7th IEEE workshop on Local and Metropolitan Area Networks
Extended Abstract

Optical Latency Storage - Unifying data communication in space and time

Tom Pfeifer
Technical University of Berlin
Franklinstr. 28, 10587 Berlin, Germany
email: pfeifer@fokus.gmd.de

The current development in optical networks moves toward their physical limit in transmission speed. 'Gigabit technology' in optical networks opens a new dimension not only for communication, but for computing in general. Several components in optical technology matured in the recent years. Among these components are:

- low attenuation, dispersion-compensated fibres
- all-optical amplification
- wavelength division multiplexing, tunable wavelength filters
- a large set of optical devices as couplers, modulators, etc.
- optical switches with electrical as well as optical control

Considered from a system point of view completely new architectural approaches become possible. While the transmission rate in the communication system is driven up step by step into the orders of Gbit and Tbit per second, one physical barrier cannot be overcome: the end-to-end signal latency, which is firmly limited to the speed of light in the medium used. In a medium of a given length, the length of a single bit becomes shorter for a higher transmission rate.

The phase velocity of light, $v_p = c/n$, in free space $v_p = c = 2.99792 \times 10^8$ m/s comes down to $v_p = 2.07 \times 10^8$ m/s in a silica fibre with n = 1.45. The latency in the fibre is then
$1/v_p = 1\ \text{s} / 2.07 \times 10^8\ \text{m} = 4.83 \times 10^{-9}\ \text{s/m} = 4.83\ \mu\text{s/km}$.

When the length of a bit becomes significantly shorter than the length of the medium, lots of other bits are stuffed into the medium before the first one reaches the receiver. The medium shows a storage effect. This effect is undesired in communication, because it causes problems in error correction, and so is the latency itself, because it just consumes time.

In this way, the transmission medium stores a certain amount of information. However, this raises the idea of utilizing the undesired latency effect, as its magnitude becomes useful in terms of storage in optical Gigabit networks. In other words, this proposal describes a latency storage, which is well known in technology for other transmission media. Magnetic latency storages could buffer some hundred bits in the era before semiconductor memories were invented, and ultrasonic delay elements store one analogue scan line in every colour TV set.

From the perspective of information theory there are not many differences between 'storage' and 'transmission': transmission works in space, while storage is a transmission in time. Furthermore, every storage in time consumes space, and every transmission in space consumes time.

When Wavelength Division Multiplexing (WDM) in optical fibres and devices multiplies the number of independently transmitted channels by 10...1000, the storage capacity of the medium is increased by the same factor. It could be further increased by purposely built latency devices (e.g., fibre loops, optical cubes).

An upper boundary of storage capacity can be calculated as follows: The three optical windows at 0.85, at 1.3 and at 1.55 μm wavelength, each approx. 200 nm wide, deliver a bandwidth of approx. 3 x 20 THz = 60 THz. Assuming a simple binary coding mechanism with 1 Hz = 1 bit/s, the storage capacity per length can not exceed:

$cap' = 60 \text{ Tbit/s} \times 4.83 \times 10^{-9} \text{ s/m} = 289.8 \times 10^{3} \text{ bit/m}$

For a fictitious single fibre with the length of the earth's circumference this would calculate as:

$u_{earth} \times cap' = 40 \times 10^{6} \text{ m} \times 289.8 \times 10^{3} \text{ bit/m} = 11 \text{ Tbit}$

However, this number will be reduced by the available technology. Most limiting is currently the number of channels in WDM.

Potential applications are manifold, embracing the simple substitution of local storage systems as well as global distributed processing, where the medium not only provides the communication space, but also the temporary and permanent shared memory between the stations [2].

One of the most important bottlenecks in computing, particularly when continuous media are involved, is the different treatment of communication and storage inside and outside the end-system. Approaches have been tested to unify the communication format inside and outside, for example in workstations configured as ATM-Desktop-Area-Networks.

On the other hand, recent experiences show that even high-end workstations rarely go beyond a communication requirement of 100 Mbit/s [3]. It becomes obvious that the classical bottleneck of communication in the area of distributed computing disappears and will be replaced by plentifulness (Figure 1).

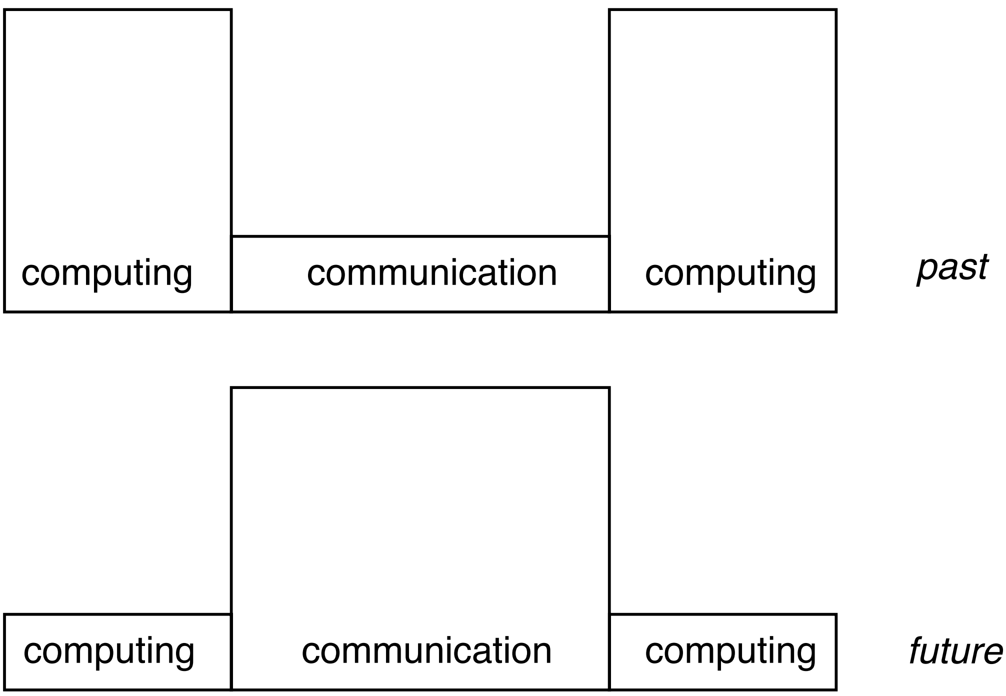

FIGURE 1 The communication bottleneck in distributed computing disappears

System Design

This paper outlines a global communication/storage system, considering which system components are technologically available now and in the near future.

The basic functional elements are shown in Figure 2. The signal is kept cycling in a fibre loop, amplified optically. The shape of the signal is regenerated physically, and on a higher level the reliability is ensured by logical error control. A switch allows the insertion of new data, the access to required items, and the removal of garbage.

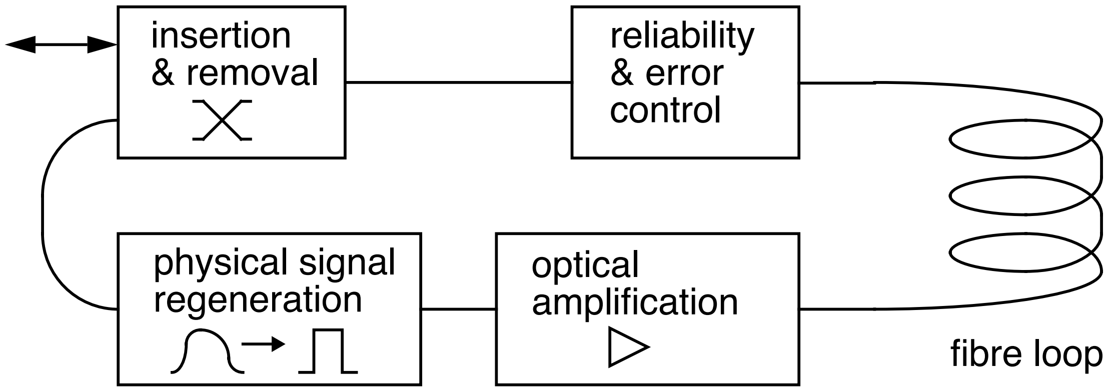

FIGURE 2 Functional elements of optical latency storage

In Figure 3, the single fibre loop expands to a large area network, with amplification, regeneration and switching located in the multiple nodes.

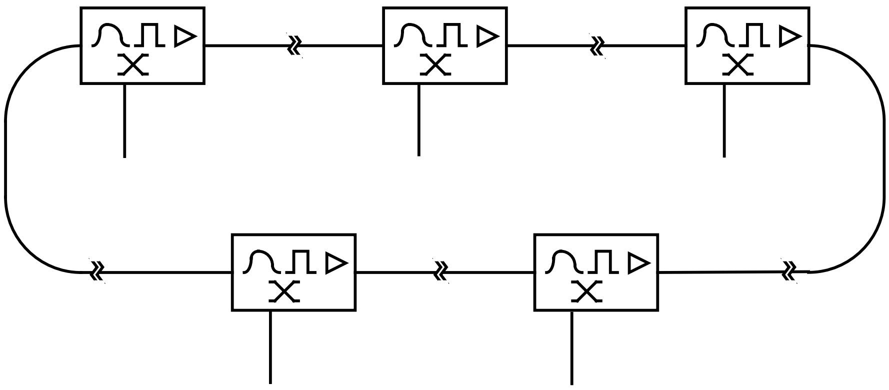

FIGURE 3 Optical latency storage in a network

Reliability

When the network is used for storage, it has to ensure the reliability and data integrity in the network itself, this task cannot be left to the end-system. It counts as an advantage that the natural error probability in fibre optical networks is very low.

While the physical signal regeneration is discussed in the technology section, a possible method of logical error control is described as follows.

Assuming the application of forward error correction (FEC) algorithms, it might be impossible to detect and correct an error in a data packet while it is transmitted through a node, and to repair or replace it at the same time. However, if the packet is stored in the network, we can take advantage of the fact that the same packet will appear again at the same place. Then, a scheme as drafted in Figure 4 comes into work:

Based on an inspection schedule, packets will be copied to the repair station selectively. If they appear to be correct, the copies are discarded. If the FEC has to be performed, the repaired packet is kept for replacement of the defective packet when the latter arrives again in the next cycle. Similarly, obsolete packets with expired life time could be identified and removed in the next cycle (garbage collection).

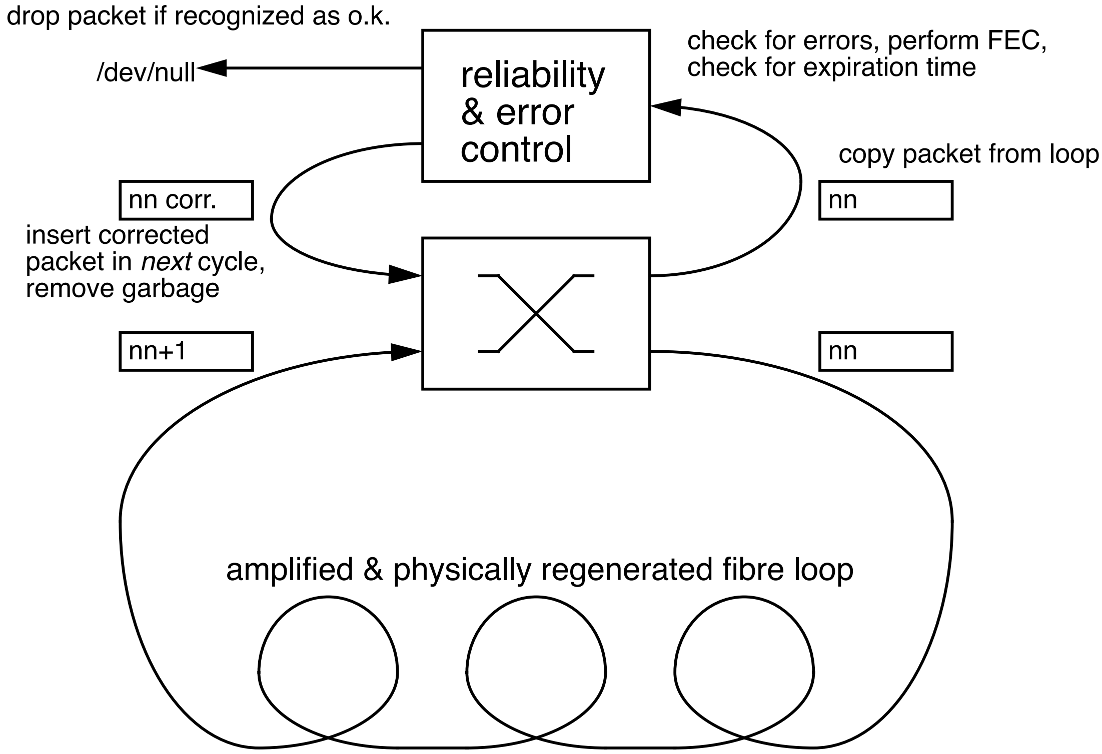

FIGURE 4 Error control with on-the-fly repair and next-cycle replacement

In addition to the periodical check and FEC on packet level, there can be a cross-interleaving of information between packets, enabling regeneration of lost or irreparable packets with redundant information in other packets.

In a meshed network, different routes are possible, allowing dynamic route re-configuration in case of partial network failures. This technique can be used to create optical self-healing rings (SHR) with a high degree of reliability.

Topology

The most essential criteria for choosing appropriate physical and virtual topologies are the possibility to create a logical (virtual) ring, allowing independent traffic in the several branches (no broadcast medium). These requirements are met by most topologies, except a broadcast bus structure, e.g.:

- linear multihop structure
- star structure

- physical ring structure
- meshed structures

Examples for the possible topologies are discussed below.

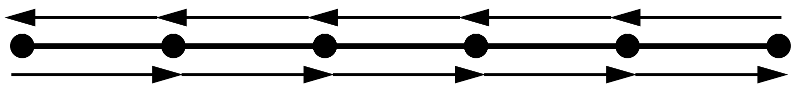

FIGURE 5 Virtual loop in a linear multihop structure

Figure 5 shows a linear multihop structure, where the load is distributed equally: two virtual paths occupy each interconnection, and two switching processes are required on each node. In the American Gigabit testbeds (Aurora, Blanca, Casa, etc.) such a linear topology were used quite often [3].

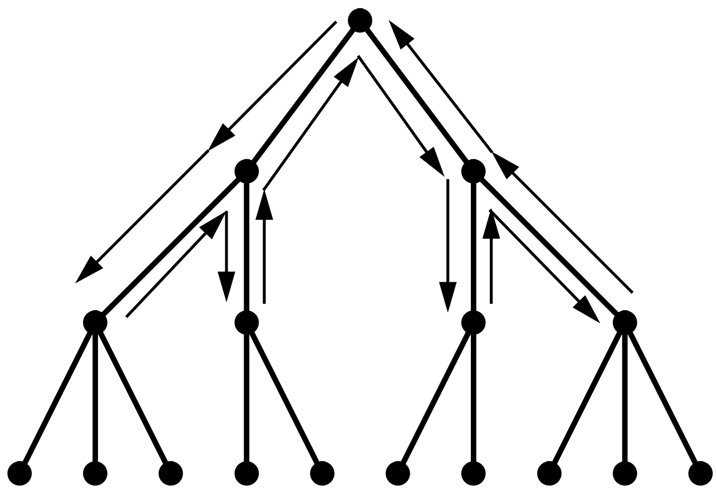

FIGURE 6 Virtual loop in a treelike distribution topology

Figure 6 shows a possible distribution network with star or tree topology, where a virtual path is defined as a loop through all nodes down to a certain level in the hierarchy. In this example, the path is not routed to the leaf nodes, because they might be considered not reliable enough, or do not belong to the provider of the main network. Each interconnection carries two virtual paths, and the intermediate nodes are loaded with one simultaneous switching process for each interconnection.

The scheme extends easily when a physical ring serves as a backbone for several star networks, as shown in Figure 7.

Multiple routes are possible in a meshed network. This allows to distribute the load dynamically for avoiding an overload of switching in central nodes, as well as re-configuration in case of partial network failures.

In all topologies, the delivery time of a packet can be fine tuned by choosing either longer routes or short-cuts, which is important for the aspect discussed next.

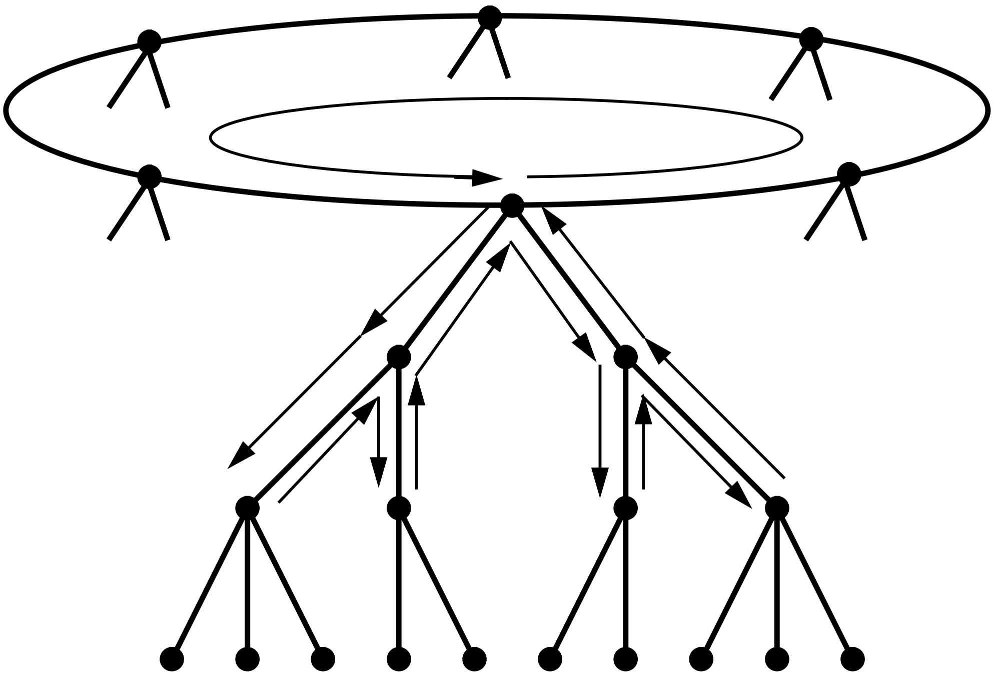

FIGURE 7 Virtual loop with a backbone ring topology

Delivery in space and time

There are two different approaches of storing data in the network. First, the data can be stored for an undetermined period of time, with unknown access requirements. In worst case, the access delay is equal to the latency of one cycle in the network.

If the storage property of the network is utilized in an environment for Global Distributed Processing, another interesting approach is possible. The data are inserted into the network not only with a destination address in space, but also with a destination in time. In this way, the interprocess data could be delivered to the remote resource “just in time” for the other process, requiring no separate buffering resources. This temporal destination could be placed into a separate header field of data packets, as shown in Figure 8. Alternatively, an expiration time would indicate that the data can be removed by the garbage collector.

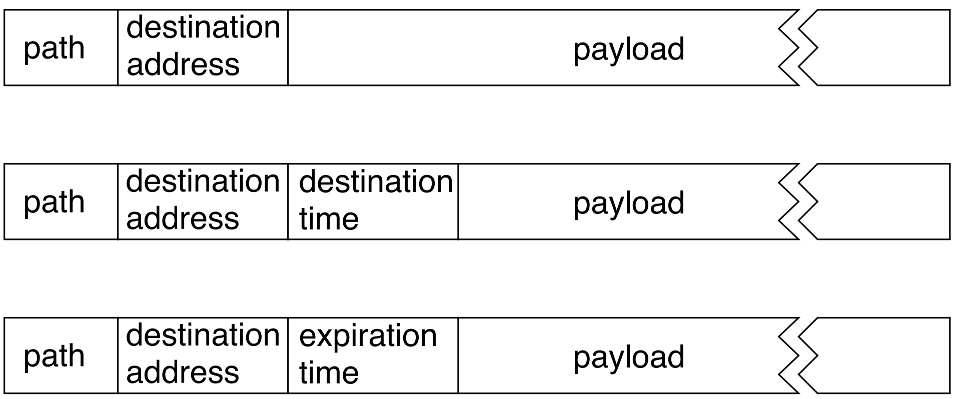

FIGURE 8 Packet headers for routing through space and time

Technology

Laboratory prototypes of all-optical fibre loop storages have been tested successfully [13]. A configuration of a laser amplifier, a switch and 160 metres of fibre allowed 150 cycles in the loop. Inserting an additional optical regenerator for noise and jitter reduction, a pulse could be stored 10^6 cycles (500 ms) [14]. Currently such storages are already used as short-term buffers in optical ATM-switches.

In the proposed system, switches are required for insertion and unloading the payload as well as for routing data packets inside the system. The majority of optical switches known today is electronically controlled. Some approaches of optical codeword detection and optical header processing utilize distinguished wavelengths pattern as header codewords [15], [16]. In [17], an all-optical switching technique is drafted, using ultra-fast optical time-division multiplexing.

WDM depends on the availability of fixed and tunable filters as well as single-frequency fixed or tunable emitters (lasers). Tunable filters have been built in great variety. The number of channels in 200 nm ranges from 10^1 to 10^3 with tuning times from 1 ms to 100 ns [4]. Tunable lasers spanning a whole 200 nm window have been developed for sub-microsecond tuning times.

Doped-fibre amplifiers are commercially available for the 1.5 μm window. It has proved difficult to build them for other windows [4]. Simultaneous amplification of wavelength-multiplexed signals has been demonstrated for a system with cascaded EDFAs [9]. In addition to fibre amplifiers, semiconductor optical amplifiers offer some advantages in simplicity and spectral properties [10].

ATM-like cell switching is one possible candidate for organizing the traffic in the system. The Quality-of-Service parameters of such a system need to be discussed for both aspects, transmission and storage. E.g., it might not be economical to keep all data available with the same access parameters all the time. The decision whether data should be possible to access in the shortest time needs to depend on the probability of request for the data in the nearest future. If this probability is not calculable, downgrading data to lower access levels can be determined by the age of the latest read access.

Application scenarios

A set of scenarios illustrates the employment of such a network memory system for open approaches of Global Distributed Processing with shared resources, as well as closed applications as e.g. Video on Demand caching. As the storage capacity of the network is limited, it is not going to replace existing storage technologies. Therefore, applications should utilize the characteristic properties of such an approach.

Video compression and decompression schemes often employ several reference frames, having pointers to moving objects in the intermediate frames. Such frames have a limited life time as well as a defined call-back time, both not exceeding e.g. 10 frame cycles. Therefore, they could be sent out with a defined return-time to the latency storage.

Furthermore, video compression is a good example for distributed processing with specialized tasks performed by remote resources, where the video data and the compressed reference frames have to be delivered just-in-time.

The possibility to create highly reliable, self healing rings leads to the possibility to store data of common importance in the network. Management data are an example for this category.

Store-and-Forward systems, for example for delivering multimedia mail, could store the information in the network until the recipient is able to take it over.

A Video on Demand server has to cope with multiple requests of new releases as well the interactivity of the users (rewind, fast forward etc.). In traditional approaches, identical copies of the files are hold on several hard disks, or cached in huge semiconductor memories. With the approach of this paper, the network which is responsible for delivering the information to the customer could also cache all recently requested slices of the videos, removing load from a central server.

Conclusion

The paper demonstrated the possibility to employ optical Gigabit networks for storage purposes as well as data delivery ‘just-in-time’, utilizing the signal latency. The basic system layout was drafted, topologies were discussed, and techniques for ensuring reliability and data integrity were shown. The technological components, which exist as singular solutions, need to be put together to a prototype system, and the access protocols have to be refined.

References

[1] Covaci, S.; Popescu-Zeletin, R.: Shaping the End-System Architecture for Global Distributed Applications. - Proc. of 1st IEEE Internat. Sympos. on Global Data Networking, Cairo, Egypt, Dec. 13-15, 1993

[2] Covaci, S.; Popescu-Zeletin, R.: The Network-Memory in a Global Distributed Processing System. - Proc. of 4th Workshop on Future Trends of Distributed Computing Systems, Lisbon, Portugal, Sep. 22-24, 1993

[3] Ferrari, Domenico: Programs for Networking Research and Testbed Activities in the U.S. - Proc. of KiVS95, Kommunikation in Verteilten Systemen, Chemnitz, Germany, Feb. 22-24, 1995, Springer, 3-540-58960-0

[4] Green, Paul. E., Jr.: Fiber Optic Networks. - Englewood Cliffs: Prentice Hall, 1993, 0-13-319492-2

[5] Palais, J.C.: Fiber Optic Communications. - Englewood Cliffs: Prentice Hall, 3rd ed.1992, 0-13-473554-4

[6] Bannister, J.A.: The Wavelength-Division Optical Network: Architectures, Topologies, and Protocols. - Dissertation, CSD-900007, Univ. of California, Los Angeles, 1990

[7] Bannister, J.A.; Gerla, M.: Design of the Wavelength-Division Optical Network. - Tech. Report CSD-890022, Univ. of California, Los Angeles, 1989

[8] Chlamtac, I.; Ganz, A.; Karmi, G.: Lightnets: Topologies for High Speed Optical Networks. - Tech. Report TR90-CSE-10, University of Massachusetts, Amherst, 1990

[9] Bayart, D.; et al.: 1.55 m fluoride-based erbium-doped fibre amplifier for WDM applications. - in [11], pp. 113-117

[10] Buus, J.: Current status of semiconductor optical amplifiers and their applications. - in [11], pp. 69-71

[11] 12th Annual Conference on European Fibre Optic Communication and Networks, Heidelberg, June21-24, 1994

[12] Abschlußbericht zum Verbundprojekt Optische Signalverarbeitung: Photonik. - Ed.: Heinrich-Hertz-Institut, Berlin, 1994

[13] Heydt, G.; Weber, H.G.; Großkopf, G.; et al.: Investigation of optical ATM-switching (in German: Untersuchung zur optischen ATM-Vermittlung). - in [12], pp. D.5-1

[14] Eiselt, M.; et al.: One million pulse circulations in a fiber ring using a SLALOM for pulse shaping and noise reduction. - IEEE Photonic Technology Letters, Vol. 4 (1993), No. 4, pp. 422-424

[15] Jajszczyk, A.; Mouftah, H.T.: Photonic Fast Packet Switching,- IEEE Communications Magazine, Feb. 1993, pp. 58-65

[16] Auracher, F.; Wiedeburg, K.: Photonic Switching with OFDM and OTDM demonstrator (in German: Photonische Vermittlung mit OFDM- und OTDM-Demonstrator). - in [12], pp. D.3-1

[17] Prucnal, P.R.: Ultrafast All-Optical Photonic Packet Switching. - Proc. of the 6th IEEE workshop on Local and Metropolitan Area Networks, San Diego, Oct. 14-16, 1993