International Journal of
DIGITAL AND ANALOG CABLED SYSTEMS

Special Issue
on
Interconnection Networks for Broadband Packet Switching
Guest Editor: A. Pattavina

CONTENTS

VOLUME 2, ISSUE No. 4 *October–December 1989*

Indexed or Abstracted by Cambridge Scientific Abstracts, Engineering Index, INSPEC and Telecommunications Abstracts

IJDSEM 2(4) 201–346 (1989)
ISSN 0894–3222

ATM SWITCHES—BASIC ARCHITECTURES AND THEIR PERFORMANCE

ERWIN P. RATHGEB, THOMAS H. THEIMER AND MANFRED N. HUBER

University of Stuttgart, Institute of Communications Switching and Data Technics, Seidenstrasse 36, D 7000 Stuttgart 1, Federal Republic of Germany

SUMMARY

The prospect of a broadband ISDN based on the ATM principle has stimulated the development of new, high-performance switching node architectures. In this paper different options for the switching networks used in such a node will be discussed in some detail. First, some generic architectures for the individual switching elements will be presented including the aspects of buffering and collision resolution, followed by a short classification of different types of switching networks and some performance figures for delta networks, a class of multi-stage, single-path interconnection networks. In the last section, the sensitivity of the performance results with respect to the traffic assumptions will be discussed.

KEY WORDS ATM BISDN Switch architectures Performance aspects

1. INTRODUCTION

CCITT has made a commitment to the asynchronous transfer mode (ATM) as the principle for the target BISDN in the recommendation I.121.[1] The main characteristics of an ATM network as defined in I.121 are

(a) information transfer in small, fixed size blocks (cells)
(b) communication based on virtual connections
(c) cells contain information field and header (identification of the virtual connections and other functions)
(d) the multiplexing scheme is asynchronous time division multiplexing (ATD)
(e) the sequence of cells within a virtual connection is guaranteed by the network
(f) user-network interfaces will be standardized at transmission speeds of about 150 and 600 Mb/s

These characteristics, together with the requirements of the envisaged services including conversational, messaging, retrieval and broadcast services, which will be used for voice, data, still pictures and video[1] set the framework for the design and the dimensioning of the network elements.

For the switching nodes this results in a total throughput requirement in the Gb/s range while keeping cross-node delays and especially cell loss due to buffer overflow at a very low level. Therefore, central control for the switching of the individual data cells has to be avoided, and highly parallel architectures have to be developed for the switch fabrics.

In this paper we try to identify the generic types for building blocks of these switch fabrics and we discuss their basic properties. Since these basic elements are limited to a maximum size that is not sufficient for a normal size switching node, this results in the need for multi-stage switch fabric structures as described in Section 3. Section 4 will provide an extension to existing performance studies by considering more realistic traffic sources and the influence of the virtual connection concept.

2. SWITCHING ELEMENT ARCHITECTURES

The switching elements are the basic modules for the construction of larger switch fabrics. They perform the actual switching functions by analysing the routeing information of the arriving cells and directing them to the correct output port. The general model of a switching element consists of an input controller (IC), an interconnection network and an output controller (OC) (Figure 1). Moreover, buffers have to be provided in the switching elements to avoid excessive cell loss in case of collisions.

The input controller has to synchronize the arriving ATM cell stream to the internal clock of the switching element. If the switch fabric is operated in a VCI-routeing mode, the virtual connection identifier (VCI) of each cell has to be translated in every switching element into a new VCI plus the internal routeing information for the interconnection network. If a self-routeing operation mode is used, the VCI translation and the addition of the complete routeing information is only performed at the entrance of the switch fabric and not in each switching element. The task of the output controller

0894–3222/89/040227–10$05.00

Received 27 May 1989
Revised 9 August 1989

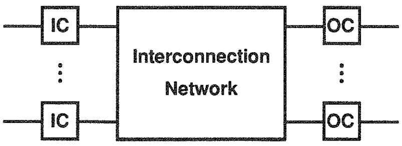


Figure 1. General model of an ATD switching element

is to transmit the cells which have been received from the interconnection network on the outgoing link.

The interconnection network is the most critical part of the switching element, because it is normally the bottleneck which limits the maximum throughput of the switch. The topologies and properties of various realizations will be described in the following.

2.1. *Matrix-type switching elements*

A matrix-type switching element is characterized by a rectangular matrix of crosspoints which is able to provide a conflict-free connection of any input controller with any output controller (Figure 2). Depending on the speed limitations of the interconnection matrix, the buffers of the switching element can be located at the input controller, at the output controller or at each crosspoint of the matrix.[2–4]

Input buffers. In these switching elements, cell buffers are located at the individual input controllers. If the speed of the incoming links is equal to the speed of the outgoing links, different blocking effects can be observed. The first is due to the fact that only one cell per cycle can be reemitted on every output link. If cells from two or more inputs compete for the same output, all but one are blocked.

Using FIFO buffers, blocked head-of-the-line cells can block other cells destined for a different, available output link. This effect can be avoided by random access memories (RAM)[5,6] which obviously require a more complex buffer control, because the correct sequence of cells destined for the same output has to be guaranteed.

If the input buffer is provided with multiple outputs (or if the buffer access time is reduced), several cells can be transferred to *different* outputs in the same clock cycle, while preserving the same speed on input and output links. The performance improvement due to these mechanisms is described in more detail in Reference 6.

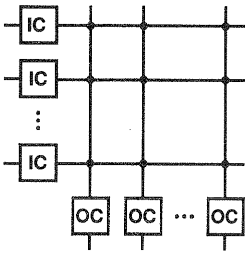


Figure 2. Matrix-type switching element

Another concept involving input buffers is called *input smoothing*,[2] where the cells arriving within a frame of b cycles at an input link are stored and offered to a matrix of size $Nb \times Nb$ at the end of the frame. Thus the speed of the matrix can be reduced by a factor of b, but internal cell losses will occur if more than b cells are destined for the same output within the duration of one frame. To keep the cell losses sufficiently low, very large frame sizes would be required. This implies a large number of buffers at the inputs and a high mean delay due to the frame assembly/disassembly and the internal speed reduction.[2]

Output buffers. In this type of switching element the buffers are allocated to the output controllers. Only the unavoidable blocking effects introduced by the contention for the output links occur.[3] Consequently, this is the optimum buffer placement, but it requires a reduced buffer access time and an internal speed which is N times the link speed for a $N \times N$ switching element. This may lead to severe technological limitations for the size of these elements.

Combined input/output buffers. The internal speed-up factor can be kept below the maximum of N for an $N \times N$ element without introducing extensive blocking effects. In this case, additional buffers at the inputs are necessary to avoid cell loss due to internal blocking.

Crosspoint buffers. The buffers can also be located at the individual crosspoints of the matrix. This buffering scheme is avoiding that cells for different outputs influence each other, and all cells arriving at the inputs can in principle be transferred to their target buffers within one clock cycle. A disadvantage from the performance point of view is that there are many small buffers which are dedicated to one input/output pair and no buffer sharing is possible. Therefore, buffers cannot be used as efficiently as in the output buffer case.

Performance comparisons for the different buffering strategies have been presented in various publications.[2–4,7,8] The main characteristics of these strategies for random input traffic are shown in Figure 3 using a 16 × 16 switching element with input/output buffers and different speed-up factors. The curve for a speed-up factor of 1 corresponds to the input buffer case, where the maximum throughput of the element is limited to about 58 per cent.[3] For a speed-up factor of 16, the behaviour is equivalent to the behaviour of an output buffer element, where the only limiting factor is the speed of the output link and the maximum throughput approaches 100 per cent. A comparable throughput can be achieved with crosspoint buffered elements. It is interesting to note that the ideal throughput curve is already reached with a speed-up factor of

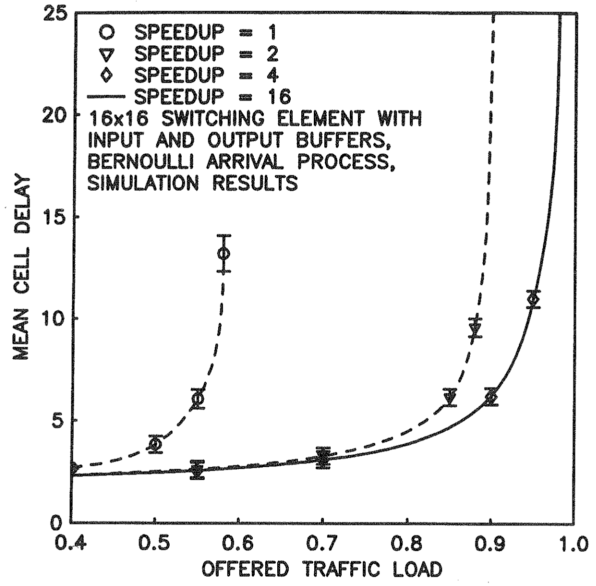


Figure 3. Influence of the speed-up factor on the maximum throughput

4, which may result in a significant simplification of the implementation.

Arbitration strategies. If several cells compete for the same output and not all of them can be transferred in one clock cycle, an arbitration has to be performed. In addition to fairness considerations, objectives for the design of the arbitration algorithm can be the minimization of the delay variations or the minimization of the cell loss.

A comparison of some frequently discussed strategies is shown in Figure 4 using a switching element with input buffers. The arrival processes at the inputs are assumed to be Bernoulli processes with identical rates, and the cells are destined to each output with the same probability. A random selection of cells can be implemented with minimum overhead, but the tail probabilities of the delay distribution indicate that this strategy introduces the highest delay variations.[9] A cyclic strategy also requires very little overhead, but the performance improvement is not significant. The optimum strategy with respect to the delay variation is a global FIFO strategy taking into account all the buffers that feed one output link (delay-dependent strategy). This implies some overhead to remember the relative order of arrival for the competing cells. To minimize cell loss, the best strategy is to choose the cell from the longest queue in the switching element (state-dependent strategy). For this algorithm, the actual queue lengths have to be evaluated and compared in case of a collision, which may be simpler to implement than the global FIFO strategy. The performance of this strategy is slightly worse with respect to the delay requirements, but is still acceptable.

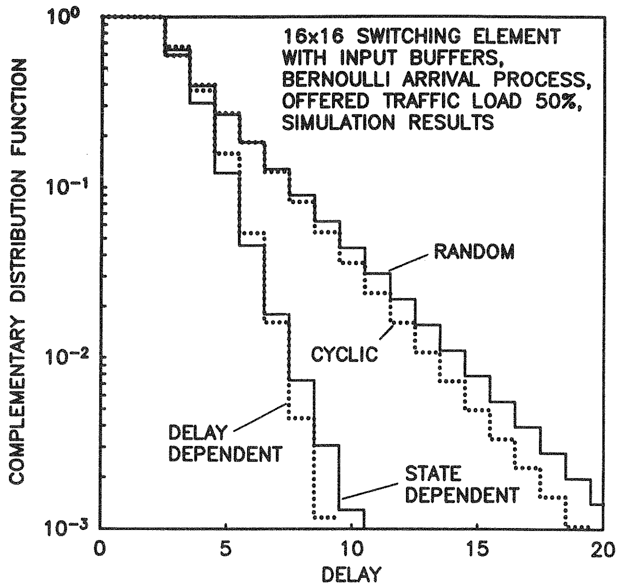


Figure 4. Influence of the arbitration strategies on the delay

2.2. *Central memory switching elements*

The input and output controllers of a central memory switching element are connected via a common memory which can be written by all input controllers and read by all ouput controllers (Figure 5). The common memory can be organized to provide logical input buffers as well as logical output buffers, but as discussed in the previous section logical output buffers will be preferred. A switching element with common memory has been implemented for the PRELUDE experiment which is described in Reference 10.

Since all buffers of the switching element share the same common memory space, a significant reduction of the total memory requirements can be achieved in comparison to physically separated buffers. On the other hand, the speed of the memory access is very critical in a central memory switching element, since the transmission speed of the ATM links is in the order of 600 Mb/s. An extremely high degree of internal parallelization (processing of up to all bits of a cell in parallel) is required in order to reduce the access time of the common memory. Consequently, the optimization of the memory requirements can only be achieved at the expense of an increased circuit complexity.

2.3. *Bus-type switching elements*

A high-speed time division multiplexing bus (TDM bus) is used to connect the input and output

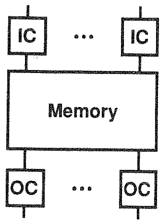


Figure 5. Central memory switching element

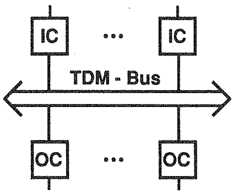


Figure 6. Bus-type switching element

controllers of a bus-type switching element (Figure 6). The capacity of this bus must be at least the sum of the capacities of all input links in order to guarantee a conflict-free transmission of cells.[11] Again, this capacity can only be achieved by a sufficient degree of internal parallelization (e.g. 16 or 32 bits in parallel).

The access to the bus is usually controlled by a time-slot allocation scheme which assigns a fixed time-slot to every input controller. Thus each input controller is able to access the bus and to transmit one cell before the arrival of the next cell is completed. Consequently, there are no collisions at the input controllers, but several cells may be entering the same output controller, whereas only one cell can be transmitted on the output link. Therefore, buffers must be provided at the output controller to store cells contending for the same output link yielding the same performance as output buffered matrix-type switching elements.

2.4. *Ring-type switching elements*

The input and output controllers of a switching element can also be attached to a ring interconnection network (Figure 7) which should be operated in a slotted fashion to minimize the overhead of the bandwidth allocation scheme. If the internal capacity allows a fixed allocation of time-slots to the input controllers, the performance of the ring is similar to the performance of a bus-type switching element. However, if the capacity of the ring is less than the total capacity of all input links, a dynamic time-slot allocation scheme is required introducing an additional overhead.

The main advantage of ring interconnection networks versus bus-type switching elements is the possibility of using a time-slot several times within one rotation. This requires the output controllers to empty a received time-slot so that it can be used immediately by one of the following controllers attached to the ring. Thus, depending on the internal traffic flow an effective utilization of more than 100 per cent can be achieved. This gain of bandwidth has to be compared with the overhead which is necessary to manage the access to the time-slots circulating on the ring.

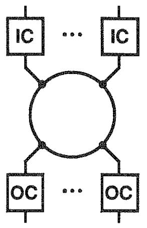


Figure 7. Ring-type switching element

A first approach to the implementation of a ring-based architecture is the ORWELL ring.[12] In this prototype the designers had to use several physical rings in parallel, forming a so-called torus of rings in order to fulfil the throughput requirements.

3. SWITCHING NETWORK ARCHITECTURES

In this section, a classification of switching networks will be given. Such a network is used within a switching node for establishing a connection between an arbitrary pair of inputs and outputs. The simplest switching network consists of a single switching element as discussed in detail in Section 2. Since a single switching element is not sufficient to satisfy the requirements of a normal size ATM switching node, it is necessary to use larger switch fabrics which are built from single switching elements. An overview of the switching networks presented in this section is given in Figure 8.

Switching networks can be subdivided into two topological classes:[35]

(i) single-stage networks
(ii) multi-stage networks

3.1. *Single-stage networks*

A single-stage network is characterized by a single stage of switching elements which are connected to the inputs and outputs of the network by a specific connection pattern. A well-known representative of these networks is the *shuffle exchange* network,[13] which is based on a perfect shuffle permutation cascaded to a stage of switching elements. This network is also called a *recirculating* network, because it may be necessary to pass the network several times before reaching the proper destination. Since the performance of single-stage networks is not very good, usually several stages are cascaded forming the so called multi-stage networks as discussed in the next section.

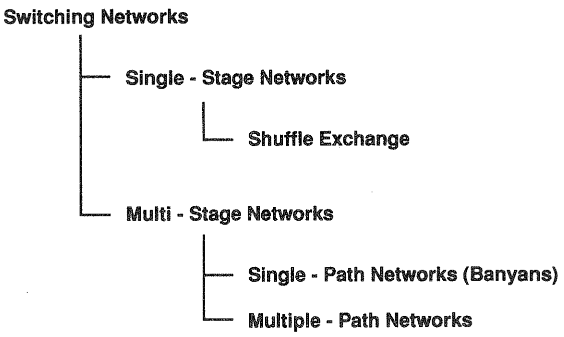


Figure 8. Switching network family tree

3.2. *Multi-stage networks*

Multi-stage networks are built of several stages which are interconnected by certain link patterns. Compared with the single-stage network no feedback is necessary so that a higher throughput can be achieved. According to the number of possible paths for reaching a dedicated output from one input, multi-stage networks can be divided into two groups, namely *single-path* and *multiple-path* networks.[14]

3.2.1. *Single-path networks.* Multi-stage networks where only one path exists for reaching the destination output from a given port are called *single-path* networks or *banyan* networks.[15] Owing to the property that an internal link can be used for simultaneous connections from different inputs blocking can occur. Single-path networks have the advantage that routeing is very simple because there exists only one path reaching the destination output. Banyan networks can be structured into several subgroups:

In (*L*)-*level* banyans only neighbouring stages are connected by links, so that each path passes exactly *L* stages. This class is subdivided into *regular banyans* and *irregular banyans*. In regular banyans only one type of switching element is used, whereas the irregular banyans are constructed from different types of basic elements. The *generalized delta* networks [16] are examples of irregular banyans.

For reasons of economical implementation regular banyans will be preferred because they are composed of identical elements. Regular banyans are split into two subclasses called *CC-banyans* and *SW-banyans*. In this paper, we will only focus on the SW-banyans because most of the existing implementations belong to this type.

SW-banyans can be constructed recursively from switching elements with *F* inputs and *S* outputs. This basic switching element can be considered as a (1)-level SW-banyan. An (*L*)-level SW-banyan is obtained by the connection of several ($L - 1$)-level SW-banyans with an additional stage of basic elements. These extra switching elements must be connected in a regular manner to the SW-banyans.

A special implementation of the SW-banyans are the *delta* networks.[17] Their definition includes the application of the so called *self-routeing scheme*. The transport of cells through this network is controlled by a destination address which is unique for every output. The destination address is a number of base *S* with *L* digits. Each digit specifies the destination output of the switching element in a specific stage. This simple routeing scheme has the advantage that it can be implemented in hardware and that central control is only involved during the set-up of a virtual connection. The switching functions are distributed over all switching elements and are done in parallel.

If the switching elements have the same number of inputs and outputs, the delta network is called *rectangular delta* network. Consequently, the number of network inputs is equal to the number of network outputs. *Bidelta* networks are delta networks with a special topological structure. They remain delta networks even if the network inputs are interpreted as outputs and vice versa.

To avoid any cell loss within the switching network, a back-pressure mechanism can be implemented informing the elements in stage *n* if a buffer in stage $n + 1$ is not able to accept cells in the current clock cycle. This blocking mechanism propagates backwards through the network and cell loss may only occur in the buffers in front of the switch fabric.

A banyan network with a given total number of inputs and outputs can be constructed from switching elements ranging in size from 2 × 2 up to a limit that is given by technological constraints. The influence of the switching element size on the maximum throughput of a 64 × 64 banyan network with input buffer switching elements and back-pressure mechanism is shown in Figure 9 for different buffer sizes within the switching elements.[9] For sufficiently large buffers, the smallest switching elements yield the highest maximum throughput, because the maximum throughput of input buffered switching elements is decreasing with the number of input ports. However, the size of output buffered switching elements is usually chosen as big as technologically possible (16 × 16 or 32 × 32) considering the total transfer time through the network and taking implementation aspects into account. This is not in contradiction with the results presented above, since the throughput of output buffer switching elements is independent of the number of input ports if the buffers are dimensioned large enough.

The influence of the size of the switch fabric on the performance is depicted in Figure 10. Owing to

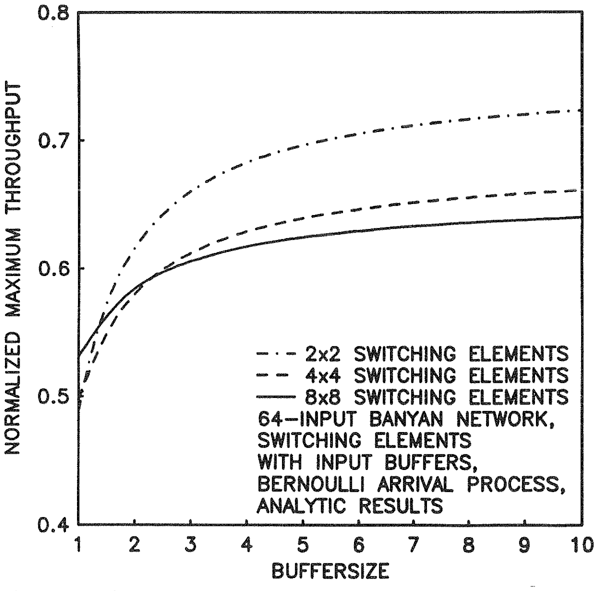


Figure 9. Influence of the switching element size on the maximum throughput

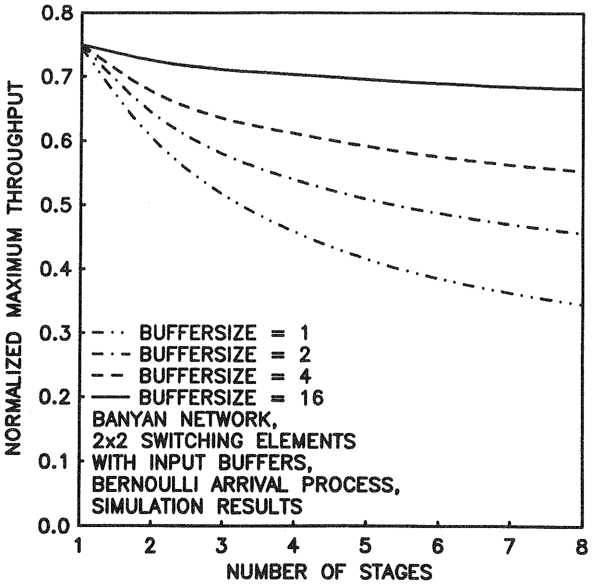


Figure 10. Influence of the switching network size on the maximum throughput

the back-pressure mechanism every additional stage of the network introduces further blocking effects so that the maximum throughput decreases if the size of the switch fabric is extended. On the other hand, without the back-pressure mechanism cell loss within the switch fabric would instead limit the throughput. These effects can only be reduced by providing adequate buffer sizes within the elements as shown in Figure 10.

Usually the cells need at least one clock cycle to cross a stage, because they are completely buffered in each switching element. For low-load situations the cell delay can be significantly reduced by a *cut through* mechanism[5] allowing a buffer to be bypassed if no collision occurs.

Owing to the cell loss requirements, the switch fabrics have to be operated well below their maximum throughput. This may result in the need to provide several parallel switching planes or other means to reduce the internal load in order to achieve a given total throughput.[18]

3.2.2. *Multiple-path networks.* The characteristic feature of *multiple-path networks* is the multiplicity of alternative paths for interconnecting a pair of inputs and outputs. Taking advantage of this property, internal blocking can be reduced or even avoided. Moreover, multiple-path networks are in general superior to single-path networks with respect to their increased fault tolerance.[19]

There are two basic alternatives to route the cells of one virtual connection within a multiple-path network. First, all cells of the connection can be routed completely independent of each other on any path leading to the desired destination. In this case, however, a resequencing mechanism is required if the correct sequence of cells belonging to one connection must be preserved by the switching node (as recommended by CCITT[1]). On the other hand, all cells of a connection can be routed on the same path within the switch fabric which is determined at call set-up. Consequently, the cell sequence is always guaranteed as long as FIFO buffers are used, but an intelligent routeing algorithm must be introduced in order to achieve an optimum utilization of all internal paths.

A straightforward approach to realizing a multiple path network is the extension of a single path network by appending additional stages with specific functions. The switch fabric described in References 5 and 20 mainly consists of a banyan network preceded by a so-called distribution network which has to distribute the arriving cells as evenly as possible over all inputs of the banyan network. Thus, the banyan network always operates under uniform traffic conditions, independent of any external traffic flow (e.g. communities of interest). The performance of this switch fabric is reported in Reference 5.

A second possibility is to use a sorting network[21] in front of the banyan network in order to arrange the cells in a monotonous sequence depending on their destination addresses. In this case, the banyan network is internally non-blocking if there is at most one cell destined for the same output link.[22,23] Therefore, the performance of this so-called Batcher–banyan network is similar to the performance of a matrix-type switch which uses input buffers to store the cells contending for the same output port.

In general, a banyan network can also be augmented with any number of additional stages which perform the same switching functions as in the original network. The *Benes* network[24] and the multipath interconnection network described in Reference 25 are constructed according to this methodology. In these networks the path of a virtual connection has to be determined at call set-up, because the complete routeing information for all stages must be appended to the cells and it is impossible to make a routeing decision for each arriving cell. A number of routeing algorithms are discussed in Reference 25, and their influence on the performance of the network is also studied.

Finally, a number of topologies with an inherent multiple path property exist which are not based on single-path networks. In principle, all topologies which are known from classical circuit-switching applications are also suitable for ATM switch fabrics, and as a first approach several folded network structures have been investigated in Reference 7. Other network topologies are the *Clos* networks,[26] the *Modified Omega* networks[27] and the PM2I networks[28] which have been previously used in various applications. However, detailed studies are still required to assess the performance characteristics of these networks in an ATM environment.

4. TRAFFIC ASSUMPTIONS FOR THE PERFORMANCE EVALUATION OF SWITCHING NETWORKS

The traffic streams that will be offered to ATM switching networks result from a superposition of single streams produced by a large number of virtual connections. These virtual connections will carry a large variety of widely unknown services differing significantly with respect to the holding time, the mean bit rate, the burstiness, the correlation structure and other characteristics. The enormous differences in the time scale between the cell level and the connection level together with the fact that the performance measures of interest include cell loss probabilities in the range of 10^{-10} make it impossible to use standard event-by-event simulation techniques for dimensioning purposes. Analysis methods, on the other hand, are restricted to comparatively simple models for systems and traffic streams. Most of the known analysis approaches[8,9,17,18,29–33] for switching networks assume the input traffic to be a memoryless Bernoulli process. In the following, we will investigate the sensitivity of these results to the coefficient of variation and the burstiness of the individual sources. Moreover, we consider the influence of the virtual connection concept on the performance characteristics of the cell level. For our investigations, we assume the virtual connections to last for the whole simulated period. We will use the burst–silence source described below to vary the different parameters. The model is in principle known from previous papers (e.g. Reference 34), where the parameters have been chosen to model a packetized voice connection.

4.1. *The burst–silence source model*

The burst–silence source can be characterized as shown in Figure 11. If the source is active on the connection level, it alternates between bursts of activity and silence phases on the cell level. During the bursts, cells are emitted with constant inter-cell distance T. A burst consists of a number of X cells, where X is a geometrically distributed random variable with mean EX. There is at least one cell in every burst. The silence phases are assumed to be negative-exponentially distributed with mean ES.

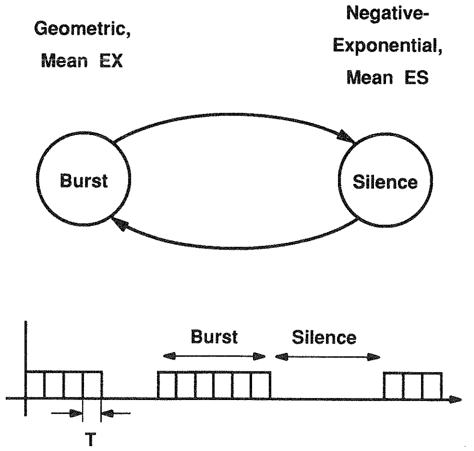


Figure 11. Burst–silence source model

The time between two cell arrivals from one source can be characterized by the mean EA, the coefficient of variation c and the burstiness B:

$$EA = T + \frac{ES}{EX},$$

$$c = \frac{\mathrm{ES}}{\mathrm{EA}}\sqrt{\left[1 - \left(\frac{EX - 1}{EX}\right)^2\right]}, B = \frac{EA}{T}$$

The burstiness gives the ratio of the maximum bit rate and the mean bit rate of the source. With this model, it is possible to keep two of the above characteristics constant while varying the third, as shown in Table I.

4.2. *Results*

The model under consideration consists of a 16 × 16 output buffer switching element with an additional buffer at each input to synchronize and multiplex the asynchronous sources feeding the link. All buffers have been dimensioned to avoid cell loss and the load is distributed symmetrically within the system.

In the following, it is assumed that 80 burst–silence sources generate an average load of 80 per cent on an input link of the switching element, which corresponds to a 1·5 Mb/s connection for a link speed of 150 Mb/s.

Figure 12 shows the effect of the source parameters on the mean delay (including transmission time) in the multiplexer queue. It is obvious that even for the superposition of 80 sources the characteristics of a single source have a significant influence on the result. The Poisson approximation, which yields a mean delay of 2·5 clock cycles, as well as a two-moment approximation using a hyper-exponential distribution (H_2) for the single sources considerably underestimate the mean delay compared with the superposition of the burst–silence sources. It is therefore not possible to use these approximations

Table I. Parameters for the burst–silence source

Type	*ES*	*EX*	*T*	*EA*	*c*	*B*
1	1000	21·71	53·94	100	3·00	1·85
2	600	7·46	19·62	100	3·00	5·09
3	550	6·18	10·98	100	3·00	9·11
4	525	5·58	5·84	100	3·00	17·12
5	200	2·49	19·62	100	1·60	5·09
6	1000	12·44	19·62	100	3·93	5·09

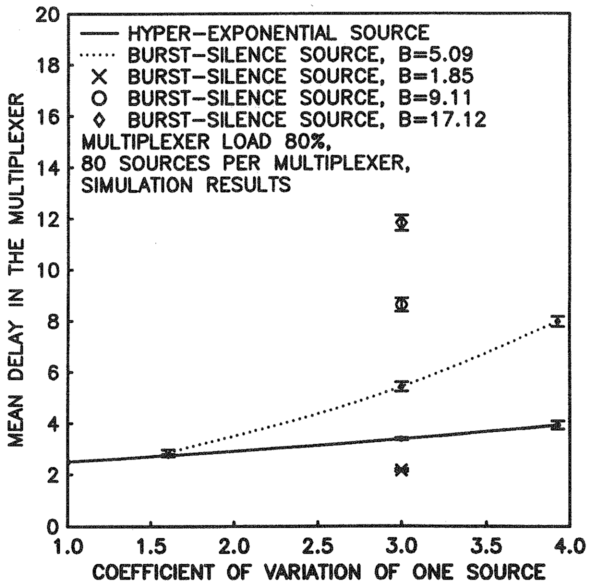


Figure 12. Influence of the source parameters on the multiplexer delay

for higher coefficients of variation and for higher burstiness values.

Another important point for the performance analysis is the consideration of the influence of the virtual connections (VCs) on the cell level. In most of the studies, the splitting of the traffic streams is modelled by a simple probabilistic decision. In reality all cells from one source have to be switched to the same output on the same path due to the virtual connection concept.

Figures 13 and 14 show the influence of the VC concept on the expected delay in the buffers of the switching element. Without the consideration of the VCs, neither the coefficient of variation nor the burstiness of a single source have an influence on the mean delay. This is mainly due to the fact, that the input process of the buffer in the switching element is formed from cells originating from all the 1280 sources so that the influence of the single source disappears.

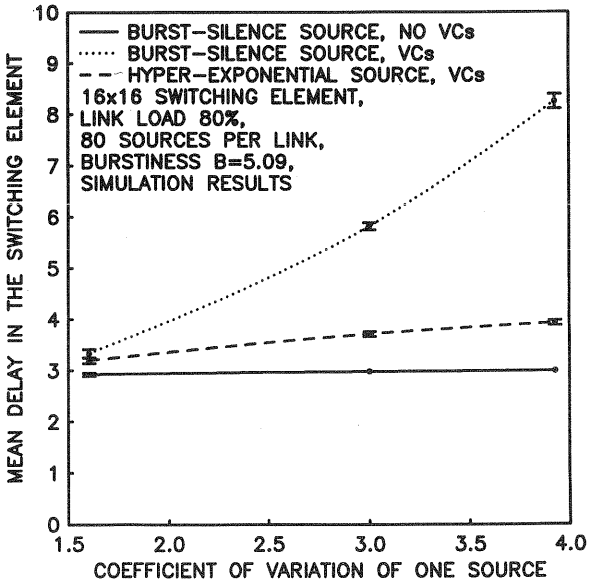


Figure 13. Influence of c on the delay within the switching element

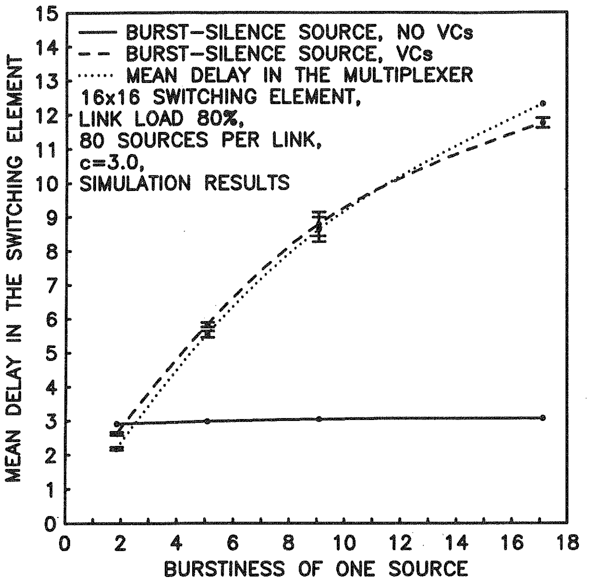


Figure 14. Influence of B on the delay within the switching element

Assuming that cells from five sources on every input link are switched to one output, we introduce the VC concept while keeping the symmetrical distribution of the load. Both parameters of the single source have a significant influence on the delay in this case, which shows a behaviour similar to the delay in the multiplexer. The multiplexer delay of only a few clock cycles is small compared to the inter-cell distance of the sources and therefore no significant smoothing effect can be observed. Taking into account the VC concept, it can be expected that the source characteristics are perceptible at every point in the network where the number of multiplexed connections is not too high.

5. CONCLUSION

In this paper we have discussed some generic architectures for ATM switching elements considering possible buffering schemes and arbitration strategies. A detailed performance study of matrix-type switching elements has shown that there is always a trade-off between the circuit complexity and the internal speed which may favour different architectures depending on technological limitations. In general, however, to obtain the best performance all blocking effects except for those introduced by the output links have to be eliminated, yielding structures which are logically equivalent to output buffer elements.

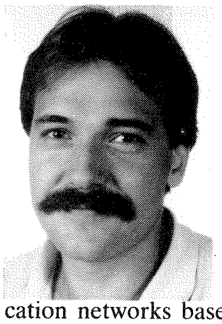
For a specific switching network, the so called Banyan network, we have evaluated the influence of various parameters on the maximum throughput. In the case of input buffer elements the highest throughput can be achieved using two-input switching elements, but for output buffer elements the limiting throughput is almost independent of the size of the switching elements. Therefore, taking into account the total transfer time through the network and some economic aspects, the switching elements are usually chosen as big as technologically possible.

Finally, we have extended our performance studies by taking into account some more realistic assumptions about the traffic streams carried by an ATM network. We have shown that the concept of virtual connections cannot be neglected in the performance evaluation of ATM switches, and we have demonstrated the influence of source characteristics such as burstiness and coefficient of variation. However, new analytical approaches as well as simulation techniques will be required in order to allow a realistic dimensioning of the network resources with reasonable effort.

REFERENCES

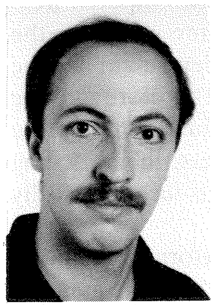
1. CCITT, Draft Recommendation I.121: 'On the broadband aspects of ISDN', *COM XVIII-R55(C)*, Seoul, February 1988.
2. M. G. Hluchyi and M. J. Karol, 'Queueing in space-division packet switching', *Proc. INFOCOM'88*, New Orleans, 1988, pp. 334–343.
3. M. J. Karol, M. Hluchyi and S. P. Morgan, 'Input versus output queueing on a space-division packet switch', *IEEE Trans. Communications,* **COM-35**, (12), 1347–1356 (1987).
4. E. P. Rathgeb, T. H. Theimer and M. N. Huber, 'Buffering concepts for ATM switching networks', *Proc. GLOBECOM'88*, Hollywood, 1988, pp. 1277–1281.
5. R. G. Bubenik and J. S. Turner, 'Performance of a broadcast packet switch', *IEEE Trans. Communications,* **COM-37**, (1), 60–69 (1989).
6. K. Rothermel and D. Seeger, 'Traffic studies of switching networks for asynchronous transfer mode (ATM)', *Proc. 12th International Teletraffic Congress (ITC)*, Torino, 1988, paper 1.3A.5.
7. K. A. Lutz, 'Considerations on ATM switching techniques', *Int. J. Digital & Analog Cabled Systems,* **1**, (4), 237–243 (1988).
8. Y. Oie, M. Murata, K. Kubota and H. Miyahara, 'Effect of speedup in nonblocking packet switch', *Proc. International Conference on Communications (ICC)*, Boston, 1989, pp. 410–414.
9. M. N. Huber, E. P. Rathgeb and T. H. Theimer, 'Banyan-networks in an ATM-environment', *Proc. International Conference on Computer Communication*, Tel Aviv, 1988, pp. 167–174.
10. P. Gonet, 'Fast packet approach to integrated broadband networks', *Computer Communications,* **9**, (6), 293–298 (1986).
11. M. De Prycker and M. De Somer, 'Performance of a service independent switching network with distributed control', *IEEE J. Selected Areas in Communications,* **SAC-5**, (8), 1293–1302 (1987).
12. J. L. Adams, 'The Orwell torus communication switch' *Proc. GSLB Seminar on Broadband Switching*, Albufeira, 1987, pp. 215–224.
13. P.-Y. Chen, D. H. Lawrie, P.-C. Yew and D. A. Padua, 'Interconnection networks using shuffles', *IEEE Computer,* **14**, (12), 55–63 (1981).
14. R. J. McMillen, 'A survey of interconnection networks', *Proc. GLOBECOM'84*, Atlanta, 1984, pp. 105–113.
15. L. R. Goke and G. J. Lipovski, 'Banyan networks for partitioning multiprocessor systems', *First Annual Symposium on Computer Architecture*, 1973, pp. 21–28.
16. D. M. Dias and M. Kumar, 'Packet switching in *n* log *n* multistage networks', *Proc. GLOBECOM'84*, Atlanta, 1984, pp. 114–120.
17. J. H. Patel, 'Performance of processor-memory interconnections for multiprocessors', *IEEE Trans. Computers,* **C-30**, (10), 771–780 (1981).
18. C. P. Kruskal and M. Snir, 'The performance of multistage interconnection networks for multiprocessors', *IEEE Trans. Computers,* **C-32**, (12), 1091–1098 (1983).
19. G. B. Adams III, D. P. Agrawal and H. J. Siegel, 'A survey and comparison of fault-tolerant multistage interconnection networks', *IEEE Computer,* **20**, (6), 14–27 (1987).
20. J. S. Turner, 'Design of a broadcast packet switching network', *IEEE Trans. Communications,* **COM-36**, (6), 734–743 (1988).
21. K. E. Batcher, 'Sorting networks and their applications', *Proc. AFIPS 1968 SJCC*, Vol. 32, AFIPS Press, Arlington, Va., pp. 307–314.
22. A. Huang and S. Knauer, 'STARLITE: a wideband digital switch', *Proc. GLOBECOM'84*, Atlanta, 1984, pp. 121–125.
23. J. Y. Hui and E. Arthurs, 'A broadband packet switch for integrated transport', *IEEE J. Selected Areas in Communications,* **SAC-5**, (8), 1264–1273 (1987).
24. V. Benes, *Mathematical Theory of Connecting Networks*, Academic Press, New York, 1965.
25. G. J. Anido and A. W. Seeto, 'Multipath interconnection: a technique for reducing congestion within fast packet switching fabrics', *IEEE J. Selected Areas in Communications,* **SAC-6**, (9), 1480–1488 (1988).
26. C. Clos, 'A study of non-blocking switching networks', *Bell System Technical Journal,* **32**, (2), 406–424 (1953).
27. K. Padmanabhan and D. H. Lawrie, 'A class of redundant path multistage interconnection networks', *IEEE Trans. Computers,* **C-32**, (12) 1099–1108 (1983).
28. T. Feng, 'Data manipulating functions in parallel processors and their implementations', *IEEE Trans. Computers,* **C-23**, (3), 309–318 (1974).
29. D. M. Dias and J. R. Jump, 'Analysis and simulation of buffered delta networks', *IEEE Trans. Computers,* **C-30**, (4), 273–282 (1981).
30. Y.-C. Jenq, 'Performance analysis of a packet switch based on single-buffered banyan network', *IEEE J. Selected Areas in Communications,* **SAC-1**, (6), 1014–1021 (1983).
31. C. P. Kruskal, M. Snir and A. Weiss, 'The distribution of waiting times in clocked multistage interconnection networks', *IEEE Trans. Computers,* **C-37**, (11), 1337–1352 (1988).
32. T. H. Theimer, E. P. Rathgeb and M. N. Huber, 'Performance analysis of buffered banyan networks', *Proc. International Seminar on Performance of Distributed and Parallel Systems,* Kyoto, 1988, pp. 57–72.
33. L. T. Wu, 'Mixing traffic in a buffered banyan network', *Proc. Ninth Data Communications Symposium, ACM SigComm Computer Communication Review,* **15**, (4), 134–139 (1985).
34. H. Heffes and D. M. Lucantoni, 'A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance', *IEEE J. Selected Areas in Communications,* **SAC-4**, (6), 856–868 (1986).
35. T. Feng, 'A survey of interconnection networks', *IEEE Computer,* **14**, (12), 12–27 (1981).

Authors' biographies:



Erwin P. Rathgeb received the Dipl.-Ing. degree in electrical engineering from the University of Stuttgart, West Germany, in 1985. Since then he has been a member of the scientific staff at the Institute of Communications Switching and Data Technics at the Unversity of Stuttgart. Currently he is involved in research projects in the field of integrated broadband communication networks based on ATM.

Thomas H. Theimer received the Dipl.-Ing. degree in electrical engineering from the University of Stuttgart, West Germany, in 1987. Since then he has been a member of the scientific staff at the Institute of Communications Switching and Data Technics at the University of Stuttgart. Currently he is involved in research projects in the field of integrated broadband communication networks based on ATM. He is a member of IEEE.



Manfred N. Huber received the Dipl.-Ing. degree in electrical engineering from the University of Stuttgart, West Germany, in 1984. Since then he has been a member of the scientific staff at the Institute of Communications Switching and Data Technics at the University of Stuttgart. Currently he works in the field of integrated inhouse communication and broadband communication. He is a member of IEEE and ITG.