

MPLS をベースとした フォトニックインターネットの展開に関する一考察

村田正幸* 北山研一**

* 〒560-0043 豊中市待兼山町 1-30
大阪大学サイバーメディアセンター
E-mail: murata@cmc.osaka-u.ac.jp

** 〒567-0871 吹田市山田丘 2-1
大阪大学 大学院工学研究科
E-mail: kitayama@comm.eng.osaka-u.ac.jp

あらまし： 本稿ではまず、MPLS の概要を振り返り、フォトニックベースへの適用形態について考察する。まず、その一形態として、WDM 技術の MPLS への適用形態を明らかにする。IP over WDM ネットワークを実現する場合(λ -MPLS)、波長ルーティングによる論理パストポロジ-の設計手法が鍵となる。本稿では、論理パストポロジ-から MPLS へのマッピングについてサーベイし、次に、フォトニック符号ラベル交換技術とその MPLS への適用(OC-MPLS) に関して実現性の検討を行う。最後に、これら 2 つの技術の長所、短所を比較検討し、それぞれの技術について今後解決すべき問題点を挙げる。特に、OC-MPLS の今後の発展の可能性について議論している。

キーワード：MPLS (Multi-Protocol Label Switching)、IP over WDM、光符号ラベル交換、波長ルーティング、光パケット交換

A Perspective on Photonic Multi-Protocol Label Switching

Masayuki Murata* and Ken-ichi Kitayama**

Cybermedia Center
Osaka University
Toyonaka, Osaka 560-0043, Japan
murata@ics.es.osaka-u.ac.jp

** Graduate School of Engineering
Osaka University
Suita, Osaka 567-0871, Japan
kitayama@comm.eng.osaka-u.ac.jp

Abstract: We first review the history of MPLS (Multi-Protocol Label Switching) in order to later introduce photonic MPLS. As an emerging technology for carrying IP traffic over photonic networks, we next discuss IP over WDM networks based on the MPLS technology (λ -MPLS). Then, an optical code label switching technique and its application to MPLS (OC-MPLS) is introduced. We extensively compare advantages and disadvantages of those two approaches, which is our main objective in this paper. By discussing remaining and challenging problems of two approaches, the possibility of a future deployment of OC-MPLS is finally addressed.

Keywords: MPLS (Multi-Protocol Label Switching), IP over WDM, photonic code label switching, wavelength routing, optical packet switching

1. Introduction

The photonic network technology is expected to provide the infrastructure of the next-generation Internet against an explosive growth of traffic demands. The recent active development of the photonic technology increases the network bandwidth. Actually, we have already had a commercially available product utilizing the WDM technology to increase the link capacity between two adjacent routers. That is, each wavelength on the fiber is treated as a physical link between the conventional IP routers. In this way, the link capacity is certainly increased by the number of wavelengths multiplexed on the fiber. It is, however, insufficient to resolve the network bottleneck against an explosion of traffic demands since it only results in that the bottleneck is shifted to an electronic router. One of the major bottlenecks in electronic routers is the longest prefix match for each incoming packet. The speed of a lookup algorithm of the routing table is determined by the number of memory accesses it requires to find the matching entry, and the speed of the memory. The memory access time typically ranges from 10ns to 60ns. If an algorithm performs eight memory-lookups with a memory access time of 10ns, 12.5 million lookups/s can be performed [Kes98]. The bit rate of the link interface with the router is, for example, only 10Gbps for a 800-bit long packet. This processing capacity is much smaller than the aggregate capacity of even WDM links of one optical fiber ever achieved, e.g., 40Gbps \times 160 wavelengths = 6.4Tbps/fiber [Ito00]. It means that rough estimation that switching speeds are 20 times greater than forwarding speeds for comparably priced hardware [Lin97] still holds even for photonic technologies.

One promising way to alleviate the capacity limit of the routers is to introduce an MPLS (Multi-Protocol Label Switching) technology [Dav98]. By MPLS, switching and forwarding capabilities are separated to fully utilize a high speed switching capability of the underlying network such as ATM. Packet forwarding to determine the destination is only performed at the edge of the MPLS domain. While MPLS needs to establish a closed domain for utilizing a new lower-layer technology, it is useful to incorporate the photonic technology for building the very high-speed Internet. Actually, a recent advancement of the WDM technology exhibits the possibility of Multi-Protocol Lambda Switching (or λ -MPLS) [Dan00], where packet forwarding is performed in the optical domain within the λ -MPLS network.

However, there are still several problems in order to deploy λ -MPLS. The most difficult problem in λ -MPLS is a capacity granularity; the unit of the bandwidth between edge node pairs of the MPLS domain is a wavelength ca-

capacity. It may be sometimes too large to accommodate the traffic between node pairs. One approach to resolving the capacity granularity problem is addressed in [Ban00], where the authors introduce *wavelength merging*, but the related technology is still immature.

Another promising technique is to utilize an optical code label switching technology recently developed in [Kit99] where the above granularity problem can be resolved in the optical domain, thus improving bandwidth efficiency. Hereafter, we will call the MPLS technology utilizing the optical code label switching method as *OC-MPLS*. Its ultra-fast photonic label processing capability is suitable for high bit rate MPLS applications.

In what follows, we will first introduce an MPLS technique based on ATM (Asynchronous Transfer Mode). We then proceed to discuss IP over WDM networks based on the MPLS technology (λ -MPLS). An optical code label switching technique is another promising method to realize IP over fiber networks, and its application to MPLS (OC-MPLS) is next introduced. We then discuss advantages and disadvantages of those two approaches thoroughly, which is our main objective in this paper. By discussing remaining and challenging problems of two approaches, a possibility of a future deployment of OC-MPLS is finally addressed.

2. Concept of MPLS

In this section, we briefly introduce an ATM-based MPLS. Then, a concept of photonic MPLS is introduced, followed by WDM-based MPLS as an example.

2.1. History of MPLS; ATM-based MPLS

The earliest motivation for MPLS was to simplify wide-area IP backbone networks by overlaying IP and the new emerging high-speed technology. During the mid-90's, the only solution was ATM in which fixed-size packets (called *cells*) are switched in hardware at nodes. It is a main reason that ATM can provide the high-speed switching capability. In IP over ATM networks, ATM is used only for providing the link-level connectivity although ATM itself had been developed in order to offer its native networking capabilities. We have a lot of excellent articles explaining ATM-based MPLS (see, e.g., [Dav98], [Arm00] and references therein). Nevertheless, we start with its introduction because OC-MPLS, which will be introduced in the next section, can provide much similar functions in the optical domain to ATM-based MPLS.

The central part of MPLS is to utilize the high-speed packet forwarding capability of the underlying network by a label-swapping forwarding algorithm (see Figure 1). A label is a short, fixed-length value carried in the packet header to

identify a Forwarding Equivalence Class (FEC). A label has only link-local significance, and it corresponds to VPI/VCI in the case of ATM.

In MPLS, packet forwarding is performed as follows.

- (1) At the ingress edge of the MPLS, which is called an Ingress LSR (Label Switched Router), the label-swapping forwarding algorithm performs label mapping from a destination address of the arriving IP packet to the initial label for injecting the packet onto the LSP (Label-Switched Path). The Ingress LSR performs a longest-prefix match routing table lookup to find an appropriate label as in the conventional router.
- (2) LSP is set up between Ingress and Egress LSRs by utilizing a VP/VC connection in ATM. That is, LSP is functionally equivalent to a virtual circuit.
- (3) Within the network, the Core LSR forwards the packet using the label-swapping forwarding algorithm. When a labeled packet arrives at the Core LSR, it uses the label and the input port number to find out the next-hop output port number and the new label by an exact match search of the forwarding table. It is a native function of ATM, and we can avoid processing burden of the longest-prefix matching of the conventional IP routers.
- (4) The role of the Egress LSR is similar to the conventional IP routers; it searches the next link by performing the longest-prefix match table lookup.

Setting up the appropriate LSP's is another concern in MPLS. It is performed by LDP (Label Distribution Protocol), which supports two styles of label distribution: independent or ordered [Dav98].

A key component of MPLS is that multiple flows can be assigned to the same label (or LSP), and a stream can range from fine to coarse granularity. Its choice balances the need to share the same label among many destinations with the need to maximizing switching capability while preserving resources. The granularity ranges from IP prefix to application-level flow.

The switched paths in the MPLS network take the form of a multipoint-to-point tree. The merging of the switched paths that occurs at a node when multiple upstream paths for a given stream are spliced to a single downstream switched path for the stream. In the case of ATM, however, merging is not always possible since much of the ATM switches was not capable of reassembling cells from multiple inbound VCs without the problem of cell interleaving. One solution to this problem is to use virtual paths to merge streams rather than VCs [Gha00]. The use of VP merging can create the tree of VPs. Cell interleaving is prevented by the assignment of unique VCIs within each VP.

MPLS realizes explicit routing by combining the

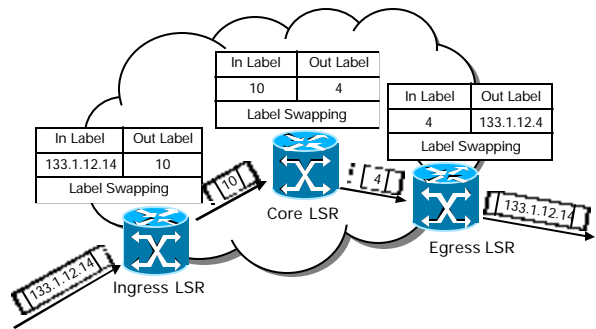


Figure 1: Label Swapping at LSRs

pre-specified label to the LSP at the time the label switched path is set up. It makes it possible to introduce several features. One example is traffic engineering in which the path selection can be performed by taking into account the network efficiency. Of course there are reasons that normal datagram networks such as IP does not adopt such a policy. One important reason is the reliability. Actually, the reliability is one active research area in MPLS. See, e.g., [Dav98].

2.2. Mapping to Photonic MPLS: λ -MPLS

Header processing and forwarding of IP packets are necessary at every router in the conventional Internet. Since packet processing in the photonic domain cannot be expected in the near future by the photonic technology, the MPLS technology can be well suited to the photonic-based Internet. In this paper, we will refer it to *photonic MPLS*. Since a recent discussion on the photonic MPLS is mainly devoted to λ -MPLS where the WDM technology is applied to MPLS, we start with a brief introduction of λ -MPLS [Gha99, Dan00, Gha00].

A key to realizing λ -MPLS is how to establish the logical topology offered to the upper layer protocol, i.e., IP in the current case. In the logical topology, wavelength paths are configured over the WDM physical network in order to carry IP packets utilizing the wavelength path. Here, the physical network means an actual network consisting of the optical nodes and optical fiber links connecting nodes. Each node has optical switches directly connecting an input wavelength to an output wavelength (which is often called as an optical cross-connect), by which no electronic processing is necessary at the node. Then, the wavelength path can be set up directly between two nodes via one or more optical switches. The wavelength path directly connecting two nodes is referred to as a *lightpath*. By utilizing the logical topology consisting of lightpaths, the physical structure of the WDM networks is embedded, and the logical topology is viewed as a underlying network by IP. In such a network, if the lightpaths are placed between every

two end-nodes, then no electronic processing is necessary within the network.

As an example, see Figure 2, showing the physical WDM network. Every link in the figure has two wavelengths λ_1 and λ_2 . Suppose that we have the traffic demand matrix $T = \{t_{ij}\}$ as,

$$T_1 = \begin{bmatrix} - & 1 & 0 & 1 \\ 1 & - & 1 & 0 \\ 0 & 1 & - & 0 \\ 1 & 0 & 0 & - \end{bmatrix},$$

where t_{ij} shows the traffic demand from nodes N_i to N_j , and the unit represents the wavelength capacity. The logical topology shown in Figure 3 is then derived. A logical view from IP is shown in Figure 4. For every traffic demand, the hop count of the lightpath is one. For example, traffic from node N_1 to node N_3 utilizes wavelength λ_1 , and does not need electronic processing at node N_5 (the Core LSR), by which high-speed packet transmission can be achieved.

Many researchers have discussed the design methods of the logical topology, which is a class of RWA (Routing and Wavelength Assignment) problem. See, e.g., [Dur00] and references therein. The problem is usually divided into two sub-problems; route selection and wavelength assignment. The objective is to maximize/minimize the cost functions such as wavelength utilization and the number of required wavelengths. The problem is formulated as integer-linear or mixed-integer programming, and computational time tends to be unacceptable. Accordingly, many heuristics have been proposed in the literature.

To illustrate a rough sketch of the optimization problem encountered in determining the logical topology, we describe a heuristic algorithm called MLDA (Minimum delay Logical topology Design Algorithm) proposed in [Ram96]. To maximize wavelength utilization, MLDA works as follows. First, it places the lightpath between two nodes if there exists a fiber directly connecting those nodes. Then, attempts are made to place lightpaths between nodes in the order of descending traffic demand on the shortest-path. Finally, if there still exist non-utilized wavelengths, lightpaths are placed randomly utilizing those wavelengths as much as possible. Among several options of MPLS, the route the packet traverses can be determined explicitly (*explicit routing*). In such a network, the lightpath should be prepared among every end node (Ingress/Egress LSRs) pairs according to the traffic demand within the MPLS domain.

The LSR in an electronic MPLS is generally able to perform various operations on packet labels. Those include label swapping, label merging, and label stacking [Gha00]. However, it has been difficult to realize those functions in

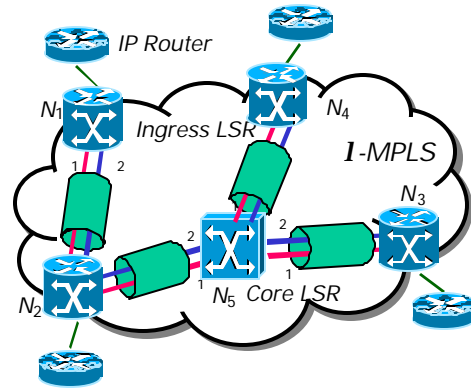


Figure 2: Example Physical Topology

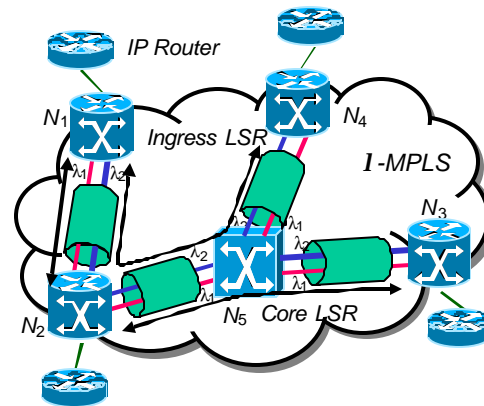


Figure 3: Logical Topology Consisting of Three Lightpaths

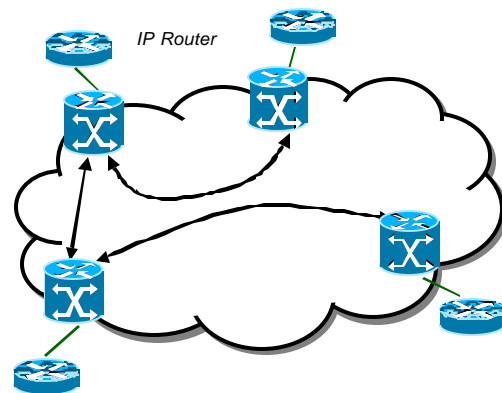


Figure 4: A Logical View Provided to IP

the optical domain. Only exceptional was that by viewing the wavelength as the label, label swapping can be performed by changing the incoming wavelength to the different wavelength at the optical cross-connect switch. However, a high-speed wavelength conversion is difficult to perform on a packet-by-packet basis by the current tech-

nology, and therefore, functionalities of Core LSRs are very limited in λ -MPLS.

Thus, the lightpath should be setup in a circuit-switched fashion between Ingress/Egress LSRs. For IP, it is very natural that all-to-all connectivity among every LSRs should be established. However, too many wavelengths are necessary for that purpose. See Figure 5 as an example, which shows that four wavelengths are necessary on the link between nodes N_2 and N_5 for all-to-all connectivity, i.e., the traffic demand is given as

$$T_1 = \begin{Bmatrix} - & 1 & 1 & 1 \\ 1 & - & 1 & 1 \\ 1 & 1 & - & 1 \\ 1 & 1 & 1 & - \end{Bmatrix}$$

To alleviate the above-mentioned problem, the lightpath within the network is intentionally split. In this approach, it may take two or more lightpaths for the packets to reach the destination Egress LSR. Also, the IP routing capability is necessary within the network. Again, see Figure 3. Suppose that we have a traffic flow between nodes N_2 and N_4 . Since

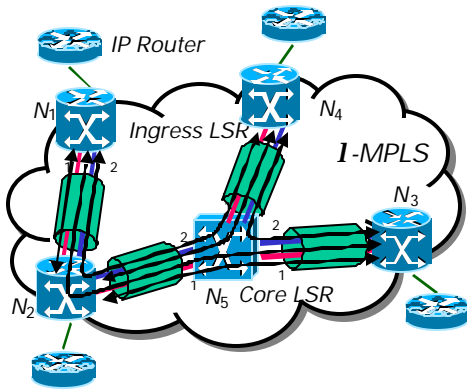


Figure 5: All-to-All Connectivity Needs Four Wavelengths

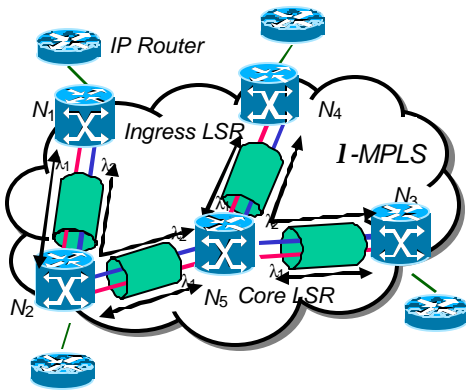
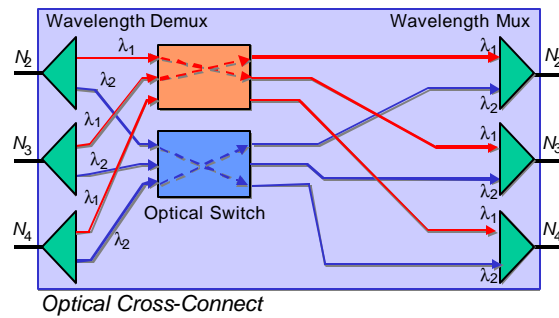


Figure 6: Lightpath Splitting

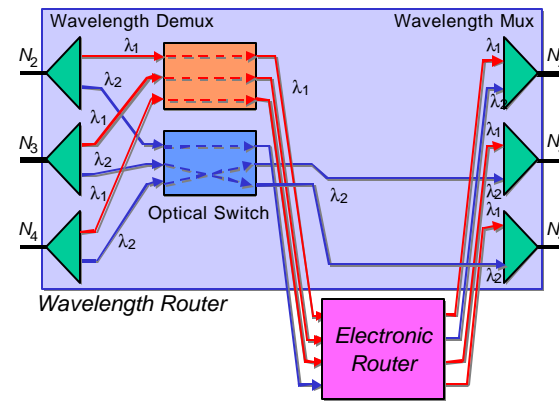
we have only two wavelengths on the WDM link and there are no more wavelengths on the fiber, we cannot setup another direct path between those nodes. Then, the lightpath should be split within the MPLS domain. See Figure 6, where traffic from nodes N_2 to N_4 should take two hops of lightpaths: from N_2 to N_5 (Core LSR) and N_5 to N_4 . This means that packets on the same lightpath from N_2 have different destinations N_3 and N_4 . The packet forwarding capability at the IP layer is therefore necessary at the Core LSR as shown in Figure 7(b). For reference purpose, we also illustrate the Core LSR with only optical cross-connect in Figure 7(a). This is another option of MPLS that Core LSRs have a capability of packet forwarding at the IP layer. It can allow the smaller number of wavelengths on the fiber, but requires additional packet processing delays in the electronic domain at the intermediate Core LSRs within the MPLS domain.

Once the logical topology is obtained, the following functions are necessary in λ -MPLS.

- 1) Ingress LSR; maps an IP address to a wavelength label.
- 2) LSP (Label-Switched Path); the labeled wavelength, i.e., lightpath.



(a) Core LSR by Optical Cross-Connect Switch



(b) Core LSR with Electronic Packet Routing Capability

Figure 7: Structure of Core LSR; Node 5

- 3) Core LSR (Core Label Switching Router); an optical cross-connect switching directly connecting input wavelength to output wavelength. Packet forwarding capability at the IP layer may be necessary if packets with different labels share the same lightpath.
- 4) LDP (Label Distribution Protocol); the logical topology design algorithm is utilized to implement the signaling protocol. However, most of the currently available design methods are not suitable to the operating λ -MPLS, which will be described in more detail in Subsection 2.4.

2.3 Optical Implementations

A successful implementation of λ -MPLS depends on how many wavelengths we can multiplex on the fiber, by which many end-to-end connections can be attained. A key optical component for realizing WDM includes optical cross-connect switch and multi-wavelength light source. A promising technology for optical cross-connect switch is micro-electro mechanical systems (MEMs) technology [Lin98]. A micromechanical space optical switch fabric is constructed by using tiny movable mirrors controlled by actuator on silicon chip based upon MEMs. Two-axis beam steering tiny mirror routes light beam between in/output fiber bundles. The mirror rotates up to 90 degrees by fine step by step by applying bias pulses to actuator. Advantages include large port count, scalability, low loss and high uniformity of loss. The switching speed is relatively slow, but the reconfiguration time of a few millisecond will meet this switching speed. A fully provisioned 112 x 112 micro-mechanical optical cross-connect switch with 35.8Tbit/s has been demonstrated [Nel00].

Another key component is a multi-wavelength light source. As the number of the WDM links of the network to which the node is interfaced increases, it seems impractical to provide the light sources as many as the WDM channel count. A supercontinuum (SC) pulse source is promising, which generates picosecond pulses at several tens of Gbps over an extremely broad spectral range [Tak00]. The seeding short pulse at a specific repetition rate is broadened continuously in its spectrum due to the fiber nonlinearities, and thus by filtering desired wavelength components out of the SC spectrum the single light source can serve a multi-wavelength pulse source. Advantages are; it uses only one pump laser, and its fixed channel spacing with accuracy equivalent to that of a microwave oscillator (MHz), enabling to lock the entire chain to absolute standard by locking just one mode of the chain. Recently, 1,010-channel with the frequency interval of 12.5GHz at 2.5Gbit/s over 100nm in 1550nm spectral region has been achieved.

2.4 Challenges for λ -MPLS

In this subsection, the remaining and challenging problems for λ -MPLS are summarized. A first and the most difficult problem is a capacity granularity problem as described in the previous subsection.

Label (wavelength) merging and splitting in optical domain is fundamentally impossible in λ -MPLS while some approach are being tried. See, e.g., [Ban00]. It is closely related to the capacity granularity problem, and it inhibits flexibly and effectively use of the large capacity of the wavelength.

The next problem is how to establish the logical topology using lightpaths suitable to deploying IP over WDM networks. The existing approaches impose

- 1) It is assumed that the traffic load matrix is known a priori.
- 2) The entire topology is determined by performing the topology design algorithm.

The assumption described in 1) is unrealistic in the data network where it is impossible to forecast the future traffic demand. The traffic load may be estimated by the traffic measurement. It implies that the traffic load can be known only after the network is established and operated. However, the existing approaches can determine the entire topology only after the traffic load is known.

The problem 2) in the above relies on a central controller that collects the long-term traffic statistics of the network, and perform an optimization problem to identify the wavelength assignment. The virtual topology control could be done in a centralized manner due to its long-term nature. However, it is a computational burden to obtain an optimal configuration. Even if the heuristic but effective optimization approach is applied, the whole network should be re-configured to realize the desired topology. More importantly, it is unrealistic whether this could be implemented in an operational network without imposing severe penalties on users [Ban00]. Thus, for establishing LDP suitable for λ -MPLS, we should seek a new approach, in which the service interruption on the operational network can be avoided or at least minimized. Only a few exceptions to counter this problem are channel rerouting using color-interchange techniques [Gha99] and virtual topology re-configuration techniques utilizing a branch-exchange method [Muk97]. A further research is necessary in this area.

The bottleneck at the Ingress LSR is another problem. One solution is to distribute packet processing by introducing WDM ring at the Ingress LSR. See Figure 8, where packet processing at the Ingress LSR is distributed among three LSRs to reduce the processing overhead. The problem is apparently that more wavelengths are necessary on the fibers [Mur00]. In the case of Figure 8, three wavelengths

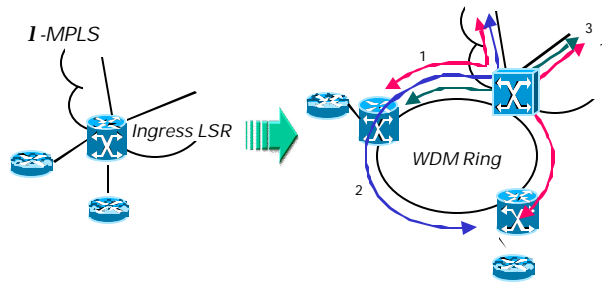


Figure 8: Packet Processing Distribution by WDM Ring

are necessary on the WDM link.

Survivability is another concern in the WDM network. The reliability can much be improved by preparing the backup path. This is realized by extending the logical topology design method (see, e.g., [Car96, Cro98]), but the deployment in the operating network without service interruption is not considered in most of works.

3. OC-MPLS Based Upon Optical Code Correlations

In photonic packet switching there has been a typical approach that switching is performed in optical domain, while, routing, forwarding, and buffering rely on electronic processing capability. Efforts are continuously being made to realize high-speed routing, forwarding, and buffering in optical domain. Recently, there have been several demonstrations of photonic MPLSs other than λ -MPLS, in which the label recognition and swapping are carried out either in opto-electronically or optically. Before we go in detail of OC-MPLS, we will briefly survey recent progresses in the other photonic label switching techniques.

First, low bit rate subcarrier multiplexed (SCM) signal is used to carry label information, while the baseband signal carries payload data [Car98, Lin00]. The subcarrier header recovery is performed using optical and microwave direct detection. While the cost of some microwave components such as local oscillator, mixer, and filter has already reached low levels, the economic advantage can be lost because the microwave detection and the electrical header processing of the header may limit the applications to moderate and relatively low bit rates. In higher bit rate applications, for example, 1,000bit long packet at 100Gbit/s, the packet duration is only 10ns, and to transmit 10bit long header its bit rate must be higher than a Gbps with the subcarrier frequency in the millimeter-wave region, well above 100GHz. This would be a significant challenge.

The next approach uses time-to-wavelength mapping [Car00]. Each bit of header bits is mapped onto a different wavelength in a different time position, and the optical de-

coding can find a matched code with the header code from the optically encoded lookup table. As the header recognition is performed in the optical domain, it has a potential of ultra-high speed operation. However, as the time-to-wavelength mapping occupies a wide wavelength range, the wavelength allocation in WDM networks could be a problem.

3.1. OC-label Processing

Recent progress in high bit rate optical code division multiplexing (OCDM) [Kit98] facilitates to utilize optical codes as the photonic labels. In OC-MPLSR at a node any connection between incoming and outgoing WDM links of different fibers can be established See Figure 9. OC-MPLS must perform the task within each WDM link of a single wavelength, that is, WDM links of different wavelengths have to be interfaced with different OC-MPLSR. This allows to reuse the same set of optical codes on different wavelengths, Here, we exclude the optical code division multiplexing (OCDM) on a WDMlink, and therefore, any packets having different OC labels on a WDM link are not overlapped in the time domain.

Architecture of OC-MPLSR is shown in Figure 10. The building blocks are photonic switch/buffer and the photonic label processor, and the photonic label swapper if required. The photonic label is a sequence of a few short optical pulses, so-called chip pulses. There are two types of optical codes; one uses bipolar phase-shift keying (BPSK) in which phase of each chip pulse takes the value of 0 or π , and the other uses On-Off keying (OOK) in which "1" and "0" are represented by the existence and non-existence of chip pulse.

The label processing is carried out as follows (see Figure 11). A photonic label is tapped from the input packet, optically duplicated by optical amplification, and power-splitting as many copies as the count of label entries in the table. Then, all the optical correlations between the copies and the label entries can be simultaneously performed in parallel. An auto-correlation peak emerges only where the exact match is obtained, while, on the contrary, all the other correlations show cross-correlation outputs. Therefore, the thresholding of the outputs can identify the input label. The photonic label recognition using 8-chip BPSK optical codes has been successfully demonstrated at the bit rate of 10Gbps [Wad00].

Unique to the photonic label processing is that no optical logic operation is involved, which has been the toughest task for photonics to deal with, but only the optical correlation is used. The most advantageous is that the optical code generation and the optical correlation can be simply performed by using passive optical device such as a special

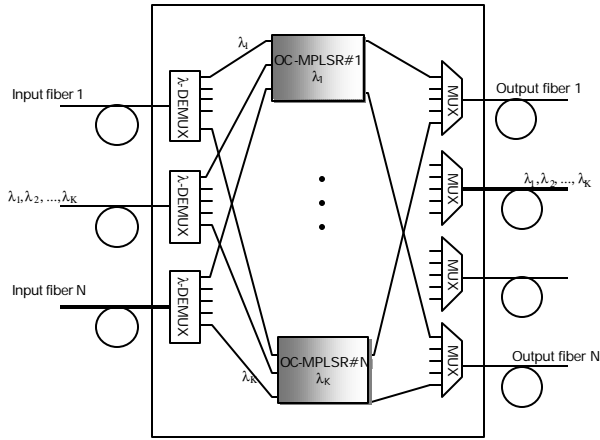


Figure 9: OC-MPLSR at a Node Configured with Wavelength MUX/DEMUX

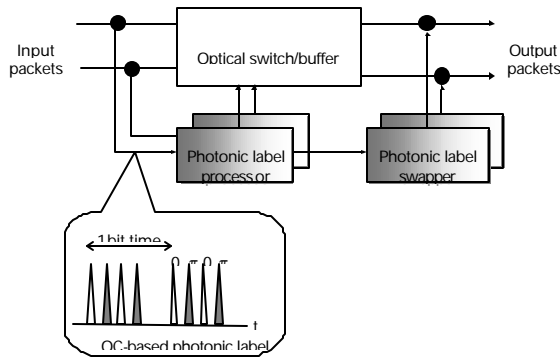


Figure 10: Architecture of OC-MPLSR

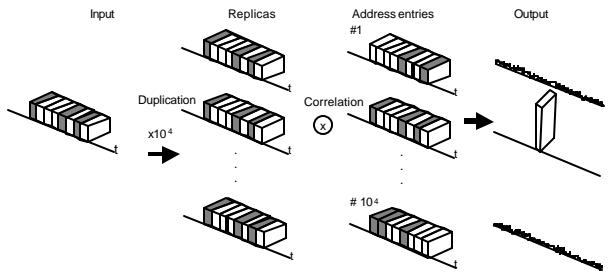


Figure 11: Parallel Optical Correlation for Photonic Label Processing

class of waveguide tapped delay line or fiber Bragg grating (FBG), and thus the processing speed is governed by the speed of light in the device.

It is crucial for OC-MPLS to transmit the photonic label for a long distance between nodes without serious waveform distortions. The chromatic dispersion and nonlinearities of optical fiber could have the deteriorating effects. The numerical simulation confirms that after 50km long propagation, the code sequence severely deforms, however, the eye opening after the time gating demonstrates a good

discrimination from the interference fiber (see Figure 12). By contrast, it was also numerically simulated that TDM 200Gbps IM-DD in RZ format pulse is severely impaired, and its eye diagram is closed only after around 5km. This will guarantee the error-free transmission over 50km but no longer than 100km. As a result, provided that an appropriate compensation scheme is introduced to overcome the dispersion effect, high bit rate photonic label can be transmitted over 50km.

Photonic label swapping for OC-MPLS at 10Gbps has been demonstrated in optical domain [Sob00]. The incoming optical code can be converted to the desired optical code by interacting with the optical control signal through cross phase modulation due to a fiber optic nonlinearity, optical Kerr effect. The quasi-instantaneous response of the nonlinearity favorably meets the demand for the ultra-fast operation.

3.2. Photonic Buffering

A conventional router in electronics technology employs a store-and-forward technique to resolve the contention of packets. Since optical random access memory (RAM) is not expected to become available in the near future, the photonic buffer must be implemented by using optical fiber delay line. Therefore, OC-MPLSR has to employ a unique buffering strategy to avoid abolishing packets when the contention occurs at the output ports. The photonic buffer is introduced inside the switch fabric in the architecture (see Figure 10). It is categorized into two; types of feed-forward and feedback. Several types of photonic buffers have been demonstrated; feedback using fiber re-circulating loops. The feed-forward buffer is structured with multistage 2x2 switch elements containing fiber delay line [Ch196] (see Figure 13(a)), while the feed-forward buffer utilizes fiber re-circulating loop, which employs optical amplifier inside the loop to compensate the round-trip loss [Hal98] (Figure 13(b)). The most important yardstick of the performance will be the buffer depth, while it maintains a good optical signal-to-noise ratio (OSNR) by coping with optical amplifier noise and compensating the optical loss. A 1,280-bit long packet (32ns) at 40Gbit/s has been stored for 320 round-trips in the experiment [Hal98]. This will guarantee the delay line type of photonic buffer be practically available.

Packet has to be of either fixed length or multiple of the fixed length because there is no way to retrieve the packet anytime earlier than the minimum delay time of the photonic buffer. The buffering strategy along with the architecture of the switch fabric may change according to the traffic pattern whether the load is high or not [Par94]. There are various designs for optical buffering, for example,

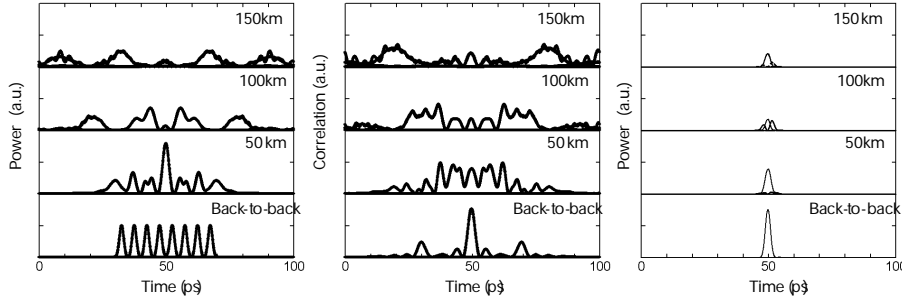


Figure 12: Temporal waveform of 8-chip equi-amplitude BPSK code sequence after the propagation in a long dispersion shifted fiber, its auto-correlation before and after time gating from the left. The calculations are made under the following conditions; wavelength dispersion $D = -0.285\text{ps/nm/km}$ and the loss $= 0.2\text{dB/km}$. Fiber nonlinearity is ignored.

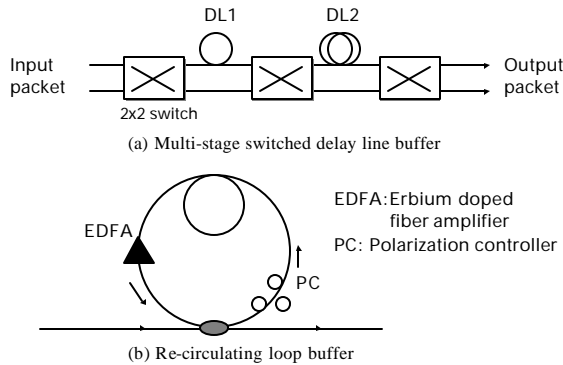


Figure 13: Implementations of Optical Buffers by Delay Lines

contention resolution delay lines (CORD) [Chl96], switch with large optical buffer (SLOB) [Hun98], and cascaded optical delay line (COD) [Cru96]. The other strategy is deflection routing, by which the overflow packets is directed to the unoccupied switch output and later re-directed to the destination via another route [Aca92].

A large space switch fabric can be constructed by incorporating these photonic buffers. For this purpose, we may be able to apply the past researches on the ATM switching techniques such as a Banyan network. The remaining problem is that since we cannot expect sufficiently large buffer by delay lines (e.g., in the order of thousand cell buffers as in the ATM switching fabric), the appropriate traffic engineering approach would be necessary to limit the traffic load at the switch, and it is left to be a future research topic.

3.3 Challenges for OC-MPLS

Let us predict the ultimate performance of OC-MPLSR. The processing time for the photonic label recognition is evaluated by $(N - 1)\Delta\tau$, where N and $\Delta\tau$ represent the number of chip pulses and the time interval between the chip pulses, respectively. It has been experimentally shown that optical correlation of 8-chip long optical code for the photonic label recognition takes 35ps [Wad00]. Ultimately, the processing time will come down to 8ps with $\Delta\tau=1\text{ps}$ and $N=8$, resulting in 125Giga lookups/s. It is reminded that the

photonic label recognition is executed only once for each incoming packet, because all the correlations are simultaneously carried out in parallel. See again Figure 11. This performance is quite a significant leap from electronic router's capability of table lookup of 12.5 million lookups/s as described in Introduction.

This prediction is not groundless because such an optical code generation should be feasible with the laser source generating 1ps optical pulse in the wavelength region of 1550nm, which is commercially available and silica-based planar lightwave circuit technology [Him98]. Note that 8-chip code represents 128 labels, although it is uncertain if this number is sufficient for the core network applications. If the overall switching time is mainly governed by the photonic label recognition, this speed is fast enough, for example, compared with the time duration of a 1,000bit long packet at 1Tbit/s is 1ns. In reality, however, the bottleneck of overall switching speed may arise from the others; one may be caused by the electronic controller which drives the switch, and the other may be due to the controller of photonic buffer. These must be taken into account in the performance evaluation, and further study will be needed.

4. Perspective on λ -MPLS and OC-MPLS

As concluding remarks, we first summarize problems of λ -MPLS.

- 1) The unit of path granularity is the wavelength capacity, which seems to be too large to accommodate the end-to-end traffic within the MPLS domain. Or an application field of the MPLS technology is very limited. The capacity increase per each wavelength does not alleviate the above problem.
- 2) The increase in the number of wavelengths may help it. However, it requires the large-scaled, for example, of 1,000x1,000 optical cross-connect, which is still under researches.
- 3) Flow aggregation at the Core LSR cannot be expected. The label exchange within the network poses

the wavelength change at an optical node, which is still difficult.

- 4) The incremental capacity dimensioning is infeasible in the current logical topology design approach, in spite of the fact that network performance is heavily dependent on the logical topology design approach.

An OC-MPLS technology has a capability of resolving above problems as we have described;

- 1) The granularity is the packet, which can lead to a flexible network structure, and simplified packet switching can offer large capacity in the optical domain, but still an ATM-based MPLS protocol suite can be applied. Traffic engineering developed for MPLS is also utilized.
- 2) OC-MPLS is capable of merging of packets by introducing optical buffering, thus attaining an ultimate bandwidth efficiency, while, on the contrary, λ -MPLS is unable to realize it due to the coarse granularity.
- 3) Unlike λ -MPLS, the length of OC photonic label could be flexible as is required. The longer label could be used as the network layer header of the packet, and it can also be used both to assigned multiple flows from IP prefix to application-level flow and distinguish its protocols such as TCP and UDP. It can also be utilized to serve provisioning various types of QoS. Optical codes are not only applicable to the exact match algorithm in the OC-MPLS but also applicable to the longest prefix match, and hence OC-based destination-based IP routing could be realizable.

Of course, OC-MPLS has its own problems. A most difficult obstacle is how to provide the packet buffer. According to the current technology, a synchronous switching at the packet level is necessary at the node. It implies that the packets are of fixed-size, and randomly arrived packets have to be time-aligned at the input ports. The possibility of handling the asynchronous and variable-size packets is discussed in [Tan00], but it requires a very complicated buffer scheduling in an electronic domain. It means that it is likely to introduce another bottleneck. More recently, an asynchronous and variable-length packet switch is proposed in [Mur01], where photonic label processing is utilized for packet processing within the switch. It is promising for achieving a lightweight packet scheduling, but it requires a careful design consideration on the buffer dimensioning.

The photonic buffer is implemented by the optical fiber delay lines as having been described in Subsection 3.2, but there is a shortage in the buffer size. Given these conditions, we need further researches:

- 1) The switch fabric, constructed with photonic space

switch and photonic buffer, has to be optimized to achieve the desired performance.

- 2) Fortunately, statistical multiplexing would work well for very huge bandwidth by OC techniques. For this purpose, however, a further research on the traffic engineering approach of MPLS is necessary under the conditions that the bandwidth is very large, but the packet buffer size is small.

There is no doubt that λ -MPLS alleviates the bottlenecks in electronic routers in IP over WDM networks in the near future, but it could not be a ultimate solution. OC-MPLS could be a breakthrough in the long run to completely resolve the bottlenecks, although, researches on OC-MPLS is still at the early stage, and there is a list of research subjects to be tackled.

References

- [Aca92] A. S. Acampora and I. A. Shah, "Multihop Lightwave Networks: A Comparison of Store-and-Forward and Hot-Potato Routing," *IEEE Transactions on Communications*, Vol.40, pp.1082-1090, 1992.
- [Arn00] G. Armutage, "MPLS: The Magic Behind the Myths," *IEEE Communications Magazine*, pp.124-131, January 2000.
- [Ban00] J. Bannister, J. Touch, A. Willner, and S. Suryaputra, "How Many Wavelengths Do We Really Need? A Study of the Performance Limits of Packet over Wavelengths," *Optical Networks Magazine*, pp.17-28, April 2000.
- [Cae98] B. V. Caenegem, W. V. Parys, F. D. Turck, and P. M. Demester, "Dimensioning of Survivable WDM Networks," *IEEE Journal on Selected Areas in Communications*, Vol.16, No.7, pp.1146-1157, September 1998.
- [Car98] A. Carena, M.D. Vaughn, R. Gaudino, M. Shell, and D.J. Blumenthal, "OPERA: An Optical Packet Experimental Routing Architecture with Label Switching Capacity," *Journal on Lightwave Technology*, Vol.16, pp.12135-2145, 1998.
- [Car00] M. C. Cardakli, D. Gurkan, S. A. Havstad, A. E. Willner, "Variable-bit-rate Header Recognition for Reconfigurable Networks using Tunable Fiber-Bragg-Gratings as Optical Correlators," *2000 Optical Fiber Conference (OFC2000)*, TuN2, Baltimore, March 2000.
- [Cha00] H. Chaskar et al. "A Framework to Support IP over WDM using Optical Burst Switching," *Workshop on Optical Networking*, Dallas, January 2000.
- [Cha96] I. Chlamtac, A. Fumagalli, L. G. Kazovsky, P. Melman, W. H. Nelson, P. Poggiolini, M. Cerisola, A. N. M. Choudhury, T. K. Fong, R. T. Hofmeister, C. -L. Lu, A. Mekikittikul, D. J. M. Sabido, C. -J. Suh, and E. W. M. Wong, "CORD: Contention Resolution by Delay Lines," *IEEE Journal on Selected Areas in Communications*, Vol.14, pp.1014-1029, 1996.
- [Cro98] O. Crochat, and J.-Y. L. Boudec, "Design Protection for WDM Optical Networks," *IEEE Journal on Selected Areas in Communications*, Vol.16, No.7, pp.1158-1165, September 1998.
- [Cru96] R. L. Cruz and J. -T. Tsai, "COD: Alternative Architectures for High Speed Packet Switching," *IEEE/ACM Transactions on Networking*, Vol.4, pp11-21, 1996.
- [Dan00] D. O. Awduche, Y. Rekhter, J. Drake, and R. Coltun,

- "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects," *IETF Internet Draft*, draft-awduche-mpls-te-optical-02.txt.
- [Dav98] B. Davie, P. Doolan, and Y. Rekhter, *Switching in IP Networks - IP Switching, Tag Switching, and Related Technologies*, Morgan Kaufmann, 1998.
- [Dur00] R. Dutta and G. N. Rouskas, "A Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks," *Optical Network Magazine*, Vol.1, January 2000.
- [Gha99] N. Ghani and S. Dixit, "Channel Provisioning for Higher-Layer Protocols in WDM Networks," *Proceedings of SPIE All Optical Networking Conference: Architecture, Control, and Management Issues*, Boston, September 1999.
- [Gha00] N. Ghani, S. Dixit, and T.-S. Wang, "On IP-over-WDM Integration," *IEEE Communications Magazine*, pp.72-84, March 2000.
- [Hal98] K. L. Hall and K. A. Rauschenbach, "All-Optical Buffering of 40-Gb/s Data Packets," *IEEE Photonic Technol. Lett.*, Vol.10, pp.442-444, 1998.
- [Him98] A. Himeno, K. Kato, and T. Miya, "Silica-based Planar Lightwave Circuit," *IEEE J. Selected Topics in Quantum Electron.*, Vol.4, pp.913-924, 1998.
- [Hun98] D. Hunter, W. D. Cornwell, T. H. Gilfedder, A. Frnzen, and I. Andonovic, "SLOB: A Switch with Large Optical Buffers for Packet Switching," *Journal of Lightwave Technologies*, Vol.10, pp.1725-1736, 1998.
- [Ito00] T. Ito, K. Fukuchi, K. Sekiya, D. Ogasahara, R. Ohhira, T. Ono, "6.4TB/s (160x40Gbit/s) WDM Transmission Experiment with 0.8bit/s/Hz Spectral Efficiency," *European Conference on Optical Communication (ECOC2000)*, PD-1.1, Munich, Sept. 2000.
- [Kes98] S. Keshav and R. Sharma, "Issues and Trends in Router Design," *IEEE Communications Magazine*, pp.144-151, May 1998.
- [Kit98] K. Kitayama, "Code Division Multiplexing Lightwave Networks based upon Optical Code Conversion," *IEEE Selected Areas in Communications*, Vol.16, pp.1309-1319, 1998.
- [Kit99] K. Kitayama and N. Wada, "Photonic IP Routing," *IEEE Photonic Technology Letters*, Vol.11, pp.1689-1691, 1999.
- [Lin97] S. Lin and N. McKeown, "A Simulation Study of IP Switching," *Proceedings of ACM SIGCOMM '97*, pp.15-24, September 1997.
- [Lin98] L. Y. Lin, E. L. Goldstein, and R. W. Tkach, "Free-Space Micromachined Optical Switches with Submillisecond Switching Time for Large-Scale Optical Cross-connect," *IEEE Photon. Technol. Lett.*, Vol. 10, pp.525-527, 1998.
- [Lin00] Y.-M. Lin, W. I. Way, Y. -M. Lin, and G. -K. Chang, "A Novel Optical Label Swapping Technique Using Erasable Optical Single-Sideband Subcarrier Label," *IEEE Photonic Technology Letters*, Vol.12, No.8, pp.1088-1090, 2000.
- [Muk96] B. Mukherjee, D. Banerjee, S. Ramamurthy and A. Mukherjee, "Some Principles for Designing a Wide-Area WDM Optical Network," *IEEE/ACM Transactions on Networking*, Vol.4, No.5, pp.684-695, October 1996.
- [Muk97] B. Mukherjee, *Optical Communications Networks*, McGraw Hill, New York, 1997.
- [Mur00] M. Murata and K. Kitayama, "IP over Thousand-Wavelength Division Multiplexing: Is It Useful and Possible for Resolving the Network Bottlenecks?," *submitted for publication*.
- [Mur01] M. Murata and K. Kitayama, "A Photonic Switch Supporting Asynchronous and Variable-Length Packets based on Photonic Label Processing," *submitted for publication*, 2001.
- [Nei00] D. T. Neilson, V. A. Aksyuk, S. Arney, N. R. Basavanahally, K. S. Bhalla, D. J. Bishop, B. A. Boie, C. A. Bolle, J. V. Gates, A. M. Gottlieb, J. P. Hickey, N. A. Jackman, P. R. Kolodner, S. K. Korotky, B. Mikkelsen, F. Pardo, G. Raybon, R. Ruel, R. E. Scott, T. W. Van Blaricum, L. Zhang, and C. R. Giles, "Fully Provisioned 112x112 Micro-mechanical Optical Crossconnect with 35.8Tb/s Demonstrated Capacity," *2000 Optical Fiber Conference (OFC2000)*, PD12-1, Baltimore, March 2000.
- [Par94] C. Partridge, *Gigabit Networking*, Addison-Wesley Publishing Company, 1994.
- [Qia00] C. Qia, "Labeled Optical Burst Switching for IP-over-WDM Integration," *IEEE Communications Magazine*, pp.104-114, Sept. 2000.
- [Ram96] R. Ramaswami and K. N. Sivarajan, "Design of Logical Topologies for Wavelength-Routed Optical Networks," *IEEE Journal on Selected Areas in Communications*, Vol.14, pp.840-851, June 1996.
- [Sob00] H. Sotobayashi and K. Kitayama, "Optical Code based Label Swapping for Photonic Routing," *IEICE Transactions on Communications*, Vol.E83-B, No.10, pp.2341-2347, 2000.
- [Tak00] H. Taara, T. Ohara, K. Mori, K. sato, E. Yamada, T. Morioka, K. -I. Sato, K. Junguji, Y. Inoue, and T. Shibata, "Over 1000 Channel Optical Frequency Chain Generation from a Single Supercontinuum Source with 12.5GHz Channel Spacing for DWDM and Frequency Standards," *European Conference on Optical Communication (ECOC2000)*, PD-3.1, Munich, Sept. 2000.
- [Tan00] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, "Optical Routing of Asynchronous, Variable Length Packets," *IEEE Journal on Selected Areas in Communications* Vol. 18, No. 10, pp.2084-2093, October 2000.
- [Vis98] A. Viswanathan, N. Feldman, Z. Wang, R. Callon, "Evolution of Multiprotocol Label Switching," *IEEE Communications Magazine*, pp.165-173, May 1998.
- [Wad00] N. Wada and K. Kitayama, "Photonic IP Routing using Optical Codes: 10Gbit/s Optical Packet Transfer Experiment," *2000 Optical Fiber Conference (OFC2000)*, WM51, Baltimore, March 2000.