Computer Science Department Technical Report
University of California
Los Angeles, CA 90024-1596

AN ALL-OPTICAL MULTIFIBER TREE NETWORK

J. Bannister
M. Gerla
M. Kovacevic

June 1992
CSD-920026

An All-Optical Multifiber Tree Network*

Joseph Bannister
The Aerospace Corporation
El Segundo, California, USA

Mario Gerla and Milan Kovačević
Computer Science Department
University of California at Los Angeles, USA

Abstract

Given that many optical fibers can be economically packaged within a single cable, it is worthwhile to consider the design of communication networks that exploit multifiber topologies. We present a network architecture that uses the multifiber tree topology to provide high-speed datagram and circuit-switching communication services to a large population of stations. Using a combination of space-, wavelength-, and time-division multiplexing, this network architecture provides all-optical transmission media to its stations and can interconnect several thousand stations in a metropolitan region without the need for optical amplification or electrooptical signal regeneration.

1 Introduction and background

It is commonplace for manufacturers to package many optical fibers in a single cable. For example, AT&T offers cable with a count of over 200 optical fibers, and Cavi Pirelli sells cable with a count of over 600 optical fibers. Given that the bandwidth of a single-mode optical fiber exceeds a terabit/s [Bra90], the aggregate bandwidth of a multifiber cable approaches a petabit/s. It is the goal of this paper to describe a network design to harness the potential of multifiber optical cables.

Although multifiber optical cable is more expensive than single-fiber cable, the marginal cost of an additional fiber is low, because packaging and other expenses dominate the cost of the cable. The cable structure is the same for a few fibers as for many fibers. Even more significant is the fact that the cost of installing and maintaining the cable is the same for a high–fiber-count cable as for a low–fiber-count cable. It has been reported [HMTS90] that the installation cost for a linear foot of cable runs as high as $23 in a campus setting—this figure is higher if conduit needs to be constructed or rights-of-way obtained. The upshot is that the acquisition and installation costs for multifiber and single-fiber cable systems differ by little.

With few exceptions (e.g. [Kar88]) the exploitation of multifiber optical cables has not been aggressively explored. We discuss multifiber optical networks that employ a combination of time-, wavelength-, and space-division multiplexing to achieve high bandwidth inexpensively. Recently, a proposal for a multiple-bus network that is based on ideas similar to those presented here appeared in [Bir92].

The remainder of the paper is organized into five more sections. To establish a set of goals to be achieved by the multifiber optical network, we review in section 2 the services required of a metropolitan area network. Section 3 outlines our assumptions about the technologies used in the implementation of the multifiber optical network. Section 4 describes the multifiber optical network and provides analyses of its characteristics. We illustrate the concepts of section 4 by discussing several concrete examples of the network in section 5. We conclude the paper in section 6 with an assessment of the advantages and disadvantages of the multifiber optical network.

2 Services required of the network

Intended for use as a metropolitan area network, the multifiber optical network must meet the needs of a range of users. Prime requirements include the following:

- transport of integrated traffic
- support for a large population of users
- low cost
- high reliability

To meet the emerging communication requirements of users, many lightwave network architectures have been proposed [Aca87, CGK89, BFG90, Ste91, GKB92]. Although these proposals have the common goal of supporting the above requirements, they make different assumptions about the future availability of lightwave and electronic technologies. A common assumption, however, is that it will be possible to multiplex a significant number of wavelengths within a single fiber. In proposing the multifiber optical network, our point of departure from earlier proposals is to assume that wavelength-division multiplexing (WDM) is limited to only a modest number of channels per fiber.

To support integrated traffic, the multifiber optical network provides for the transport of stream-oriented real-time and bursty non–real-time traffic. Real-time traffic

*This work was partially supported by State of California–Pacific Bell MICRO grant 4-541121-19907 KC7F.

requires a fixed bandwidth and must be delivered within an allotted amount of time. The required throughput of real-time traffic ranges from low (e.g. the transmission of voice traffic) to high (e.g. the transmission of full-motion video traffic). For this reason a real-time conversation is usually given a dedicated (real or virtual) circuit for the duration of the conversation. Non–real-time traffic is not subject to strict delivery deadlines and can be bursty. Although non–real-time traffic can tolerate variation in delay, its throughput requirements can also range from low to high. Since minor delays are acceptable, many bursty non–real-time traffic streams can share a channel.

A large population of users in a lightwave network makes demands on bandwidth and optical-signal power. Since each individual user also demands greater throughput, the network must have the capability of carrying high aggregate traffic loads. Large populations also place exacting demands on the cable plant, since a greater optical-power budget is usually necessary when a signal must be delivered to many users that share the medium.

Given users' demonstrated sensitivity to cost and their intolerance of service interruptions, cost and reliability are obviously important requirements for the multifiber optical network. This has implications for laser-based solutions, because of the relatively high cost and low reliability of lasers.

3 The components of the multifiber network

In this section we discuss the lightwave technologies that can be used to realize the multifiber network. The lightwave components needed include optical fibers, couplers, optomechanical crossbar switches, tunable light sources and detectors, and wavelength-division–multiplexing transceivers. These components are described next.

The backbone of the multifiber network is the multifiber cable. Manufacturers commonly package a large number of single-mode optical fibers within one cable. Typically, optical-fiber cable, such as AT&T's AccuRibbon, can be purchased in multiples of 12 fibers. Groups of 12 fibers are embedded in a thin ribbon, which can be stacked on top of each other and fitted into a polyethylene tube. AT&T sells cable with up to 216 single-mode fibers that have attenuation losses that approach 0.2 dB/km.

Two optical fibers are permanently joined by a splice. The object is to pass all optical signals from the inbound to the outbound fiber. Splicing, which is a well-understood aspect of optical waveguide technology [Pal88], merely abuts one fiber to another so that as much light as possible passes from one fiber to the other. Low-loss, permanent fiber splices can be made with losses below 0.1 dB.

Coupling—or joining three or more fibers in a light-splitting or -combining configuration—is also well-understood, and many off-the-shelf coupling products can be purchased. The basic coupler used to join single-mode fibers is the fused biconical-taper coupler (FBTC), which is produced by heating and stretching two adjacent fibers to produce an input taper,† a coupling region, and output tapers [Tek90]. Light enters the input taper, is coupled from one fiber to the other in the coupling region, and the split beam exits via the two output tapers; this can be used as a 1×2 power divider. Conversely, two optical signals from the output tapers are combined in the coupling region and jointly exit via the input taper; this can be used as a 2×1 power combiner. By adjusting the geometry of the taper, the FBTC can be produced with a range of coupling ratios. We assume the use of the common 3-dB coupler, which splits optical signals evenly between the two output tapers. When two optical signals are combined, each beam undergoes a 3-dB loss, since the FBTC's symmetric structure causes half the combined signal's power to exit via the isolated port. By using fibers that have slightly different core radii, one can construct an essentially wavelength-independent FBTC, i.e., a device with the same coupling ratio over a range of wavelengths. The FBTC can be cost-effectively manufactured in quantity with low excess loss and consistent, wavelength-independent coupling ratios. Couplers constructed by the fused biconical-taper process are available with excess losses of 0.1 dB or less.

The reflective star coupler (RSC) is an n-port device that couples light from any single port to all other ports. Light that enters through one port is evenly split into n beams which exit via the n ports. We are primarily concerned with the two-port RSC, which can be constructed by looping an optical fiber back on itself and joining a segment of the fiber using the fused biconical-taper process. Thus can one manufacture inexpensive wavelength-independent RSCs in the same manner as FBTCs. We most commonly use RSCs in which the power of an incoming signal is reduced by a factor of $1/n$. The optical-power loss in an ideal n-port RSC is thus $10 \log_{10} n$ dB. A typical 2-port RSC would have a 3-dB power-splitting loss and a 0.1-dB excess loss.

An optical switch performs the function of switching light from an incoming fiber to one of the output fibers. Photonic switches, such as those based on lithium niobate devices or free-space optics [Hin87], and optomechanical switches, such as mirror-, prism-, and solenoid-based designs, are widely available. Figure 1 illustrates the design of an $n \times n$ optomechanical crossbar that uses retractable mirrors to reflect light from an input fiber to any output fiber. This principle can be used to design a $1 \times n$ optomechanical switch that switches the input fiber to one of the n output fibers. Optomechanical switches can be made with insertion losses of less than 1.5 dB.

Given that the usable bandwidth of an optical fiber is several terahertz, one would like to employ WDM to partition this bandwidth into many high-speed optical channels. An optical carrier can be modulated by means of the

†The other input taper is an isolated port that is normally not used.

same techniques used in radio-frequency transmission, e.g. amplitude, frequency, or phase modulation. Many modulated carriers can be simultaneously multiplexed onto an optical fiber, and receivers can use heterodyne or homodyne detection to demodulate the desired signal. Tunable optical transmitters and receivers are needed to modulate and demodulate light of specific wavelengths. WDM systems have been demonstrated in the laboratory and their capabilities continue to improve steadily, e.g. [TONT89] describes an experimental WDM system with 16 622-megabit/s intensity-modulated channels spaced at 5 gigahertz that achieves a bit-error rate of 10^{-9} at an average received power of -40 dBm.

Despite the continuous improvement in WDM technology, there are still problems to overcome before high-capacity WDM systems can be brought to market. Much of the difficulty in developing WDM systems revolves around the challenge of developing a laser with the required capabilities. Lasers today represent a technology bottleneck: research and engineering must develop, package, and commercialize a narrow-linewidth, frequency-stable, rapidly tunable laser that is cost-effective, reliable, field-deployable, and safe. We also note that it is easier to implement an n-channel WDM system on multiple fibers than on a single fiber, because the channel spacings can be less demanding in the multifiber system than in the single-fiber system.

Erbium-doped fiber amplifiers (EDFAs) promise to make it possible to boost optical signals at any point on the fiber [GDT+89]. However, because EDFAs require lasers to amplify signals, they are costly and prone to failure. The failure of an EDFA has serious consequences, since it can affect a large number of users. Furthermore, few EDFA product offerings exist, and those currently on the market are expensive. Certainly it would not be advantageous to use a large number of EDFAs in the network, e.g. in a configuration that incorporates an EDFA into the receiver. For these reasons we do not propose to use EDFAs in our design, but they could be used at such time as EDFAs are widely available at competitive prices.

On the one hand, passive optical technology appears to be well-developed, while, on the other hand, coherent lightwave technology to implement WDM is still developing. It seems possible to use passive optical technology to implement a space-division–multiplexing multifiber network. Furthermore, we can also achieve modest levels of WDM with existing technology (tens versus hundreds of high-speed channels). It is also cost-effective to obtain a large set of channels by using a modest level of WDM on each of a collection of fibers, since the cost of implementing a modest level of WDM on a single fiber is reasonably low.

In the next section we describe a network architecture that uses the components just described to realize a high-performance network based on space- and wavelength-division–multiplexing.

4 A multifiber optical network architecture

In this section we present a multifiber optical network architecture based on the principle of embedding a large number of independent "fiber" plants within a single cable plant, as illustrated in figure 2. Each *fiber plant* is a fully broadcast medium that can distribute an optical signal among a number of stations without the benefit of optical amplification. Each station of the network can access a subset of the fiber plants via an optomechanical switch, thereby achieving connectivity to all other stations attached to the chosen fiber plant.

Although optical-fiber networks have been implemented in several topologies, we focus on the tree topology, such as the TreeNet architecture of [GF88]. TreeNet provides connectivity for a modest number of stations, supports a reasonable optical-power budget, and requires no optical amplification. Furthermore, we can embed fiber plants within a large cable plant in a natural way. Given a large arboriform multifiber cable plant, we connect groups of stations by smaller subtrees embedded within the supertree. Thus, several subtrees can be effectively overlaid onto the multifiber supertree. Figure 3 shows an example of an eight-leaf subtree embedded within the 32-leaf supertree. Since $2^N - N - 1$ distinct subtrees of two or more leaves can be embedded within a complete N-leaf supertree, it is clear that we can embed many different fiber plants within a single cable plant, if its fiber count is high enough. The number of fiber plants that can be embedded within a cable plant is limited by the number of fibers contained within the cable, i.e, the number of fiber plants that share a segment of cable cannot exceed the count of fibers in this segment. Another constraint is on the size of the fiber plant, since the worst-case optical-power loss inherent to the fiber plant increases as we increase the number of stations that have access to that plant.

We next present a specific network design that is based on the concept of embedding subtree fiber plants within a common supertree cable plant.

4.1 A description of the multifiber optical network

The multifiber optical network uses a multifiber cable plant connected by couplers, splices, and optomechanical switches. The network is designed to allow any two stations to communicate directly and without intermediate packet or circuit switches, i.e., in a single-hop manner—although multihop communication is also allowed for users that can tolerate variation in delay. Topologically, the network is a binary tree that consists of nodal "closets" joined by multifiber optical cables, as shown in figure 4. The closets house the couplers and splices that interconnect the optical fibers of the cable plant. Consisting of an array of couplers whose input and output ports are connected to external fibers in a set pattern, closets can be manufactured

to specification by automated techniques and efficiently installed in the field by ribbon splicing. The stations of the network are grouped into clusters. The network cable plant can be viewed as two tiers of trees, the intercluster and the intracluster cable plants. All stations of a given cluster are connected by an intracluster tree, and the intercluster tree connects groups of clusters.

Next we describe how several fiber plants can be embedded within the multifiber cable plant shown in figure 4. Specifically, we address the nontrivial problem of how to provide a collection of fiber plants, so that there is complete connectivity among a large population of stations and the fiber plants are capable of meeting a reasonable optical-power budget.

We assume that there are N clusters of stations and that each cluster contains M stations. There are B fiber plants to be embedded within the cable plant, and each fiber plant is to serve a group of $K \leq N$ clusters. The cable has a count of F optical fibers. For simplicity we assume that M, N, and K are powers of two, with $M = 2^m$, $N = 2^n$, and $K = 2^k$.

The intercluster tree consists of n levels of closets. The closets serve as hubs for connecting incoming and outgoing fibers. Within a closet pairs of fibers can be joined with a splice in a pass-through configuration, triplets of fibers can be joined by a three-port fused biconical-taper coupler (FBTC) in a splitter–combiner configuration, and pairs of fibers can be joined by a two-port reflective star coupler (RSC) in a root-node configuration. Thus, the root closet houses only RSCs, while nonroot closets can house RSCs, FBTCs, and splices.

The intracluster tree consists of m levels of closets. The top closet (i.e., the root of the subtree) houses one RSC and several FBTCs, while the inferior closets house only FBTCs. In addition to providing intercluster connectivity, this tree serves as the principal medium through which stations of the cluster communicate. The RSC at the top of the tree, the $m - 1$ levels of FBTCs, and the optical fibers connecting them comprise a fiber plant that physically connects only those stations of the cluster. We use a multiaccess protocol to allow the stations of a cluster to share the intracluster network for the exchange of datagram traffic. Since the root of the intracluster subtree is situated close to the stations of the cluster, propagation delay is small, so the token-bus scheme would be a simple, efficient solution.

A station of the network has two transceivers, one for intercluster and one for intracluster communication, as shown in figure 5. The intercluster transceiver connects to one of many optical fibers by means of a $1 \times B$ optomechanical switch (there are B intercluster fibers). Once connected to a fiber, the station is able to communicate with any other cluster also connected to the cable plant. Not every fiber in the cable can be used by the station, as some fibers are "dead" by virtue of not being connected to other fibers in the lowest-level closet (these fibers are not shown in the figure). The intercluster transceiver is tunable over a limited range of wavelengths. Although not shown in figure 5, other tunable lasers can be incorporated into the station so that multiple circuit-switched sessions with the station could exist. Also, tunable lasers could be modulated at different speeds to support the simultaneous transmission and reception of real-time data with different throughput requirements. Lasers modulated at a particular speed would be tunable over a band of wavelengths allocated for WDM channels of that speed. Stations therefore use a combination of space- and wavelength-division multiplexing. That is, once a source station in one cluster and a destination station in another cluster have mutually selected a fiber plant through which they will communicate, the two stations must agree on a common wavelength.

4.2 The optical-power budget

The size of the network is influenced by the optical-power-loss characteristics of the overall cable plant. Each fiber plant has an optical-power loss that depends on the network parameters M and K. A station in one cluster of a given intercluster fiber plant transmits to a station in another cluster of the fiber plant by propagating its optical signal up its intracluster tree, up the intercluster tree, and back down again. Since the intracluster tree has m levels of couplers and the intercluster tree has a k-level subtree embedded within the n-level supertree, an optical signal must pass through $m + k - 1$ power-combining couplers on its way up to the root and $m + k$ power-splitting couplers on its way back down from the root. Letting ℓ_{PC}, ℓ_{PS}, and ℓ_{XS} represent the power-combining, power-splitting, and excess losses, respectively, of a coupler, we see that the total loss in an intercluster fiber plant is at least

$$(m + k - 1)(\ell_{\mathrm{PC}} + \ell_{\mathrm{PS}} + 2\ell_{\mathrm{XS}}) + \ell_{\mathrm{PS}} + \ell_{\mathrm{XS}} \qquad (1)$$

The optical-power loss in the intracluster fiber plant is similarly computed, except that the signal is always contained within the intracluster cable plant:

$$(m - 1)(\ell_{\mathrm{PC}} + \ell_{\mathrm{PS}} + 2\ell_{\mathrm{XS}}) + \ell_{\mathrm{PS}} + \ell_{\mathrm{XS}} \qquad (2)$$

The optical-power budget for a transmitter–receiver pair on the same intercluster or intracluster fiber plant, respectively, must exceed the losses represented by equations (1) or (2), respectively. The optical-power loss values of equation (1) is a lower bound on the actual worst-case loss characteristics of the fiber plants, because the equation does not take into account the effects of fiber splices or of the attenuation that results when the signal propagates over the length of the fiber. These losses are difficult to evaluate exactly, but they can be bounded since no fiber plant has a root-to-leaf path with more than $n - k$ closet splices (i.e., the fiber plant has k levels of couplers embedded within the n-level supertree) and the longest path length of any fiber plant is dominated by the longest path length of the overall cable plant. If the fiber-attenuation loss is ℓ_{FA} per unit distance and the longest path length

of the cable plant is d, then the total attenuation loss is bounded from above by $d\ell_{FA}$. Likewise, the loss attributable to splices is bounded by $2(n-k)\ell_{FS}$, where ℓ_{FS} is the insertion loss of a fiber splice. Thus, an upper bound on ℓ_{WC}, the worst-case optical-power loss in a fiber plant, is given by

$$\begin{aligned}\ell_{WC} &= (m+k-1)(\ell_{PC}+\ell_{PS}+2\ell_{XS})+ \\ &\quad \ell_{PS}+\ell_{XS}+2(n-k)\ell_{FS}+d\ell_{FA} \qquad (3)\end{aligned}$$

If the optical-power budget of a transmitter–receiver pair exceeds the value of this expression, then the pair can communicate dependably over any fiber plant.

The embedding of several fiber plants within a single multifiber cable plant is restricted by the number of fibers available in a cable. We note that we can always embed at least F fiber plants in the cable plant, since each fiber plant could use a dedicated fiber. However, in a cable plant of fiber count F, it is usually possible to embed more than F fiber plants. We shall see an example of such an embedding below.

4.3 A construction based on pair coverings

Can we choose a group size that permits complete pairwise connectivity, meets a reasonable optical-power budget, supports a large station population, and can be implemented with the given number of optical fibers in a cable? To answer this question, we must make a small digression into a discipline of combinatorics that deals with set coverings. A *covering* of pairs of $S=\{1,2,\ldots,N\}$ by K-sets is a collection of K-element subsets $S_1,S_2,\ldots,S_L$ such that every pair of elements over S is in at least one of the subsets S_i. It is well-known that

$$L \geq \left\lceil \frac{N}{K}\left\lceil \frac{N-1}{K-1}\right\rceil\right\rceil \qquad (4)$$

but it is, in general, unknown whether this lower bound can be achieved, except for small values of K, e.g. L exceeds this bound by 2 or less when $K \leq 4$ [Mil72, Mil73]. Moreover, it is unfortunate that proofs of the existence of a covering seldom give procedures for constructing the covering. Below we give a procedure to construct a specific covering, although the size of the covering is not minimal. The size of the covering is, however, small enough to be useful in many multifiber networks with a reasonable fiber count.

For a given N and K we now show how to construct a covering of pairs by K-sets. Define $B=N/K$. For $1 \leq i \leq B$ we define the $2B$ $(K/2)$-sets $T_{2i-1}=\{K(i-1)+1, K(i-1)+2,\ldots,K(i-1)+K/2\}$ and $T_{2i}=\{K(i-1)+K/2+1, K(i-1)+K/2+2,\ldots,K(i-1)+K\}$. It is clear that these disjoint sets comprise all integers from 1 to N. We take $S_{ij}=T_i \cup T_j$, where $1 \leq i < j \leq 2B$. Since, given any pair (x,y), we can find an i and a j such that $x \in T_i$ and $y \in T_j$, it is clear that $x \in S_{ij}$ and $y \in S_{ij}$. Therefore, the K-sets $S_{1,2}, S_{1,3},\ldots,S_{1,2B}, S_{2,3},\ldots,S_{2B-1,2B}$ form a covering of all pairs over the integers from 1 to N, and the number L of sets in the covering is given by

$$L = 2B^2 - B = 2N^2/K^2 - N/K \qquad (5)$$

The number of sets in this covering is considerably higher than the lower bound of equation (4), although it is less than twice as great as the bound. For example, with $N=64$ and $K=8$, $L=120$ while the lower bound is 72. Even though the covering is not optimal, it is still useful in implementing networks, as we shall see.

The idea behind this construction is that we divide the set of the first N positive integers into $2B$ blocks of $K/2$ successive numbers. We then combine all pairs of these blocks to create a collection of new subsets, each of which contains K numbers. Since there are $2B(2B-1)/2 = 2B^2-B$ pairs of these blocks, this is the number of subsets in this (nonoptimal) covering, as we note in equation (5). An example of how the construction can be embedded in an arboriform cable plant is illustrated in figure 6 for $N=64$, $K=8$, and $B=8$; we show in the figure only four of the 120 8-sets that comprise the covering. These 8-sets of clusters are connected by subtrees with three levels of couplers and varying numbers of splices. This embedding is accomplished by representing the network's 64 clusters as the first 64 positive integers and then grouping them into 16 four-element blocks. Pairs of the blocks are combined in accordance with the covering constructed above by linking them together via a single fiber plant. The nodes of the arboriform fiber plant are either spliced (pass-through) or coupled (splitting–combining) connections.

4.4 An analysis of the construction

Since there are N clusters of M stations, the total population of stations is MN. Communication between stations in different clusters of the same group must be coordinated so that the stations select the group mechanically and tune their transceivers to a free wavelength. The groups of clusters can be designated as $S_1,S_2,\ldots,S_L$, and if these sets cover all pairs of clusters, then any pair of stations in the network can communicate. Conversely, if all pairs of stations in the network are to be able to communicate, then it is clear that the K-sets $S_1,S_2,\ldots,S_L$ that correspond to the cluster groups form a covering of all pairs over the first N positive integers.

The covering described above yields a collection of embedded fiber plants that can be used to implement a multifiber optical network. Given that the parameters K, M, and N have been chosen so that the optical-power budget exceeds the worst-case fiber-plant loss of equation (3), the cable plant must be designed to provide fiber plants with fully pairwise connectivity. In addition to these given parameters, we are also constrained by the fiber count F of the cable plant, i.e., the embedding of fiber plant within the cable plant cannot require more than F optical fibers in any segment of cable between two closets.

Recalling from the construction above that $B = N/K$, we can see that the root of the intercluster tree contains B^2 RSCs, since each of the B $(K/2)$-station blocks at the leaves of the left subtree is joined with each of the B $(K/2)$-station blocks at the leaves of the right subtree by a RSC at the root. Likewise, a closet immediately below the root contains $(B/2)^2$ RSCs, and so forth, down to the closets at level $n-k$, each of which contains one RSC.[‡] In general, the number of fiber plants rooted at a closet on level i $(0 \le i \le n-k)$ is $(B/2^i)^2$, which is also the number of RSCs in that closet. Furthermore, each of the closets at the root of the N intracluster trees contains one RSC that is used for the intracluster fiber plant. Thus, the total number of RSCs in the cable plant is

$$N + \sum_{i=0}^{n-k} 2^i \left(\frac{B}{2^i}\right)^2 = N + B^2 \sum_{i=0}^{n-k} 2^{-i} = N + B^2(2 - 2^{k-n})$$

which simplifies to $N + 2B^2 - B$, since $B = N/K = 2^{n-k}$. Therefore, the number N_{RSC} of RSCs in a multifiber optical network based on the above covering is given by

$$N_{\text{RSC}} = N + 2N^2/K^2 - N/K \tag{6}$$

The number of FBTCs in the intercluster cable plant can be calculated by observing that each intercluster fiber plant in the above construction consists of $2^{k+m} - 2$ FBTCs, i.e., the number of nonleaf nodes of a full $(k+m)$-level binary tree minus the root. Clearly, the number of FBTCs used in an intracluster fiber plant is $2^m - 2$, since these fiber plants correspond to full m-level binary trees. In equation (6) N represents the number of RSCs used in intracluster trees while $2(N/K)^2 - N/K$ represents the number of RSCs used in intercluster trees. Therefore, the number N_{FBTC} of FBTCs in a multifiber optical network based on the above covering is given by

$$\begin{aligned} N_{\text{FBTC}} &= [2(N/K)^2 - N/K](2^{k+m} - 2) + N(2^m - 2) \\ &= 2MN^2/K - 4N^2/K^2 - 2N/K - 2N \end{aligned} \tag{7}$$

Next we consider the number of optical-fiber splices required in the multifiber optical network. Although fiber splices would certainly be used at several points in the cable plant because of the need to connect segments of cable, we consider only those closet-housed splices that are inherent to our construction. As noted above, our construction houses all splices in levels 1 through $n-k$ of the intercluster tree. A fiber plant with its root on level i $(0 \le i < n-k)$ includes $n-k-i$ levels of spliced pass-through connections. Since the fiber plant has left and right subtrees, the total number of splices in the fiber plant rooted on level i is $2(n-k-i)$. As discussed earlier, there are $2^i(B/2^i)^2 = B^2/2^i$ fiber plants rooted at level i. Hence the number N_{SPLICE} of closet-housed splices in the cable plant is given by

$$\begin{aligned} N_{\text{SPLICE}} &= \sum_{i=0}^{n-k-1} (B^2/2^i)2(n-k-i) \\ &= B^2 \left[\sum_{i=0}^{n-k-1} 2^{1-i}(n-k) - \sum_{i=0}^{n-k-1} i(1/2)^{i-1} \right] \\ &= B^2 \left[(n-k)(4 - 2^{k+1-n}) - \frac{\mathrm{d}}{\mathrm{d}x} \sum_{i=0}^{n-k-1} x^i \right]_{x=1/2} \\ &= 4(n-k-1)N^2/K^2 + 4N/K \end{aligned} \tag{8}$$

Calling an optical fiber used in a fiber plant *active* (to distinguish it from dead or inactive fibers in the cable), we define $F_{\max}$ be the maximum number of active optical fibers in any segment of cable between two closets. Our cable plant accommodates such an embedding as long as the fiber count F per cable exceeds $F_{\max}$, i.e., $F \ge F_{\max}$. To compute $F_{\max}$ in the embedding of our construction, we proceed by observing that the B^2 RSCs at the root of the intercluster tree connect a total of B^2 fibers coming up from each cable of the left and right subtrees. Furthermore, we observe that closets at levels 1 through $n-k$ of the intercluster tree contain only RSCs and splices when the fiber plants are embedded according to the construction above. To see this more concretely the reader can refer to the example of figure 6—an FBTC at levels 1–3 would imply that more than two four-cluster blocks are linked by a fiber plant, which is contrary to the way the covering was constructed. We have already seen that a closet at level i $(0 \le i \le n-k)$ contains $(B/2^i)^2$ RSCs, while those closets below level $n-k$ contain no RSCs, except for the single RSC at the root of every intracluster tree. At level 1 there are B^2 active fibers entering the "upper" end of the closet from the root, and these fibers will be spliced so that half of them feed down to the left subtree rooted at level 1 while the other half go to the right subtree. Also, the fibers rooted at the closet's RSCs feed down to both subtrees. Inductively, we see that at level i there are $B^2(2^i - 1)/2^{2i-2}$ active optical fibers entering the "upper" end of the closet and $B^2(2^{i+1} - 1)/2^{2i}$ active optical fibers leaving the "lower" end of the closet. At levels below $n-k$ either $2B$ or $2B-1$ active optical fibers are required (one extra fiber for intracluster communication). Clearly, then, the maximum number of active optical fibers is found at the root of the intercluster tree and is equal to B^2, so that

$$F_{\max} = N^2/K^2 \tag{9}$$

An immediate consequence of equation (9) is that we can implement a network based on our construction with no more than B^2 optical fibers per cable.

4.5 Principles of operation

So far we have described a network that provides complete pairwise connectivity to all stations by means of a com-

[‡] We number the levels of the tree from the root, starting at level 0.

bination of space-, wavelength-, and time-division multiplexing. Next we describe the protocols by which users exchange information over this network.

As mentioned earlier, intracluster datagram communication is controlled by a token-bus protocol, similar in spirit to the one used in the IEEE 802.4 local-area–network standard. The token-bus protocol can be used with any broadcast medium. A token is passed among the cluster's stations according to a predetermined sequence, allowing token-holding stations to transmit waiting packets for the duration of the stations' token-holding periods. The next station to receive the token is specified in the successor-station field of the packet and seizes the token as soon as it recognizes its address in this field of the packet. The algorithm that governs token-holding periods can range from a simple "one-shot" rule that permits no more than one transmission per token-holding period to the sophisticated timed–token-rotation protocol specified in the IEEE 802.4 standard. Because the stations of the cluster are normally situated in a compact geographical region, the overhead of token passing remains small, and thus the efficiency of the token-bus protocol is high. The token-bus protocol is adequate for the exchange of datagrams within a cluster to support non–real-time communication. Multicast intracluster communication is readily accomplished using the token-bus protocol, since each intracluster fiber plant is a fully broadcast medium. Thus are all stations within a cluster able to share the intracluster fiber plant by time-division multiplexing.

Although one might expect that a reservation time-division multiple-access protocol would be more efficient for the intracluster network, the token-bus protocol yields acceptable performance yet is simple to implement. In an intracluster token bus the walk time W is given by $W = (r + p)M$, where r is the average round-trip propagation delay between root and leaf, p is the average token-processing and -generation time at a station, and M is the number of stations in the cluster. If u is the normalized network load, then the average cycle time T is [Eis72]

$$T = \frac{W}{1 - u}$$

For a cluster with a 2-km radius, the worst-case round-trip propagation time is $r = 20$ μs. The token-processing and -generation time is dominated by the packet transmission time, since a new token cannot be released until the entire packet has been received. Assuming a 100-megabit/s transmission speed and an average packet size of 1000 bits, we have $p = 10$ μs. With $M = 32$ active stations per cluster (which is the largest cluster size that we ever consider), the average walk time is 960 μs. If the traffic load factor is $u = 0.5$, then the average cycle time T is less than 2 ms, which is satisfactory for non–real-time datagram traffic.

To permit the exchange of non–real-time datagrams between stations in different clusters, the network can use intercluster routers or bridges designed to forward packets from one cluster to another. Such routers can be interconnected by spare fibers in the cable plant, or they can be interconnected via the circuit-switching service soon to be described. These routers would normally consist of multiple interfaces, so that they could attach simultaneously to more than one fiber plant.

To support real-time communication between stations that require dedicated bandwidth and guaranteed response time, the network offers a circuit-switching service. The circuit-switching service is provided by means of the intercluster network. A circuit is established between two stations via their tunable transceivers. The two stations select a common fiber plant and tune their transceivers to an agreed-upon, unused WDM channel of the fiber plant. Having thus established a common circuit, the stations have the entire bandwidth of the channel at their disposal, since no other stations share this channel. Thus, these stations use a combination of space- and wavelength-division multiplexing to communicate with each other.

To support non–real-time traffic, the station uses its fixed-tuning transmitter and receiver to communicate with both intra- and intercluster stations. This communication is via WDM channels and routers that are shared by other stations. Although the circuit-switching service could also be used for non–real-time traffic, this tends to be an inefficient use of resources, given the bursty nature of non–real-time traffic.

Real-time traffic can be classified as low-, medium-, or high-speed. It is possible to divide the optical spectrum of the fiber into low-, medium-, and high-speed WDM channels. Stations would then be equipped with either low-, medium-, and/or high-speed transceivers, according to the requirements of their users. Of course, it would also be possible for all real-time traffic to use only high-speed WDM channels, but this would waste communication resources.

To support the establishment of circuits between stations, the network employs a connection-management protocol. Using the datagram services of the packet-switched network, a calling station can make a request for a connection. Given that the called station can satisfy the request, the call request is positively acknowledged and the chosen fiber plant and WDM channel to be used in the call are also confirmed.

Focusing our attention on a single fiber plant, we are interested in determining how many WDM channels are required to support the stations attached to the fiber plant. Defining P to be the number of stations served by the given fiber plant, we see immediately that $P = KM$. Given that P stations are attached to the fiber plant, there can be a maximum of $P/2$ distinct two-party conversations within the fiber plant. Thus, $KM/2$ WDM channels are sufficient to allow all possible communication within the fiber plant without blocking as a result of exhaustion of channels. Of course, blocking can occur because destination stations are busy with other conversations, but we are now interested only in the effects of the supply of WDM channels on the blocking probability. The following analysis, therefore, will be a pessimistic evaluation of the number of WDM chan-

nels required in the fiber plant to achieve a desired level of blocking.

We assume that each of the $P/2$ possible conversations occurs according to a Poisson process with parameter λ. The duration of a conversation is assumed to be an exponentially distributed random variable with normalized mean 1. Given that the "population" of conversations is $P/2$ and the number of WDM channels is c, the c-server finite-population Markovian queue with blocked calls lost—the M/M/c/c/(P/2) queueing system—provides a queueing model for the system. The steady-state M/M/c/c/(P/2) queue has the following expression for the probability that a call is blocked because no channels are available [Saa83, p. 304]:

$$I\!P\{\text{blocking}\} = \frac{\binom{P/2-1}{c}\lambda^c}{\sum_{i=0}^{c}\binom{P/2-1}{i}\lambda^i} \tag{10}$$

Equation (10) describes the Engset probability distribution, values of which have been tabulated for many combinations of parameters.

In practice, it is unlikely that all P stations would utilize one fiber plant. Therefore, the value of c determined from equation (10) would be higher than the value actually required. A more typical case might be where all MN stations of the network establish circuits with each other. In this case $MN/2$ circuits would be needed. If the circuits were established between random pairs of stations, then each of the fiber plants would host an equal number of circuits. Recalling that the number L of fiber plants in our construction is given in equation (5), we see that the number of circuits per fiber plant is equal to the smallest integer greater than or equal to

$$\frac{MN}{2L} = \frac{MK^2}{4N - 2K} \tag{11}$$

Unlike equation (10), this latter expression for the number of required channels per fiber plant does not depend explicitly on the traffic rate. Rather, the expression is valid when all conversations are evenly balanced across all fiber plants.

In planning the network we can use equations (10) and (11) to estimate the number c of WDM channels required. Given an initial first-order estimate of c, we could then explore designs for transceivers, couplers, and fibers that support the required degree of multiplexing. As discussed earlier $KM/2$ is the maximum number of channels that can be used by a fiber plant. By finding a value of c that—when substituted into equation (10)—yields an acceptably low probability of blocking, we can determine a worst-case level of multiplexing for the given traffic load. Similarly, equation (11) represents more of an ideal-case analysis of c, i.e., one in which the traffic load is perfectly balanced. The disparity between these estimates can be seen by comparing their values for a specific configuration: when $K = 8$, $M = 16$, $N = 64$, and $\lambda = 0.5$, the maximum number of channels needed is 64, the number of channels needed to ensure a blocking probability of 0.01 is 30, and the number of channels needed to handle evenly distributed conversations is five.

5 Network Performance

Because the multifiber optical network provides to each station access to many individual (fully broadcast) fiber plants and—within each fiber plant—many WDM channels, the network can achieve very high aggregate throughput. Given that MN stations are served by the multifiber optical network and that it is possible for all of them to establish simultaneous conversations among themselves with a small number of channels [as specified in equation (11)], we see that the network can support $MN/2$ conversations if the endpoints of the conversations are randomly distributed across the fiber plants. Thus, in the best case the throughput of the network is given by $MNC/2$, where C is the capacity of a single WDM channel. With typical values of M, N, and C, this represents an enormous bandwidth. For example, in the next section we shall see examples of networks that—in the best case—can achieve a throughput of over a terabit per second.

In practice the multifiber optical network does not achieve the level of performance described above, because conversations are not typically distributed randomly over the network. It is common, however, to compare network architectures on the basis of how they handle uniformly distributed traffic. By way of comparison, using 1-gigabit/s channels to interconnect 2048 stations, ShuffleNet achieves a throughput of about 400 gigabits/s [Aca87], the deflection-routing bidirectional Manhattan street network about 250 gigabits/s [DTZ92], and the multifiber optical network about 1 terabit/s. Clearly, the multifiber optical network offers superior performance, especially when one considers that the multifiber optical network does not require amplifiers, store-and-forward buffers, or a high degree of WDM.

6 Examples

To illustrate how to implement the multifiber optical network described in the previous section, we now present an example. We design a network that uses an optical cable carrying 264 optical fibers, i.e., $F = 264$. The network comprises 128 16-station clusters, i.e., $N = 128$ and $M = 16$. The population of stations is therefore equal to $MN = 2048$. We design fiber plants to provide access for groups of eight clusters, i.e., $K = 8$. It is clear that $B = 16$. We assume that the network is to cover a geographical area with a diameter of 25 km, centered about the root. Thus, the longest path length is also 25 km, i.e., $d = 25$ km.

We assume that power-splitting and -combining losses of couplers are 3 dB and that excess loss can be kept to 0.1 dB, i.e., $\ell_{\text{PS}} = 3$ dB, $\ell_{\text{PC}} = 3$ dB, and $\ell_{\text{XS}} = 0.1$ dB.

High-precision splices can be made with loss of about 0.1 dB, i.e., $\ell_{FS} = 0.1$ dB. High-quality single-mode optical fibers can achieve attenuation loss of around 0.2 dB/km in the 1550 nm window, i.e., $\ell_{FA} = 0.2$ dB/km. Substituting the values above into equation (3), we determine the worst-case optical-power loss of the network to be 46.1 dB, which leaves a margin of nearly 4 dB when the receiver sensitivity is −50 dB.

Applying equation (9), we note that at least $F_{max} = 256$ fibers per cable are required to implement the construction. Hence, the 264-fiber cable plant can comfortably accommodate the prescribed embedding. The construction results in the embedding of 496 fiber plants in the 264-fiber cable plant, as one can verify from equation (5).

The cost of the network in terms of closet-housed equipment can be determined from equations (6)–(8). There are 624 RSCs required in the network and 64224 FBTCs. Also, 3136 splices are required in the closets for pass-through connections. Although large numbers of devices are needed to implement this network, the acquisition and installation costs of the actual devices are small relative to the overall cost of installing and maintaining a network of this size.

We use equation (10) to help determine the number of WDM channels required per fiber plant. Figure 7 plots the blocking probability as a function of the call arrival rate for three degrees of multiplexing. With only 30 WDM channels, the blocking probability can be kept low (i.e., below 1 percent) when the mean time to generate a call is twice as long the mean call duration. Referring to equation (11), we see that, on the average, only three channels suffice when the conversations are evenly balanced across all fiber plants. Thus, it appears that we can provide good quality of service for our example network with a reasonable number of WDM channels.

Besides the example network just discussed, other network configurations can be implemented. In table 1 we show the attributes of interest (e.g. population, optical-power budget, component counts, and number of WDM channels) for several choices of the design parameters N, M, and K. The entries of the table are calculated using the same device characteristics, network geography, and maximum blocking probability as in the detailed example above. The chosen entries comprise all multifiber optical networks that can achieve a population of 512 or more stations given a worst-case optical power budget of 50 dB or less using a 264-fiber cable. The final column shows three estimates of the required degree of multiplexing, i.e., the absolute maximum number of channels needed, the pessimistic estimate determined from equation (10) for $\lambda = 0.5$, and the ideal-case estimate determined from equation (11).

Table 1 illustrates tradeoffs faced by the network designer. Each of the measures of interest (e.g. columns 4–10 of table 1) varies as a function of the design parameters N, M, and K. The population increases with M and N; the required power budget increases with K, M, and N; and the level of multiplexing increases with K and M. The component counts, however, increase with M and N but decrease with K. Thus, if we increase M and N to support larger networks, then we also increase the required power budget, component counts, and degree of multiplexing, trading additional hardware complexity for expanded service. On the other hand, if we increase K to reduce component counts, then we increase the required power budget and degree of multiplexing, trading additional hardware complexity of one type (components) for additional hardware complexity of another type (transceivers, lasers, etc.). Applying these tradeoffs, the network designer can choose the most appropriate set of design parameters.

In summary, we have outlined how to implement an all-optical multifiber network that serves without optical amplification more than 2000 stations in a geographical area of diameter 25 km. The technology for the network is not exotic and can be obtained cost-effectively in today's marketplace.

7 Conclusion

Having observed the proliferation of multifiber optical cable, we propose a network architecture based on such cables to achieve high throughput for a large population of users. The scheme is attractive from several points of view: it

- can provide service to many customers in a metropolitan region, given a reasonable optical-power budget;
- does not require the wide use of high-performance lasers, which are costly and failure-prone;
- supports the integration of real-time and non–real-time traffic with a range of throughput requirements; and
- is based on currently obtainable products and technology.

The multifiber optical network combines space-, wavelength-, and time-division multiplexing to deliver integrated communication services to its users. The network uses a mix of single- and multihop routing to deliver messages.

The multifiber optical network suffers from a few disadvantages. Chief among them is that its circuit-switching service can support only one connection at a time, i.e., a calling station might find its destination station busy with another call and have to call back. This situation can be partially mitigated by providing stations with multiple transceivers. As with other circuit-switching approaches, it is difficult to provide a general multicasting service. Multicast transmission of non–real-time traffic can be accomplished by known routing techniques [Dee88]. However, multicast transmission of real-time circuit-switched traffic is more difficult to achieve. Establishing a multicast circuit among a small collection of stations is often feasible in

the multifiber optical network, if all stations share a fiber plant. This, however, cannot always be guaranteed with the covering described in section 4, but other coverings can be devised that guarantee that any group of stations shares at least one fiber plant. It remains a matter of future research to explore the implementation of such coverings. Examining table 1, one might be concerned about the large numbers of couplers needed to implement the multifiber optical network. Components such as FBTCs are very reliable, being implemented entirely from optical fibers with no active devices. Moreover, the cost of an individual FBTC—although not low but steadily falling—drops dramatically with the number of parts ordered. Even with the thousands of FBTCs required by the network, the total cost of these couplers is quite low in relative terms.

References

[Aca87] A. S. Acampora. A multichannel multihop local lightwave network. In *Proceedings of GLOBECOM '87*, pages 37.5.1–37.5.9, Tokyo, Japan, November 1987.

[BFG90] Joseph A. Bannister, Luigi Fratta, and Mario Gerla. Topological design of the wavelength-division optical network. In *Proceedings of IEEE INFOCOM '90*, volume 3, pages 1005–1013, San Francisco, California, June 1990.

[Bir92] Yitzhak Birk. Fiber-optic bus-oriented single-hop interconnections among multi-transceiver stations. In *Proceedings of IEEE INFOCOM '92*, pages 2358–2367, Florence, Italy, May 1992.

[Bra90] Charles A. Brackett. Dense wavelength division multiplexing networks: Principles and applications. *IEEE Journal on Selected Areas in Communications*, 8(6):948–964, August 1990.

[CGK89] I. Chlamtac, A. Ganz, and G. Karmi. Purely optical networks for terabit communication. In *Proceedings of IEEE INFOCOM '89*, volume 3, pages 887–896, Ottawa, Canada, April 1989.

[Dee88] Stephen E. Deering. Multicast routing in internetworks and extended LANs. In *Proceedings of the ACM SIGCOMM '88 Symposium*, pages 55–63, Stanford, California, August 1988.

[DTZ92] M. Dècina, V. Trecordi, and G. Zanolini. Performance analysis of deflection routing multichannel-metropolitan area networks. In *Proceedings of IEEE INFOCOM '92*, pages 2435–2443, Florence, Italy, May 1992.

[Eis72] Martin Eisenberg. Queues with periodic service and changeover time. *Operations Research*, 20:440–451, 1972.

[GDT+89] C. R. Giles, E. Desurvire, J. R. Talman, J. R. Simpson, and P. C. Becker. 2-Gbit/s signal amplification at $\lambda = 1.53$ μm in an erbium-doped single-mode fiber amplifier. *Journal of Lightwave Technology*, 7(4):651–656, April 1989.

[GF88] Mario Gerla and Luigi Fratta. Tree structured fiber optic MAN's. *IEEE Journal on Selected Areas in Communications*, SAC-6(6):934–943, July 1988.

[GKB92] Mario Gerla, Milan Kovačević, and Joseph Bannister. Multilevel optical networks. In *Proceedings of IEEE ICC '92*, Chicago, Illinois, June 1992.

[Hin87] H. Scott Hinton. Photonic switching technology applications. *AT&T Technical Journal*, 66(3):41–53, May/June 1987.

[HMTS90] Gerald J. Herskowitz, Donald N. Merino, Demetrios Tselios, and Vijay C. Shroff. Fibre optic LAN systems and costs. *Computer Communications*, 13(1):37–47, January/February 1990.

[Kar88] Mark J. Karol. Optical interconnection using ShuffleNet multihop networks in multi-connected ring topologies. In *Proceedings of the ACM SIGCOMM '88 Symposium*, pages 25–34, Stanford, California, August 1988.

[Mil72] W. H. Mills. On the covering of pairs by quadruples I. *Journal of Combinatorial Theory, Series A*, 13:55–78, 1972.

[Mil73] W. H. Mills. On the covering of pairs by quadruples II. *Journal of Combinatorial Theory, Series A*, 15:138–166, 1973.

[Pal88] Joseph C. Palais. *Fiber Optic Communications.* Prentice Hall, Englewood Cliffs, New Jersey, second edition, 1988.

[Saa83] Thomas L. Saaty. *Elements of Queueing Theory with Applications.* Dover, New York, New York, 1983.

[Ste91] Thomas E. Stern. A linear lightwave MAN architecture. In Guy Pujolle, editor, *High-Capacity Local and Metropolitan Area Networks*, pages 161–179. Springer-Verlag, Berlin, Germany, 1991.

[Tek90] V. J. Tekippe. Passive fiber-optic components made by the fused biconical taper process. *Fiber and Integrated Optics*, 9:97–123, 1990.

[TONT89] Hiromu Toba, Kazuhiro Oda, Kiyoshi Nosu, and Norio Takato. Optical FDM information distribution systems using channel selective receivers. In *Proceedings of IEEE ICC '89*,

pages 14.5.1–14.5.6, Boston, Massachusetts, June 1989.

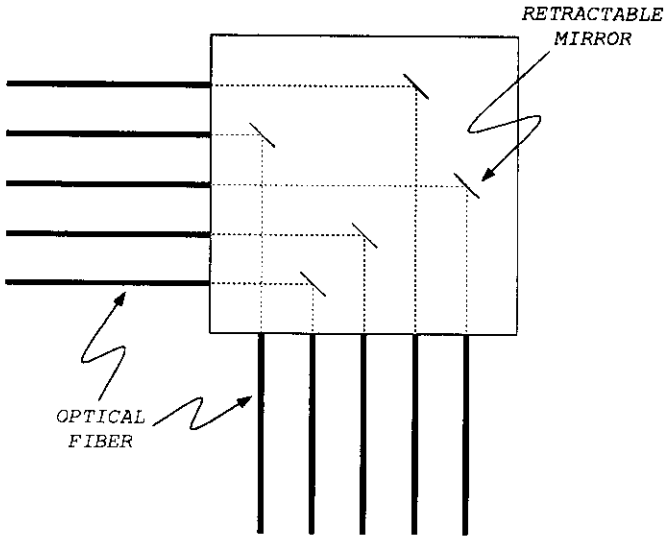

Figure 1: The $n \times n$ optomechanical crossbar

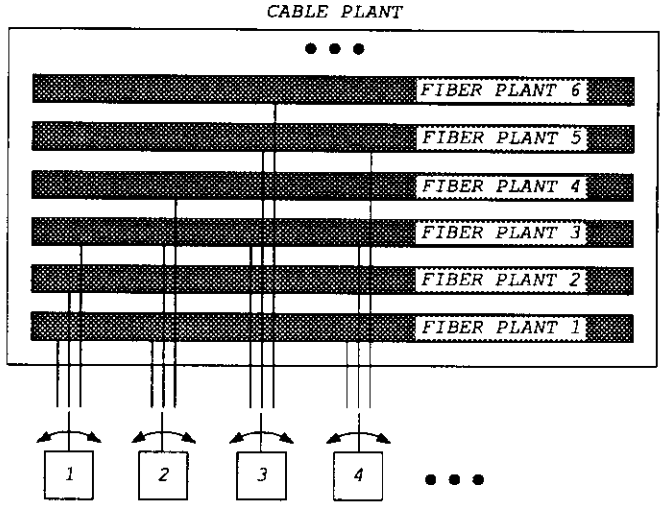

Figure 2: Fiber plants embedded within a cable plant

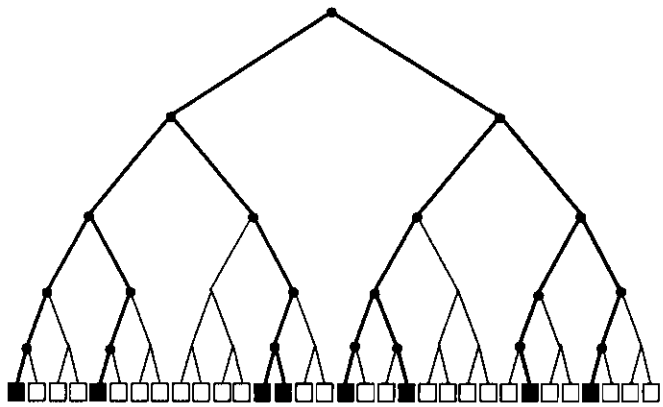

Figure 3: A subtree embedded within a supertree

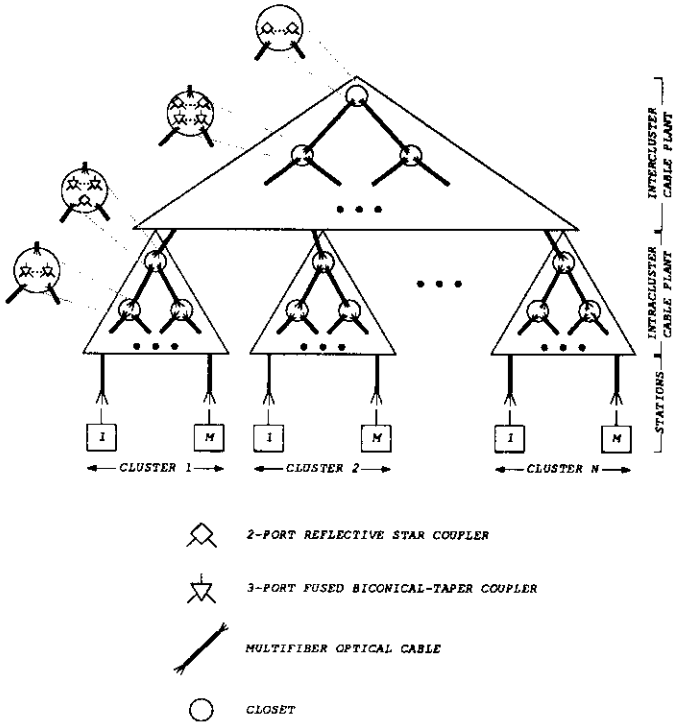

Figure 4: Two-level cable plant of the multifiber optical network

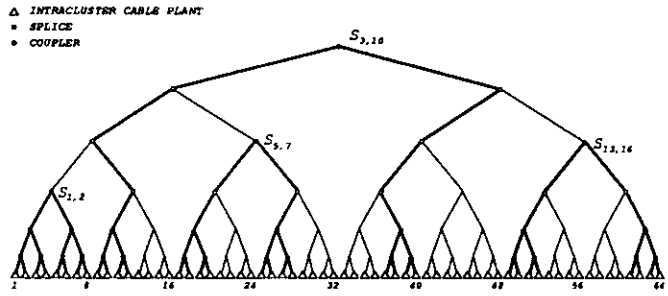

Figure 6: Several fiber plants embedded within a cable plant

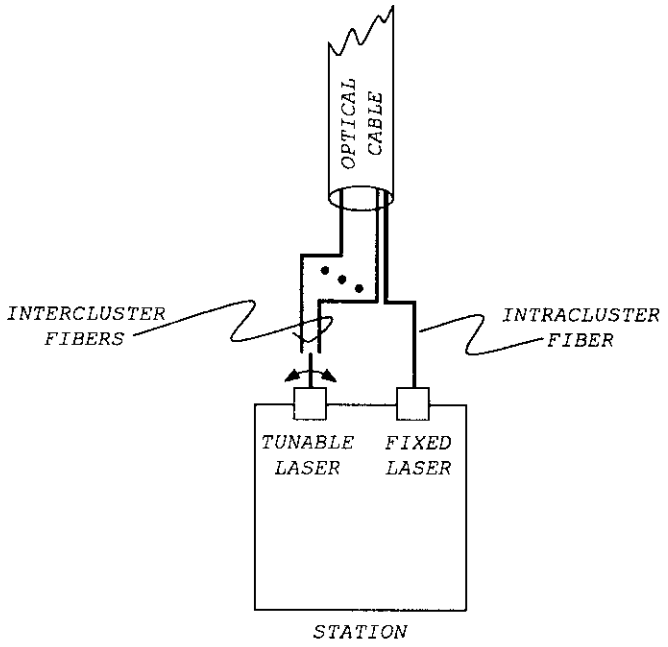

Figure 5: The structure of the station

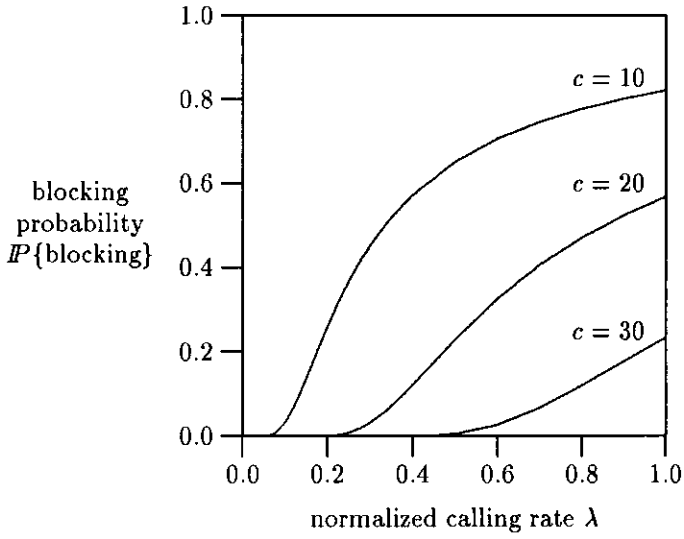

Figure 7: Blocking probability versus calling rate and number of WDM channels for a 2048-station network

Table 1: Network attributes as a function of the design parameters

Design Parameters			Population	Optical Power Budget	Component Counts				Degree of WDM Required
N	M	K			$F_{\max}$	N_{RSC}	N_{FBTC}	N_{SPLICE}	
128	16	8	2048	46.1 dB	256	624	64224	3136	64, 30, 3
64	32	4	2048	46.1 dB	256	560	64352	3136	64, 30, 3
64	16	8	1024	45.9 dB	64	184	15984	544	64, 30, 5
64	16	4	1024	39.9 dB	256	560	31584	3136	32, 17, 2
32	32	4	1024	45.9 dB	64	152	16048	544	64, 30, 5
32	32	2	1024	39.9 dB	256	528	31648	3136	32, 17, 2
16	32	4	512	45.7 dB	16	44	3992	80	64, 30, 10
16	32	2	512	39.7 dB	64	136	7888	544	32, 17, 3