WUSEAS-2002-12

Terabit Burst Switching Progress Report (10/01-3/02)

Authors: Turner, Jonathan

May 14, 2002

Abstract: This report summarizes progress on Washington University's Terabit Burst Switching Project, supported by DARPA and Rome Air Force Laboratory. This project seeks to demonstrate the feasibility of Burst Switching,
a new data communication service which can more effectively exploit the large bandwidths becoming available in WDM
transmission systems, than conventional communication technologies like ATM and IP-based packet switching. Burst switching systems dynamically assign data bursts to channels in optical data links, using routing information
carried in parallel control channels. The project will lead to the construction of a demonstration switch with throughput
exceeding 200 Gb/s and scalable to over 10 Tb/s.

Terabit Burst Switching
Progress Report (10/01-3/02)

Jonathan S. Turner
jst@cs.wustl.edu

WUCS-2002-12

May 15, 2001

Department of Computer Science
Campus Box 1045
Washington University
One Brookings Drive
St. Louis, MO 63130-4899

Abstract

This report summarizes progress on Washington University's *Terabit Burst Switching* Project, supported by DARPA and Rome Air Force Laboratory. This project seeks to demonstrate the feasibility of *Burst Switching*, a new data communication service which can more effectively exploit the large bandwidths becoming available in WDM transmission systems, than conventional communication technologies like ATM and IP-based packet switching. Burst switching systems dynamically assign data bursts to channels in optical data links, using routing information carried in parallel control channels. The project will lead to the construction of a demonstration switch with throughput exceeding 200 Gb/s and scalable to over 10 Tb/s.

This work is supported by the Advanced Research Projects Agency and Rome Laboratory (contract F30602-97-1-2703).

Terabit Burst Switching
Progress Report (10/01-3/02)

Jonathan S. Turner
jst@cs.wustl.edu

This report summarizes progress on the Terabit Burst Switching Project at Washington University for the period from October 1, 2001 through March 31, 2002.

1. Prototype Burst Switch Progress

During this period, a $300,000 budget reduction was imposed on this project, removing about half of the final year funding. This reduction made it impossible to complete the burst switch prototype, making it necessary to shut down this component of the project.

2. 160 Gb/s ATM Switch

The following paragraphs summarize status and progress on the various components being developed for the 160 Gb/s ATM switch being constructed as part of this project. Figure 1 shows the overall structure of the prototype and details the location of each component in the overall architecture.

- *PC Boards and Physical Design*. The designs for all the printed circuit boards required for the system have been completed, and all boards have been fabricated. The quad-OC-12 line cards\, dual G-link line card and OC-48 line card have all been fabricated and tested. The IO Modules have been assembled and are undergoing testing to check for manufacturing flaws. The Switch Element board will be assembled and integrated into the prototype in the second quarter.
- *ATM Switch Element* (ASE). This chip is a revised version of a chip that was developed in an earlier project. The new chip implements four priority classes, doubles the cell buffering of the previous chip and corrects timing flaws that limited the operational frequency of the original chip. As reported earlier, the chip has a design flaw that disables one of the eight input ports, but we do not expect this to prevent us from achieving our primary objectives.
- *ATM Input Port Processor* (IPP). The IPP is a modified version of a component developed for an earlier project. The new chip provides a larger VPI/VCI lookup table (4096 entries instead of 1024) and allocates those entries more flexibly. It also implements features for reliable multicast and provides more extensive support for traffic monitoring. The IPP has been implemented in a .35 micron ASIC process. As reported earlier, the circuit has a design flaw that prevents us from using the reliable multicast feature. We have concluded that there are no realistic options available to us for correcting the design, so are proceeding with the current chips.

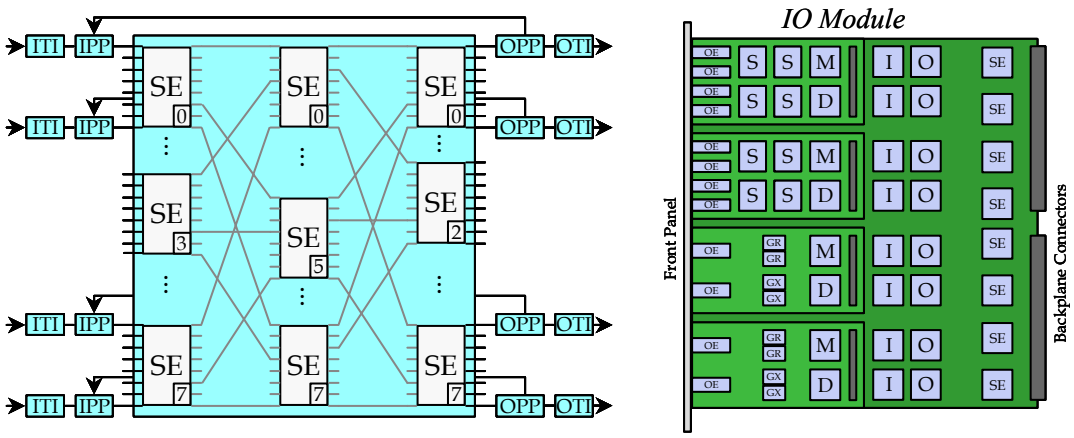

Figure 2. 160 Gb/s ATM Switch

- *ATM Output Port Processor* (OPP). This chip was developed in an earlier project. The required die (fabricated in a .7 micron ASIC process) have been packaged in ball grid array packages (rather than the original pin grid array package) to make them compatible with other components in the system. The repackaged chips have been tested and work correctly.
- *Dual G-link Line Card*. This card multiplexes a pair of 1 Gb/s links onto a single core switch port, using an FPGA to perform the input-side multiplexing and output-side demultiplexing. This card is now complete, tested and all 24 planned units have been produced.
- *Quad OC-12 Line Card*. This card multiplexes four OC-12 links onto one switch port. It uses an FPGA to do the required multiplexing and demultiplexing. This board is now complete, tested and all 24 planned units have been produced.
- *OC-48 Line Card*. This card terminates a single OC-48 link. The board has been completed, fabricated and tested.

3. Time-Sliced Optical Packet Switching

The burst switching architecture requires wavelength conversion to enable good statistical multiplexing performance. Unfortunately, even the best methods for wavelength conversion require a device that is at least as expensive as a laser plus a modulator. This makes it difficult to compete economically with routers that use electronics for switching, since the optics constitutes a large fraction of the parts cost of such a router. For this reason, we have begun to explore alternatives ways to obtain good statistical multiplexing performance in a system with an all optical datapath.

The obvious alternative is to use buffering, as with conventional routers. Unfortunately, the amount of buffer space required makes this infeasible. A link in a wide-area network router generally needs a buffer capacity that is several times the product of the link bandwidth and the network round trip delay experienced by packets traveling over that link. (This large buffer requirement is a consequence of the use of adaptive congestion control in the TCP transport protocol, which causes highly correlated traffic fluctuations on network links.) For wide-area

networks with a round-trip delay of say 100 ms, this translates to about 500 Mbytes of buffer space for a 10 Gb/s link. Storing this many bits in an optical delay line (still the most cost-effective method of optical storage) requires over 400,000 miles of fiber. Even if wavelength division multiplexing is used to increase the storage density by a factor of 200, we still need 2000 miles of fiber, which at a cost of a penny a foot, costs over $100,000. Electronic storage, on the other hand, costs less than $1 per megabyte, meaning that the cost of the same amount of memory in an electronic router is just $500, or half of one percent of the cost of optical storage. Clearly, the optical alternative cannot be competitive using this approach.

We have recently been considering a third alternative, which appears much more promising. The idea is to use statistical time-division multiplexing to get the statistical multiplexing gain needed for packet switching. Each link in the network would still contain multiple WDM channels, but within each wavelength, data would be organized in a time-division format with repeating frames carrying a fixed number of time-slots. Packets would be switched in space and in time at switches along the path, but would not change wavelengths, eliminating the expense of wavelength conversion. As an example, a 40 Gb/s wavelength channel could be organized into 400 time channels of 100 Mb/s each, to match the natural rate of fast Ethernet access channels. Packets arriving at a switch output link would be dynamically time-switched to one of the 400 channels on the same wavelength as the packet, using a *Time Slot Interchanger* (TSI). If no channel were available for an arriving packet, the packet would be discarded. For 400 channels and average link loads up to about 85%, the probability of packet loss is less than 10^{-6}. If each time slot is 500 ns in duration, a frame takes 200 μs and the amount of storage needed is only about 400 μs worth of storage capacity. This is one-thousandth the comparable number for conventional packet switching, and brings the cost of optical storage for this approach well below the cost of the memory needed for an electronic router.

We are now exploring switch architectures that could make this approach practical. The primary issue with this type of architecture appears to be the number of switching operations needed to implement the time switching function. Optical TSIs work by switching time slots through delay lines of varying length, using a dynamically determined switching schedule. To switch data from an input time slot to a different output time slot may require that the data be switched through several different delay lines. The most storage-efficient TSI design requires up to 30 switching operations in the worst-case when configured for 256 time channels. Since each switching operation degrades the optical signal, it's necessary to regenerate a signal after a certain number of switching operations have been performed on it. If regeneration is required too often, then the approach won't be practical. Alternative TSI designs can reduce the worst-case number of switching operations from 30 to 2 at the cost of substantially more storage. We are now evaluating designs that keep the storage requirements low and allow the number of switching operations to vary, with the objective of minimizing the average number of switching operations. In this approach, the number of switching operations that a given packet has been subjected to would be tracked and recorded in the packet header, to enable regeneration on an as-needed basis. It appears likely that the average number of switching operations per TSI can be reduced to between 2 and 4, making it possible for packets to pass through as many as 10 optical packet switches without requiring regeneration.

Work on this approach is just beginning and will continue throughout the remainder of the project. A more complete summary will be included in our final report.

REFERENCES

[EA99] Ramamirtham, Jeyashankher and Jonathan Turner. *Design of Wavelength Converting Switches for Optical Burst Switching*. Submitted to *Infocom 2002*, 7/01.

[TU98a] Turner, Jonathan S."Terabit Burst Switching," Washington University Technical Report, WUCS-98-17, 1998.

[TU98b] Turner, Jonathan S. "Terabit Burst Switching Progress Report (12/97-3/98)," Washington University Technical Report, WUCS-98-16, 1998.

[TU98c] Turner, Jonathan S. "Terabit Burst Switching Progress Report (3/98-6/98)" Washington University Technical Report, WUCS-98-22, 1998.

[TU98d] Turner, Jonathan S. "Terabit Burst Switching Progress Report (6/98-9/98)" Washington University Technical Report, WUCS-98-30, 1998.

[TU98e] Turner, Jonathan S. "Terabit Burst Switching Progress Report (9/98-12/98)" Washington University Technical Report, WUCS-98-31, 1998.

[TU99a] Turner, Jonathan S. "Terabit Burst Switching," *Journal of High Speed Networks*, vol. 8, no. 1, 1999.

[TU99b] Turner, Jonathan S. "WDM Burst Switching," *Proceedings of INET*, San Jose, CA, 6/99.

[TU99c] Turner, Jonathan S. "WDM Burst Switching for Petabit Capacity Routers," *Proceedings of Milcom*, Atlantic City, NJ, 11/99.

[TU99d] Turner, Jonathan S. "Terabit Burst Switching Progress Report (1/99-6/99)" Washington University Technical Report, WUCS-99-21, 1999.

[TU99e] Turner, Jonathan S. "Terabit Burst Switching Progress Report (7/99-12/99)" Washington University Technical Report, WUCS-99-32, 1/2000.

[TU00a] Turner, Jonathan S. "WDM Burst Switching for Petabit Data Networks" *Proceedings of the Optical Fiber Conference*, 3/2000.

[TU00b] Turner, Jonathan S. "Terabit Burst Switching Progress Report (1/00-6/00)" Washington University Technical Report, WUCS-00-18, 7/2000.

[TU00c] Turner, Jonathan S. "Terabit Burst Switching Progress Report (7/00-9/00)" Washington University Technical Report, WUCS-00-28, 10/2000.

[TU01a] Turner, Jonathan S. "Terabit Burst Switching Progress Report (10/00-3/01)" Washington University Technical Report, WUCS-01-09, 5/2001.

[TU01b] Turner, Jonathan S. "Terabit Burst Switching Progress Report (4/01-6/01)" Washington University Technical Report, WUCS-01-23, 7/2001.