Design and Evaluation of an Input Buffered Packet Switch

Azeddine Bilami[1], Mustapha Lalam[2], Mehammed Daoui[2], and Mohamed Benmohammed[3]
[1] Department of Computing Science, University of Batna, Algeria
[2] Department of Computer Science, University of Tizi Ouzou, Algeria
[3] Department of Computer Science, University of Constantine, Algeria

Abstract: *Many architectures of internet routers, ATM and ethernet switches have been proposed and analysed in literature. Theoretically reliable and valid solutions have been developed to achieve high performances but a lot of them are not feasible in practice for commercial and technological reasons. Few papers develop the implementation and simulation aspects. The objective of this paper is the design of a packet switch with a minimum cost and hardware complexity. We propose an input-queuing architecture using a multistage interconnection network and a simple cell selection policy implemented by hardware. The switch is described and simulated using a VHDL language. Performances in terms of throughput and cell loss are evaluated.*

Keywords: *Routing, switch, multistage interconnection network, Benes network, self routing, VHDL.*

Received June 23, 2004; accepted October 30, 2004

1. Introduction

A major challenge concerned to high-speed switching is related today to switch design that requires the best possible compromise between ease of implementation and goodness of performances.

Switch functionality is twofold:

- Managing packet buffering while selecting packets to transmit each time to avoid contentions and cell loss.
- Routing packets from their incoming ports to their destination ports.

To avoid contentions and cell loss, the incoming packets are stored in buffers. These buffers can be in inputs, in outputs, in inputs and in outputs or shared by inputs and outputs. So, a choice that a designer may have is where to place the buffers. Although, a lot of existing switches use the shared buffers technique, it has been shown through several publications that the method using input buffers is the only one which can constantly answer to the increasing needs of large switches and to the high rates of present and future communication lines. With such solution, the cost of the switch remains acceptable and the management of the queues in the buffers less complex. Nevertheless, it introduces the well known problem of the Head Of Line (HOL) blocking which is usually solved by means of scheduling algorithms. Appropriate scheduling algorithms are adopted depending on the application environment and some constraints such as line rate, number of ports to connect, cost, expected performances.

To route packets from input ports to output ports, Multistage Interconnection Networks (MINs) are very attractive. They can route in parallel the incoming packets. They have a relatively low cost, and are better adapted for VLSI implementation. MINs have been used initially in multiprocessors architectures to connect the processors to memory banks. Recently, and regarding to their characteristics, another interest is granted to these networks: They are used in the Internet routers, in the ATM and Ethernet switches and are appropriate to be used to construct electro-optic switches.

Using Benes network which presents the best cost among non blocking MINs, and input buffers technique with a simple selection policy of maximum cells to transmit without conflicts in a cycle, we propose in this paper a switch with a low cost introducing a minimum hardware complexity.

This paper is organised as following. In the next section, we present a self routing algorithm for Benes network. In section 3, we describe and compare different buffering techniques. The fourth section discusses performances of a variety of scheduling algorithms presented in literature. In section 5, we describe the selection procedure adopted in this paper and explain how it operates through an example. In section 6, simulation results are presented and compared to those of other solutions. Finally, we give some concluding remarks in section 7.

It may be useful to define the following terms that are used frequently in this paper.

Permutation: Is a one-to-one I/O mapping, where all inputs and outputs are active. It is called a partial permutation, if any I/O are not active.

Cell: We consider packets of a fixed size. We prefer using in the continuation of this article, the term 'cell' instead of 'packet'.

Throughput: Is the number of cells arriving to their destinations divided by the total number of departed cells from their sources in a unit time (a cycle).

2. Interconnection Network

2.1. Benes Network

A (NxN) Benes network (Figure 1) is a network with N inputs and N outputs. Its dimension is r = $\log_2 N$. It is composed of 2 ($\log_2 N$) - 1 stages of 2^{r-1} switching elements (SE for short) and presents with Waksman network, the best cost among all the non blocking multistage networks [1]:

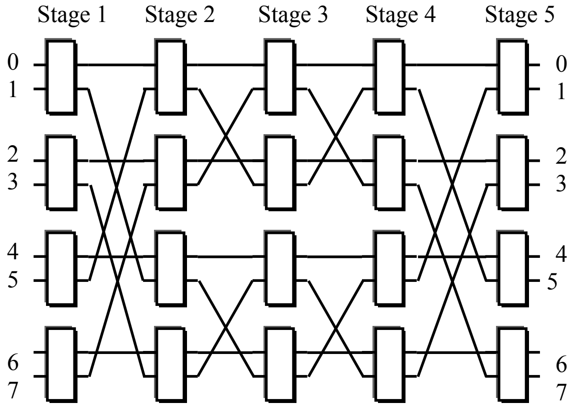


Figure 1. A Benes network for N = 8.

Benes network is dynamic and rearrangeable. It presents N/2 possible paths to establish a link between a free input and a free output. It is for this fact fault tolerant since we can always establish a path even if some switches are out of use. It offers a constant latency for all couples (input,output). The Only drawback that is known about it, is the complexity of its routing control algorithms. The solution usually used is centralized: It needs a global controller, which configures the network before the transfers. This solution requires O (N . r) sequential time to position all the SE(s).

2.2. Self Routing

One of the main features that this network has is the self-routing property: Every 2x2 switch can decide to which of its outputs the incoming cell will be transmitted, depending only on the cell destination address. This means that the implementation of a complex routing algorithm, either centralized or distributed, is not needed.

It has been proved that some permutations families representing a subset of all possible permutations (N! for a NxN Benes network), can be routed automatically in only one pass, consuming O ($\log_2 N$) in parallel time. Lenfant in [17], has defined the Frequently Used Bijections (FUB) family. Nassimi and Sahni have proposed in [23], an algorithm for the Bit Permute Complement (BPC) class. Boppana and Raghavendra [3] have resolved the problem for the Linear Complement (LC) family. These families are regrouped in the same class Bit Permut CLosure (BPCL) of self routing permutations.

Four strategies for the BPCL self routing permutations have been defined in [8]: TCR, BCR, LCR, HCR. In each of these strategies a particular bit of the destination address called the Routing Bit (R-Bit), determines the action to perform at the SE level. The switching command can be in this case, implemented locally at the 2x2 switch.

- Top Control Routing (TCR): R-Input is the superior input.
- Bottom Control Routing (BCR): R-Input is the lower input.
- Least Control Routing (LCR): R-Input is the smallest input.
- Highest Control Routing (HCR): R-Input is the biggest input.

The R-bit (routing bit) is defined:

$$R\text{-}Bi\ t = \begin{cases} X_i \ for\ 1 <= i <= r\text{-}1 \\ X_{2r-i} \ for\ r <= i <= 2r\text{-}1 \end{cases}$$

Where: $X_r, X_{r-1}, \ldots, X_1$ is the destination address of the R-Input

2.3. Routing Algorithm

The adopted routing algorithm in this paper, is one of the most effective self routing algorithms for the Benes network [25].

In a first time, the load is fairly distributed on the 2x2 switches of the r-1 first network stages using LCR self routing strategy: The SE(s) are positioned according to the R-Bit of the input with the smallest destination address. We note here, that HCR strategy can be used instead LCR for the r-1 first stages. The performances will be the same: In [8], it has been shown that $|HCR| = |LCR|$.

For the r remaining stages, the switches are positioned using the classic routing algorithm applied in the omega network: If the R-bit is a 0, than the cell will be routed to the upper output of the SE, else it will be routed to the lower output.

An example of routing using this algorithm is illustrated in section 5.3.

3. Cell Buffering

The buffering operation consists in queuing the cells to transmit. The performances of the switch can be affected differently according to the way that is done. Different strategies are used depending on the physical site of the queues: In inputs, in outputs, in both inputs and outputs, or in shared buffers.

3.1. Output Buffers

Although output buffers give the optimal delay-throughput performance, switches that use them are difficult to achieve. In this method, every output can receive simultaneously during the same cycle, N cells from the N inputs. Thus, the switch must be able to put in the same queue and during one cycle, the N cells destined to the same output. The operation of setting in queue must therefore operate N times quickly than the rate of cell arrivals (speed up). If this solution is feasible in case of small capacity switches, it should be not possible for switches of big capacities (the N-times speed-up in the switch limits the scalability of this architecture) [4, 7].

3.2. Shared Buffers

The approach using the shared buffers is a closer solution to the previous. Instead of associating a separated queue to every output, a common memory buffer is used by all outputs. Only one queue is used to hold all cells. This method is more economic in space. It has been implemented in a lot of switches: ATLANTA of Lucent Technologies, ATLAS [15], and adopted in a lot of papers [16, 19, 26].

The disadvantages of this solution are related to:

- The complexity of the logic to control such memory since the management of a single global queue is difficult.
- The limit of the bandwidth of Dynamic RAM (DRAM) that are commonly used in order to provide large buffer storage space [10].

3.3. Input Buffers

To design high-speed switches and routers, alternatives that present now a big interest are input buffers adopted in our switch, and Combined Input Output Queues (CIOQ), proposed in several recent publications [6, 7, 27]. These two techniques do not require some queues operating at multiple speed of the communication lines.

The most realistic solution is the one using input buffers, because it simplifies the implementation of big capacity switches. Input buffers switches require only O (N) buffers and O (N) controllers (a buffer is associated with each input for queuing the incoming cells). Also, this technique is able to overcome the technology limitation and complexity management of a single large multi queue memory. In this strategy, the memory can be accessed simultaneously by a read and a write operations. The memory must therefore, present an access time at least two times faster than the line rate (Speed up S = 2). This limitation can be a serious problem only in case of very high rates (several hundreds of Gb/s). With less speeds, such OC48c (2.5 Gb/s) or OC192c (10 Gb/s), many solutions can be proposed to surmount it:

- Using multi-ports memories where such operations can take place in parallel.
- Combined fast static SRAM with big capacity dynamic RAM [10].
- Doubling the memory bandwidth using some memory management techniques as: “Ping Pong” [11].
- Queuing in a parallel manner the different packets [18].

A major drawback of input buffers is related to queue managing while selecting cells to transmit at every cycle. The simplest way consists in storing the incoming cells in FIFO queues. The cells at the head of queues are the first served. This approach introduce the problem of HOL blocking (the cells in the queues can be blocked by the cells in conflicts at the heads of queues). It has been demonstrated in [12], that in this case only $2-\sqrt{2} \approx 58\%$ of the input cells are served.

Several solutions have been adopted to palliate to the problem of the HOL blocking:

- Operating the switch fabric at *S* times faster than the input lines (speedup). A speedup by a factor of S can remove *S* cells from each input port within each cycle. This scheme removes completely the HOL blocking effect. Unfortunately, and like output queuing, a speedup of $S = N$ in the switch, is not feasible actually for high values of *N*. For low values of *S*, HOL blocking phenomenal is only reduced.
- In [2], the authors propose a copy of the routing network for the control. Using the self routing ability of Banyan network in order to avoid the network set up operation before transfers (usually adopted in such networks). The control network is topologically similar to the routing network. But, it only routes the labels (destination addresses) and not the data during each cycle, to find the best matching to submit to the routing network during the next cycle. In the test phase, and in case of conflicts, a random selection of other cells in the queues that generate the conflicts is performed. This procedure is repeated until obtaining the maximum matching or until the end of cycle. The use of two networks, one for the control and the second for the routing is advantageous for its simplicity and allows a high throughput, but it is expensive to implement,

since it needs more SEs than other solutions based on non blocking multistage networks.

- Another approach is that which increases the number of queues while adopting the solution of virtual channels or virtual queues Virtual Output Queues (VOQ). In this scheme, *N* virtual output queues (corresponding to the *N* outputs) are associated to every queue in input. A cell that arrives at an input is queued depending on its destination in the appropriate virtual queue. At each cycle, only one cell is selected from every input (i. e., from all the associated VOQ). Some scheduling algorithms are required to find the optimal solution (Maximum Matching) of maximum inputs to transmit without conflicts. Some authors [13, 21] have oriented their researches toward this solution to achieve 100% of throughput. Nevertheless and for big values of *N*, it needs a large memory to implement the big number of buffers: O (N^2) queues for an NxN Benes network.
- A simple and economic solution but with less performances than the previous has been proposed in [14]. The number of queues at each input port is limited to two. A queue containing the cells destined to the outputs with even addresses and the second, the cells destined to the outputs with odd addresses (from where the name of the proposed switch: 'odd-even switch'). The selection is done in a first time from the even queues. In a second time, cells of the 'odd' queues, corresponding to the same inputs will replace cells that loose in the first round. Although, the performances offered by this switch are modest, this solution represents an improvement comparatively to a simple FIFO technique while offering a throughput of about 70% without introducing an important hardware complexity.
- Using Window Based Scheduling Algorithms [5, 22]. This solution improves the throughput while choosing in a window (a part of the different queues) the cells to transmit. The selected cells are not necessarily the ones in the heads of queues. When a cell in a head of queue is in conflict, the algorithm selects another one of the same queue depending on some criteria, in a window of W depth. During a cycle, a set of a maximum N cells (a cell from each queue) is selected to be transmitted.

4. Scheduling Algorithms

The research of the optimal solution in the queues can be translated to a research of the maximum matching in a bipartite graph. The algorithm establishes from the set *X* of input addresses and the set *Y* of destinations addresses a bipartite graph *G* in such a way that every edge of *G* joins a vertex in *X* to a vertex in *Y*. A sub set *M* of edges is called matching if there are not two edges of *M* that have a common vertex. An example of bipartite graph, correspondent to the treated example in this article is given in section 5.3.

To find the optimal solution some scheduling algorithms as Maximum Size Matching (MSM), and Maximum Weight Matching (MWM) [21] use different metrics such as the length of the queue (Longest Queue First (LQF)), the age of the cell in the queue (Oldest Cell First (OCF)). It has been proved that used in a combination with virtual output queuing (VOQ), these algorithms can achieve high throughput ≈ 100% in case of uniform traffic. However, this is only a theoretical result. They consume respectively O ($N^{5/2}$) and O ($N^3 \log_2 N$) and are too complex to be implemented in hardware, since they need big comparators to compare the different ages, lengths or weights of the queues.

Several other algorithms with less complexity have been proposed and implemented in hardware. Matrix Unit Cell Scheduler (MUCS) [9] finds the optimal solution by computing a traffic matrix. Iterative algorithms, such iSLIP use multiple iterations to converge on a maximal matching [20]. Round Robin Matching (RRM) algorithm [28] affects a rotating priority 'round robin' to the different queues to avoid that some of them could be not served during a long time (since their lengths or weights are less competitive) and converge toward the optimal solution on average in O ($\log_2 N$) iterations.

The optimal solution can be also obtained by the algorithms based on simple heuristics, but without guarantee of an optimal throughput. A simple example of these algorithms is the next one: At every beginning of a cycle, the cells at the head of queues are in competition to be transmitted, those that loose, let the possibility to the cells that follow and so forth on the whole width W of a window. This process can be interrupted before, if the optimal solution is reached. No metrics are used. This algorithm based on the simple heuristic consume O (N . W) in sequential time [22].

5. Proposed Switch

Our solution is based on the use of Input buffers to queue the incoming cells. The general structure of the proposed switch is shown in Figure 2.

5.1. Cell Buffers

We consider a switch operating at maximum rate of 3,2 Gb/s. The memory access time does not represent a major problem since the required memory access time in case of input buffers and single ported memory is estimated to L / (2 * R) = 10ns (with data word length L = 64 bits and line rate R = 3.2 Gb/s). Although, it corresponds to a faster access time than typical access time of DRAM memories (varying from 12 ns to 70 ns), with the later development of SDRAM, more larger memory bandwidth becomes possible without cost increasing. The recent Double Data Rate 266-DDR SDRAM with an 8 byte wide bus, offers a bandwidth of 266 MHz x 8 Bytes = 2.1GB/s. In case of dual ported memory (simultaneous read and write operations are possible), which is very attractive for input queuing systems, we can slow down the memory access rate by half and hence achieve more than the required bandwidth.

The logic to control the input queues is simple as only a read and write pointer need to be maintained for each queue.

5.2. Selection Process

The proposed solution is a deterministic window based algorithm. To determine the cells that can be self routed through the routing network, in one pass without conflicts, the algorithm searches a match in a window of W width. The algorithm describe in [22] establishes the optimal solution by consulting all the NxW cells. Such policies, based on the search of the maximum matching, are not appropriated for high values of N (since an exhaustive search is time consuming). We opt for a solution with less time complexity, while looking just for cells destined to outputs not addressed by cells in the HOL vector which contains the destination addresses of cells to transmit during every cycle.

The algorithm presents at the inputs of the routing network the permutation or the partial permutation, where the maximum of outputs are solicited, to be then routed automatically. This objective is reached using a counter for every output, that indicates the number of the output occurrences in the HOL vector.

In a first time, an output with an empty counter (= 0) does not appear in the HOL vector. A cell addressed to this output is then looked for, in the queues on the width W of the window. If found, it will replace a cell destined to an output that appears more than once in the HOL vector.

Our algorithm runs in O (W. (N^2 - N)) sequential time in worst case (all the N cells are addressed to the same output), and O (N) sequential time in best case (all the outputs are addressed in the HOL vector).

The selection process operates in three steps. The principle is the following:

1. The cells are directed (using AIG component) according to the highest weight bit of the destination address toward a buffer in high half (HM) if the bit is 0 (corresponds to outputs 0 to N/2-1), or toward a buffer in the low half (LM) if the bit is 1 (corresponds to outputs N/2 to N-1).
2. Compose the vector HOL1 (in parallel, HOL2) from the N cells in the head of HM queues (respectively from cells in the head of LM queues). The number of occurrences of every destination Nocc (output i) is counted. Verify that: $\forall$ i, 0 < i < N / 2 - 1, Si $\in$ HOL1 (in parallel, $\forall$ i, N / 2 – 1 < i < N - 1, Si $\in$ HOL2), that means there is no occurrence counter equal to 0 or all N/2 destinations exist in HOL1 (respectively in HOL2). If this condition is satisfied, pass to step 3. Otherwise for each output *i* with Nocc = 0, search in W for a cell destined to an output of which Nocc (i) = 0, if it exists it will take a place in HOL1 (respectively in HOL2), in replacement of a cell addressed to an output of which Nocc (i) > 1.
3. Iteratively, constitute the HOL vector to transmit while selecting cells from HOL1 and HOL2. At each iteration, HOL1 (i) or HOL2 (i) are in competition to take place in HOL vector. The cell wth the less counter Nocc in the HOL is selected, and its correspondent counter updated. If the two cells have been already selected, no cell will be transmitted from the input. A partial permutation is so obtained.

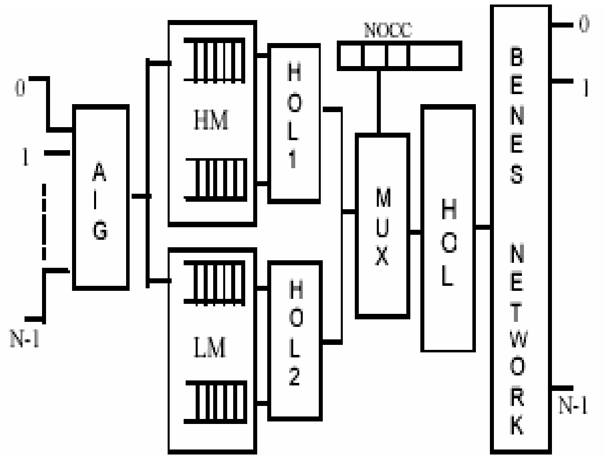
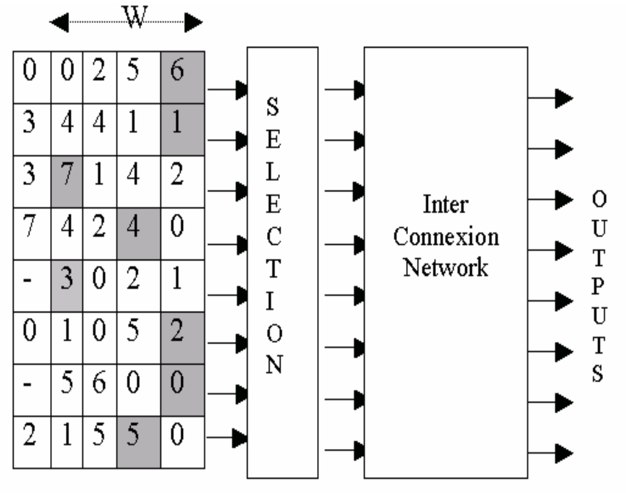


Figure 2. Architecture of the proposed switch.

5.3. Example

Our algorithm is illustrated through the following example:

In case of this example, the permutation selected corresponds to the optimal solution. Yet, its research was not the objective of the algorithm. We noted through several simulations that on 100 permutations generated randomly, the solution obtained coincides with the optimal solution (maximum matching) in more than 20%. Figure 4 shows the bipartite graph corresponding to the cells of the window (W = 4). The bold edges correspond to the graph of the maximum matching.



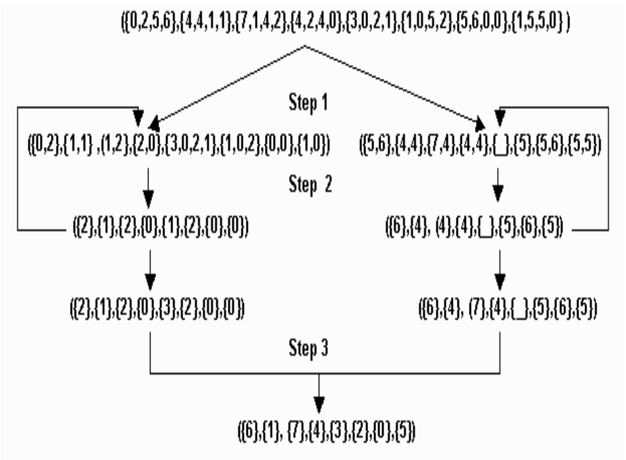


Figure 3. An example of cell selection.

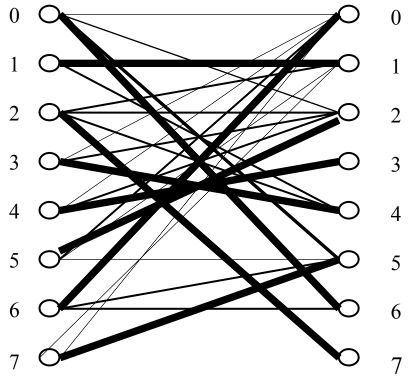


Figure 4. Bipartite graphe.

The permutation (6, 1, 7, 4, 3, 2, 0, 5) at the output of the selection circuit is submitted to the routing network. The permutation is routed using the algorithm defined in section 2.3.

6. Simulation and Evaluation

6.1. VHDL Simulation

The switch design is based on VHDL entries. We use the standard Very-high speed integrated circuits Hardware Description Language (VHDL) to describe and verify the good functioning of the proposed switch.

A recursive construction of Benes network is used. A 2x2 SE has been described according to the algorithm given in section 2.3; it performs differently depending on its physical emplacement, either in the first r-1 stages or in the r following stages.

After that,we have specified the selection module in the architecture level as discussed in the previous section. We verify the functionality of each component, then we grouped all the components to form a complete switch including selection module and routing module. A functional simulation has been performed to verify the behaviour of the proposed switch. Examples of complete functioning have been verified via VHDL simulation with uniform arrivals.

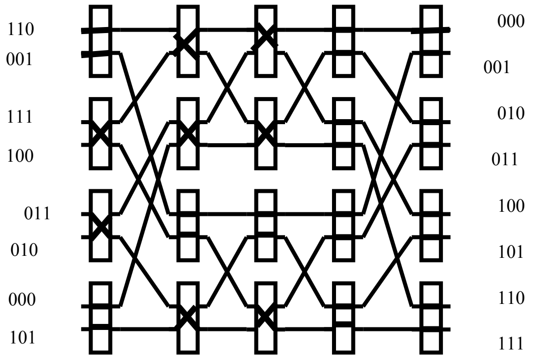


Figure 5. Routing of P: (6, 1, 7, 4, 3, 2, 0, 5).

6.2. Experimental Results

Extensive simulations have been conducted to evaluate the performance of the proposed switch. The evaluation is made for an 8x8 switch with different sizes of the window W = 2, W = 4, …, W = 10. The circuit operates at a frequency of 100Mhz, with a 64 bit data bus. There is a maximum bandwidth of 3,2 Gb/s. On 800 cells with uniform traffic and destination cells randomly distributed among all the outputs, a set of maximum N = 8 cells is delivered each cycle (= 10ns). Different performance factors are measured.

The switch delay defined (for one cell) as the average time spent between a cell arriving at the input port and departing from the switch, is shown in Figure 6. The throughput and cell loss simulation results are given in Figure 7.

Figure 8 shows the latency at two different levels (selection circuit and routing network). We notice that the total latency is especially affected by the routing network latency. The selection procedure is more faster. Thus, the research of the optimal solution could be done on larger widths of W, while exploiting the waiting times of the selection circuit. It results, an improvement of the throughput (Higher than 95% for W = 10).

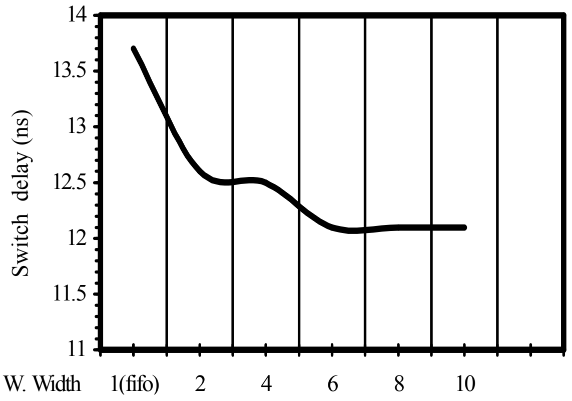


Figure 6. Average delay per cell.

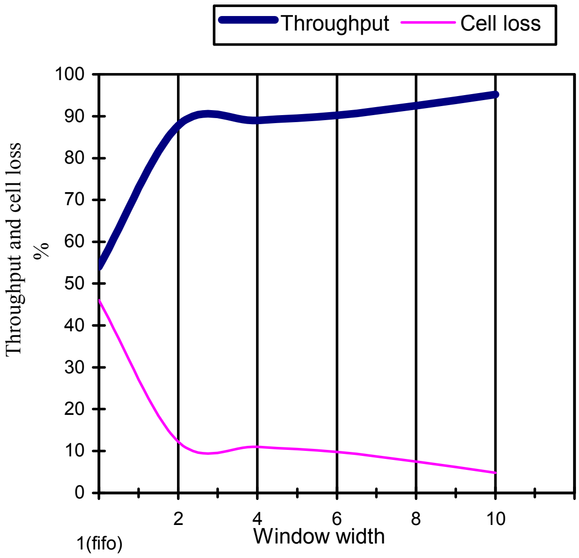


Figure 7. Throughput and cell loss with various window sizes.

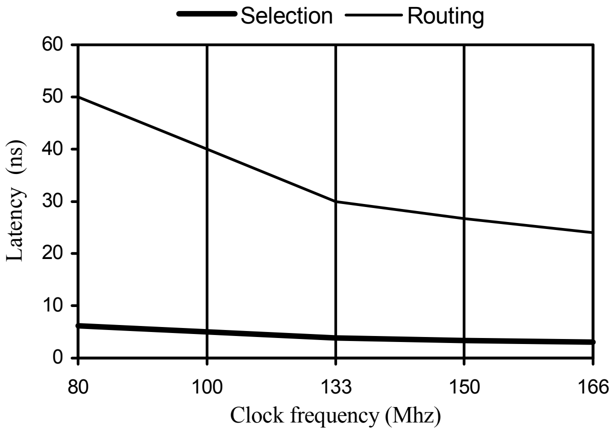


Figure 8. selection circuit and routing network latencies under different internal clock rates

6.3. Cost Evaluation

The cost of a switch is closely related to the hardware complexity. The cost of input buffer switch depends on the interconnection network cost, the size and the number of buffers that it uses, and the hardware complexity of the selection policy or scheduling algorithm. The cost of interconnection network is proportional to the number of switching elements and the number of links between them. Traditional techniques of finding the total cost of a chip, take for granted that the cost of interconnecting subsequent stages is negligible. This means that we can assume that the cost depends only on the number of switching elements. Therefore, the cost of Benes network expressed in number of 2x2 SE is estimated to $N\log_2 N$-N/2.

The switch cost remains relatively low since we adopted the following choices:

1. A Benes network that presents one of the lowest costs among the non blocking multistage interconnection networks. ($N\log_2 N - N / 2$, against $N\log_2 N$ for Banyan networks and N^2 for crossbar).
2. No global controller is needed.
3. A simple selection policy that does not require some big comparators, or other complex circuits doing some operations on a big number of bits.
4. The number of queues or buffers used is limited to O (N), comparatively to O (N^2) what it is used in other solutions adopting virtual queues (VOQ).

The throughput and time complexity of our solution are compared to others, especially FIFO, simple and optimal heuristics, and odd-even switch, because among the studied approaches in this paper, they present the minimum hardware and time complexity.

The throughput results given in table 1, correspond to the solution using window based scheduling algorithms for w = 2, which is analogous to the Odd-Even model where contention resolution also consists of two rounds (one for polling the odd and one for polling the even queues). With higher values of W, it is obvious that the obtained throughput applying these algorithms will be better.

Table 1. Comparison with other solutions.

Method	Time Complexity	Throughput
FIFO	O (N)	0,586
Simple Heuristic	O (WN)	0,68
Optimal	O ($WN^{3/2}$)	0.71
MWM	O ($N^3\log_2 N$)	1.00
Odd-Even	O (2N)	0,71
Our solution	O ($W(N^2$-N))	0,877

The experimental results show that a throughput improvement of about 50% and 23% is achieved in comparison respectively with the FIFO strategy, Odd-Even and optimal heuristics under uniform traffic.

Future simulations are projected to study the functioning of the switch under other arrival models, especially under burst traffic.

7. Conclusion and Future Work

In this paper, we have attempted to show that to design high performance packet switches, theoretical solutions related to buffering, selection, and routing switch functionalities could be defined. However, in practice many of these solutions providing 100% of throughput are not foreseeable.

In fact, the designer has to find the best compromise between actual implementation and performances. Our proposed switch has a better throughput comparatively to studied switches involving minimum hardware and time complexity. In our design, we avoided every choice that can improve throughput and switch delay, but will have a significant impact on cost.

Finally and as future work, we plan to synthesise the switch and implement it using FPGA technology. The expected result will be a low cost high performance packet switch in a single chip.

References

[1] Beauquier B. and Darot E., "On Arbitrary Waksman Networks and their Vulnerability," *Technical Report*, INRIA, no. 3788, 1997.

[2] Boppana R. V. and Raghavendra C. S., "Designing Efficient Benes and Banyan Based Input-Buffered ATM Switches," *in Proceedings of ICC'99,* Vancouver, Canada, 1999.

[3] Boppana R. V. and Raghavendra C. S., "Optimal Self-Routing of Linear-Complement Permutations in Hypercubes," *in Proceedins of the 5th Distributed Memory Computing Conference (DMCC'5)*, pp. 800-808, 1990.

[4] Brown T., "A High Performance Two-Stage Packet Switch Architecture," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1792-1795, December 1999.

[5] Cam H., "Preventing Internal and External Conflicts in an Input Buffering Reverse Baseline ATM Switch," *International Journal of Communication Systems*, vol. 13, no. 4, pp. 317-334, 2000.

[6] Cessa R., Oki E., Jing Z., and Chao H. J., "CIXB-1: Combined Input-Once-Cell-Crosspoint Buffered Switch," *IEEE Workshop on High Performance Switching and Routing*, Dallas, TX, 2001.

[7] Chuang S. T., Goel A., McKeown N., and Prabhakar B., "Matching Output Queuing with a Combined Input Output Queued Switch," *in Proceedings of INFOCOM'99*, New York, USA, 1999.

[8] Das N., Mukhopadhyaya K., and Dattagupta J., "Self Routing in Benes Network," *Technical Report No. E/02/92*, Indian Statistical Institute, Calcutta, 1992.

[9] Duan H., Lockwood J. W., and Kang S. M., "Matrix Unit Cell Scheduler (MUCS) for Input-Buffered ATM Switches," *IEEE Communications Letters*, vol. 2, no. 1, 1998.

[10] Garcia J., Corbal J., Cerda L., and Valero M., "Design and Implementation of High-Performance Memory Systems for Future Packet Buffers," *IEEE Proceedings of the 36th International Symposium on Microarchitecture*, 2003.

[11] Joo Y. and McKeown N., "Doubling Memory Bandwidth for Network Buffers," *IEEE INFOCOM*, vol. 2, pp. 808-815, San Francisco, 1998.

[12] Karol M., Hluchyj M., and Morgan S., "Input Versus Output Queuing on Space Division Switch," *IEEE Transactions on Communications*, pp. 1347-1356, 1987.

[13] Keslassy I. and McKeown N., "Analysis of Scheduling Algorithms that Provide 100% Throughput in Input-Queued Switches," *in Proceedings of the 39th Annual Allerton Conference on Communication Control and Computing*, Monticello, Illinois, 2001.

[14] Kolias C. and Kleinrock L., "The Odd-Even Input Queuing ATM Switch: Performance Evaluation," *in Proceedings of ICC'96*, 1996.

[15] Kornaros G., Pnevmatikas D., Vatsolaki P., Kalokerios G., Xanthaki C., Mavroidis D., Serparos D., and Katerimis M., "Implementation of ATLAS 1: A Single Chip ATM Switch with Backpressure," *in Proceeding of IEEE Hot Interconnects VI Symposium*, Standford University, California, USA, 1998.

[16] Lauer H. C., Ghosh A., and Shen C., "A General Purpose Queue Architecture for ATM Switch," *Technical Report 97-17,* Mitsubishi Electric Research Laboratories, 1994.

[17] Lenfant J., "Parallel Permutations of Data: A Benes Network Control Algorithm for Frequently Used Permutations," *IEEE Transactions on Computers*, vol. 27, no. 7, pp. 637-647, 1978.

[18] Lyer S. and McKeown N., "Analysis of the Parallel Packet Switch Architecture," *IEEE/ACM Transactions on Networking*, vol. 11, no. 2, 2003.

[19] Lyer S. and McKeown N., "Techniques for Fast Shared Memory Switches," *HPNG Technical Report*, Stanford University, 2001.

[20] McKeown N., "iSLIP: A Scheduling Algorithm for Input-Queued Switches," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 188-201, April 1999.

[21] McKeown N., Mekkittikul A., Venkat A., and Walrand J., "Achieving 100% throughput in an Input- Queued Switch," *IEEE Transactions on Communications*, vol. 47, no. 8, August 1999.

[22] Moh W. M. and Chung Y. F., "Design and Evaluation of Cell Scheduling Algorithms for

ATM Switches," *in Proceedings of IEEE Singapore International Conference on Networks*, pp. 355-369, World Scientific, 1997.

[23] Nassimi D. and Sahni S., "Parallel Permutation and Sorting Algorithms and New Generalized Connection Network," *Journal of the ACM*, pp. 642-667, 1982.

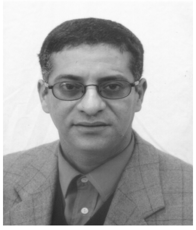
[24] Park Y. K., Cherkassky V., and Lee G., "ATM Cell Scheduling for Broadband Switching Systems by Neural Network," *in Proceedings of International Workshop on Applications of Neural Networks to Telecommunications (IWANNT)*, Princeton, pp. 112-118, 1993.

[25] Raghavendra C. S. and Boppana V., "On Self Routing in Benes and Shuffle-Exchange Networks," *IEEE Transactions on Computers*, vol. 40, no. 9, 1991.

[26] Tutsch D., Hendler M., and Hommel G., "Multicast Performance of Multistage Interconnection Networks with Shared Buffering," *in Proceedings of ICN'2001*, in Lorenz P. (Ed), pp. 478-487, 2001.

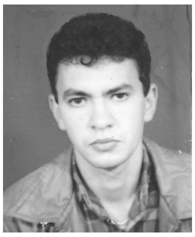
[27] Yang M. and Zheng S. Q., "An Efficient Scheduling Algorithm for CIOQ Switches with Space-Division Multiplexing Expansion," *IEEE INFOCOM*, 2003.

[28] Yihan Li., Shivendra P., and Chao H. J., "On the Performance of a Dual Round-Robin Switch," *IEEE INFOCOM*, pp. 1688-1697, 2001.

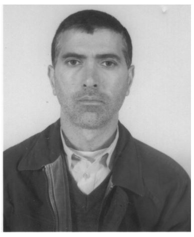


Azeddine Bilami received his BSc degree from the High School of Computer Science (CERI) Algiers, Algeria, in 1983, and the MSc degree in computer science from the University of Batna, Algeria, in 1996, where he is currently an assistant professor. His current research interests are interconnection networks, parallel architectures, and wireless networks.

Mustapha Lalam received the MSc degree in computer architecture from the High School of Computer Science, Algiers, Algeria in 1980. He also received the PhD degree in computer science from University of Toulouse, France in 1990. He joined University of Tizi Ouzou, Algeria in 1993. He has been engaged in the research and development of computer architecture, distributed systems and mobility management for wireless mobile computing and communications. He is the dean of the Faculty of Electrical Engineering and Computing in Tizi Ouzou.



Mehammed Daoui received his MSc in computing systems from the University Mouloud Mammedi, Tizi Ouzou, Algeria in 2001. He is an assistant professor at the same university. He has been engaged in the research and development of computer architecture, distributed systems and mobility management for wireless mobile communications



Mohamed Benmohammed received his BSc degree from the High School of Computer Science (CERI) Algiers, Algeria, in 1983 and the PhD degree in computer science from the University of Sidi Belabbes, Algeria, in 1997. Currently, he is an assistant professor at Constantine University. His research interests include parallel architectures and high level synthesis.