Optical tree topologies:
Access control and wavelength assignment

Mario Gerla \textsuperscript{a}, Milan Kovačević \textsuperscript{*,a}, Joseph Bannister \textsuperscript{b}

\textsuperscript{a} UCLA Computer Science Department, Los Angeles, CA 90024-1596, USA
\textsuperscript{b} The Aerospace Corporation El Segundo, CA 90245-4691, USA

Abstract

Fiber optic tree topologies are emerging as attractive candidates for Passive Optical Network (PON) implementations, next to the more traditional star and bus topologies. In this paper, we first address the "physical" properties of the optical tree topology (like layout cost, geographical coverage, power budget and fault tolerance) and compare them to similar properties in buses and stars. Then, we discuss the access schemes which can be implemented on a tree. We discuss the introduction of Wavelength Division Multiplexing (WDM) in the PON and describe single-hop and multi-hop "virtual" topologies which can be supported. Finally, we present MONET, a Multilevel Optical Network architecture which can be efficiently implemented on a tree topology.

1. Introduction

Fiber Optic Local Area Networks (FOLANs) have become increasingly popular for applications involving the exchange of data at very high rates. The applications are as diverse as the transfer of large files, the handling of real-time control data, the interconnection of LANs, and the integration of data, voice and video services. It is to this last application (i.e. integration of services) that the high bandwidth of the FOLAN is ideally suited. Therefore, much of the current research on FOLANs is directed at providing an environment supporting both real-time traffic (voice, video and image) as well as the more traditional computer data traffic (interactive communication and file transfers).

The first generation of FOLANs evolved from previous LAN designs on a bus, ring or star. The most important examples are the Fiber-Distributed Data Interface (FDDI) \cite{27} and the Distributed Queue Dual Bus (DQDB) \cite{25}. First generation architectures, however, cannot be economically operated at speeds beyond a gigabit/second (Gb/s), because each station's interface represents an electro-optical bottleneck that limits network speed. Thus, though adequate for the near term, they will not scale up to aggregate speeds beyond tens of Gb/s and will not be able to provide service to the several thousands of users who are expected on a FOLAN in the future.

In an effort to overcome these limitations, a second generation of FOLANs has been proposed during the past few years. These networks
can achieve terabit/second (Tb/s) capacity by exploiting Wavelength Division Multiplexing (WDM) techniques [7]. These second generation FOLANs are, for the most part, based on Passive Optical Networks (PONs), i.e. networks in which the data stream does not undergo any processing (like control field inspection or storing and forwarding) along the transmission path on the optical topology. Only all-optical amplification and wavelength filtering are allowed. Thus, the topology can be viewed as a “mass of glass” [1].

In order to provide in a PON full interconnectivity among the attached stations, without traversing multiple optical paths, the physical topology must be restricted to three choices: star, bus or tree (or proper combinations of thereof). In contrast, the “virtual” topology, which is the interconnection graph induced by the channels (wavelengths) multiplexed on the fiber, can take many different forms. The ring, the ShuffleNet and the de Bruijn graph are only a few examples [3]. Likewise, a great variety of schemes have been proposed for channel access and fiber medium sharing in PONs (e.g. TDMA, WDMA, combined time and wavelength division, code division) [16]. More recently, techniques for the dynamic sharing of multiple fibers installed in the same conduit have also been proposed [5]. It has been observed that the choice of topology layout is to a great extent independent of the choice of the virtual topology and the access scheme, whereas the latter two are much more closely interrelated. That is, many virtual topologies and access schemes can be implemented equally well on a bus, star or tree.

In this paper we focus on the investigation of passive tree topologies. Exploiting the relative independence between physical topology and other system considerations, we structure the paper in two parts. In the first part we address the “physical” properties of the optical tree topology and compare them with the corresponding properties of buses and stars. The key aspects examined are power budget, geographic coverage, layout cost, and fault tolerance.

In the second part, we investigate the “system” aspects (namely, access scheme and wavelength allocation) related to the tree topology. We point out that the tree (and star) topology does not allow the “attempt-and-defer” access mode supported by the bus topology. It does, however, provide other efficient forms of access synchronization. As for wavelength allocation, we review the two classes of “virtual” topologies (single-hop and multi-hop) which have been proposed in the past for a tree-shaped PON. We identify some limitations on either a single- or a multi-hop scheme. These limitations prompt us to propose a hybrid architecture, MONET (Multilevel Optical Network).

2. Physical properties

We begin by investigating power budget aspects. Specifically, we are interested in computing the number of stations that can be supported by a symmetric binary tree configuration of the type shown in Fig. 1, using state-of-the-art transmitters and receivers. Assume a binary coupler loss equal to 4 dB (i.e., 3 dB power splitting plus 1 dB extra loss which includes single mode fiber attenuation loss on the links of the tree). Let $M$ be the power margin (defined as the ratio between laser transmit power and the minimum power detectable at the receiver in dB. A typical value

![Fig. 1. A tree with 8 stations.](image-url)
of $M$ is 40 dB for a single mode fiber, direct signal detection and 1 Gb/s signal rate (e.g., 0 dBm transmit power and −40 dBm receiver sensitivity). Let $N$ be the number of stations.

To comply with the power margin, the maximum power loss from a station to the root is allowed to be half of the total budget, thus $M/2$. Since at each level a 4 dB coupler loss is incurred, the maximum number of levels in the tree is $[M/8]$. Therefore, it follows that the maximum number of stations $N$ is:

$$N = 2^{\lceil M/8 \rceil}$$

For $M = 40$ dB this gives $N = 32$.

Station connectivity can be significantly improved by introducing an all-optical amplifier at the root. Assuming an amplifier gain of 40 dB, the number of stations that can be supported now becomes:

$$N = 2^{\lceil M/41 \rceil}$$

For $M = 40$ dB this results in $N = 1024$. Another way to increase station connectivity $N$ is by interconnecting several trees with a modular $P \times P$ star coupler (a $4 \times 4$ star coupler example is shown in Fig. 2). The modular star coupler is built by interconnecting several stages of single mode, $2 \times 2$ couplers. The number of stages in the modular star coupler is $\log_2 P$, where $P$ is the number of ports. Each stage introduces 4 dB attenuation. Using the previous assumptions, a system with a 4-stage star and 3-level trees can support 128 stations even without amplification. If amplification is used, the modular star permits a dramatic increase in the total number of stations. For example, consider a $32 \times 32$ star (i.e. 5 stages) to which we interconnect 32 subtrees, each with 1000 stations. To provide a comfortable power margin, amplifiers with 30 dB gain are installed at both the inputs and the outputs of the star. The total number of stations is thus 32,000. Configurations of up to 1,000,000 stations can be obtained with hybrid tree/star schemes [12].

For the sake of comparison, let us now consider a linear, folded bus. With three taps per stations, uniform coupling ratio (i.e., 50% splitting, or 3 dB loss, for each coupler), and negligible coupler extra loss, the maximum station connectivity is $N = 4$. With very careful, position-dependent, optimal adjustment of the coupling ratios, and assuming negligible extra loss, up to $N = 20$ stations can be connected without amplification. This result, however, is very difficult to achieve in a practical setting, with off-the-shelf couplers. To support 1000 stations on a folded bus, from 50 to 250 amplifiers are thus required (with optimized and uniform coupling ratios, respectively), as opposed to a single amplifier at the root of a tree!

In comparing tree and star, we note that the star performs best if no amplification is allowed. In fact, based on our assumptions, a star can support up to 1000 stations (with a $1000 \times 1000$ coupler) without amplification. However, such large star couplers are difficult to fabricate. Furthermore, a very large number of amplifiers (one per port) is required to go beyond $N = 1000$ in pure star configurations. The tree, and more generally the hybrid tree/star topology requires much fewer amplifiers, which can be conveniently installed at the roots of the subtrees.

Another important concern in PONs is power calibration. Note that in all of the above schemes the power received by a station may vary because of unequal transmit power, topology and geographical asymmetry, as well as nonuniform optical amplifier gains. For proper reception, transmit-power calibration is required to assure that stations receive the same power regardless of which station transmits. In a tree topology, this is
easily accomplished, because each station compares its own received power with that received from a reference station, and it adjusts its output power (using an attenuator for example) until the received power is identical to that of the reference station [12].

One of the most important applications of a PON will be the metropolitan area network. Thus, cost effective coverage of a large geographical area is an important requirement. It is intuitive that a tree topology leads to lower-cost layouts than a star or bus. This intuition is confirmed by quantitative results based on the comparison of a variety of topologies for a vast sample of node locations [3]. Those results clearly show the cost advantage of trees over stars. In a large, multi-level tree it may be practical to cluster several layers of the tree in a central office and in distribution centers, in order to simplify the installation [12].

The last (but not least) concern of the PON topology is tolerance to faults, such as unintentional fiber cuts. This becomes a critical issue if PONs are deployed in metropolitan areas and are therefore subject to physical abuse beyond the control of network operators. For example, a fiber conduit could be inadvertently damaged by a street maintenance crew. Similar incidents affecting metropolitan, underground cable plants are routinely reported in the news. It is likely that the same will occur to fiber-based cable plants. Tree topologies are clearly very vulnerable to cuts. One single cut can disconnect a major section of the network. In contrast, the F OLANs currently proposed for metropolitan deployment (e.g., DQDB and FDDI) are based on double loop topologies, which are protected against single cable failures. In order to be competitive, the PON must survive at least single cuts. Preferably, it should achieve this objective while preserving its “passive” nature. This means that there are no active switches inside the network. A simple solution can be obtained by overlaying two redundant, diverse routed trees (see Fig. 3) [13]. The “left” tree and the “right” tree are combined at the root. Thus, each station has two separate transmission paths to and from the root. Initially, the station transmits and receives on one of the two paths. If a failure occurs along the path, the station immediately detects the failure because of the loss of the received signal. It then switches to the alternate path.

It has been shown [13] that this redundant tree scheme is protected from all single failures and from a large number of multiple failures. Most importantly, it allows the network operator to quickly locate the point of failure based on failure reports from the various stations. The impact on power budget is minimal (4 dB loss at the root to account for the coupling of the two trees). The additional layout cost is typically much less than 50%.

In contrast, a redundant star requires a diverse-routed fiber for each station, with a 100% cost increase. A passive linear bus topology also requires full fiber redundancy and diverse routing (with a 100% installation cost increase) in addition to active bypass switches. Moreover, the fault may be still difficult to locate. If the bus topology is configured as a loop, one can easily show that

![Fig. 3. Redundant trees: (a) left tree, (b) right tree.](image-url)
single-failure protection can be obtained with a single, reconfigurable bus with no additional cost for redundancy [27]. Note, however, that two failures may bring down the loop network entirely; whereas two failures in the redundant tree cause only limited damage, and, in most cases, no damage at all.

3. Access schemes

In this section, we examine the various schemes that can be used to access the tree-shaped broadcast medium. We then compare trees to stars and buses with respect to the variety of access schemes available for them. For the moment, we assume single-fiber, single-wavelength operation. Clearly, many of the schemes proposed in other broadcast media, such as LAN, radio, and satellite, may be applied here as well. It is helpful, however, to review first some of the properties that make the tree topology unique.

The fact that each transmission passes through the root of the tree before being broadcast to all stations (including the transmitting station) permits us to define a global time reference, much the same way as is done in satellite multiple access channels or optical star networks. Namely, each station measures its roundtrip delay to the root and synchronizes its local time to that of a predefined reference station. This permits us to define a reference frame at the root of the tree, similar to the reference frame in a satellite channel.

To pursue further the satellite analogy, we observe that a typical MAN is characterized by a very high value of the ratio $a$, defined as

$$ a = \frac{\text{roundtrip propagation delay}}{\text{transmission time}} $$

For a MAN with a 50 km radius, 1000 bit packets, and 1 Gb/s channel speed, one finds that $a = 500$. This value is comparable to those observed in satellite channels and is several orders of magnitude higher than typical LAN values, which range from 0.01 to 0.1. The large value of $a$ in high speed, metropolitan tree networks precludes the use of carrier sensing and collision detection. Thus, access schemes previously proposed for other, lower speed, tree shaped fiber networks, such as Hubnet [24] are not suitable for our application.

Keeping the above properties in mind, we now review and compare the main candidate access schemes. In this comparison the principal, and often conflicting, performance criteria are low delay and high efficiency under heavy load. We will subdivide the access schemes into four classes: random access, token mode, reservation TDMA (time division multiple access), and integrated TDMA frame.

(1) Random access. In the absence of carrier sensing, the simplest random access scheme available on a tree topology is ALOHA. The slotted ALOHA (S-ALOHA) version is generally used, since slot synchronization can be implemented easily in the tree. The main advantage of ALOHA is low latency, at the expense of poor channel efficiency or large delay variance, and instability under heavy load. For these reasons, ALOHA has been proposed mostly for transmission of low volume, low latency, control traffic in tree topologies. The performance under more sustained traffic load can be improved by using a collision resolution scheme such as the Capetanakis tree algorithm [8]. To this end, the structure of the optical tree topology naturally defines the subtrees to be probed in the tree-algorithm, although any other binary partition could also be used. Given the large value of $a$ in a MAN, an implementation of the tree algorithm similar to that proposed for satellite channels must be used [9].

(2) Token mode. The problems of large delay variance and instability found in the random access mode can be overcome by using a token protocol to enforce ordered access. While the most popular token protocols have been implemented on rings and buses, it is actually possible to implement several versions of the token protocol on the tree topology as well. The simplest scheme is the token bus scheme similar to that defined by the IEEE 802.4 standard for local area networks. This token bus protocol can be used with any broadcast medium and therefore with a passive tree. A token is passed among the stations according to a predetermined sequence,
allowing token-holding stations to transmit waiting packets during the token-holding period. The algorithm that governs token-holding periods can range from a simple “one packet per shot” rule to the sophisticated timed-token rotation protocol of the IEEE 802.4 standard. A well-known drawback of the token-bus protocol is latency, caused by propagation delay and token processing time. In small, geographically compact trees, latency is tolerable. For instance, in a tree with 32 stations, 100 Mb/s channel rate, 1000 bit packet size, 2 km radius and 50% channel utilization, the latency is 1 ms, a value which is adequate for non real-time, datagram traffic. For larger populations and geographies, latency can be reduced by implementing a clustering scheme, as shown in Fig. 4, where clusters in the form of bus extensions are created at the bottom of the tree. In the cluster configuration, a new scheme, the “implicit token” scheme, originally developed for high-speed buses [12,30], can be used. A token sweeps in a predefined sequence all the clusters (e.g., buses in Fig. 4) at the bottom of the tree. The end station on each bus must reissue a new token. The stations aligned along the bus simply attach their packets to the token, or to another packet, thus forming a “train” of packets. The train is broadcast to the entire tree. The next end station in the predefined order issues a new token upon sensing the end of the train and so on. The term “implicit token” derives from the fact that no explicit token (i.e., separate token frame) is required to schedule transmissions of stations within each cluster, thus reducing token latency. Still, latency is a concern in trees with a large number of clusters, which must all be polled even if some have no packets to transmit. Several schemes were proposed in [12] to overcome this problem. In selective polling the main idea is to dispatch the token only to those buses that have backlogged stations (i.e., stations with packets to transmit). In order to allow backlogged stations to identify themselves, we exploit the global synchronization feature offered by the tree topology (but not available in buses), and define a channel frame consisting of reservation subframe and data subframe. The reservation subframe is subdivided into many minislots, one per cluster. Each backlogged station transmits a short burst in its minislot to announce that it needs to be “polled”. In the data subframe, the token sweeps only the clusters with backlogged stations. This scheme (which is very similar to the bit-map scheme described in [29]) drastically reduces the latency overhead, at the expense of reservation subframe overhead. Another solution (inspired by the BuzzNet protocol [14]) is the hybrid random access/token scheme. This scheme attempts to eliminate the reservation subframe latency introduced by selective polling. Under light load, the hybrid scheme operates in random access mode. If a collision occurs, the scheme switches to selective polling until the backlog is cleared. Then, random access mode is resumed. This scheme is superior to selective polling under light load, where collisions are rare, and, therefore, the data flow is not interrupted by the transmission of a reservation subframe.

(3) Reservation TDMA. The basic scheme is inspired by the R-TDMA scheme proposed for satellite channels [6]. It assumes that a fixed frame size is used. The frame is subdivided into two subframes, a reservation subframe and a data subframe. The reservation subframe is further subdivided into minislots, one per station (or
cluster); the reservation minislot is used to broadcast the number of packets backlogged in the station (or cluster). The data subframe is also slotted. Each slot can carry a maximum size packet. Slots are dynamically assigned to the backlogged stations in a round robin manner. Since each station knows the backlog of all the other stations, this assignment can be carried out in a totally distributed manner. The potential drawback of the R-TDMA scheme is the overhead of minislot reservation, which increases linearly with the number of stations. An alternative is the use of S-ALOHA, instead of TDMA, for placing reservations. This way, the reservation overhead is proportional to the traffic load, rather than the number of stations, thus allowing for much larger station populations. The S-ALOHA reservation scheme can be further enhanced by dynamically adjusting the boundary between the reservation subframe and the data subframe. For instance, if the data subframe is empty (i.e. the global queue is empty), the reservation subframe is expanded to occupy the entire frame [20].

(4) Integrated, movable boundary TDMA frame. One of the key requirements in a high-speed network is the integration of data and real-time (voice and video) traffic. This integration can be efficiently accomplished in a tree topology by sharing a TDMA frame among various services. For example, the frame may consist of a video/voice subframe and a data subframe. The video/voice subframe is subdivided into video slots (where each video slot can be further subdivided into voice slots as in Fig. 5). The data subframe is small, but can be expanded by moving its boundary if there are unused voice/video slots. To initiate a video or voice call, a station must submit a call request (in the data portion of the channel). The call is accepted only if video/voice slots are available. In the data subframe data packets can be transmitted with any of the access schemes discussed above. The simplest scheme is the basic token scheme. The complete polling cycle would, of course, be spread over several frames. That is, a station must suspend its transmission at the end of the data subframe, and resume it at the beginning of the next data subframe.

After this brief review of tree topology access schemes, we pause to establish a comparison with other topologies. For the star, one easily recognizes that all the tree topology schemes can be directly applied to it as well. For the bus, on the other hand, one notices that the linear structure permits operations which are not possible in a tree or star. For simplicity, we will base our comparison on a linear folded bus; however, our considerations can also be extended to dual buses.

First, we note that slot or frame synchronization is imposed in a bus by the end station and is acquired by the other stations on the uplink (or transmit) bus. This procedure is easier than the corresponding (slot) synchronization acquisition in trees or stars. Closely related to the above property is the ability to implement the implicit token scheme (based on the “attempt and defer” access mode). That is, each station can probe the bus until it finds the end of a packet train, and then attach its packet to the train. This implicit token operation is obviously possible only in a linear bus structure. It cannot be implemented in a tree topology (except for the eventual bus cluster at the bottom of the tree, as shown in Fig. 4). A third advantage is the random access to free slots. Assuming slotted bus operation and station clock synchronization (imposed by the end station), a station can detect any free slot that passes by, mark it as busy, and use it. A drawback of this approach is its unfairness, which can be corrected using Distributed Queueing and Bandwidth Bal-

---

![Fig. 5. Integrated video, voice and data frame.](image-url)
ancing, as in DQDB [17], or reservations as in CRMA-II [31]. In contrast, one may note that tree topology access protocols are perfectly fair, in that they do not favor any particular station on the basis of its geographical position in the tree. Linear bus and, more generally, ring topologies also permit the reuse of slots freed by the destinations [10]. This however requires active interfaces (e.g., shift registers) along the bus, and thus violates our initial assumption of strictly passive interfaces in the PON.

In balance, buses allow more channel access options than trees or stars. However, these options often require additional, complex procedures to make them work (e.g., distributed queueing and fairness algorithms in DQDB). Furthermore, these options cannot be easily transferred to the multiple wavelength environment. For example, to implement DQDB on multiple wavelengths on the same bus, one should install a passband filter at each tap, and a band suppressor filter on the passthrough between the receive and transmit taps. These filters adversely impact the power budget (which is already a critical issue in linear buses). Furthermore, they limit the ability to reconfigure the "virtual topology" by retuning the wavelengths. For these reasons, bus topologies cannot fully exploit the very important benefits offered by wavelength division (to be discussed in the next section).

4. Wavelength division

Wavelength division is a form of optical bandwidth sharing which can be effectively exploited in PONs to achieve enormous aggregate capacities (in the Tb/s range). Different wavelengths can be used to interconnect stations (in a point-to-point or shared channel scheme), defining a virtual topology which is often radically different from the original physical topology (corresponding to the fiber layout). Because of the relative independence between the virtual and physical topologies, most of the wavelength-division concepts that follow are applicable to any physical topology (tree, star, or bus). However, power budget and filtering cost considerations make bus topologies less desirable than stars and trees, as discussed in the previous section.

A chief distinction between virtual topologies is whether a transmission must traverse one or many hops, (i.e., single- or multi-hop network). We briefly review these alternatives below:

(a) Single-hop networks. A single-hop lightwave network can be defined as a network where a direct transmission can be achieved between each source and destination pair. For instance, all inputs are combined in a star coupler and broadcast to all outputs. WDM can be implemented in various ways, depending on whether transmitters, receivers or both are tunable. Two representative WDM network examples are Bellcore's LAMB-DANET [15] and IBM's Rainbow [11]. Single-hop networks can support both circuit-switched (C/S) and packet-switched (P/S) service. The simplest way to provide C/S service is by dedicating an entire wavelength to each connection, as is done in Rainbow. Separate controls (and possibly, a separate control channel) are required to arrange for the transmitter and receiver to rendezvous on the same wavelength. Recently, more efficient C/S schemes have been proposed, which combine time- and wavelength-division multiplexing [19,23].

For P/S traffic, the dedication of wavelength for the entire session is clearly very inefficient because of the bursty nature of P/S applications. Thus, for P/S support, time slotting is generally used in conjunction with dynamic wavelength assignment [7]. The rendezvous protocol now becomes much more complex than in the C/S case, since the retuning of transmitters and/or receivers must be done on a slot-by-slot basis. A variety of rendezvous protocols and access schemes have been proposed (requiring different transmitter/receiver capabilities and configurations at each station) [26]. The simplest access scheme is S-ALOHA, stabilized with retransmission time out controls, which can be implemented with a single laser/detector pair per station, without a separate control channel [28].

(b) Multi-hop networks. Instead of using a direct path from the source to the destination, multi-hop networks require some packets to travel across several hops. Basically, the multi-hop net-
work is a store-and-forward network embedded in the passive optic network. The switches of the multi-hop network are represented by the user stations (thus, they are located at the periphery of the passive broadcast medium); the links consist of dedicated wavelength channels established between pairs of stations [1]. The multi-hop network differs from a conventional P/S mesh network in that it uses streamlined network protocols especially designed for high speed operations. It features simple network controls, and efficient packet buffering strategies at the station. Namely, packet forwarding is accomplished with a single output buffer per channel and is often carried out by cut-through [22]. Deflection routing is used to prevent queueing within the network [4]. Clearly, each hop incurs a penalty (in terms of additional packet delay and processing overhead) if an electro-optical conversion is required; for this reason all-optical multihop networks have also been proposed [2]. Thus, the virtual topology should be designed to minimize the number of hops. The most prominent example of virtual multi-hop network is ShuffleNet, proposed in [1]. ShuffleNet exploits WDM to embed a perfect-shuffle interconnection within a fully broadcast physical topology. Fig. 6 shows an example of an eight-station perfect-shuffle virtual topology embedded in a physical tree topology.

Both the single-hop and the multi-hop schemes, however, have some limitations. In single-hop networks, frequency agile transmitters and/or receivers with submicrosecond tuning times are required to achieve good T/WDMA protocol efficiency in bursty traffic environments. These devices are not yet commercially available, although rapid progress of the technology in this direction is reported [7]. Furthermore, T/WDMA involves substantial control overhead to coordinate packet-by-packet transmissions and to assign slots dynamically in time and wavelength. For these reasons, the single-hop system does not scale up well to a large network size. The multi-hop system also has its own limitations on account of hop-by-hop store-and-forwarding, namely: congestion sensitivity, inadequate support for real-time traffic, and the inability to provide efficient broadcast/multicast services.

![Diagram](image-url)

Fig. 6. The eight-station ShuffleNet implemented as a physical tree.
One way to overcome these limitations is to implement a hybrid scheme in which both single- and multi-hop transmissions are included and used to support the services for which they are best suited. This approach is prompted by the observation that these schemes have complementary features: for example, single-hop networks can easily support broadcast/multicast (but do not scale up well with network size), whereas multi-hop networks can handle large numbers of nodes (but have difficulties in providing broadcast/multicast and real-time traffic support). By combining features of both architectures, we could enjoy the benefits of both. Hybrid optical-network approaches for packet and circuit integration have already been reported in the literature [21]. In the next section we investigate the hybrid concept more systematically than in previously reported studies and present a hybrid, multilevel optical network (MONET) which effectively combines elements from both single- and multi-hop schemes, and which is implemented on a physical-tree topology.

5. MONET: A multilevel optical network

MONET is an all optical network architecture based on the following concepts:
- hybrid single-hop and multi-hop approach,
- clustering and multilevel structure,
- efficient exploitation of multifiber cables.

Each one of the above concepts has already been proposed or demonstrated separately in previous optical networks. But, in MONET these concepts are for the first time jointly and systematically applied to the same architecture. As we shall see, MONET is actually a family of network architectures. It is a framework in which the designer can choose among various options (e.g. channel access protocol, wavelength assignment policy, etc.) at various levels in the network. Some versions of MONET can also be implemented on an optical star topology. However, the full benefits of MONET can be obtained only in an optical tree topology, as will be shown in the following sections.

5.1. Basic MONET

To set the stage, we introduce a simple, two-level version of MONET. We recall that the physical topology is a tree. Further, we assume that the stations are partitioned into clusters. Each cluster is assigned a separate wavelength and is organized as a single-hop virtual subnet. The virtual subnets are embedded in the physical subnet. In the example of Fig. 7, we show four clusters (each with four stations) derived from a 16-node physical tree. In this case, each cluster is associated with a separate physical subtree. More generally, the clusters may be spread out and interleaved with each other in the tree. One such

Fig. 7. Virtual topologies of the two-level architecture.
example is shown in Fig. 8. One of the stations in each cluster is designated as the gateway, and the gateways are interconnected to each other via a *multi-hop* backbone network, e.g. a ShuffleNet. The backbone links are implemented with dedicated wavelengths and are also embedded in the physical topology. We have thus created a hybrid, two-level structure, with *single-hop* local clusters interconnected by a *multi-hop* network at the higher level. As a difference from conventional two-level architectures, line cost minimization is not the prevailing criterion in the selection of clusters. This is because in MONET line cost is determined exclusively by the physical topology (a tree) and is therefore totally independent of the specific cluster arrangements.

We now proceed to describe the P/S service support in the basic MONET. We will discuss circuit switching in the next section. In the two-level MONET just described, *intracluster* packet transmissions are carried by a single wavelength channel and thus can be handled with any of the access schemes described in Section 3. *Intercluster* traffic is intercepted by gateways and forwarded to the multi-hop backbone. Several extensions to this simple operation can be immediately suggested. For instance, a gateway can be shared by two or more clusters, thus saving backbone wavelengths. Carrying this to the extreme, we could even propose to use a single gateway to interconnect all the virtual subnets. The problem with this, of course, is potential gateway overload and lack of redundancy. One can also go the other way, and assign several gateways to each subnet, thus improving redundancy and providing more alternate paths at an increased cost.

The basic P/S service requires only one laser and one detector per user station. Multiplexing on the intracluster channel is strictly on a time-division basis. Gateways, on the other hand, are equipped with multiple lasers tuned to different (fixed) wavelengths, namely, one intracluster wavelength and two intercluster wavelengths. Note that the number of required wavelengths $W$ increases linearly with the number of clusters $C$. More precisely, with one gateway per cluster, we find:

$$W = 3C$$

The number of wavelengths can be reduced by allowing sharing of the intercluster wavelengths among the gateway channels via TDMA. This approach, however, reduces backbone link capacity, which may in turn cause congestion side effects.

To increase the capacity of the intracluster P/S service, multiple wavelengths can be assigned to each subnet and can be accessed using T/WDMA, for example. This, of course, in-

---

**Fig. 8.** Example of a cluster spread out across the leaves of the tree.
creases the cost and complexity of the subnet implementation, since it requires each station to be equipped with a tunable laser and/or receiver. A subchannel must also be dedicated for the control of the time/wavelength division channels.

5.2. Circuit switched service support

The C/S service in MONET is used to support stream traffic (e.g. voice, video, visualisation) as well as high-rate file transfers. In the basic MONET, the C/S service implementation is keyed to the two-level structure. Within a cluster, C/S connections can be established in several different ways. If only one wavelength is available within a cluster, and it must be shared by both P/S and C/S service, then, the most-effective solution is the “integrated, movable boundary TDMA frame” scheme described in Section 3. A circuit connection between two stations is established by dedicating a certain number of slots (proportional to the data rate) in the C/S subframe (see Fig. 5). If, on the other hand, one or more wavelengths are allocated exclusively to the C/S service within the cluster, then these wavelengths can be demand-assigned to C/S connections. The use of on-demand C/S channels introduces two extra requirements: a tunable laser/receiver pair at each station, to rendez-vous on the desired wavelength (in addition to the fixed laser/receiver pair used for datagram traffic), and; a control procedure for call set-up and management of on-demand channels. The control procedure usually relies on the packet switched channel for the exchange of control packets. It should be noted that, if a user station is allowed to engage in only one C/S call at a time, the laser/receiver pair must be tuned only at call setup time. Thus, a slow tuning time (fractions of a second) is adequate. If on the other hand the station must maintain several calls simultaneously, then fast tuning (on the order of nanoseconds) is required to switch wavelength from one slot to the next within the T/WDMA frame, unless multiple receivers are provided and used with special scheduling algorithms (e.g. pipelining) [23].

Intercluster circuit connections can be handled in a way similar to the intracluster connections, by setting aside a pool of C/S wavelengths. Thus, a pair of stations in separate clusters can communicate with each other by tuning to one of the free intercluster wavelengths. These on-demand wavelengths can also be multiplexed using T/WDMA (as in the intracluster case), so as best to exploit the available bandwidth.

It should be noted that if the stations within a cluster are equipped with only one laser/receiver pair tuned to a fixed wavelength (as is the case when only one wavelength is available within the cluster), then intercluster circuit connections must be established through the gateways. Namely, within a cluster, a station sets up a connection to the local gateway by reserving slots within the TDMA frame. The local gateway then sets up an intercluster connection to the remote gateway using the on-demand channels. Gateways basically operate as circuit switches in this case. Clearly, we are trading off station complexity (tunable laser/receiver pairs) for gateway complexity (additional circuit switching functions). Taking this concept one step further, we can envision two classes of stations which coexist in the MONET cluster: (a) small stations, equipped with only a fixed transmitter/receiver pair; and (b) regular stations, equipped with a fixed as well as a tunable transmitter/receiver pair. A small-station configuration would be suitable for a single workstation, for example. A regular-station configuration may be required for a mainframe, or for a traffic concentrator supporting several multimedia workstations.

5.3. Geographically separate subtrees

One drawback of virtual subnet clustering is the fact that each subnet requires separate wavelengths. If direct detection is used, the number of separate wavelengths that can be tapped using state of the art devices in the 1.5 µm window is on the order of a few dozens. Recalling that the number of wavelengths required just for the P/S service is \( W = 3C \), where \( C \) is the number of clusters, we see that this restriction severely limits the number of clusters that can be configured in
MONET. To alleviate this problem, we can map clusters into geographically separate (as opposed to overlapped) virtual subtrees, so that wavelengths can be reused. This solution would actually be quite natural when a large geographical area must be covered (e.g., metropolitan distribution), since clusters would then be formed based on geographical proximity and would correspond to disjoint physical subtrees. The subtrees could then be interconnected at their roots via a star coupler (see Fig. 9). The wavelength window can then be divided into two regions, one for intracluster, and one for intercluster communications.

A filter at the star coupler blocks the intracluster wavelengths. Thus, intracluster wavelengths can be reused from cluster to cluster to reduce the overall wavelength requirements. Intercluster wavelengths, on the other hand, are passed by the filter and, therefore, must be shared among all subnets (they cannot be reused): these include the "pool" of wavelengths assigned on demand, as well as the multi-hop backbone channels. Note that the benefits of wavelength reuse come at the expense of extra optical-power loss caused by the filters. Optical amplifiers are therefore required at the star’s outputs.
5.4. Multiple hierarchical levels

Although all the cases reviewed so far are based on two-level MONETs, it is also possible to implement MONET on multilevel structures. Namely, several levels of clustering are defined, and intra- and intercluster communications are handled by different channels, as previously discussed (see Fig. 10). The advantage of the multilevel structure becomes apparent when in the MONET tree topology the clusters at the various levels are implemented with geographically separate subtrees. In this case, it is possible to reuse wavelengths within clusters at each level, leading to substantial savings. For example, in Fig. 10, the clusters $A$ and $B$ may use the same wavelength $\lambda_1$ for internal communications. Likewise, clusters $a$, $b$, $c$, and $d$ can use $\lambda_2$ for internal communications.

It should be observed that the major benefits of multilevel clustering derive from the ability to capture traffic "localities". If there is strong locality in the system, and it is reflected in the network's hierarchical structure, then at higher levels the intercluster channels (for either packet or circuit traffic) will require less bandwidth (for user pairs) than at lower levels. At the same time, the number of potential pairs communicating with each other over an intercluster channel clearly increases with the level. In the ideal case, these two effects compensate each other, and therefore the product of (bandwidth per user) $\times$ (number of user pairs) is a constant for each level. If this property is verified, and wavelengths are reused, then the number of wavelengths increases linearly with the number of levels, and logarithmically with the number of end users. This is a major improvement with respect to the case in which all clusters share the same physical tree. In that case, the number of required wavelengths grows linearly with the number of end users.

5.5. Multifiber MONET

As mentioned earlier, one of the limitations in scaling up MONET is the number of available wavelengths. Geographically separate subtrees allow us to overcome in part this limitation by making wavelength reuse possible. Another solution to this problem is the use of multifiber cables, i.e., cables which contain hundreds of fibers. This solution is attractive because, although multifiber cable is more expensive than single-fiber cable, the marginal cost of an additional fiber is low, compared with installation and packaging costs. The payoff of the multifiber plant is very high, in that the number of available wavelengths can now be amplified by a factor of hundreds.

There are many different ways in which a multifiber MONET can be configured. A possible scheme based on a two-level MONET was presented in [5]. Following this scheme, the station population is subdivided into clusters. Each clus-
ter is allocated a certain number of intracluster “fiber plants”, where a fiber plant is a single fiber subtree, embedded in the cable plant. The cable plant is the tree formed by the multifiber cables. We assume that the cable plant is equipped at its internal nodes with binary couplers, reflective star couplers and optical crossbar switches. By properly manipulating these components, any arbitrary fiber plant can be constructed within the cable plant. Intracluster communications (packet and circuit switched) are thus carried out using these intracluster fiber plants, in any of the possible ways described in Sections 5.1 and 5.2. Note that since the cluster fiber plants are typically disjoint (i.e., they do not share common cables), the above strategy achieves both wavelength reuse, as well as fiber reuse (within the cable).

For intercluster communications, the solution proposed in [5] distinguishes between P/S and C/S traffic. To support P/S traffic, a multi-hop backbone is used. The links of the multi-hop backbone are implemented with wavelength channels carried by various intercluster fibers. More specifically, we establish, within the MONET cable plant, a sufficient number of intercluster fiber plants so that each fiber plant covers a (small) number of clusters, and yet for any pair of clusters there is at least one fiber plant to interconnect them. Fig. 11 shows an example with four fiber plants \(S_1, S_2, S_3, \text{ and } S_4\) embedded in a cable plant. Each fiber plant covers eight clusters. To provide interconnectivity among all clusters in this case, the network uses a total of 496 fiber plants embedded in a 264-fiber cable [5]. The rationale for keeping the fiber plants small in size has to do with optical power budget; in fact, an important goal of the proposed solution is to avoid or reduce the need for optical amplification (at the root of the cable plant), by placing a constraint on the number of levels (and therefore size) of the intercluster fiber trees. Once the intercluster fiber plants have been established (by properly configuring the switches inside the cable plant), the “links” of the multi-hop topology are then established on them.

For intercluster C/S connections, the approach proposed for the multifiber MONET is similar to that discussed in Section 5.2. A pool of fibers and wavelengths is made available on demand. When a connection between two remote stations must be established, the stations select a common fiber plant and then tune their transceivers to an agreed-upon, unused WDM channel in that plant. Control information to establish the connection is exchanged over the multihop backbone using the P/S service. At the station, the connection to the common fiber plant is accomplished by connecting (through an optical switch) the transceiver to the proper fiber in the cable. Thus, the stations use a combination of space- and wavelength-division multiplexing to communicate with each other.

To illustrate the efficacy of a multifiber MONET implementation, we cite an example reported in [5]. In that example, the multifiber cable carries 264 fibers. The network comprises 128 clusters, of 16 stations each (thus, 2048 stations). For intercluster packet communications, 496 fiber plants (each covering 8 different clusters) are embedded in the cable plant. Even after taking into account the power losses of all the optical components, one finds that this very large MONET can be operated to provide more than 1 Tb/s throughput without optical amplification!

One may have noticed an analogy between the use of multiple fibers in a cable and the use of multiple wavelengths in a fiber. Namely, both schemes permit one to multiplex (via space division and wavelength division, respectively) several channels on the same physical carrier. The question of whether we can in general trade multiple fibers for multiple wavelengths is definitely an intriguing and challenging one, but it is beyond the scope of this paper. One observation is worth making, however, in regard to spatial reuse. Both wavelengths and fibers can be reused in geographically separate sections of the tree. But, wavelength reuse requires wavelength selective filters, which are costly, complex to manufacture, and very lossy (4 to 5 dB extra loss); whereas the reuse of a fiber within the cable requires optomechanical switches, which are technologically simple, inexpensive, and introduce only 0.5 dB loss. In balance, the multifiber technology seems to be more desirable (at least with respect to spatial reuse) for large network applications where opti-
cal power budget is a major concern. In practice, wavelength and fiber multiplexing technologies will be used jointly, as we did in MONET, exploiting the best features of each.

5.6. MONET channel access protocols

As shown in previous sections, MONET is a family of network architectures which allows the designer to exercise different options at different levels. Consequently, different channel access protocols may be used in different versions of MONET. In this section we review and, when necessary, elaborate on the various access schemes which have been proposed for MONET.

First, we consider the P/S service and separately examine the intra- and intercluster cases. We recall that, for intracluster packet communications, each cluster is allocated a dedicated wavelength (or fiber). The stations in the cluster can thus access this channel using any of the schemes presented in Section 3. However, if the cluster contains stations which are equipped only with fixed (i.e. nontunable) transceivers, then the "integrated TDMA frame" is preferred. This choice in fact will permit these "small" stations to use also the C/S service, and to set up C/S connections in the C/S portion of the frame. For intercluster packet communications, the scheme proposed for MONET is the multi-hop scheme. Namely, a multi-hop network interconnects the gateways in the various clusters. The multi-hop choice was motivated by equipment cost considerations (only two transceivers per gateway are required) and by throughput considerations (the throughput of a degree-2 multi-hop network, such as ShuffleNet, for example, far exceeds the data rate of a single wavelength channel). Here again, however, in some situations it may be desirable to replace the multi-hop network with a single-hop network. Namely, a single, multiaccess, broadcast channel can be used to interconnect all the gateways and to provide all intercluster communications. This scheme may become attractive if there is strong locality of traffic within clusters, so that intercluster traffic is relatively low. The MONET framework can clearly accommodate also the single-hop intercluster strategy, as well as various possible extensions (e.g. multiple, superimposed single-hop networks; multiple, partially overlapped single-hop networks, etc.). Next, we discuss support for the C/S service. Here again, we distinguish between intracluster and intercluster service. The intracluster service is typically supported by a pool of dedicated wavelengths. To permit the transmission of broadcast/multicast traffic, we assume that each station is equipped with a fixed transmitter and one or more tunable receivers [7]. In the simplest case, the (simplex) connection is established by tuning the receiver to the transmitter wavelength and dedicating that wavelength for the entire duration of the session. To improve the efficiency of the C/S service (e.g., in supporting multiple, low-rate connections not requiring the full channel bandwidth or supporting multiple simultaneous calls into or out of the same station, etc.) it is necessary to use combined time and wavelength division multiplexing (i.e. T/WDMA). In T/WDMA the receiver must be retuned to different wavelengths, in different time slots, within a time division frame. Slot-by-slot retuning typically requires rapidly tunable, very expensive receiver technology. Fortunately, we have shown in [23] that the use of pipelined, multiple filters combined with subframe scheduling allows us to achieve very good results even with relatively inexpensive, slowly tunable, off the shelf filters.

The above schemes can also be applied to intracluster C/S communications in the multifiber MONET case, where entire fibers (in addition to wavelengths) may be allocated to the C/S service. In this case the access scheme includes also the space division component (i.e. choice of fiber) and thus becomes S/T/WDMA. Transmitter and receiver requirements at each station are the same as in T/WDMA. In addition, fast lithium niobate switches can be used to switch from one fiber to another (subnanosecond switches have been recently reported [18]). We should also mention that in a multifiber MONET the C/S service can be provided using exclusively space and time division (i.e., S/TDMA), without wavelength division. In this case, all stations can operate on the same wavelength.

Intercluster C/S connections are handled in a
way very similar to intracluster connections. A pool of wavelengths and/or fibers is dedicated to the intercluster C/S service; WDMA, T/WDMA or S/T/WDMA can be used to share the pool among users. We recall that the typical station is equipped with two sets of transceivers: a fixed pair, to be used for packet communications; and a set with one fixed transmitter and one or more tunable receivers to be used for both intra and intercluster circuit switched communications. Thus, there is no difference in the intracluster and intercluster C/S access protocols, apart from the fact that separate wavelength and/or fiber pools are used.

The setup of a C/S connection requires coordination between the origin and destination and, possibly, the intervention of a centralized manager. This coordination implies the exchange of control traffic between the various parties. In MONET, the C/S-related coordination and control traffic can be conveniently and efficiently carried by the packet network. This is made possible by the fact that in MONET the packet subnetwork is totally independent of the circuit subnetwork.

5.7 Benefits of the MONET architecture

By combining various network concepts and technologies in a hybrid, tree structured architecture, MONET provides a variety of benefits that cannot be found all at the same time in any other network. The benefits are:

Efficient support for C/S service. This support is mainly provided by the single-hop component of MONET.

Efficient support for P/S service. This support is provided by the single-hop component within each cluster, and by the multi-hop component for intercluster communications.

Efficient broadcast/multicast. For the C/S service and the intracluster P/S service, the broadcast/multicast capability is a natural byproduct of the single-hop implementation. Intercluster broadcast/multicast P/S traffic can be efficiently carried on dedicated wavelengths (or time slots within a wavelength channel, or entire fiber plants in a multifiber cable plant) which interconnect a specific set of users (e.g. a closed user group).

Wavelength reuse. MONET alleviates the limitation in the number of available wavelengths by allowing “spatial” reuse of wavelengths in geographically separate clusters.

Multifiber utilization. The MONET tree topology makes it possible to fully exploit multifiber cables, and in particular to benefit from the “spatial” reuse of optical fibers.

Cost effective topology layout. The MONET tree topology permits us to minimize fiber installation costs in metropolitan areas.

Favorable power budget. The tree structure yields a more favorable power budget than the bus. Moreover, it can easily and efficiently incorporate all-optical amplifiers at its root.

Resilience to fiber cuts. The tree topology can be made robust to fiber cuts with moderate extra cost (for redundancy) and in a way that is totally transparent to the access protocols.

Growth flexibility. In MONET, a virtual cluster can be expanded incrementally. Growth in traffic can also be easily met in MONET by introducing new wavelengths, splitting large clusters and adding more gateways in the backbone.

Static/dynamic reconfigurations. In a pure multi-hop network, a virtual-topology reconfiguration tends to be quite pervasive, that is, it affects a large portion of the network. It may imply large control-traffic overhead, and may cause the disruption of service to many users (an issue particularly critical in dynamic reconfigurations). In MONET, changes are more limited by virtue of the two-level structure. Either we split/merge/expand a cluster, without affecting other clusters, or we modify the backbone, without affecting intracluster traffic.

6. Conclusions

In this paper we have investigated the use of tree topologies for PON installations. We have evaluated the physical properties of optical tree networks, as well as the options available for channel access and wavelength allocation.
We have shown that trees enjoy distinctive advantages over other topologies in the areas of power budget, low-cost metropolitan coverage, and fault tolerance. As for channel access, we have recognized the fact that the options available in a tree are more limited with respect to those available in a bus. However, the options unique to the bus topology are often associated with high implementation costs and cannot scale well to a WDM environment. With respect to wavelength allocation and virtual-topology design, we basically have found that physical tree topologies can be dealt with in the same way as buses and stars. We have also recognized that the limitations of the current leading strategies: single-hop and multi-hop. To overcome such limitations we have proposed a novel architecture (MONET) which is particularly suited to tree topologies.

In MONET, by partitioning the network into clusters, we can greatly improve the efficiency of the network which can operate with fewer wavelengths, thus alleviating the dynamic tuning problems and can exploit traffic locality. The use of separate optical media (i.e., multiple fibers in the same conduit, or geographically separate clusters) enables wavelength reuse in the clusters, thus overcoming, in part, the scarcity of available wavelengths. For intercluster communications, two schemes are provided in parallel: (1) the multi-hop backbone, which offers good statistical multiplexing properties and low access delay, and is thus ideal for control and datagram traffic; and (2) the pool of on-demand channels, which provide an efficient transport mechanism for real-time and broadcast traffic. The advantages of traffic locality and wavelength reuse are further amplified by multilevel configurations.

References

[21] L. Kazovsky and P. Poggiolini, STARNET: A multigbit/s...


Milan Kovačević was born in Belgrade, Yugoslavia, in 1960. He received his B.S.E.E. and M.S.E.E. degrees from the School of Electrical Engineering, University of Belgrade, Yugoslavia in 1985 and 1987, respectively. From 1985 to 1987 he was a Research Engineer at the Department of Electronics and Computer Engineering, University of Belgrade, where he worked on a fault-tolerant multiprocessor system project. He is currently working towards the Ph.D. degree in computer science at the University of California, Los Angeles. His current research interests include optical networks, multiaccess protocols and performance modeling and analysis. He is the recipient of UCLA Chancellor’s Dissertation Year Fellowship Award for 1992–1993.

Joseph Bannister manages the High Performance Computing Section at the Aerospace Corporation, El Segundo, CA. He has held positions in research and development at Sytek, Unisys, and Xerox. After earning the B.S. degree in mathematics from the University of Virginia, he obtained the M.S. degree in electrical engineering, the M.S. degree in computer science, and the Ph.D. degree in computer science, all from the University of California, Los Angeles. He has conducted research and development in high-speed networking, network design, network management, performance evaluation, and dependable computing. His current interests are in the area of optical networks. Dr. Bannister is a member of the Institute of Electrical and Electronics Engineers and the Association for Computing Machinery’s Special Interest Group on Communications.