflicts between cells requesting the same link are handled by deflecting one of the cells over the wrong link and making it recommence its routing from the resulting location. Provided there are sufficient remaining switching stages in the network the cell may still arrive at the required output. The opportunity to remove cells can be given after every stage [16, 11],

after the first $log_2 N$ stages [76], or after each complete banyan network if multiple banyan networks are connected in series [73].

Switch Planes in Parallel: Another approach to introducing multiple paths into an interconnection network is to connect multiple switch planes in parallel, fig. 5(e). This approach offers increased reliability as well as improved performance since the loss of a complete switch plane will reduce the ca-
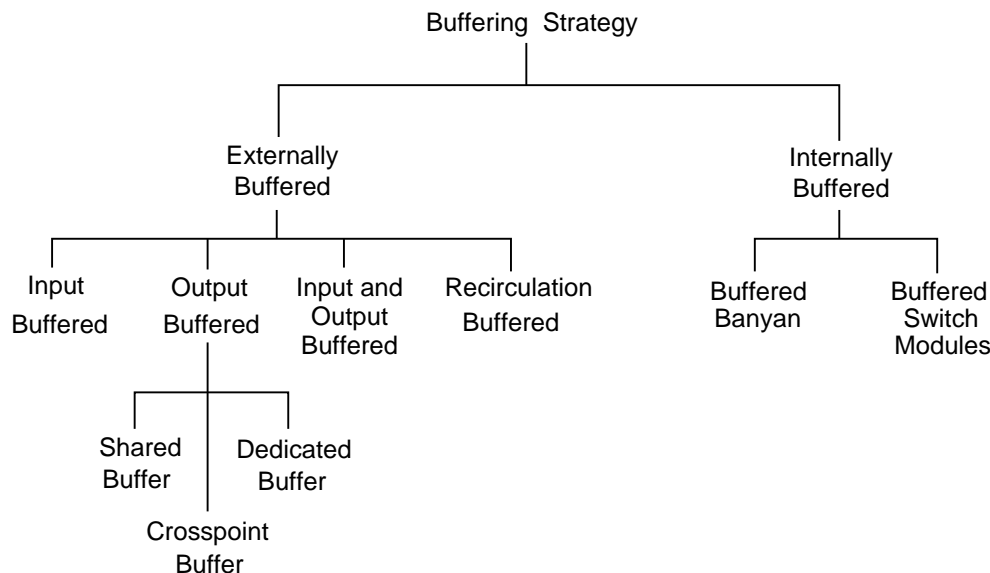
Figure 6: Classification of buffering strategy.

pacity but not the connectivity of the network. Switch designs that investigate this approach include: [2, 25, 29, 43, 63, 13, 55].

Clos Networks: The two-sided Clos network is shown in fig. 5(c). A Clos network constructed with non-blocking switch modules will have m paths between each input output pair and will be strictly non-blocking if $m \geq 2n - 1$. The Clos network can also be constructed in the folded form 5(d). In this case all of the interconnection links are bidirectional and cells pass through the first stage of the folded network in both directions. Thus the first stage performs the functions of both the first and last stages of the two-sided structure. The Clos structure has been proposed in a number of very large switch designs. The two-sided structure is used in: the large switch proposals from Bellcore [50, 48]; the Generalized Knockout Switch from ATT [21]; the NEC ATOM Switch [37]; [69] from Toshiba; and also in [68]. Examples of switch designs using a folded structure include: [4] from Alcatel; [23] from Siemens; [24] from the RACE program; and [65]. The ATOM Switch and the Alcatel projects are implementing large switch prototypes based upon a Clos structure of intercon-

nected shared medium or shared memory switching elements.

Recirculating Networks: Yet another approach to providing multiple paths through an interconnection network is to recirculate cells that have failed to reach the required output fig. 5(f). Some recirculating designs require the size of the interconnection network to be increased to accommodate the ports required for recirculation [34, 75, 29]. Other designs permit cells to exit from the internal stages of the network [16, 73, 11].

Load Sharing Networks: Finally, a multiple path network may be achieved by adding extra paths to interconnect switching elements in the same stage of a banyan (or delta) network while preserving the self-routing property. This approach is investigated in [46, 42].

Buffering

A simple classification of buffering strategies proposed for an ATM switch is given in fig. 6. The most important distinction between different buffering strategies is whether any cell queues are located within the switch fabric (internally buffered) or not
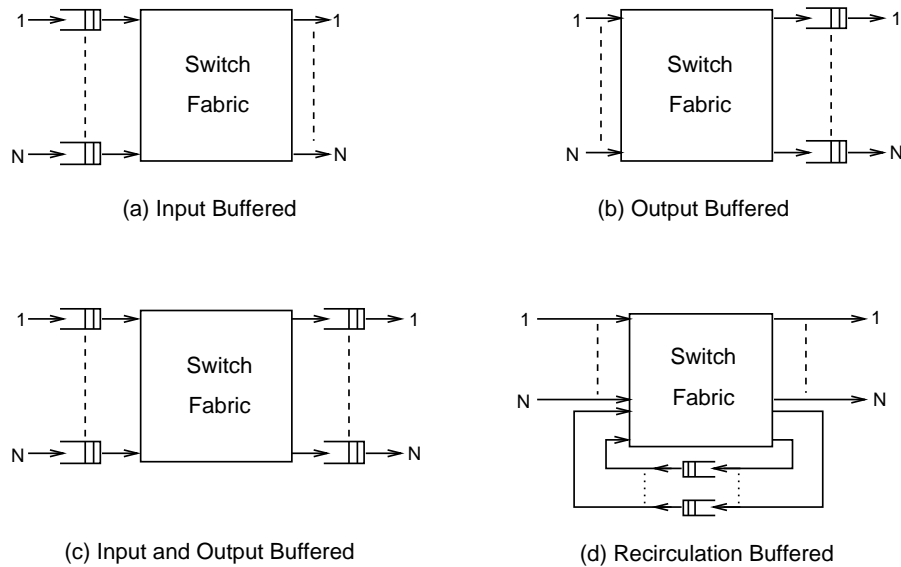
(a) Input Buffered

(b) Output Buffered

(c) Input and Output Buffered

(d) Recirculation Buffered

Figure 7: External Buffering.

(externally buffered). Fig. 7 illustrates the major externally buffered configurations.

Internal Buffering

If cell queues are located within the switch fabric, all cells belonging to the same virtual connection must travel the same path across the switch fabric if sequence errors are to be avoided. This implies that if a multiple path switch fabric is used, a path across the switch fabric must be selected for each virtual connection at call setup. To do this the switch must keep a record of the estimated traffic load on each link within the switch fabric and base its call acceptance decision on the estimated load across each possible path. This will complicate the call establishment process significantly for a large switch. The alternative is to allow each cell to take any path through the switch fabric and resequence the cells on exit from the switch, e.g. [4]. Multicast operation may be supported within the switch fabric of an internally buffered switch and is greatly simplified if a label routing approach is employed for the multicast traffic.

A switch formed from a single shared memory switch module (fig. 3(b)) may be considered as in-

ternally buffered with respect to its physical construction although it offers the performance of an output buffered switch. The shared memory design permits a single buffer to be shared by many input and output ports. This sharing of buffers substantially reduces the number of cell buffers required to support a given performance [45]. This is a significant advantage if the cell buffers are implemented within a custom integrated circuit (i.e. there is limited space available for buffers). It also gives a much better performance for bursty traffic if the total number of cell buffers is limited but can make the support of a large number of priority levels across the switch fabric more difficult. Also, when congestion occurs, the sharing of the buffers between many ports can make it more difficult to locate the source of the congestion by monitoring the queue occupancy.

One class of internally buffered designs locates the buffers on the input side of every switching element. These designs generally adopt a banyan (or possibly delta) network for the switch fabric and are often referred to as buffered banyan designs [62, 74, 75, 2, 25, 70]. In this class of design the switching element is simple to construct but the performance suffers from blocking within the switch
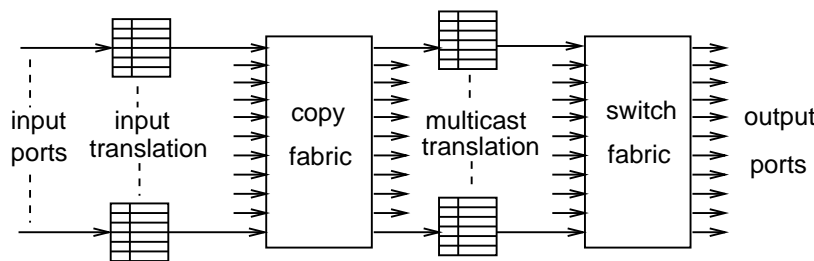
Figure 8: Multicast operation using an external copy fabric.

fabric. Higher performance internally buffered designs are obtained if a multiple path, multistage switch fabric is employed with shared memory or output buffered switch modules [23, 43, 4, 37, 69]. A number of large switch designs propose this approach.

External Buffering

External buffering allows the cell queues to be located close to the switch ports that they serve. Each switch port may monitor its cell queues and perform load monitoring to support congestion control. Furthermore, each queue may easily be separated into multiple classes of service and each port controller may implement a dynamic scheduling policy based upon queue occupancy to best serve the delay and loss requirements of each traffic class. The absence of cell queues within the switch fabric eases the support of multiple levels of priority across the switch fabric to support the different classes of traffic. Also, if a multiple path switch fabric is employed, a random selection between the alternative paths may be implemented to distribute the traffic evenly across the switch fabric. Hence the switch need not keep a record of the estimated traffic load on every internal link within the switch fabric. This considerably simplifies the call acceptance process. Cells on the same virtual connection will not suffer sequence errors even if they take different paths across the switch fabric as there are no buffers within the switch fabric.

In an externally buffered space switch, high bandwidth multicast operation is generally supported with a copy fabric prior to the switch fabric, fig. 8.

The copy fabric replicates the required number of copies of each multicast cell and the switch fabric routes them to the required output ports. A label translation operation is necessary between the copy fabric and the switch fabric. Copy fabric designs for this approach to multicast operation are presented in [74, 47, 56].

Input Buffering

In an input buffered switch the bandwidth between each input port and the switch fabric, and between the switch fabric and each output port, need only be slightly greater than that of the port itself. This permits the input queues to be located separately from the switch fabric, simplifies the implementation of the switch fabric, and avoids the need for buffers operating at some multiple of the port speed.

A large non-blocking switch with first-in-first-out input buffers, saturated with uniform random traffic, has a throughput of about 58% compared to that of the ideal output buffered switch [39, 33]. The performance for an input buffered delta or banyan network is degraded even further [55]. The performance may be improved by a technique known as input queue bypass if access is permitted to other cells in the input queues besides the cell at the head of the queue [59, 66, 53]. However, this technique complicates the implementation of the input buffers and requires a more complex contention resolution scheme. Examples of input buffered switch designs include: [35, 46, 71, 53, 3].

Output Buffering

In an ideal output buffered switch every output port must be able to accept a cell from every input port simultaneously (or at least within a single timeslot). In anything larger than a small switch module it is unreasonable to expect the switch fabric and output buffers to have sufficient capacity to achieve ideal output buffered operation. Thus, in an output buffered switch of reasonable size, there is always the possibility that more cells will request access to a particular output port than the switch fabric or output buffer can support. In this case the excess cells must be discarded. It is the task of the switch designer to ensure that the cell loss probability is sufficiently low for all reasonable patterns of incident traffic and acceptable operating loads. An output buffered switch can be much more complex than an input buffered switch because the switch fabric and output buffers must effectively operate at a much higher speed than that of each switch port to reduce the probability of cell loss. If each output buffer can receive up to 8 cells within any single timeslot, a cell loss probability of less than 10^{-6} may be achieved for a traffic load of 90% with uniform random traffic [79].

A single stage shared memory design may be considered as output buffered from the viewpoint of its performance. It has a single buffer shared by all input and output ports [17, 44, 69]. A switch with dedicated output buffers has a separate buffer on each output port (fig. 7(b)). Each output buffer is shared by all input ports wishing to access that output port [79, 14, 1, 21, 12, 29, 5, 16, 73, 77]. Output buffered switch designs based upon a matrix interconnection network (fig. 4(c)) use crosspoint buffers. A separate crosspoint buffer is required for each input output pair resulting in N^2 crosspoint buffers [57, 40, 41].

Input and Output Buffering

An input and output buffered switch combines the two approaches of input buffering and output buffering. Cell loss within the switch fabric of an output buffered switch due to transient traffic patterns is undesirable. Therefore instead of discard-ing cells that cannot be handled during the current timeslot they are retained in input buffers. The input buffers need not be large to substantially reduce the probability of cell loss for reasonably random traffic even at very high loads. This is an approach favored by a number of large switch designs [50, 63, 13, 61, 48, 55, 11].

Recirculation Buffering

In this approach, output port contention is handled by recirculating those cells that cannot be output during the current timeslot back to the input ports via a set of recirculation buffers fig. 7(d). It offers the performance of an output buffered switch but will discard cells in the switch fabric if more cells require recirculation during a particular timeslot than the number of recirculation ports. Recirculation may cause out of sequence errors between cells in the same virtual connection unless steps are taken to prevent it. Recirculation buffering has been suggested in the Starlite switch [34] and in the Shuffle-out design [16]. Also, the Sunshine [29] and Tandem Banyan [73] switch designs combine recirculation buffering with output buffering.

Contention Resolution

Contention occurs when two or more cells compete for a single resource such as an internal link or an output port. A simple classification of contention resolution schemes is given in fig. 9. In an internally buffered switch, contention is handled by placing buffers at the point of contention. In an externally buffered switch a contention resolution mechanism is required. Three basic actions can be taken once contention is detected in an externally buffered switch: backpressure; deflection; and loss. Input buffered switch designs typically use backpressure from the point of contention to the input buffers. Pure output buffered designs use a loss mechanism where cells that cannot be handled are discarded at the point of contention. Switch designs based upon recirculation use both deflection and loss. A deflection mechanism will route the cells that lose contention over a path other than the shortest path to the requested destination. This path may either
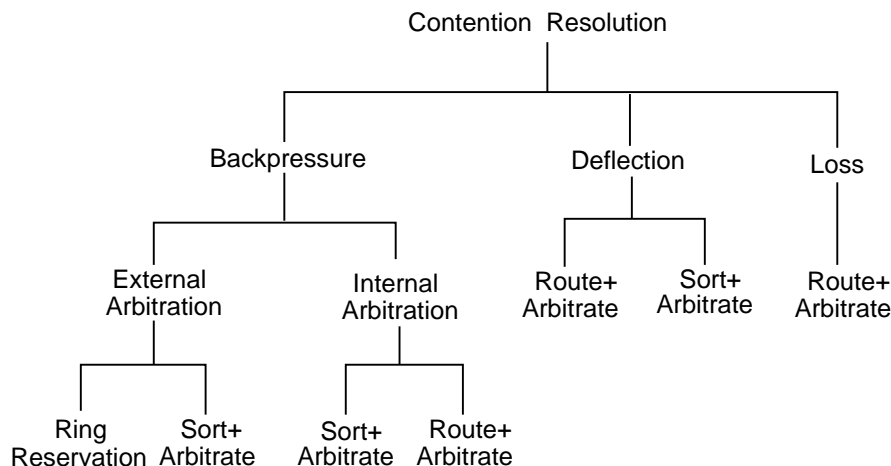
Figure 9: Classification of contention resolution mechanisms.

be a recirculation path or an alternative path in the forward direction from the point of contention.

The decision as to which cells to accept and which to reject is made by an arbiter. The arbitration decision may be based upon cell priority, upon a timestamp within the cell, or it may be random. The arbiter may be centralized and implemented externally to the switch fabric, or distributed and implemented internally within each switching component that forms the switch fabric. Most arbiter designs give a simple accept or reject decision in response to cell transmission requests. However, in [53] the arbiter schedules each request and replies with the delay before the next free timeslot to each requested output port.

Three basic arbitration mechanisms have been proposed: ring reservation; sort and arbitrate; and route and arbitrate. In ring reservation the input ports are interconnected via a ring which is used to request access to the output ports. Switches that use a ring reservation arbitration mechanism include: [8, 22, 48]; while a logical ring within a centralized external arbitration device is presented in [13].

For switches that employ a sorting mechanism in the switch fabric, all cells requesting the same output port will appear adjacent to each other after sorting. Thus an arbitration mechanism may be

implemented by comparing the destination requests of each cell to those of its neighbors on exit from the sorting network. Switch designs that include a sort and arbitrate mechanism include: [3, 53] which use an external arbitration mechanism; [35, 61] which use an internal probe and acknowledgement approach; and [34, 29] which use deflection via recirculation.

In the route and arbitrate approach cells are routed through the switch fabric and arbiters detect contention at the point of conflict. Switch designs that include a route and arbitrate mechanism include: [46, 63, 71, 55] which use an internal acknowledgement; [76, 16, 11, 73] which use deflection; and [79, 21, 12, 77] which use a pure loss approach.

ATM Networking Issues

Recent advances in technology have enabled the implementation of packet switching in hardware, at high speed, to realize the ATM switch. While the design and implementation of an ATM switch is far from trivial it is no longer shrouded in mystery. Sufficient research and experimental prototypes have been reported in the literature to assure us of the feasibility of constructing an ATM switch. However, the networking of ATM switches is not as well understood and much current research is now di-

rected at the problem of ATM networking.

The majority of the issues that remain to be solved in ATM networking concern the application of ATM technology to the public broadband network, B-ISDN. The application of ATM technology to the private networking environment presents networking issues that permit much simpler solutions than is possible in the public domain. The private networking environment is much less hostile than that of the public network. We can assume that terminals connected to the network will cooperate with the network — that requests to throttle back will not be ignored and that traffic shaping parameters may be applied at the source. Also, in the private domain, system administration under the control of a single entity is available for network management with monitoring and control of long-term parameters.

Traffic Characterization

A virtual connection in broadband ISDN is a contract between the network and the customer to deliver traffic of specified statistical characteristics to the destination with a specified grade of service. The need for such a contract between customer and network is a natural result of offering a public service. The specification of the traffic characteristics permits the preventive congestion control techniques of admission control and bandwidth allocation to be applied to each virtual connection. However, for many classes of traffic the specification of the traffic characteristics and the required grade of service is a difficult problem and is a subject of current research [67]. Most current data traffic is extremely variable [49] and will become increasingly so at higher speeds. This makes it difficult to characterize with a small set of simple parameters [28].

The majority of current data networks operate on a very different model of the relationship between host and network. The network is a resource shared by all. In the local area, bandwidth is assumed to be plentiful and excess bandwidth may be traded for simple control mechanisms. Best effort delivery is offered by the network and simple mechanisms are available to detect and recover from congestion, e.g.

the slow-start enhancement to TCP/IP [38]. None but the simplest assumptions are made about the traffic from any particular source and no estimation of traffic characteristics is expected of the user.

A private ATM network may employ this model of operation to offer a basic high-speed data communication service if a simple reactive congestion control mechanism is implemented in the local area. No guarantee is given by the network for the basic data service and the network does not expect sources to declare their traffic characteristics in order to use this service. Other classes of traffic that impose more stringent loss and delay requirements may be carried at a higher priority using simple approximations of their traffic characteristics for admission control and bandwidth allocation [31]. Any bandwidth not consumed by higher priority traffic is dynamically available to the basic data communication service.

Congestion Control

The preventive congestion control techniques of admission control and bandwidth allocation must be applied to virtual connections that require a guarantee of service from the network. However, a basic high-speed data communication service may be offered, at least in the local area, with only a simple reactive congestion control scheme. A forward congestion notification scheme has been proposed for B-ISDN and investigated in [52]. If a small amount of additional bandwidth is available a backward congestion notification scheme is very effective over a distance of up to a few tens of kilometers. Both schemes require a traffic source to throttle back when the network indicates that it is experiencing congestion. In a private network it is acceptable to assume that traffic sources will comply with such requests from the network. In a public network no such assumption can be made although the network may ensure that a source will receive better overall service if it does comply with such requests.

Reactive congestion control allows a local area or campus area ATM network to survive periods of high utilization without loss of traffic. Reactive

control of local data traffic also permits a limited amount of wide area data traffic to transit the local area without loss if it is carried at higher priority. For higher levels of wide area data traffic, packet discard may be employed on the trunk interfaces. Packet discard alleviates congestion on outbound trunks by discarding complete packets or bursts of packets on the basic data communication service. Each packet typically consists of at least several cells. So discarding entire packets is preferable to the random loss of cells that would otherwise result in a much larger number of packets being corrupted. Detection of packet loss may be used to trigger higher layer reactive congestion control mechanisms [38]. To implement packet discard the outbound trunk interface must be able to recognize the beginning and end of packets. This implies it must know the adaptation layer protocol in use on each data connection which is acceptable in a private network. Packet discard is not acceptable in a public B-ISDN network as the adaptation layer is an end to end matter and is not visible to the switch at the ATM layer.

Source Policing

In the public broadband network the contract between the customer and the network needs to be enforced on each virtual connection to protect the network and to avoid interference with other traffic. This function is known as bandwidth enforcement, source policing, or usage parameter control. It ensures that the agreed traffic parameters are not violated by the customer on access to the network. Various simple mechanisms are available to perform this function but traffic with a high degree of variability can be difficult to enforce accurately [64]. This is particularly so for traffic arriving with random queueing delay variations or for the enforcement of statistical parameters such as average rate or burst length.

Source policing is less critical in a private network. Local area data traffic using the basic data communication service need not be subject to source policing at all if reactive congestion control is implemented. For other classes of traffic and for

wide area traffic, sources of approved design may implement traffic shaping. This permits an upper bound on the statistical parameters of the source to be accurately established directly at the source according to the requirements of the network. The traffic shaping function may be implemented together with the adaptation layer hardware support and the rate control mechanism for reactive congestion control. Thus the source policing function on the access ports of a private network is a tool of network management to protect against faulty terminals and intentional abuse rather than a precise per connection traffic management function.

Accounting and Interworking

Accounting and interworking are issues facing the introduction of the public broadband ISDN [15, 28]. Charging on a per cell basis for each virtual connection would seem to introduce additional and unnecessary complexity to the user-network interface. Charging on the basis of negotiated traffic parameters is likely to be insufficient for very bursty traffic sources such as the majority of interactive data applications. Detailed accounting is generally not required in private networks.

Compatibility of new broadband services is required with existing services and customer premises equipment. This forms a much greater challenge for public networks than for private networks. Equipment in the private domain is amortized more rapidly so there is less of a requirement to interwork with aging technology. Also, useful services may be delivered by ATM technology in the private domain without requiring universal and immediate interworking with all existing telecommunications equipment.

Multipoint Connections

Data networking in the local area has developed from the capabilities of the shared medium LAN. A corporate ATM network must interwork with existing LANs, bridges, routers, and protocol suites. One of the essential characteristics of a LAN, employed in many current protocols and network applications, is the ability to broadcast or multicast to

a group of stations in the local environment. This capability must be supported by an ATM network in the local area.

A limited implementation could be achieved with a multicast server in each ATM switch but a much more efficient approach is to support multicast operation directly within the switch. A multicast group can be formed with a point-to-multipoint connection from each group member to all other members of the group. However, this approach consumes a large number of VCIs and incurs a great deal of administrative overhead when members are added to and deleted from the group. A more direct approach is to support a multicast group with a single many-to-many connection. Many-to-many connections require cell to packet reassembly at the destination capable of handling interleaved cells from different sources arriving on the same multipoint virtual channel.

Multipoint connections are awaiting further study for the public B-ISDN. Many issues such as accounting, signaling, and traffic management require study before multipoint capabilities may be introduced into the public network. Simple many-to-many connections for the basic data service may be introduced into a private network without great difficulty, especially if the majority of such connections remain within the campus or metropolitan area.

Conclusion

ATM switches designed for corporate networks will very soon become available. The initial commercial application will be a high capacity backbone to relieve the bandwidth bottleneck that is beginning to constrain current solutions. Today's data networking services will use tomorrow's bandwidth to remove congestion, offer high-speed data services, and eliminate many of the current problems of managing large data networks. Within the campus or local area, the physical structure of the network will no longer constrain the communication bandwidth available. Corporate wide area ATM networks will soon arise as local ATM sites are interconnected via

leased lines. Given sufficient bandwidth this will erode the distinction between the local area network and the wide area network from the user's perspective. The ATM LAN will expand to include remote sites participating in network applications as though they were locally connected, ultimately limited only by the speed of light.

The emergence of ATM networks in the commercial environment will stimulate the development of new communication services. Multimedia applications are an obvious candidate, as is high-speed data networking, but advanced services as yet unknown will be enabled by the introduction of ATM technology. The availability of such applications will encourage direct ATM connection to the workstation. So ATM switch designs for corporate networks must have the ability to grow to a large number of ports and the cost per connection must fall to permit the direct attachment of individual workstations.

The majority of existing ATM switch prototypes constructed by the leading telecommunication equipment manufacturers are based upon a time division switch module. Each design has developed a chip set in 0.8 μm CMOS or BiCMOS technology to implement the switch module. These switch modules are interconnected in a Clos structure to achieve a large internally buffered switch. A time division switch module can be very flexible as it may also be designed for use as an access multiplexer and a traffic concentrator. One manufacturer has chosen a space division matrix structure with crosspoint buffers for the basic switch module. A switch designed for a campus network application has selected a buffered banyan approach. The construction of a large Batcher-banyan switch is also currently under investigation. Most of the other designs reported in the literature tend more towards research investigations rather than manufacturing prototypes. While there is much research value in investigating new self-routing space division structures, it appears that the safe money is behind internally buffered switch modules — at least with current implementation and interconnection technology.

However, the ATM switch prototypes from the telecommunication manufacturers have been designed to satisfy the requirements of the public broadband ISDN. There are a number of significant differences between broadband ISDN and the application of ATM technology to the corporate network environment. In the corporate network, ATM technology must interwork with current LANs, bridges, routers, and protocols. Existing data traffic sources are not going to declare their traffic characteristics before submitting traffic to the network. Nor have appropriate simple traffic descriptors, or algorithms to compute the effective bandwidth, yet been defined. A simple reactive congestion control scheme with reasonably large buffers will allow an ATM network to offer a campus-wide high-speed data service with the characteristics of a LAN. The support of multicast groups will assist existing protocols to operate over an ATM LAN. Wide area connectivity can employ traffic shaping at the source and employ packet discard on outbound trunks to alleviate congestion. A substantial number of classes of priority within the switch will be helpful as the network matures to support new classes of loss and delay sensitive traffic.

An ATM switch is of no use if there are no products with ATM interfaces. An ATM network interface chip is required for high-speed data applications with properties similar to today's LAN controllers. This device must perform the adaptation layer functions of packet to cell segmentation and reassembly to offer a packet level interface to the host. It would be very useful if the device also supported rate control, for the congestion control scheme, and traffic shaping to improve wide area network performance and to allow the accurate definition of worst case traffic characteristics.

The use of large buffers and queue monitoring for reactive congestion control, with multiple queue priorities and a flexible queue service algorithm, suggests that an externally buffered switch may be more appropriate for corporate networking applications. Internally buffered designs tend to have a limited buffer space and place greater reliance upon the preventive congestion control mechanisms

of accurate traffic characterization, source policing, call admission, and bandwidth allocation. External buffering also permits traffic to be evenly distributed across a multiple path switch fabric without knowledge of the traffic characteristics and without requiring resequencing. However, the design of the switch itself is not the most critical issue in the successful application of ATM technology to the corporate network. Cost per port, capacity, growth capability, and the range of existing products for which ATM interfaces are offered will of course play their part. But the solution of the ATM networking issues, and the support that the switch and the network interface provide to implement these solutions, will have a critical impact upon the success of any offering in the corporate ATM networks market.

References

[1] H Ahmadi, W E Denzel, C A Murphy, and E Port. A high-performance switch fabric for integrated circuit and packet switching. In *Proc. IEEE Infocom*, pages 9–18, New Orleans, Mar. 1988.

[2] G J Anido and A W Seeto. Multipath interconnection: A technique for reducing congestion within fast packet switching fabrics. *IEEE J. Select. Areas in Commun.*, 6(9):1480–1488, Dec. 1988. (Reprinted in [18]).

[3] N Arakawa, A Noiri, and H Inoue. ATM switch for multimedia switching system. In *Proc. Int. Switching Symp. (ISS '90)*, volume 5, pages 9–14, Stockholm, May 1990.

[4] T R Banniza et al. Design and technology aspects of VLSIs for ATM switches. *IEEE J. Select. Areas in Commun.*, 9(8):1255–1264, Oct. 1991.

[5] P Barri and J A O Goubert. Implementation of a 16×16 switching element for ATM exchanges. *IEEE J. Select. Areas in Commun.*, 9(5):751–757, Jun. 1991.

[6] K E Batcher. Sorting networks and their applications. In *Proc. Spring Joint Computer Conf.*, pages 307–314, 1968. (Reprinted in [7]).

[7] A Bhargava, editor. *Integrated Broadband Networks*. Artech House, 1991.

[8] B Bingham and H Bussey. Reservation-based contention resolution mechanism for Batcher-banyan packet switches. *Electronics Letters*, 24(13):772–773, June 1988.

[9] G Broomell and J R Heath. Classification categories and historical development of circuit switching topologies. *Computing Surveys*, 12(2):95–133, Jun. 1983. (Reprinted in [18]).

[10] R G Bubenik and J S Turner. Performance of a broadcast packet switch. *IEEE Trans. Commun.*, 37(1):60–69, Jan. 1989.

[11] P Campoli and A Pattavina. An ATM switch with folded shuffle topology and distributed access. In *Proc. IEEE Int. Conf. Commun.*, volume 2, pages 1021–1027, Denver, Jun. 1991.

[12] H J Chao. A recursive modular terabit/second ATM switch. *IEEE J. Select. Areas in Commun.*, 9(8):1161–1172, Oct. 1991.

[13] A Cisneros. Large packet switch and contention resolution device. In *Proc. Int. Switching Symp. (ISS '90)*, volume 3, pages 77–83, Stockholm, Sweden, May 1990.

[14] M De Prycker and M De Somer. Performance of an independent switching network with distributed control. *IEEE J. Select. Areas Commun.*, SAC-5(8):1293–1301, Oct. 1987.

[15] M Decina. Open issues regarding the universal application of ATM for multiplexing and switching in the B-ISDN. In *Proc. IEEE Int. Conf. Commun.*, volume 3, pages 1258–1264, Denver, Jun. 1991.

[16] M Decina, P Giacomazzi, and A Pattavina. Shuffle interconnection networks with deflection routing for ATM switching: The closed loop Shuffleout. In *Proc. IEEE Infocom*, volume 3, pages 1254–1263, Apr. 1991.

[17] M Devault, J Y Cochennec, and M Servel. The "Prelude" ATD experiment: Assessments and future prospects. *IEEE J. Select. Areas in Commun.*, 6(9):1528–1536, Dec. 1988. (Reprinted in [18]).

[18] C Dhas, V K Konangi, and M Sreetharan, editors. *Broadband switching architectures, protocols, design and analysis.* IEEE Computer Society Press, 1991.

[19] D M Dias and J R Jump. Analysis and simulation of buffered delta networks. *IEEE Trans. Computers*, C-30(4):273–282, Apr. 1981.

[20] D M Dias and M Kumar. Packet switching in $N \log N$ multistage networks. In *Proc. IEEE Globecom*, pages 114–120, 1984. (Reprinted in [18]).

[21] K Y Eng and M J Karol. The growable switch architecture: A self-routing implementation for large ATM applications. In *Proc. Int. Conf. Commun.*, volume 2, pages 1014–1020, Denver, Jun. 1991.

[22] K Y Eng, M J Karol, and I Chih-Lin. A modular broadband (ATM) switch architecture with optimum performance. In *Proc. Int. Switching Symp. (ISS '90)*, volume 4, pages 1–6, Stockholm, May 1990.

[23] W Fischer et al. A scalable ATM switching system architecture. *IEEE J. Select. Areas in Commun.*, 9(8):1299–1307, Oct. 1991.

[24] D G Fisher et al. A flexible network architecture for the introduction of ATM. In *Proc. Int. Switching Symp. (ISS '90)*, volume 2, pages 35–44, Stockholm, May 1990.

[25] A Forcina, T DiStefano, and E Taormina. A multicast broadband switching module in a hybrid ATM environment. In *Proc. IEEE Int. Conf. Commun. (ICC '89)*, volume 1, pages 111–117, Boston, Jun. 1989.

[26] A G Fraser. Designing a public data network. *IEEE Commun. Mag.*, pages 31–35, Oct. 1991.

[27] I Gard and J Rooth. An ATM switch implementation — Technique and technology. In *Proc. Int. Switching Symp. (ISS '90)*, volume 4, pages 23–27, Stockholm, May 1990.

[28] J Gechter and P O'Reilly. Conceptual issues for ATM. *IEEE Commun. Mag.*, pages 14–16, Jan. 1989.

[29] J N Giacopelli et al. Sunshine: A high-performance self-routing broadband packet switch architecture. *IEEE J. Select. Areas in Commun.*, 9(8):1289–1298, Oct. 1991.

[30] L R Goke and G J Lipovski. Banyan networks for partitioning multiprocessor systems. In *Proc. First Annual Symp. Computer Architecture*, pages 21–28, Dec. 1973.

[31] R Guerin, H Ahmadi, and M Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE J. Select. Areas in Commun.*, 9(7):968–981, Sep. 1991.

[32] J J Hickey and W S Marcus. The implementation of a high speed ATM packet switch using CMOS VLSI. In *Proc. Int. Switching Symp. (ISS '90)*, volume 1, pages 75–84, Stockholm, May 1990.

[33] M G Hluchyj and M J Karol. Queueing in high-performance packet switching. *IEEE J. Select. Areas in Commun.*, 6(9):1587–1597, Dec. 1988. (Reprinted in [7]).

[34] A Huang and S Knauer. Starlite: A wideband digital switch. In *Proc. IEEE Globecom*, pages 121–125, Nov. 1984. (Reprinted in [7, 18]).

[35] J Y Hui and E Arthurs. A broadband packet switch for integrated transport. *IEEE J. Select. Areas Commun.*, SAC-5(8):1264–1273, Oct. 1987. (Reprinted in [7, 18]).

[36] A Itoh. A fault-tolerant switching network for B-ISDN. *IEEE J. Select. Areas in Commun.*, 9(8):1218–1226, Oct. 1991.

[37] A Itoh et al. Practical implementation and packaging technologies for a large-scale ATM switching system. *IEEE J. Select. Areas in Commun.*, 9(8):1280–1288, Oct. 1991.

[38] V Jacobson. Congestion avoidance and control. In *Proc. ACM SIGCOM*, pages 314–329, 1988. (Reprinted in [7]).

[39] M J Karol, M G Hluchyj, and S P Morgan. Input versus output queueing on a space-division packet switch. *IEEE Trans. Commun.*, COM-35(12):1347–1356, Dec. 1987.

[40] Y Kato et al. A VLSIC for the ATM switching system. In *Proc. Int. Switching Symp. (ISS '90)*, volume 3, pages 27–32, Stockholm, May 1990.

[41] U Killat et al. A versatile ATM switch concept. In *Proc. Int. Switching Symp. (ISS '90)*, volume 4, pages 127–134, Stockholm, May 1990.

[42] H S Kim and A Leon-Garcia. A self-routing multistage switching network for broadband ISDN. *IEEE J. Select. Areas in Commun.*, 8(3):459–466, Apr. 1990.

[43] T Koinuma et al. An ATM switching system based upon a distributed control architecture. In *Proc. Int. Switching Symp. (ISS '90)*, volume 5, pages 21–26, Stockholm, May 1990.

[44] T Kozaki et al. 32×32 shared buffer type ATM switch VLSIs for B-ISDNs. *IEEE J. Select. Areas in Commun.*, 9(8):1239–1247, Oct. 1991.

[45] H Kuwahara et al. A shared buffer memory switch for an ATM exchange. In *Proc. IEEE Int. Conf. Commun. (ICC '89)*, volume 1, pages 118–122, Boston, June 1989.

[46] C T Lea. Design and performance evaluation of unbuffered self-routing networks for wideband packet switching. In *Proc. IEEE Infocom*, volume 1, pages 148–156, San Francisco, Jun. 1990.

[47] T T Lee. Nonblocking copy networks for multicast packet switching. *IEEE J. Select Areas in Commun.*, 6(9):1455–1467, Dec. 1988. (Reprinted in [18]).

[48] T T Lee. A modular architecture for very large packet switches. *IEEE Trans. Commun.*, 38(7):1097–1106, Jul. 1990.

[49] W E Leland. LAN traffic behavior from milliseconds to days. In *Proc. 7th Int. Teletraffic Congress Seminar*, New Jersey, Oct. 1991.

[50] S C Liew and K W Lu. A three-stage architecture for very large packet switches. *Int. J. Digital and Analog Cabled Systems*, 2(4):303–316, Oct. 1989.

[51] G W R Luderer and S C Knauer. The evolution of space division packet switches. In *Proc. Int. Switching Symp. (ISS '90)*, volume 5, pages 211–216, Stockholm, May 1990.

[52] B A Makrucki. On the performance of submitting excess traffic to ATM networks. In *Proc. IEEE Globecom*, Dec. 1991.

[53] H Matsunaga and H Uematsu. A 1.5 Gb/s 8×8 cross connect switch using a time reservation algorithm. *IEEE J. Select. Areas in Commun.*, 9(8):1308–1317, Oct. 1991.

[54] B Monderer, G Pacifici, and C Zukowski. The Cylinder Switch: An architecture for a manageable VLSI giga-cell switch. In *Proc. IEEE Int. Conf. Commun. (ICC '90)*, volume 2, pages 567–571, Apr. 1990.

[55] P Newman. A fast packet switch for the integrated services backbone network. *IEEE J. Select. Areas in Commun.*, 6(9):1468–1479, Dec. 1988. (Reprinted in [18]).

[56] P Newman and M Doar. A slotted ring copy fabric for a multicast fast packet switch. In *Proc. Int. Switching Symp. (ISS '90)*, volume 5, pages 205–210, Stockholm, May 1990.

[57] S Nojima et al. Integrated services packet network using bus matrix switch. *IEEE J. Select. Areas Commun.*, SAC-5(8):1284–1292, Oct. 1987. (Reprinted in [18]).

[58] Y Oie. Effect of speedup in nonblocking packet switch. In *Proc. IEEE Int. Conf. Commun. (ICC '89)*, volume 1, pages 410–414, Boston, Jun. 1989.

[59] Y Oie et al. Survey of the performance of non-blocking switches with FIFO input buffers. In *Proc. IEEE Int. Conf. Commun. (ICC '90)*, volume 2, pages 737–741, Apr. 1990.

[60] J H Patel. Performance of processor-memory interconnections for multiprocessors. *IEEE Trans. Computers*, C-30(10):771–780, Oct. 1981.

[61] A Pattavina. A multiservice high-performance packet switch for broad-band networks. *IEEE Trans. Commun.*, 38(9):1607–1615, Sep. 1990.

[62] G Perucca. Research on advanced switching techniques for the evolution to ISDN and broadband ISDN. *IEEE J. Select Areas Commun.*, SAC-5(8):1356–1364, Oct. 1987.

[63] R J Proctor and T S Maddern. Synchronous ATM switching fabrics. In *Proc. Int. Switching Symp. (ISS '90)*, volume 4, pages 109–114, Stockholm, May 1990.

[64] E P Rathgeb. Modeling and performance comparison of policing mechanisms for ATM networks. *IEEE J. Select. Areas in Commun.*, 9(3):325–334, Apr. 1991.

[65] Y Sakurai et al. Large scale ATM multistage switching network with shared buffer memory switches. *IEEE Commun. Mag.*, pages 90–96, Jan. 1991.

[66] K W Sarkies. The bypass queue in fast packet switching. *IEEE Trans. Commun.*, 39(5):766–774, May 1991.

[67] Y Sato and K-I Sato. Virtual path and link capacity design for ATM networks. *IEEE J. Select. Areas in Commun.*, 9(1):104–111, Jan. 1991.

[68] K Sezaki, Y Tanaka, and M Akiyama. The Cascade Clos broadcast switching network — A new ATM switching network which is multiconnection non-blocking. In *Proc. Int. Switching Symp. (ISS '90)*, volume 4, pages 143–147, Stockholm, May 1990.

[69] Y Shobatake et al. A one-chip scalable 8×8 ATM switch LSI employing shared buffer architecture. *IEEE J. Select. Areas in Commun.*, 9(8):1248–1254, Oct. 1991.

[70] Y Shobatake and T Kodama. A cell switching algorithm for the buffered banyan network. In *Proc. IEEE Int. Conf. Commun. (ICC '90)*, volume 2, pages 754–760, Apr. 1990.

[71] Q Ta and J S Meditch. A high speed integrated services switch based on 4×4 switching elements. In *Proc. IEEE Infocom*, pages 1164–1171, San Francisco, Jun. 1990.

[72] F A Tobagi. Fast packet switch architectures for broadband integrated services digital networks. *Proc. IEEE*, 78(1):133–167, Jan. 1990.

[73] F A Tobagi, T Kwok, and F M Chiussi. Architecture, performance, and implementation of the Tandem Banyan fast packet switch. *IEEE J. Select. Areas in Commun.*, 9(8):1173–1193, Oct. 1991.

[74] J S Turner. Design of a broadcast packet switching network. *IEEE Trans. Commun.*, 36(6):734–743, June 1988. (Reprinted in [18]).

[75] H Uematsu and R Watanabe. Architecture of a packet switch based on banyan switching network with feedback loops. *IEEE J. Select. Areas in Commun.*, 6(9):1521–1527, Dec. 1988. (Reprinted in [18]).

[76] S Urushidani. Rerouting network: A high-performance self-routing switch for B-ISDN. *IEEE J. Select. Areas in Commun.*, 9(8):1194–1204, Oct. 1991.

[77] W Wang and F A Tobagi. The Christmas Tree Switch: An output queueing space-division fast packet switch .... In *Proc. IEEE Infocom*, volume 1, pages 163–170, Apr. 1991.

[78] S C Yang and J A Silvester. A reconfigurable ATM switch fabric for fault tolerance and traffic balancing. *IEEE J. Select. Areas in Commun.*, 9(8):1205–1217, Oct. 1991.

[79] Y S Yeh, M G Hluchyj, and A S Acampora. The Knockout switch: A simple modular architecture for high-performance packet switching. *IEEE J. Select. Areas Commun.*, SAC-5(8):1274–1283, Oct. 1987. (Reprinted in [7, 18]).