# A Novel Tbit/sec Switch Architecture for ATM/WDM High-Speed Networks

Itamar Elhanany and Dan Sadot
Electrical and Computers Engineering Department
Ben-Gurion University,
Beer-Sheva 84105, Israel
Tel: +972-7-6461581
Fax: +972-7-6472949
E-mail: itamar@ieee.org, sadot@ee.bgu.ac.il

**Abstract**

A new high capacity switching scheme for ATM/WDM networks is presented. The proposed architecture is contention-free, flexible and highly scalable. Switching performance is studied for data rates of up to 10 Gbit/sec/port, providing aggregated capacity of over 1 Terabit/sec. Simulation results show that low latency is achieved, yielding a powerful solution for high-performance packet-switch networks.

## 1. Introduction

Wavelength division multiplexing (WDM) is widely recognized as a promising technology for high-capacity, scalable, cost effective optical networks. Considerable attention has been paid to WDM contention resolution and wavelength assignment issues, commonly targeted at packet-switched networks such as asynchronous transfer mode (ATM).

Recently, several switching schemes were designed in order to support high capacity, large number of ports and low latency requirements [1-4]. Here, we introduce a new WDM packet-switching scheme that offers Terabit/sec capacity along with support of over 100 ports and switching latency of tens of microseconds. The new architecture is contention-free and has the advantages of low implementation complexity and high scalability. In addition, the architecture can easily be adapted to comply with diverse quality of service (QoS) requirements.

The remainder of this paper is organized as follows. Section 2 describes the proposed switch architecture. Section 3 focuses on the channel (wavelength) allocation discipline. Sections 4 present simulation results and section 5 draws the main conclusions.

## 2. Switch Architecture

Figure 1 depicts the proposed switch architecture. The nodes, corresponding to the switch ports, have bi-directional optical data links interconnected via an optical passive star coupler. Each transmitter can be tuned to any of the $N$ wavelengths, while each receiver is assigned a fixed and distinct wavelength. ATM traffic received at each port is distributed to various buffers within the node on a cell-by-cell basis, where each buffer relates to a single wavelength according to the desired destination node. All nodes are connected to a central wavelength reservation scheduler via a common electronic wavelength reservation bus and individual control lines. The $N$ bus lines are accessible to all nodes and indicate the reservation status of each of the $N$ wavelengths. The individual control lines are used by the scheduler to signal each node, in turn, to commence the wavelength reservation procedure.

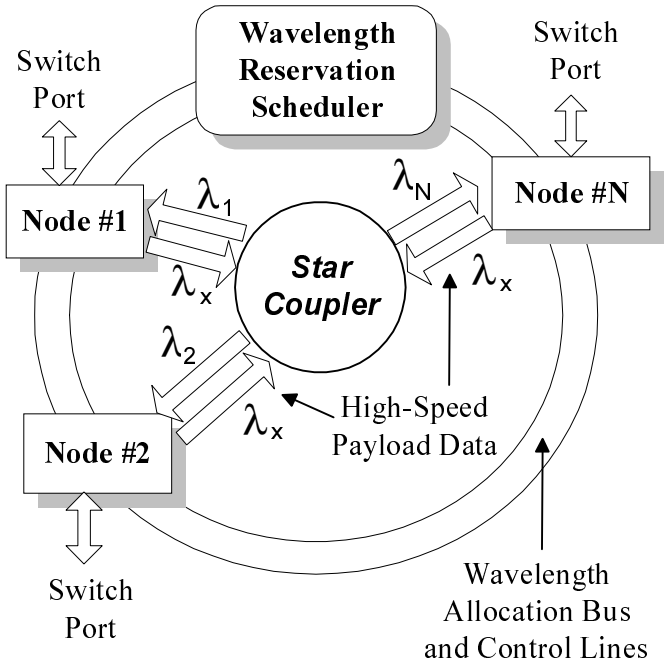


Figure 1: Terabit/sec ATM over WDM switch architecture

## 3. Wavelength Reservation

Upon receiving a control signal from the wavelength reservation scheduler, each node performs wavelength reservation according to two major guidelines: (a) global switch resources status, i.e., available wavelengths at the reservation time, and (b) local considerations, i.e., the status and priorities of the node's internal buffers. Figure 2 illustrates a block diagram of the wavelength reservation hardware of a single node in an 8×8 switch.

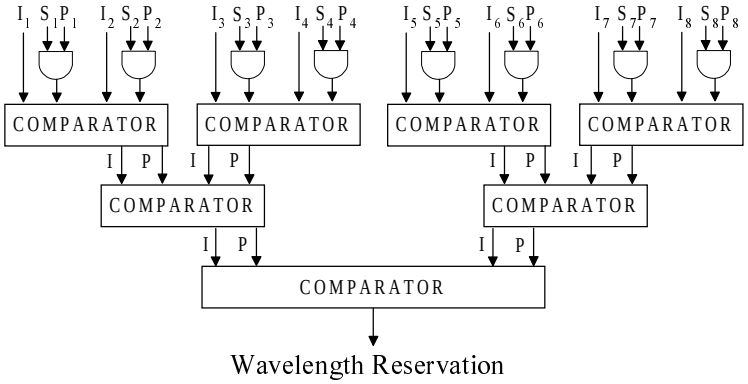


Figure 2: Block diagram of node wavelength reservation logic for an 8×8 switch.

At the highest level, the wavelength reservation status lines, denoted by $S_i$, either grant or discard buffer priorities, denoted by $P_i$, via designated AND logic. Consequently, only buffer indices, denoted by $I_i$, relating to available wavelengths advance to the lower levels. Each level consists of a set of comparators, which concurrently receive as input a pair of indices, along with their respective priorities, and output the higher

priority and its corresponding index. Figure 3 depicts the logic comprising each comparator unit. The output of the last comparator determines the "prevailing" buffer, which held the highest priority out of the subset of buffers relating to unreserved wavelengths. Any number of parameters, such as buffer load, accumulated delay and required QoS can affect the buffer priority metrics. Node wavelength selection is instantly followed by assertion of the relevant line within the wavelength reservation bus. At that point, utilizing a weighted Round Rubin procedure, the central scheduler signals the next node to begin wavelength reservation. Contention is inherently avoided, since at any given time only one node attempts to reserve a wavelength. After all $N$ nodes complete wavelength reservation, high-speed data is optically transmitted via the star coupler.

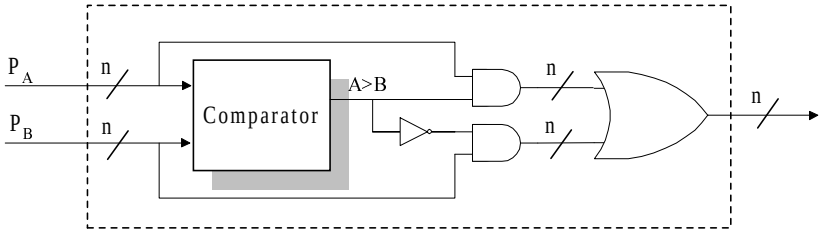


Figure 3: Single-level comparator logic with n-bits per priority value

The wavelength reservation and data transmission are conducted in a time slot discipline. Nodes transmit data concurrently at wavelengths reserved during the previous time slot. Assuming $N$ nodes, the time slot period can be calculated as

$$t_{ts} = N \cdot \log_2(N) \cdot t_c \tag{1}$$

where $t_c$ is the propagation delay of a single-level comparator logic. Accordingly, $t_{ts}$ dictates the minimal number of cells required to be transmitted during each time slot and hence the queuing time delays.

Utilizing current 0.35-μm VLSI technology $t_c$=1 *ns* is achieved. Consequently, extremely short processing time for resource allocation is attained, yielding high switching performance. Figure 4 shows the relationship between the minimal number of cells required to be transmitted during each time slot and the switch aggregated throughput. Various port bitrates are presented as a parameter. The marks on each curve represent standard number of ports, e.g., 8, 16, 32, 64 and 128.

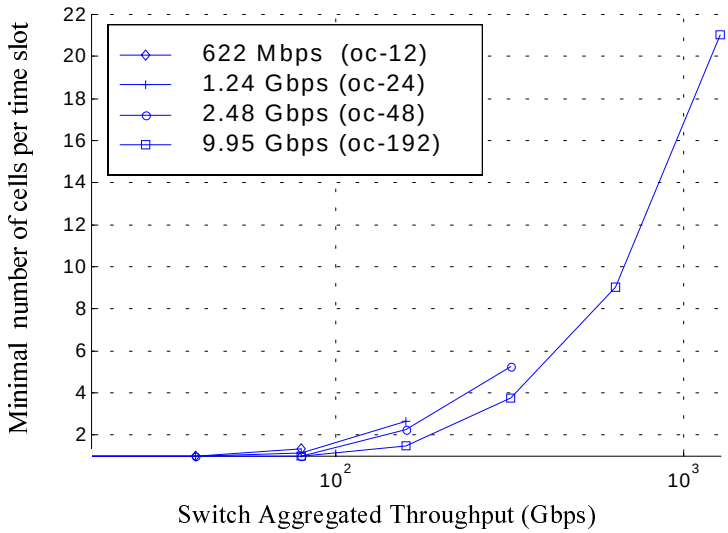


Figure 4: Switch aggregated throughput for different port bit rates

## 4. Simulation Results

Simulations were carried out for various traffic loads of an 100×100 switch ($N$=100), where each node receives data at 10 Gbps (oc-192). Figure 5 presents typical results for the mean cell delay versus traffic load. 100% traffic load corresponds to 1 Tbit/sec switch aggregate throughput. As shown in figure 5, utilizing the proposed architecture, mean cell delay of less than 25 μsec is achieved for traffic loads exceeding 95%. In our simulation, uniform destination distribution is assumed, i.e., each arriving cell has equal probability $1/N$ of being addressed to any of the $N$ destinations. In addition, a binomial cell arrival process is assumed, meaning that at any given cell slot, the probability that a cell will arrive on a particular node is $p$, where $p$ obeys a binomial distribution. Under these assumptions, our simulation results can be compared to other high-speed buffer management approaches such as in [1] and [2]. In such architectures, various types of real time Round Robin algorithms are employed. Correspondingly, contention is avoided using backpressure mechanisms or by locally updating input and output queues scheduling pointers leading to sub-optimal performance and scalability.

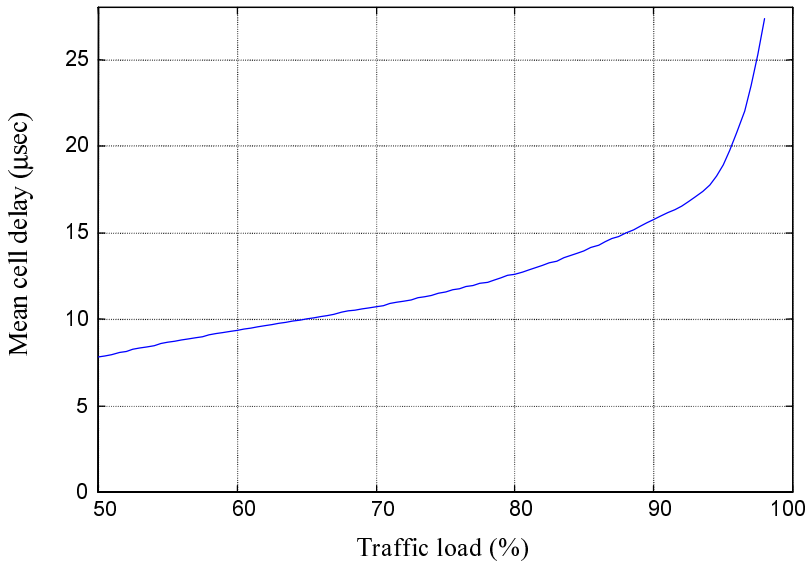


Figure 5: Mean cell delay for various traffic loads

Although in this paper we discuss WDM as the physical layer switching fabric and fixed-size packets, the proposed method can be extended to other multiplexing technology, e.g., optical SDM (space division multiplexing), and to variable-size packets, e.g., IP traffic.

## 5. Conclusions

A novel Tbit/sec switch architecture for ATM over WDM packet-switched networks has been proposed. By applying an ultra-high speed scheduling discipline, extremely high capacity and low latency switching is achieved for fabrics of up to 100×100. The design is simple, scalable and flexible to support diverse traffic characteristics. The method can be adapted to IP traffic and to various multiplexing technologies. Simulation results indicate that Tbit/sec throughput is achieved with mean cell delay as low as 25 μsec.


### Acknowledgement

This work was supported by the Israeli Ministry of Science under contract no. 9404-1-97, the Wolfson foundation and the Rector's office at the Ben-Gurion University of the Negev, Israel.

## References


1. F. M. Chiussi, J. G. Kneuer, and V. P. Kumar, “Low-Cost Scalable Switching Solutions for Broadband Networking: The ATLANTA Architecture and Chipset,” *IEEE Communications Magazine*, vol. 25, no. 12, pp. 44-53, Dec. 1997.
2. N. McKeown, M. Izzard, A. Mekkittikul, B. Ellersick, and M. Horowitz, "The Tiny Tera: A packet switch core," *IEEE Micro*, vol. 17, pp. 27-40, Feb. 1997.
3. V. Sivaraman, G. N. Rouskas, “HiPeR-1: A High Performance Reservation Protocol with Look-Ahead for Broadcast WDM Networks,” *Proc. IEEE INFOCOM '97*, vol. 3, pp. 1270-7, April 1997.
4. C. Salisbury, R. Melhem, “A High Speed Scheduler/Controller for Unbuffered Banyan Networks,” *Proc. IEEE ICC'98*, Session S18-6, June 1998.