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

DESIGN OF THE WAVELENGTH-DIVISION OPTICAL NETWORK

Joseph A. Bannister
Mario Gerla

May 1989
CSD-890022

Design of the Wavelength-Division Optical Network*

Joseph A. Bannister
Mario Gerla
University of California
Los Angeles, CA 90024-1596

May 3, 1989

Abstract

Fundamental advances in fiber-optic and digital electronic technologies will result in new applications such as wavelength-division multiplexing and fast packet-switching. The Wavelength-Division Optical Network (WON) [Ban88], a generalization of the ShuffleNet concept [Aca87, AKH87, HK88], uses these new applications of technology to achieve substantial increases in the network population and aggregate bandwidth possible in conventional local and metropolitan area networks of comparable cost. We describe the basic architecture of the WON and consider extensions of the basic design, including channel sharing for hop-count reduction and wavelength agility for reconfigurable interconnection of stations.

Motivated by the potential drawback of large delays in poorly designed WONs, we introduce and use a queueing-network model of the WON that provides a general framework for analyzing the performance of these networks. We have used this performance model to formulate a number of significant problems in the large-scale design of the WON, one of which—the Virtual-Topology Design Problem—we consider in detail.

Using the optimization techniques of simulated annealing [KGV83] and genetic algorithms [Hol75, Gol89], we have solved a number of instances of the Virtual-Topology Design Problem. The results have been very encouraging and have yielded significant—and sometimes dramatic—improvement in delay and throughput, compared to previously proposed, unoptimized design approaches such as ShuffleNet and the Manhattan Street Network [Max85]. We discuss our experiences with the optimization algorithm and outline continuing efforts in this area.

*This work was partially supported by a State of California - Pacific Bell MICRO Grant, 4-541121-19907 KC7F.

Contents

List of Tables

List of Figures

1 Introduction

The performance characteristics of fiber-optic and digital electronic technology continue to advance at a rapid pace. We are, however, arriving at a point where the performance potential of fiber optics threatens to overtake that of digital electronics. While it is possible to transmit and receive lightwave signals at the rate of Terabits per second (Tbps), cost-effective interfaces that convert electrical signals to lightwave signals (and vice versa) at such speeds are currently unavailable. It would be extremely difficult to build the electronics needed to write and read data to and from communication media operating in the Tbps range. A computer[1] needing to communicate with another computer is thus presented with a vast "sea of bandwidth" in the fiber-optic communication subsystem but is incapable of driving the interface at the necessary speed. One option is to package the bandwidth of the communication subsystem so that the slower, electrically limited interfaces of the computer may effectively tap in. Although the performance of fiber optics, as gauged by the speed-distance product, is growing steadily, we require the introduction of new techniques for multiplexing several high-speed lightwave signals onto a single optical fiber to realize the huge improvement in the potential bandwidth of computer communication systems. This type of multiplexing, called wavelength-division multiplexing (WDM), also makes possible the use, on a single optical fiber, of several signaling channels that operate at rates compatible with the computer's electrical interface. This is, at least in principle, a means of providing each computer with a manageable portion of an enormous aggregate bandwidth.

This fundamental shift in technologies and their tradeoffs forces the computer communication network designer to consider new architectures for connecting com-

[1] Although we speak primarily of computer-to-computer communication, the network user can be any entity requiring high-performance communication services.

puters. Recognizing the advantages of fast packet-switching using high-speed digital electronics made possible by very large-scale integrated circuits, researchers have proposed local and metropolitan area network (LAN and MAN) architectures based on transmitting a packet from source to destination with possible switching via intermediate stations, e.g. the Manhattan Street Network (MSN) of Maxemchuk [Max85]. This approach, termed multihop, stands in contrast to conventional LAN and MAN architectures, which typically allow delivery in a single hop by means of complete sharing of the transmission medium. The multihop approach can result in networks that have, in comparison with conventional LANs and MANs, much higher total throughput, wider geographical dispersion, and larger populations of stations. One of the most attractive and exciting proposals for large, high-speed LANs and MANs, ShuffleNet [Aca87, AKH87, HK88], is based on the premise that fiber optics of the future will offer essentially infinite aggregate bandwidth that can be unbundled and offered to individual network users in bite-size chunks. In this paper we propose the *Wavelength-Division Optical Network (WON)*. The WON, based on the combination of WDM and multihop approaches, generalizes networks such as the MSN and ShuffleNet by permitting a wide range of station interconnections. Although the WON clearly enables one to construct networks with plentiful aggregate bandwidth, it is a nontrivial problem to structure these networks in a way that yields acceptable cost, performance, and dependability. The richness of possible interconnections in the WON also imposes on the designer a number of complex design decisions to be resolved. We intend to undertake in this research a careful study of some of these design decisions and their impact on the operating characteristics of the WON. The principal issue addressed in this paper is the design of WONs with the goal of optimizing specific performance requirements. Although the original architects of the multihop approach have considered certain fundamental design problems, we believe that our research goes considerably beyond these pioneering steps and deals with the design and analysis of such networks in a realistic and original way.

The remainder of this paper is organized as follows. Section 2 provides the reader with essential background information, including a description of the architecture of the WON, a queueing-network model for evaluating the performance of the WON, and the problem of virtual-topology design in the WON. In section 3 we propose optimization techniques for designing the virtual topology of the WON and present the results of a number of experiments intended to compare the performance of unoptimized and optimized WONs. In section 4 we discuss the significance of our results, draw appropriate conclusions, and outline future work.

2 Background

In this section we provide the reader with a description of the WON, a model of its performance, and a formulation of the Virtual-Topology Design Problem (VTDP).

2.1 Architecture of the Wavelength-Division Optical Network

The WON generalizes the ShuffleNet architecture originally proposed by Acampora [Aca87] and subsequently refined and studied in [AKH87, HK88]. The WON is a store-and-forward communication network that uses an optical waveguide to transmit and receive lightwave signals. As a MAN, the WON would be expected in the extreme case to serve several thousands of users in a metropolitan region spanning a radius of a few hundred kilometers.

An optical fiber medium has the potential to carry signals at the rate of several Terabits per second (Tbps), but, since the digital electronic circuits that attach to this medium operate at rates far below this capacity, it is necessary to divide the medium's bandwidth into several independent channels that operate at slower electronic speeds. For the sake of discussion we assume that the upper speed limit of the digital electronic interface to the fiber-optic system is 1 Gbps, so that the system can be divided up into a very large collection of 1-Gbps WDM channels. This partitioning of the medium's bandwidth constructs WDM channels corresponding to different wavelengths of light which can be independently and simultaneously modulated and demodulated by stations tuned to these channels. The situation is analogous to cable television-based broadband networks that use frequency-division multiplexing in the high- to very-high–frequency bands, except, of course, the number of channels and the bandwidth of each channel are substantially greater in WDM lightwave systems.

The WON is composed of stations arranged on a single fiber-optic distribution system that is capable of carrying the full complement of WDM channels. Signals in each of the WDM channels may be received by any station that is tuned to the channel, and, likewise, transmission is permitted to any station tuned to the channel; but it will usually be necessary to insure orderly and nonconflicting access if multiple stations' transmitters are tuned to a given channel. The medium may be topologically laid out in a variety of configurations: proposed topologies include the bus, star, tree, and ring. It is conceptually easiest to think of the WON as a linear bus, and some of our examples will be presented using the bus topology, but, given the bus's poor signal loss characteristics, we would normally implement the large WON in a tree topology with a wideband amplifier at the headend, such as Tree-Net [Ger88, GF88], which has been shown to accommodate greater numbers of stations with better signal loss characteristics than the other topologies. Stations passively attach to the medium by means of multiple fiber-optic transmitters and receivers which must be permanently or semi-permanently tuned to one of the WDM channels; there is no restriction on the tuning of the transmitters and receivers, but, as we shall see later, the chosen tuning can have a substantial impact on the network's performance.

Each station is a store-and-forward packet switch with p transmitters and p re-

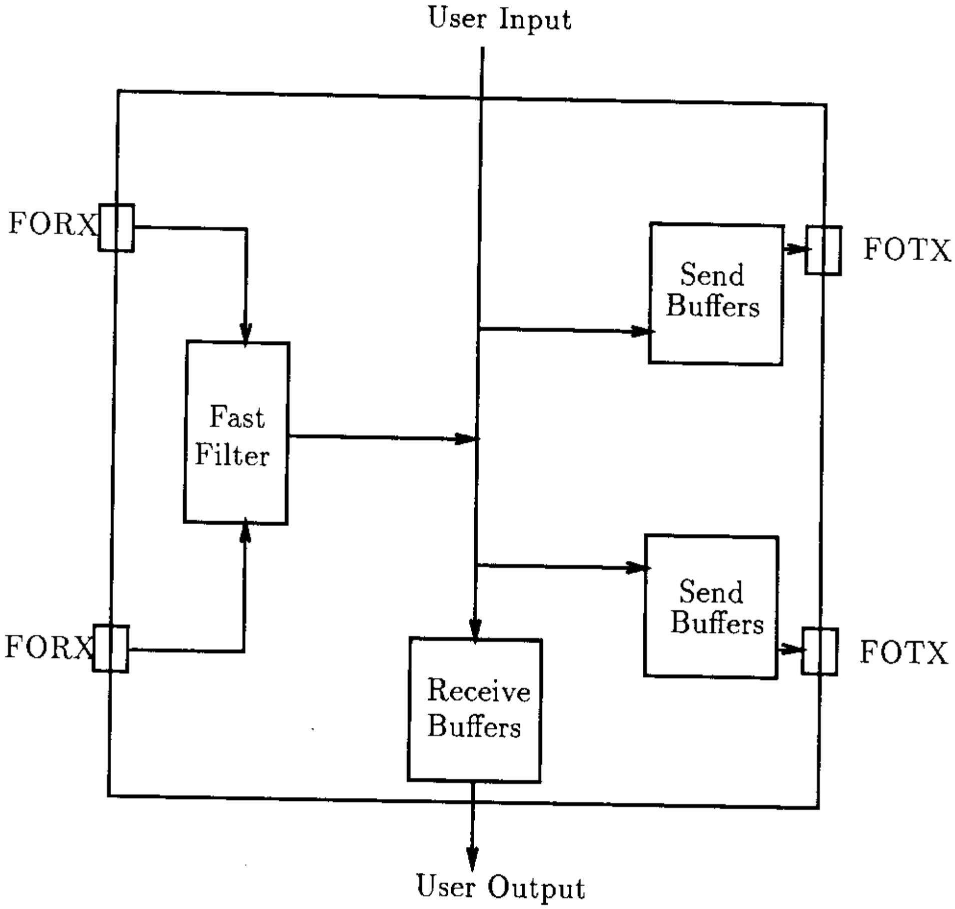

Figure 1: The WON Station Architecture.

ceivers attached to the optical fiber medium. To keep the cost of a station low, p is usually a small number; it will be common to choose $p = 2$, resulting in a dual-transceiver station. A port is also available for connecting users (e.g. computers, local networks, gateways, and other equipment) to the station. The station architecture is shown in figure 1. A packet arrives to the station through the incoming user port or receiver and is buffered in fast memory, its header is examined and the station makes a routing decision to send the packet through a specific transmitter or the outgoing user port. Routing decisions may be based on fast table lookup, hardcoded structural information about the network, or other state information—the main requirement is that it be fast since message transmission and reception occur at Gbps rates.

Each station of the WON must tune its transmitters and receivers to WDM channels in a way that permits efficient data transfer between all stations. The *interconnection graph* or *virtual topology* of a WON is a directed graph whose nodes correspond to the stations of the WON; the correspondence will usually be indicated by numbering both stations and nodes 0, 1,..., $N-1$. The interconnection graph contains an arc from node i to node j if and only if one of station i's transmitters and one of station j's receivers are tuned to the same channel. Paths in the interconnection graph represent multihop routes that a message takes in traveling from its source station to its destination station. Clearly, one of the requirements of the interconnection graph is that it permit connectivity between all pairs of stations. It is also desirable that no message traverse "too many" WDM channels in its journey from source to destination. For example, we could achieve complete connectivity by using a ring as the interconnection graph, but for the WON with a large number of stations this type of interconnection graph would imply a large number of hops for the typical message, which will obviously affect performance, especially when we recall that each hop may cover hundreds of kilometers.[2]

Given the impact of hop count on performance, it is important to choose an interconnection graph that provides short paths between pairs of nodes. The ShuffleNet architecture specifies the recirculating perfect p-ary shuffle as its interconnection graph because this type of graph is well known for its small diameter, viz. the maximum number of hops in any shortest path of the graph. The ShuffleNet interconnection graph consists of $N = mp^m$ nodes, each of which has p incoming arcs and p outgoing arcs. If we number the nodes of the interconnection graph from 0 to $N-1$ and the incoming and outgoing arcs of each node from 0 to $p-1$, then we can express the

[2] We are discussing a *logical*ring in which logical neighbors may be geographically separated from each other by large distances.

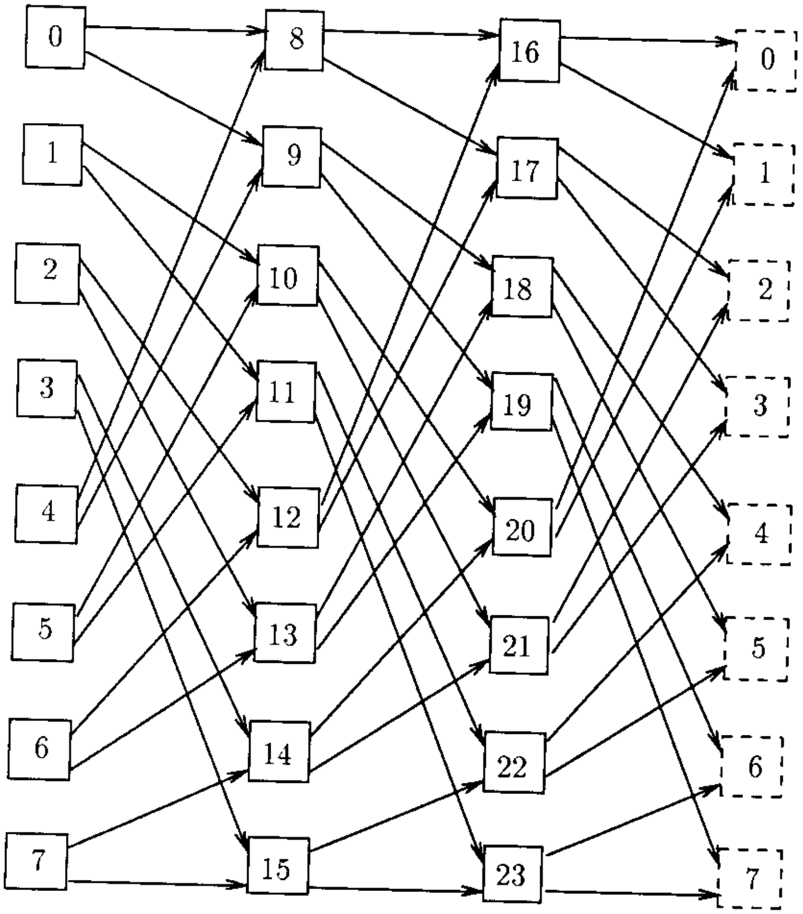

Figure 2: The 24-Node Binary ShuffleNet Interconnection Graph.

adjacency relation of the graph by specifying that node i's jth outgoing arc is the lth incoming arc of node k, where

$$l = \left\lfloor \frac{i \bmod p^m}{p^{m-1}} \right\rfloor$$

$$k = \left\{ p^m \left\lceil \frac{i}{p^m} \right\rceil + p \left[(i \bmod p^m) \bmod p^{m-1} \right] + j \right\} \bmod N$$

We can see from figure 2, which depicts the interconnection graph of the binary ShuffleNet of 24 nodes, that the general ShuffleNet interconnection graph is arranged in m stages, and the nodes of adjacent stages are connected according to the p-way shuffle. One of the most attractive properties of ShuffleNet is its small diameter relative to its size (such a graph is said to be dense): no two nodes are separated by more that $2m - 1$ hops, and the average number of hops between two nodes is [AKH87]

$$E[\text{hops}] = \frac{kp^m(p-1)(3m-1) - 2m(p^m - 1)}{2(p-1)(mp^m - 1)}$$

Unfortunately, this equation does not apply if we wish to model a WON with nonuniform traffic, or if the different hops of the WON have different lengths. These shortcomings highlight the need for more sophisticated and realistic models of the WON, one of which we shall present and develop later.

In addition to the recirculating perfect p-ary shuffle, we can construct the WON from other interconnection graphs such as the p-dimensional torus, the p-ary de Bruijn graph, the p-ary tree, etc. These graphs, since each node has exactly p incoming arcs and p outgoing arcs, are called p-regular directed graphs. One of the goals of our research is to understand what types of regular graphs provide good virtual topologies for the WON.

If some node of the interconnection graph has more than K outgoing arcs, then the corresponding station of the WON must be transmitting to more than one station on a single WDM channel, and such a WON is called a *shared-channel* WON. If none of the WON's channels are shared, then it is a *dedicated-channel* WON. To permit orderly and efficient communication over channels shared by more than one transmitter in the shared-channel WON, we must use a multiple-access protocol on those channels.

Channel sharing in the WON is used for two distinct purposes. In [Aca87] it was shown that ShuffleNet could be implemented by using single-transceiver stations that shared channels, which provides a means of making a more cost-effective network. The other justification for channel sharing that has been pointed out in [Ban88] is performance improvement that can result from the reduction in hop count since shared channels provide more shortest paths in the network's interconnection graph.

It is this latter use of channel sharing that we shall concentrate on in later sections.

It is instructive to contrast the WON with two other multihop network architectures, Maxemchuk's MSN [Max85] and the Store-and-Forward with Integrated Frequency-Time (SWIFT) network [CG87, CGK88]. Like the WON, these networks achieve high throughput by allowing the concurrent use of many channels, which is made possible by providing multihopping capabilities in the networks' store-and-forward stations. The MSN is physically laid out as a two-dimensional torus with each station connected by optical fibers to its four geographically nearest neighbors so that messages can be easily routed from source to destination in a small number of hops; the WON, in contrast, has several WDM channels on a single (albeit long) length of optical fiber and is restricted to neither the physical grid topology nor the toroidal interconnection of the MSN. In the SWIFT architecture all store-and-forward stations share a common medium partitioned into a small number of channels, but each station dynamically tunes its transmitter and receiver according to a fixed schedule which provides multihop paths between all source-destination pairs; the WON, with its abundance of channels, can use a fixed tuning of stations to channels and thus can avoid the need to coordinate the dynamic tuning of transmitters and receivers, which would anyways be difficult since light sources and detectors typically require tuning times that are orders of magnitude greater than message delivery times.

2.2 Performance Model

The design of a network requires well defined criteria for deciding when one solution performs better than another. There are two basic performance metrics of interest in the WON: mean packet delay, which is the time that a typical packet spends in the WON, and maximum throughput, which is the amount of traffic that can be offered to the network without the occurrence of congestion, buffer overflow, or excessive queueing.

We next outline the queueing-network model of the WON, which is similar to the well known model originally proposed by Kleinrock [Kle64, Kle76] to evaluate the performance of wide area packet-switch networks. Our queueing-network model, however, can be used to analyze both dedicated- and shared-channels WONs.

We model the WON as an open queueing network whose nodes correspond to components of the interconnection graph in a way that will be described below. For the sake of analytical tractability we must make the traditional assumptions about the WON. All packets offered to the WON are assumed to have their lengths (in bits) chosen from an exponential distribution with mean $1/\mu$. A packet originating at the user input port of station i travels through the network to its final destination via a series of stations (packet switches) and channels in accordance with some routing procedure. The specifics of the routing procedure do not concern us at this point except that we assume that the next hop for a packet is chosen from a set of alternatives probabilistically and independently of other events in the network. Furthermore, as the packet completes a hop we suppose that its length is independently chosen anew from an exponential distribution with mean $1/\mu$; this is Kleinrock's celebrated Independence Assumption, and, as long as there is adequate mixing of traffic at routing points, it is usually a reasonable approximation. Packets destined for station j arrive to the user input port of station i as a Poisson process of intensity γ_{ij} packets per second, and the $N \times N$ matrix (γ_{ij}) is called the network *traffic matrix.*

An $N \times N$ *distance matrix* (δ_{ij}) specifies the distance (in kilometers) between stations i and j. The distance is measured as the length of optical fiber from station i to station j. The transmission signal propagates along this optical-fiber path at the speed of light through glass, which we take to be $\tilde{c} \approx 2 \times 10^5$ kilometers per second. Thus the time for a bit to propagate from station i to station j is $\delta_{ij}/\tilde{c}$ seconds.

The queueing-network model of the WON falls into the Baskett-Chandy-Muntz-Palacios (BCMP) class of open product-form queueing networks. The service centers of the queueing network are either infinite-server (IS) centers, which model propagation delay, or first-come–first-serve (FCFS) centers, which model queueing delay. The customers of the queueing network represent packets of the WON and move through the queueing network according to a routing chain for each source-destination pair in the network. There may be as many as N^2 IS centers in the network, each denoted $\mathcal{C}_{ij,IS}$ where $1 \leq i \leq N$ and $1 \leq j \leq N$. Service center $\mathcal{C}_{ij,IS}$ models the propagation delay experienced by a packet as it hops directly from station i to station j. Thus service center $\mathcal{C}_{ij,IS}$ has deterministic service time $\delta_{ij}/\tilde{c}$. There are K FCFS centers in the network (one for each channel of the WON), each denoted $\mathcal{C}_{k,FCFS}$ where $1 \leq k \leq K$. Service center $\mathcal{C}_{k,FCFS}$ models the queueing delay experienced by a packet from the time that it queues for transmission at channel k until it is successfully placed on the channel. The service center $\mathcal{C}_{k,FCFS}$ has a state-dependent service rate $\mu_k(n)$ that depends upon both the channel index k and the number n of packets queued for transmission at a particular instant in time. Thus the time to transmit a packet of length $1/\mu$ when there are n packets in the queue is $1/\mu B \mu_k(n)$ seconds, where B is the channel speed in bits per second.

The use of the state-dependent FCFS service center to model channel-access delay allows us to incorporate a number of different channel-access schemes into the network. The dedicated channel is straightforward to model: $\mathcal{C}_{k,FCFS}$ is the ordinary fixed-rate FCFS center with $\mu_k(n) = 1$ for all n. By choosing the appropriate form of the service-rate function $\mu_k(n)$ we can model other access schemes; for example, in [LZGS84, 339–341] a service-rate function is proposed for modeling access delay in the carrier-sense–multiple-access protocol. Results provided by the use of the state-dependent FCFS center as a channel model are in general only approximations. This is because multiple-access protocols do not usually provide service in FCFS order (or, for that matter, according to any of the other BCMP service disciplines). Moreover, multiple-access protocols usually provide nonconservative service in which the server may remain idle even though there is work to be done in the queue. Despite these shortcomings, the state-dependent service center can still be used as a reasonable approximation to the queueing delay in shared channels [GPL83]. At any rate, our principal concern is not to achieve a high degree of accuracy in the prediction of packet delay but rather to compare the relative performance of different designs.

With each pair of stations we associate a routing chain consisting of the paths used by packets exchanged between the pair. For the purposes of simplifying the discussion we can assume that the WON uses fixed single-path routing. When a packet from a particular source-destination pair is routed over a given channel, it makes a contribution to the traffic loading of that channel. Given the routes of all source-destination pairs, we can determine the throughput at each service center of the queueing-network model. The quantities $\lambda_{ij,IS}$ and $\lambda_{k,FCFS}$ give the throughputs (in packets per second) at centers $\mathcal{C}_{ij,IS}$ and $\mathcal{C}_{k,FCFS}$, respectively.

The service centers of the queueing network are connected in a manner that naturally reflects the virtual topology of the WON. The IS center $\mathcal{C}_{ij,IS}$ is directly connected to the FCFS center $\mathcal{C}_{k,FCFS}$ if station j transmits on channel k. If station l transmits on channel k and station m receives on channel k, then service center $\mathcal{C}_{k,FCFS}$ is directly connected to service center $\mathcal{C}_{lm,IS}$. This situation is depicted in figure 3.

In a product-form queueing network the mean queue length at each service center has a convenient expression: if we let the random variables $L_{ij,IS}$ and $L_{k,FCFS}$ represent the number of customers at centers $\mathcal{C}_{ij,IS}$ and $\mathcal{C}_{k,FCFS}$, respectively, and there is some value g for which $\mu_k(n) = \mu_k(g)$ for all $n \geq g$, then

$$E[L_{ij,IS}] = \lambda_{ij}\delta_{ij}/\tilde{c} \tag{1}$$

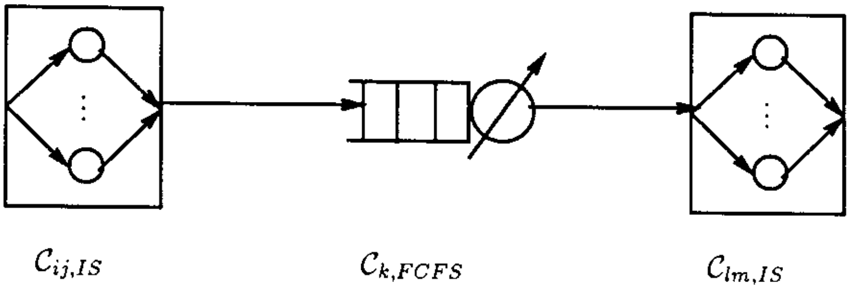

Figure 3: A Fragment of the Queueing-Network Model of the WON.

and [LS83, pages 153–155]

$$\begin{aligned} E[L_{k,FCFS}] &= p_0 \tilde{\rho}_k \left\{ \frac{\mu_k(g)}{\mu_k(g-1)} \frac{1}{(1-\rho_k)^2} + \right. \\ & \left. \sum_{i=0}^{g-2} i \tilde{\rho}_k^i \left[\frac{1}{\prod_{j=2}^{i+1} \mu_k(j)} - \frac{1}{\mu_k(g-1)[\mu_k(g)]^{i-1}} \right] \right\} \end{aligned} \tag{2}$$

where $\rho_k \triangleq \lambda_k / \mu B \mu_k(g)$, $\tilde{\rho}_k \triangleq \lambda_k / \mu B$, and

$$p_0 = \left\{ \frac{\mu_k(g)}{\mu_k(g-1)} \frac{1}{1-\rho_k} + \sum_{i=0}^{g-2} i \tilde{\rho}_k^i \left[\frac{1}{\prod_{j=2}^{i+1} \mu_k(j)} - \frac{1}{\mu_k(g-1)[\mu_k(g)]^{i-1}} \right] \right\}$$

In particular, if $C_{k,FCFS}$ is a fixed-rate FCFS center with $\mu_k(n) = 1$ for all n, then

$$E[L_{k,FCFS}] = \frac{\lambda_k/\mu B}{1 - \lambda_k/\mu B} \tag{3}$$

We may now give an explicit expression for the mean packet delay in the WON. Let the random variable L stand for the steady-state number of customers in the network. We may use Little's Result to express the mean packet delay as

$$E[T] = \frac{1}{\gamma} E[L] \tag{4}$$

where $\gamma \triangleq \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij}$ is defined to be the total of exogenous traffic offered to the network. Since the total number of packets in the network is the sum of packets at each service center, we may write

$$E[L] = \sum_{i=1}^{N} \sum_{j=1}^{N} E[L_{ij,IS}] + \sum_{k=1}^{K} E[L_{k,FCFS}]$$

Substituting this last equation back into equation (4), we obtain the basic formula for mean packet delay:

$$E[T] = \frac{1}{\gamma}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N} E[L_{ij,IS}] + \sum_{k=1}^{K} E[L_{k,FCFS}]\right\} \tag{5}$$

where we can use equations (1) and (2) to evaluate the expressions within the summations.

The formula for mean packet delay given in equation (5) has a queueing-delay component which is a function of the traffic load on the channels, and a propagation-delay component which is a function of the distance that a signal must travel from source to destination. When used as a MAN the dedicated-channel WON will have packet delays in which propagation delay is the dominant component. For example, a 1000-bit packet will require 1 microsecond for transmission and 500 microseconds for propagation on a 100-kilometer link operating at 1 Gbps. These ratios of queueing delay to propagation delay will be preserved even when traffic is increased—by the time queueing delay begins to approach propagation delay the problem of packet-buffer depletion will predominate. Thus it is often reasonable to approximate mean packet delay in the dedicated-channel WON by considering only propagation delay. When the shared-channel WON is operated as a MAN, however, we must take queueing delay into account unless the WON is very lightly loaded. Multiple-access protocols (e.g., token passing) typically have access delays comparable to the end-to-end propagation delay on the channel, so that queueing effects will certainly form a nonnegligible component of the overall packet delay. Even in schemes—like Aloha [Abr70]—with nearly instantaneous access to the channel, there is a high probability of packet collision even at moderate traffic loading, and such collisions make a significant contribution to delay since this normally entails retransmission of the collided packets. The mere presence of multiple stations on a shared channel also suggests that the channel might be more heavily loaded because more traffic is being placed on the channel. For these reasons we will generally consider queueing and propagation delays to be prime components of the mean packet delay in the shared-channel WON.

2.3 Virtual-Topology Design

The WON has both a physical and a virtual topology. The physical topology, which is the network topology in the conventional sense, is the geographical layout of the WON's stations and the optical fiber connecting them. We have already mentioned different physical topologies for the WON, including the bus, tree, star, and ring. The virtual topology, on the other hand, is not constrained by the physical topology and is represented by the interconnection graph of the WON. Whereas the physical topology of the WON is determined when the network is laid out, the virtual topology is uncommitted until the stations' receivers and transmitters are properly tuned to produce a given interconnection graph. Thus, by the proper tuning of transmitters and receivers, one can assign any virtual topology independent of the physical topology of the WON. This "protean" characteristic of the WON gives the network designer a great deal of flexibility in designing the WON and is one of its most attractive features. The network designer can choose the virtual topology of the WON to optimize a given attribute of the network.

In [Ban88] we identified and formalized three major problems in the large-scale design of the WON: the Physical-Topology Design Problem, the Virtual-Topology Design Problem, and the Flow Assignment Problem. Of these three problems, we will now focus exclusively on the Virtual-Topology Design Problem.

Given a physical topology, we would like to identify a virtual topology that is, by some measure, the best one possible. If we accept mean packet delay or throughput as the principal performance metrics for the WON, then we must select that virtual topology that minimizes one, or possibly both, of these metrics. As we have seen, the overwhelming component of mean packet delay for the typical dedicated-channel WON comes from the propagation delay that a packet experiences on its route from source to destination. We will therefore make the assumption that by minimizing mean propagation delay, and ignoring mean packet queueing delay, we can effectively minimize the overall mean packet delay. This assumption is certainly true in the lightly loaded WON since queueing will be negligible. In the heavily loaded WON we may have to take steps to reduce queueing, not only because of queueing's impact on delay, but also because we wish to reduce the probability that a packet gets dropped should it arrive to a packet switch whose buffers are full.

We can deal with the case of the heavily loaded WON by first solving the VTDP and then, once the virtual topology has been determined, solving the Flow Assignment Problem, which seeks to minimize mean packet delay (including both queueing and propagation delay components) by routing traffic flows.

The virtual topology of the WON is defined by tuning transmitters and receivers to specific channels so as to form the desired interconnection graph. Seeking to design the network with maximal performance, we would choose a virtual topology based upon parameters such as the traffic matrix and the network geography. Although the design of the virtual topology of the WON would be completed prior to the installation of the WON, it may be desirable to change the virtual topology of the WON at some point after its installation, especially when network conditions change. Given that the assumptions underlying the initial design of the virtual topology may have changed (e.g., the appearance of new stations, changes in the traffic matrix), the capability of retuning transmitters and receivers would permit us to define new virtual topologies during the WON's operational lifetime. The use of such wavelength-agile transmitters and/or receivers, though more costly than fixed-wavelength components, would provide the WON's operators with considerable flexibility in adapting to evolving conditions in the network. For instance, a network management system could collect and analyze long-term traffic statistics in the WON and take action to reconfigure its virtual topology when conditions warrant this. Such a network management system would require the facilities to determine which virtual topologies are optimal for the prevailing network conditions. It is this design optimization problem that we now turn to.

Assuming that mean packet delay is our principal performance metric, we can formulate the VTDP as a nonlinear zero-one mathematical program. The decision variables are comprised of the tuning matrix, (κ_{ijk}), an entry of which specifies whether station i transmits to station j over channel k, and the routing matrix $(\pi_{ij}^{(lm)})$, an entry of which specifies the probability that a packet will be routed from station i to station j, given that the source of the packet is station l and its destination is station m. Both sets of decision variables κ_{ijk} and $\pi_{ij}^{(lm)}$ assume discrete values from the set $\{0, 1\}$. By making the routing matrix take discrete values, we are stipulating that the network use fixed single-path routing, in which there is only one path from any source to its destination. The stipulation of single-path routing is not necessary—we could allow alternate routing and it would not affect the mean propagation delay because queueing effects are ignored. The basic problem is to arrange the network in a way so that the routing procedures can deliver packets in the least amount of time. The problem may thus be expressed as a minimization of the mean packet delay, subject to connectivity and flow-conservation constraints. Let us define $\gamma_i \triangleq \sum_{j=1}^{N} \gamma_{ij}$ as the rate at which network traffic is generated by station i and $\eta_i \triangleq \sum_{j=1}^{N} \gamma_{ji}$ as the rate at which network traffic is consumed by station i. For the sake of illustration we assume that the form of the service-rate function at each FCFS center is $\mu_k(n) = 1$ so that we can formulate the VTDP using equation (3) rather than the more cumbersome formula of equation (5). Formally, we propose to minimize the mean packet delay:

$$\min E[T] = \frac{1}{\gamma}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij}\delta_{ij}/\tilde{c} + \sum_{k=1}^{K}\frac{\lambda_k/\mu B}{1-\lambda_k/\mu B}\right\} \tag{6}$$

subject to

$$\eta_i + \sum_{j=1}^{N}\lambda_{ij} = \gamma_i + \sum_{l=1}^{N}\lambda_{li} \quad \forall i \tag{7}$$

$$\lambda_k = \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij}\kappa_{ijk} \quad \forall k \tag{8}$$

$$\lambda_{ij} = \sum_{l=1}^{N}\sum_{m=1}^{N}\pi_{ij}^{(lm)}\gamma_{lm} \quad \forall i\,\forall j \tag{9}$$

$$\sum_{j=1}^{N}\pi_{ij}^{(lm)} \leq 1 \quad \forall i\,\forall l\,\forall m \tag{10}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa_{ijk} = 1 \quad \forall k \tag{11}$$

$$\sum_{k=1}^{K}\max_{j=1}^{N}\kappa_{ijk} = p \quad \forall i \tag{12}$$

$$\sum_{k=1}^{K}\max_{i=1}^{N}\kappa_{ijk} = p \quad \forall j \tag{13}$$

$$\lambda_{ij} \geq 0 \quad \forall i\,\forall j \tag{14}$$

$$\lambda_k \geq 0 \quad \forall k \tag{15}$$

$$\kappa_{ijk} \in \{0,1\} \quad \forall i\,\forall j\,\forall k \tag{16}$$

$$\pi_{ij}^{(lm)} \in \{0,1\} \quad \forall i\,\forall j\,\forall l\,\forall m \tag{17}$$

In subsection 2.2 we have noted that queueing delay in the dedicated-channel WON is negligible in comparison to propagation delay. By the time queueing delay approaches the level of propagation delay, there is a high likelihood that one or more stations of the network will have experienced serious problems with buffer overflow (and, consequently, packet loss). In the case of the shared-channel WON we have therefore taken the approach of first minimizing mean path length (which is directly proportional to propagation delay), leaving the task of finding routes with minimal congestion as an optimization to be performed later. Thus in solving the VTDP for the dedicated-channel WON we will ignore the second summation of equation (6) However, in the formulation of the VTDP for the shared-channel WON the objective function in equation (6) must be modified [by using equation (5)] to account for queueing delay. This is because propagation delay in the shared-channel WON is minimized by simply placing all stations on one common channel, which is clearly undesirable. Furthermore, the multiple-access scheme may be very sensitive to factors such as traffic loading or station population on the channel. We must therefore penalize against this situation by using queueing delay to drive up the value of the objective function whenever a channel is unfavorably loaded.

There is a large collection of constraints in the VTDP, and we now explain the significance of each constraint in detail. First note that, in addition to the primary decision variables κ_{ijk} and $\pi_{ij}^{(lm)}$, we have introduced secondary decision variables λ_{ij} and λ_k which represent the traffic flows through stations and channels. These traffic flows are related to the routing variables in terms of the system of linear equations shown in equation (9). Equations (7) and (8) are conservation-of-flow conditions which specify that the traffic flowing into a station or a channel balances with the traffic flowing out of the station or channel. Equation (9) states that a routing chain contributes flow only to those hops over which the chain passes, and (10) guarantees that the routing probabilities are valid probabilities. Equations (12) and (13) say that every station is assigned to precisely p receive and transmit channels, while equation (11), which applies only to the dedicated-channel WON, says that each channel has exactly one transmitting and one receiving station. Finally, equations (14)–(17) specify the possible values over which the primary and secondary decision variables may range.

In the next section we will address the issue of how to solve the VTDP defined by equations (6)–(17).

3 Solving the Virtual-Topology Design Problem

The large-scale design of the WON addresses a number of problem areas: the selection of a topology, the selection of an interconnection graph, and the assignment of traffic flows to channels. These three problems were formally proposed in [Ban88], but we will now focus exclusively on the Virtual-Topology Design Problem, which is the problem of determining the interconnection graph that minimizes mean packet delay, given the physical network topology and the traffic pattern between all pairs of stations.

The VTDP, as defined by equations (6)–(17), is a combinatorial optimization problem for which there are no known efficient solution procedures. Although problems related to the VTDP have been investigated, such as the task of finding the minimum-diameter regular graph of a given degree and size [TS79], the approaches used in these problems have been based upon heuristics or *ad hoc* techniques. For example, the approach used in [TS79], which was based on local search, is not guaranteed to yield a global minimum. Moreover, these techniques often require extensive computation, which underscores the need to balance the quality of the solution against the amount of computation used to obtain this solution.

We have chosen to solve the VTDP by means of generic optimization techniques that asymptotically find the global optimum. For this purpose we have adapted simulated annealing [KGV83] and the genetic algorithm [Hol75, Gol89]. These algorithms, which can be used in the unconstrained optimization of an arbitrary function (including those with multiple local optima), are shown in figures 4 and 5.

The solution of the VTDP is influenced by the parameters of the problem, specifically the traffic and distance matrices (γ_{ij}) and (δ_{ij}). In what follows we treat these

1. *Initialization.* Select an initial temperature T and an initial state S.

2. *Epoch Initiation.* Start a new "epoch".

3. *Perturbation.* Randomly choose a neighbor state S' of S. $\Delta \leftarrow cost(S') - cost(S)$.

4. *Acceptance/Rejection.* $S \leftarrow S'$ with probability $\min(e^{-\Delta/T}, 1)$.

5. *Temperature Reduction.* If the "epoch" has expired then $T \leftarrow rT$, else go to 3.

6. *Convergence Test.* If the state is "frozen" then halt, else go to 2.

Figure 4: The Simulated Annealing Algorithm.

parameters as random variables that can change from network to network. For any set of experiments we choose different values for the entries of a matrix by selecting the values from a specified distribution. To illustrate, when we choose the interstation distances δ_{ij} from a deterministic distribution with mean 100 kilometers, we are using a model in which all stations are equidistant from the headend. Likewise, we can independently choose each value of δ_{ij} from the uniform distribution with mean 100 kilometers, which results in stations that are uniformly scattered over a disc of radius 200 kilometers, and the typical station is located 50 kilometers from the headend. We assume that stations are scattered over the plane and that the length of optical fiber between station i and station j is given by δ_{ij}, as shown in figure 6. In the following experiments the values of δ_{ij} are chosen by first scattering the stations of the WON over the plane so that their mean distance from the headend is 50 kilometers; thus the mean distance between any pair of stations is 100 kilometers, since a lightwave signal must travel from the first station to the headend and from there to the second station. We call the amount of variability in the distribution of interstation distance

1. *Initialization.* Randomly select a population of M graphs and evaluate $cost(G_1), \ldots, cost(G_M)$.

2. *Selection.* Randomly select a subpopulation of K low-cost graphs.

3. *Crossover/Mutation.* Randomly "mate" pairs of the population to produce $K/2$ new offspring graphs; allow mutation. Evaluate the cost of each offspring.

4. *Ranking.* Include the offspring graphs into the population and "kill off" $K/2$ highest-cost graphs.

5. *Convergence Test.* If the stopping condition is satisfied then halt, else go to 2.

Figure 5: The Genetic Algorithm.

scatter; a scatter of 0 corresponds to a WON with equidistant stations, and higher scatter values imply that the stations are scattered more randomly over the plane. We treat the traffic matrix similarly: we fix the overall mean amount of traffic exchanged between pairs of stations, but vary the amount exchanged between specific pairs of stations according to a given distribution. The variability of traffic is referred to as *skew*, and a skew value of 0 corresponds to a uniform traffic matrix in which all stations exchange the same amount of traffic.

We performed the experiments to be described in the sequel assuming a mean packet length of 1000 bits and 1-Gbps channel speeds.

The VTDP falls into two different subproblems, depending on whether the WON has dedicated channels or shared channels. Each problem is discussed below.

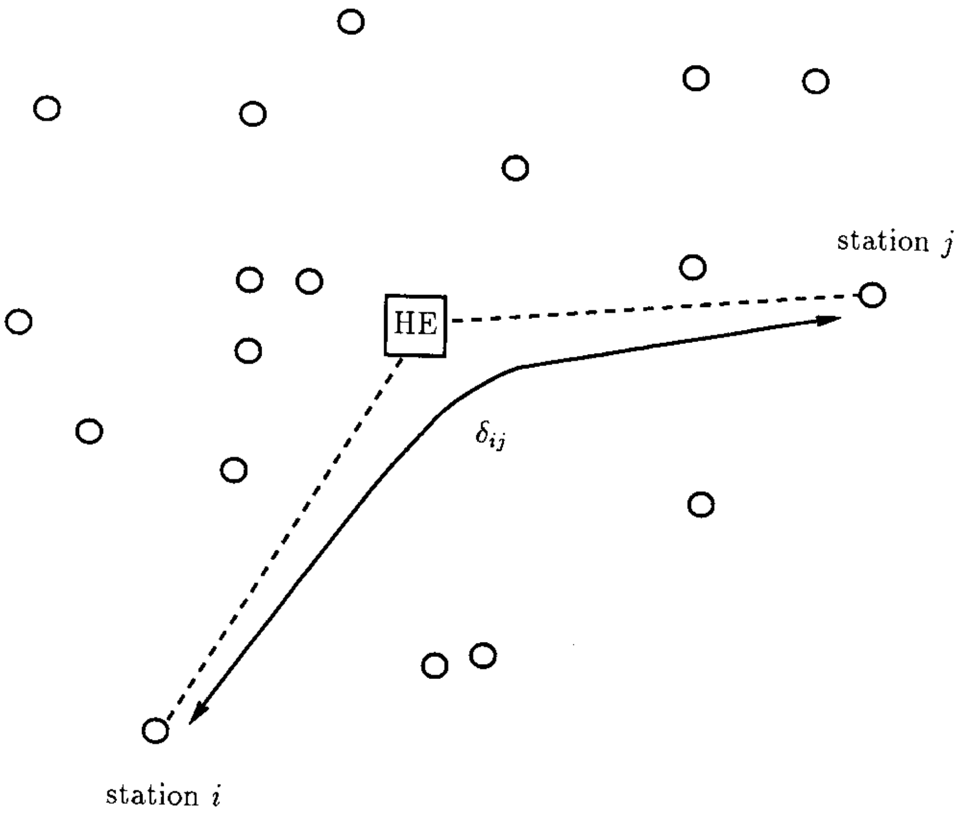

Figure 6: Station Geography and Interstation Distance.

3.1 The Dedicated-Channel Virtual-Topology Design Problem

We address the problem of finding the interconnection graph for a dedicated-channel WON that provides a randomly chosen packet with the shortest path (in terms of time or distance) from its source to its destination. This problem was addressed by means of the simulated annealing algorithm shown in figure 4.

In the simulated annealing algorithm for the VTDP a state corresponds an interconnection graph, so that finding a "frozen" (or minimal energy) state corresponds to finding a least-cost interconnection graph. The initial state (interconnection graph) is one of the regularly structured interconnection graphs, such as the binary Shuf-

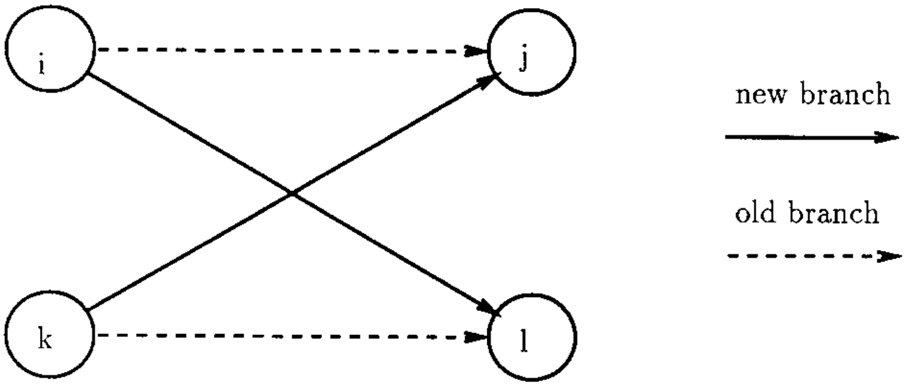

Figure 7: The Branch-Exchange Operation.

fleNet or the MSN, and we start with a temperature of 100 "degrees", reducing it by 5 percent at the end of each epoch. A single epoch consists of $8N$ distinct state perturbations, which appears to be sufficient to reach steady state. Given that a state corresponds to an interconnection graph, we choose to perturb it by means of the branch-exchange operation in which a new graph is produced by swapping the targets of two arcs as shown in figure 7. The branch-exchange operation is attractive because a branch exchange applied to a regular graph produces another regular graph. The stopping criterion requires that the optimization not continue for more than five epochs without an improvement in cost; when this criterion is satisfied it is deemed that further improvement is unlikely, and hence the state is "frozen".

The calculation of the cost function, since it involves finding all shortest paths in the graph being evaluated, is computationally intensive. Given that simulated annealing attempts to find a global minimum by traversing promising regions of the entire search space, we decided to try to speed up the optimization as much as possible. The simulated annealing algorithm was therefore implemented on a Sequent Symmetry parallel processor. Instead of choosing to parallelize the basic annealing loop, as in [CRSV86, DKN87], we parallelized the computation of all shortest paths in the cost function by assigning to different processors the computation of all shortest paths from different source nodes. Traffic between a particular source-destination pair is routed on a single shortest path without regard for balancing traffic flow along channels. This is known to be nonoptimal, but the need for fast packet switching and routing in the WON suggests that a simple, fixed single-path routing procedure (possibly source routing) be used.

We applied the simulated annealing algorithm to a number of instances of the VTDP. We applied the algorithm to WONs of size 8, 24, 64, and 160, and—with the exception of the 160-station WON—varied both the skew and scatter for each of the sizes.

The first set of experiments is meant to compare how a regularly structured interconnection graph, such as ShuffleNet, performs against an interconnection graph explicitly optimized for the given traffic matrix and station geography. For each value of N we allow five values for the skew ranging from 0 (uniform traffic) to 10 (highly asymmetric traffic). We also allow four values for the scatter parameter ranging from 0 (all stations equidistant from the headend) to 3 (two-thirds of the stations clustered near the headend). Throughout the set of experiments the traffic and distance matrices are randomly chosen in such a way as to keep the mean headend-to-station radius and the offered load fixed. The results are shown for WONs of 8, 24, and 64 stations in tables 1, 2, and 3. This data shows that by selecting a new virtual topology for the WON we are able to improve its performance in all cases tested. The improvement over the ShuffleNet interconnection graph ranges from as little as 5 percent to as much as 77 percent, depending on the specific problem parameters, and on average there is a 27-percent improvement in propagation delay. We can observe that the simulated annealing algorithm is not very effective in optimizing performance when skew and scatter are low. The small 5-percent improvement over ShuffleNet in the case of the WON with uniform traffic (skew = 0) and stations that are equidistant from the headend (scatter = 0) suggests that ShuffleNet performs quite well in such

			Propagation Delay (ms)		
N	skew	scatter	intial	best	gain
8	0	0	0.996	0.943	5 %
8	0	1	0.910	0.834	8 %
8	0	2	0.758	0.687	9 %
8	0	3	0.858	0.596	31 %
8	1	0	1.007	0.863	14 %
8	1	1	0.560	0.475	15 %
8	1	2	1.224	1.035	15 %
8	1	3	1.328	0.816	39 %
8	2	0	0.974	0.778	20 %
8	2	1	0.587	0.508	14 %
8	2	2	1.015	0.856	16 %
8	2	3	1.004	0.657	35 %
8	3	0	0.840	0.685	19 %
8	3	1	0.583	0.488	16 %
8	3	2	0.862	0.647	25 %
8	3	3	0.771	0.490	36 %
8	10	0	1.159	0.497	57 %
8	10	1	0.743	0.526	29 %
8	10	2	0.886	0.368	59 %
8	10	3	0.781	0.265	66 %

Table 1: Comparison of Delays in the 8-Station Dedicated-Channel WON.

a case—this is, in fact, the case for which Shufflenet was explicitly designed. As we increase the skew and scatter, however, ShuffleNet performance becomes less attractive, and the effectiveness of optimization, as evidenced in the dramatic improvement afforded by the simulated annealing algorithm, becomes significant. The explanation for the dramatic improvement in propagation delay when the scatter is increased lies in the fact that a greater proportion of the WON's stations are situated close to the headend. The virtual topology chosen by the simulated annealing algorithm provides shortest paths that are not necessarily minimum-hop paths: a packet traveling from its source to its destination will take shortcuts created by the presence of intermediate

N	skew	scatter	Propagation Delay (ms)		
			intial	best	gain
24	0	0	1.630	1.552	5 %
24	0	1	1.471	1.235	16 %
24	0	2	1.344	1.013	25 %
24	0	3	1.179	0.571	52 %
24	1	0	1.614	1.484	8 %
24	1	1	1.532	1.318	14 %
24	1	2	1.054	0.807	23 %
24	1	3	0.558	0.436	22 %
24	2	0	1.665	1.396	16 %
24	2	1	1.576	1.223	22 %
24	2	2	1.067	0.778	27 %
24	2	3	0.576	0.443	23 %
24	3	0	1.718	1.279	26 %
24	3	1	1.625	1.126	31 %
24	3	2	1.086	0.745	31 %
24	3	3	0.589	0.446	24 %
24	10	0	1.741	0.845	51 %
24	10	1	1.602	0.837	48 %
24	10	2	1.112	0.404	47 %
24	10	3	0.584	0.407	30 %

Table 2: Comparison of Delays in the 24-Station Dedicated-Channel WON.

stations near the headend. The packet may use a number of the centrally located stations, but the overall distance (and therefore time) covered will be small. Thus the WON essentially takes advantage of shortcuts created by stations near the headend, instead of using a smaller number of intermediate hops that could take the packet over a longer distance. This creates a kind of "distributed switching center" that consists of centrally located stations and their abbreviated hop distances.

The next set of experiments examines the effects of optimization on both propagation delay and throughput. Recalling that the ratio of carried load to offered load is equal to the mean number of hops in a network [Kle76], we hypothesize that by

N	skew	scatter	Propagation Delay (ms)		
			intial	best	gain
64	0	0	2.317	2.191	5 %
64	0	1	2.087	1.671	20 %
64	0	2	1.551	1.075	31 %
64	0	3	0.869	0.550	37 %
64	1	0	2.318	2.169	6 %
64	1	1	2.214	1.579	29 %
64	1	2	1.456	0.912	37 %
64	1	3	0.814	0.531	35 %
64	2	0	2.324	2.083	10 %
64	2	1	2.228	1.563	30 %
64	2	2	1.435	0.882	39 %
64	2	3	0.770	0.501	35 %
64	3	0	2.318	2.031	12 %
64	3	1	2.216	1.553	30 %
64	3	2	1.439	0.862	40 %
64	3	3	0.771	0.499	35 %
64	10	0	2.364	1.701	28 %
64	10	1	2.234	1.327	41 %
64	10	2	1.465	0.762	48 %
64	10	3	0.773	0.485	37 %

Table 3: Comparison of Delays in the 64-Station Dedicated-Channel WON.

minimizing the weighted mean number of hops in the WON we will also tend to increase the maximum throughput achievable by the WON. As in the previous set of experiments we use the simulated annealing algorithm to minimize the number of hops traveled by a typical packet, without regard for the distance covered in a single hop. Such a procedure will result in decreased propagation delay, but the improvement is not as impressive as in the previous set of experiments, where minimizing the actual time delay was the prime objective. Thus we do not vary the scatter parameter, since interstation distance would not affect the mean number of hops in the WON. The experiments are performed on WONs with 8, 24, 64, and 160 stations, and we examine five skew values ranging from 0 to 10, as in the previous experiments. To determine the maximum throughput that a particular network would support we "pump up" the initial traffic load by a scaling factor and observe when the network reaches saturation, which is defined to be when the average buffer occupancy in some station exceeds a specified threshold—at this point the WON would be dropping too many packets. From each optimization run we are able to collect the following four items of information:

1. mean propagation delay for the initial ShuffleNet interconnection graph
2. mean propagation delay for the resulting optimized interconnection graph
3. maximum throughput for the initial ShuffleNet interconnection graph
4. maximum throughput for the resulting optimized interconnection graph

The results of the experiment are shown in table 4. Besides the improvement in propagation delay that we saw in the previous set of experiments, we note that we can achieve an average improvement in throughput of 99 percent. We can also observe an interesting phenomenon in the data of table 4: a small improvement in propagation delay yields a large improvement in the maximum sustainable throughput of the WON. This observation reinforces the need for optimization since it not only gives better delay performance but makes for a WON that will carry heavier traffic loads.

If we study tables 1–4 we observe that, for a fixed scatter, increasing the skew does not significantly degrade performance in ShuffleNet. This finding corroborates the analysis of [EM88] which, using several mathematical models of traffic skew, concluded that the throughput of ShuffleNet will not degrade by more than about 50 percent when the assumption of uniform traffic is relaxed.

		Propagation Delay (ms)			Throughput (Gbps)		
N	skew	intial	best	gain	intial	best	gain
8	0	0.996	0.943	5 %	5.10	7.00	37 %
8	1	1.007	0.863	14 %	3.81	6.44	69 %
8	2	0.974	0.778	20 %	3.53	6.38	81 %
8	3	0.840	0.685	19 %	4.65	6.22	34 %
8	10	1.159	0.497	57 %	1.90	5.60	195 %
24	0	1.630	1.552	5 %	8.24	12.55	52 %
24	1	1.614	1.484	8 %	7.42	12.81	73 %
24	2	1.665	1.396	16 %	7.09	12.32	74 %
24	3	1.718	1.279	26 %	6.13	10.82	77 %
24	10	1.741	0.845	51 %	6.13	11.04	80 %
64	0	2.317	2.191	5 %	12.10	21.37	77 %
64	1	2.318	2.169	6 %	10.89	22.18	104 %
64	2	2.324	2.083	10 %	11.69	23.79	104 %
64	3	2.318	2.031	12 %	11.69	21.37	83 %
64	10	2.364	1.701	28 %	10.08	16.93	68 %
160	0	3.035	2.852	6 %	17.81	45.79	157 %
160	1	3.037	2.832	7 %	17.81	43.25	143 %
160	2	3.032	2.803	7 %	17.81	43.25	143 %
160	3	3.031	2.757	9 %	17.81	45.79	157 %
160	10	3.022	2.553	15 %	15.26	40.70	166 %

Table 4: Joint Improvement of Delay and Throughput in the Dedicated-Channel WON.

3.2 The Shared-Channel Virtual-Topology Design Problem

In section 2.1 we alluded to the use of channel sharing for improving the performance of the WON. On a shared channel a given station can transmit to more stations that it could have on a dedicated channel, and this increased fanout opens up the possibility of defining denser interconnection graphs. As the extreme case we point out that assigning *all* stations to *one* channel would allow any packet to get to its destination in one hop. These considerations motivate us to study how much improvement can be expected when channel sharing is used in the WON.

The shared-channel VTDP, which seeks to find the virtual topology that permits minimal packet delay in the shared-channel WON, differs from the dedicated-channel VTDP in both its formulation and solution. As discussed in subsection 2.2, we can not ignore the queueing component of packet delay in the shared-channel WON unless the WON is very lightly loaded. Indeed, ignoring the queueing component would imply that the solution of the VTDP in the shared-channel WON is best accomplished by collecting all stations on one channel, a solution that is in general unacceptable because of its tendency to saturate the single channel. We must therefore account for queueing delay by means of penalty functions or constraints in the formulation of the VTDP. For example, we could introduce the constraint that no channel should host more than a fixed, small number of stations. We have chosen to use the penalty-function approach: we explicitly use channel-access delay as our penalty function.

The objective function we use is the formula for mean packet delay given in equation (5). For this study we assume that shared channels operate as pure Aloha channels [Abr70], which is not a very efficient access scheme but is fairly simple to implement. The principal drawback of Aloha is that work is wasted whenever packet collisions occur, and these can happen frequently if the channel is even moderately loaded. We would therefore prefer to operate the Aloha channel at a traffic loading that results in a fairly low rate of collisions. A rate of packet loss of more than, say, 3 percent could adversely affect packet delay by increasing the rate of retransmission. Using the classical formula $S = Ge^{-2G}$ [Abr70, Kle76], which relates channel throughput S to offered traffic load G in pure Aloha, we find that to keep retransmissions below 3 percent we must have $S > 0.97G$. Thus we must have $Ge^{-2G} > 0.97G$, or $G < -(\ln 0.97)/2 \approx 0.02$. For our shared-channel model we therefore use a service-rate function that drops to 0 once a channel shared by two or more stations is more than 2 percent utilized; if the channel is less than 2 percent utilized, then the service rate is 1. Of course, when only one station is assigned to the channel, we use the simpler fixed-rate FCFS model, which, in principle, allows the channel to be perfectly utilized without any packet loss due to collisions.

To solve the VTDP in the shared-channel WON we employed the genetic algorithm shown in figure 5. The reason for using the genetic algorithm, rather than the simulated annealing algorithm used in the dedicated-channel VTDP, was that there appeared to be no satisfactory method to generate neighbor graphs—the genetic algorithm, on the other hand, provides a number of natural ways to derive new graphs by combining old ones. Furthermore, using a new algorithm allows us the opportunity to compare the solution quality and execution time of different algorithms, which is especially important since the simulated annealing algorithm requires long running times on problems of large size.

In applying the genetic algorithm to the shared-channel VTDP we have used two different crossover and mutation mechanisms. We began by using a "graph-splicing" mechanism that "mates" two graphs by taking a set of nodes from one parent graph and the complementary set from the other parent graph and placing both sets of nodes and their arcs together to construct a new "pseudograph". The "pseudograph", which could contain dangling arcs, is then transformed into an offspring graph by a heuristic that connects dangling arcs to nodes without incoming arcs; random mutation consisting of the merging or splitting of channels could also be applied to the offspring graph. This "graph-splicing" crossover and mutation mechanism, while performing well on the lightly loaded shared-channel WON, is ineffective when higher traffic loads are encountered. We therefore applied a second "graph-overlaying" crossover and mutation mechanism that produces offspring graphs by taking the first parent graph and adding to it all arcs of the second parent graph. Since this overlaying produces a graph in which nodes have twice as many "transmitters" as possible, half of the outgoing arcs at each node must be pruned. The pruning is based on a breadthfirst search of the graph which builds a tree of all shortest paths from a randomly chosen root node to every other node in the graph. Arcs are chosen to produce the "bushiest" tree possible so that mean path length would be kept short. We allow random mutation in a manner identical to the "graph splicing" case. The results for higher loads gotten by the "graph-overlaying" mechanism are substantially better than those gotten by the "graph-splicing" mechanism.

To start the genetic algorithm we require an initial population of graphs with a fair amount of "genetic diversity". These graphs are mated and pass their characteristics on to their offspring so that characteristics with high survival value proliferate throughout the population. The initial population is constructed by taking a small set of regularly structured graphs, such as ShuffleNet, the MSN, the de Bruijn graph, the Moore graph, etc., and applying random mutation to them. This produces graphs with a good degree of variation and a fairly high overall survival value (survival value equates to low cost).

The criterion for stopping the genetic algorithm is that further search would be unlikely to yield a better solution than has already been found. If the algorithm can not find an improvement for 50 consecutive generations, then the algorithm is stopped.

The genetic algorithm was parallelized in a very natural way by allowing different processors to independently manage the crossover and mutation of pairs of graphs. The runing times of the genetic algorithm are considerably shorter than those of the simulated annealing algorithm. The genetic algorithm also seems to find low-cost graphs easily and consistently, though the quality of the solutions is typically not as good as those found by simulated annealing.

We show the data from our experiments in tables 5–7. The tables, given for WONs of size 24, 64, and 160, show the mean packet delays for a range of skew and scatter parameters. In each table we show mean packet delays for the lightly loaded

			Mean Packet Delay (ms)					
			Light Load			Moderate Load		
N	skew	scatter	initial	best	gain	initial	best	gain
24	0	0	1.630	0.501	69 %	1.618	1.576	3 %
24	0	1	1.471	0.538	63 %	1.535	1.456	5 %
24	0	2	1.344	0.452	66 %	1.019	1.002	2 %
24	0	3	1.179	0.453	62 %	0.632	0.514	19 %
24	1	0	1.614	0.522	68 %	1.611	1.559	3 %
24	1	1	1.532	0.554	64 %	1.549	1.429	8 %
24	1	2	1.054	0.516	51 %	1.143	0.900	21 %
24	1	3	0.558	0.500	11 %	0.979	0.567	42 %
24	2	0	1.665	0.501	68 %	1.618	1.510	7 %
24	2	1	1.576	0.543	64 %	1.537	1.352	12 %
24	2	2	1.067	0.539	50 %	1.209	0.895	26 %
24	2	3	0.576	0.528	9 %	1.016	0.652	36 %
24	3	0	1.718	0.515	70 %	1.640	1.452	11 %
24	3	1	1.625	0.532	67 %	1.534	1.170	24 %
24	3	2	1.086	0.561	48 %	1.239	0.839	32 %
24	3	3	0.589	0.556	6 %	1.025	0.613	40 %
24	10	0	1.741	0.521	69 %	1.504	1.169	22 %
24	10	1	1.602	0.583	64 %	1.601	1.157	28 %
24	10	2	1.112	0.432	61 %	1.083	0.675	38 %
24	10	3	0.584	0.427	27 %	0.878	0.474	46 %

Table 5: Mean Packet Delays in the Lightly and Moderately Loaded 24-Station Shared-Channel WON.

			Mean Packet Delay (ms)					
			Light Load			Moderate Load		
N	skew	scatter	initial	best	gain	initial	best	gain
64	0	0	2.317	0.602	74 %	2.388	2.236	6 %
64	0	1	2.087	0.527	75 %	3.380	3.014	11 %
64	0	2	1.551	0.486	69 %	3.267	2.922	11 %
64	0	3	0.869	0.490	44 %	2.680	2.330	13 %
64	1	0	2.318	0.575	75 %	2.453	2.449	0 %
64	1	1	2.214	0.542	76 %	3.481	3.181	7 %
64	1	2	1.456	0.494	66 %	3.624	3.173	12 %
64	1	3	0.814	0.525	36 %	3.212	2.263	30 %
64	2	0	2.324	0.650	72 %	2.470	2.224	10 %
64	2	1	2.228	0.534	70 %	3.283	3.053	7 %
64	2	2	1.435	0.572	60 %	3.533	3.097	12 %
64	2	3	0.770	0.557	28 %	3.280	2.486	24 %
64	3	0	2.318	0.585	75 %	2.353	2.265	4 %
64	3	1	2.216	0.528	76 %	3.225	2.854	12 %
64	3	2	1.439	0.516	64 %	3.598	2.769	23 %
64	3	3	0.771	0.559	28 %	3.302	2.566	22 %
64	10	0	2.364	0.578	76 %	2.913	2.655	9 %
64	10	1	2.234	0.519	77 %	3.460	3.013	13 %
64	10	2	1.465	0.531	64 %	3.577	3.084	14 %
64	10	3	0.773	0.588	24 %	3.425	2.185	36 %

Table 6: Mean Packet Delays in the Lightly and Moderately Loaded 64-Station Shared-Channel WON.

			Mean Packet Delay (ms)					
			Light Load			Moderate Load		
N	skew	scatter	initial	best	gain	initial	best	gain
160	0	0	2.904	1.533	47 %	2.905	2.873	1 %
160	0	1	2.628	0.745	72 %	2.630	2.261	14 %
160	0	2	2.065	0.884	57 %	2.112	1.999	5 %
160	0	3	1.001	0.510	49 %	41.254	0.676	98 %
160	1	0	2.906	1.047	64 %	2.907	2.904	0 %
160	1	1	2.330	0.683	71 %	2.331	2.188	6 %
160	1	2	2.640	0.673	75 %	2.686	2.535	6 %
160	1	3	1.612	0.742	54 %	2.666	1.810	32 %
160	2	0	2.900	1.717	41 %	2.901	2.865	1 %
160	2	1	2.328	1.955	16 %	2.329	2.288	2 %
160	2	2	2.633	0.659	75 %	2.683	2.631	2 %
160	2	3	1.616	0.612	62 %	15.643	1.923	88 %
160	3	0	2.985	1.466	51 %	2.896	2.853	1 %
160	3	1	2.325	0.806	65 %	2.327	2.308	1 %
160	3	2	2.640	1.477	44 %	2.687	2.510	7 %
160	3	3	1.618	0.617	62 %	15.582	1.777	89 %
160	10	0	2.888	1.909	34 %	2.889	2.877	0 %
160	10	1	2.333	0.860	63 %	2.335	2.212	5 %
160	10	2	2.631	0.957	64 %	2.684	2.222	17 %
160	10	3	1.617	0.627	61 %	28.194	2.404	91 %

Table 7: Mean Packet Delays in the Lightly and Moderately Loaded 160-Station Shared-Channel WON.

and the moderately loaded shared-channel WON. In the light-load scenario there is almost no traffic offered to the WON, and in the moderate-load scenario there is just enough traffic to produce queueing delays, but not so much that congestion is prevalent. Given each network size, we offer a fixed amount of traffic to the WON, regardless of the skew or scatter parameter; for example, we perform all the experiments with the moderately loaded 160-station WON assuming that approximately 7.6 million 1000-bit packets are offered to the network every second. This means that sometimes a channel of the WON will be saturated, depending on the parameters of the experiment; for example, four of the entries in table 7 show initial networks with comparatively high mean packet delays, which essentially means that the network is congested in those configurations. It can also be seen that we are able in all cases to eliminate the congestion by optimizing the virtual topology of the WON. We notice also that in all cases the congestion occurs when the scatter is at its peak value (viz. scatter = 3), which we explain by hypothesizing that the routing procedure is using minimum-distance shortcuts provided by stations located near the headend, and this overutilizes a few specific channels in the WON.

We see in the data of tables 5–7 the familiar drop-off in delay as scatter is increased. As in the dedicated-channel WON this is because the larger proportion of stations located near the headend can be used to provide shortcuts for traffic. The genetic algorithm is able to improve upon the initial interconnection graphs—which either are or are directly derived from regularly structured graphs such as ShuffleNet—but the results are different depending on the traffic load scenario. The average improvement in the lightly loaded WON is about 61 percent, compared with a 20-percent average improvement in the the moderately loaded WON. It would appear easier to find the optimal virtual topology in the lightly loaded shared-channel WON, perhaps because the goal of conglomerating as many communicating stations as possible on a single channel without overloading that channel is comparatively simple. It is interesting to note that in the moderately loaded shared-channel WON channel sharing is essentially not used: as traffic load increases it becomes infeasible to share pure Aloha channels because of their sensitivity to load, and thus the optimization favors designs with a single transmitter per channel. These results seem to suggest that even after designing a satisfactory virtual topology for the WON it is still necessary to attempt to perform optimal routing in the WON.

4 Discussion

In this paper we have proposed the Wavelength-Division Optical Network, a generalization of the ShuffleNet concept, and argued that the network designer will want to optimize system performance by judiciously choosing the best virtual topology for the given network traffic, geography, and physical topology. After developing a basic but versatile analytical model of WON performance, we presented a mathematical formulation of the Virtual-Topology Design Problem. Because of the potential for high delays, the solution of the VTDP plays a prominent role in the design of the high-performance WON.

We have studied several versions of the VTDP, including variants with both dedicated and shared channels. We have shown the usefulness of both the simulated annealing and genetic algorithms, performing over 200 individual experiments to demonstrate the degree of improvement that we can expect.

A list summarizing our major findings is given below:

1. Using optimization we can improve the performance in all of the regularly structured interconnection graphs that we studied, e.g. ShuffleNet, the Manhattan Street Network, de Bruijn graphs.

2. We can significantly improve performance when the geography of the network

places a high proportion of stations close to the headend.

3. A small improvement in delay may be accompanied by a large improvement in throughput.

4. We can significantly improve performance by sharing channels in the case of light traffic loading.

5. When the traffic loading is increased to even moderate levels in the shared-channel WON (using pure Aloha channels) we can still obtain reasonable performance improvement, but little of the gain can be attributed to the use of channel sharing.

The VTDP is only one part of the design process and further research needs to be conducted into other aspects of WON design. Other problems such as the Physical-Topology Design Problem and the Flow Assignment Problem have been identified as important in WON design [Ban88] and will be investigated in further depth. In particular, it is important to understand the interplay among these design problems, especially from the perspective of how the solution to one problem affects the solution to another. Intuitively, the design of the physical topology precedes the design of the virtual topology which in turn precedes the determination of optimal routing; given this sequence of tasks—which will generally produce a suboptimal design—how does the resulting network compare to an optimally designed one? Furthermore, there are a number of important research areas in the design of protocols for the WON, including congestion control, broadcast services, and network management.

References

[Abr70] N. Abramson. The ALOHA system—another alternative for computer communications. In *Proceedings of the Fall Joint Computer Conference, AFIPS Conference 37*, pages 281–284, 1970.

[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.

[AKH87] Anthony S. Acampora, Mark J. Karol, and Michael G. Hluchyj. Terabit lightwave networks: The multihop approach. *AT&T Technical Journal*, 66(6):21–34, November/December 1987.

[Ban88] Joseph A. Bannister. The WOMAN (wavelength-division optical metropolitan area network): Architectures, topologies, and protocols. Unpublished Ph.D. Prospectus, UCLA Computer Science Department, Los Angeles, California, September 1988.

[CG87] Imrich Chlamtac and Aura Ganz. Toward alternative high speed networks: The SWIFT architecture. In *Proceedings of IEEE INFOCOM '87*, pages 1102–1108, San Francisco, California, March 1987.

[CGK88] I. Chlamtac, A. Ganz, and G. Karmi. Circuit switching in multi-hop lightwave networks. In *Proceedings of the ACM SIGCOMM '88 Symposium*, pages 188–199, Stanford, California, August 1988.

[CRSV86] Andrea Casotto, Fabio Romeo, and Alberto Sangiovanni-Vincentelli. A parallel simulated annealing algorithm for the placement of macro-cells. In *Proceedings of the 1986 IEEE International Conference on Computer-Aided Design*, pages 30–33, Santa Clara, California, November 1986.

[DKN87] F. Darema, S. Kirkpatrick, and V.A. Norton. Parallel techniques for chip placement by simulated annealing on shared memory systems. In *Pro-*

ceedings of the 1987 IEEE International Conference on Computer Design, pages 87–90, Rye Brook, New York, October 1987.

[EM88] Martin Eisenberg and Nader Mehravari. Performance of the multichannel multihop lightwave network under nonuniform traffic. *IEEE Journal on Selected Areas in Communications*, 6(7):1063–1078, August 1988.

[Ger88] Mario Gerla. Tree-Net, a multi-level fiber optics MAN. In *Proceedings of IEEE INFOCOM '88*, pages 4B.2.1–4B.2.10, New Orleans, Louisiana, March 1988.

[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.

[Gol89] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, Reading, Massachusetts, 1989.

[GPL83] Arthur Goldberg, Gerald Popek, and Steve Lavenberg. A validated distributed system performance model. In A. K. Agrawala and S. K. Tripathi, editors, *Proceedings of Performance '83*, pages 251–268, Amsterdam, The Netherlands, 1983. North-Holland.

[HK88] Michael G. Hluchyj and Mark J. Karol. ShuffleNet: An application of generalized perfect shuffles to multihop lightwave networks. In *Proceedings of IEEE INFOCOM '88*, pages 4B.4.1–4B.4.12, New Orleans, Louisiana, March 1988.

[Hol75] John H. Holland. *Adaptation in Natural and Artificial Systems.* The University of Michigan Press, Ann Arbor, Michigan, 1975.

[KGV83] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.

[Kle64] Leonard Kleinrock. *Communication Nets: Stochastic Message Flow and Delay.* McGraw-Hill, New York, New York, 1964.

[Kle76] Leonard Kleinrock. *Queueing Systems, Volume II: Computer Applications.* John Wiley and Sons, New York, New York, 1976.

[LS83] Steven S. Lavenberg and Charles H. Sauer. Analytical results for queueing models. In Steven S. Lavenberg, editor, *Computer Performance Modeling Handbook*, chapter 3. Academic Press, New York, New York, 1983.

[LZGS84] Edward D. Lazowska, John Zahorjan, G. Scott Graham, and Kenneth C. Sevcik. *Quantitative System Performance: Computer System Analysis Using Queueing Network Models.* Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[Max85] N. F. Maxemchuk. Regular mesh topologies in local and metropolitan area networks. *AT&T Technical Journal*, 64(7):1659–1685, September 1985.

[TS79] Sam Toueg and Kenneth Steiglitz. The design of small-diameter networks by local search. *IEEE Transactions on Computers*, C-28(7):537–542, July 1979.