Topological Design and Implementation of Optical Packet Switching Networks Using Kautz Graph

By

Mushi Jin

A thesis submitted to
the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Master of Applied Science

CCNR Lab
School of Information Technology and Engineering
University of Ottawa
September 2002

0-612-76595-4

Acknowledgments

I would like to express my sincere thanks to my supervisor, Professor Oliver W. W. Yang, for his continuous encouragement, professional competence and patience. I would also like to thank the members in the Computer Communication Network Research (CCNR) group in SITE of University of Ottawa, for their continuous cooperation, encouragement and meaningful discussion.

We would also like to acknowledge the financial support of an OGSST/OPTENIA scholarship, a Nortel research grant and NSERC funding.

To my parents.

Abstract

In optical communication systems, messages must be delivered fast and efficiently in order to achieve high quality of service and fully exploit the resources. Two elementary issues that affect this objective are the network topology and the need for fast switching node. The network topology affects the routing and the device cost, while switches affect the exchange of messages from one link to another. Today, these two areas have attracted substantial amount of research interest.

In this thesis, a multilayer optical network topology based on the Kautz graph is analyzed first. This multilayer Kautz graph has been proven very useful in the optical network in comparison with some known architectures. This topology has also been implemented with optical passive star couplers and optical switches. Three contention resolution schemes have been examined. Next this multilayer idea is used in the design of a multistage optical switch. This switch architecture has inherited some properties of the known topologies. With a low device cost, it can provide a remarkable advantage in optical packet switching. Finally, we have also discussed some critical issues in the design and implementation of the proposed architectures with available technologies and components.

Table of Contents

Figures

Tables

Table of Abbreviations and Acronyms

		Page of first appearance
ATM	Asynchronous Transfer Mode	16
DBG	De Bruijn Graph	9
DWDM	Dense Wavelength Division Multiplexing	3
FDL	Fiber Delay Line	13
FIFO	First In First Out	53
GMPLS	Generalized Multiprotocol Label Switching	5
IP	Internet Protocol	4
KEOPS	Keys to Optical Packet Switching	13
KG	Kautz graph	9
MEMS	Micro-Electro-Mechanical System	60
MKG	Multilayer Kautz Graph	19
MPLS	Multiprotocol Label Switching	4
MSN	Manhattan Street Network	9
OADM	Optical Add-Drop Multiplex	7
OAMP	Operation, Administration, Maintenance and Provisioning	3
OPS	Optical Passive Star	6
OXC	Optical Cross Connect	7
O/E/O	Optics-Electronics-Optics	9
POPS	Partitioned Optical Passive Star	6
QoS	Quality of Service	3
RWA	Routing and Wavelength Assignment	7
SCM	Sub-carrier Multiplexing	27
SDH	Synchronous Digital Hierarchy	2
SDM	Space Division Multiplexing	13
SE	Switching Element	63
SKG	Stack Kautz Graph	10
SLOB	Switch With Large Optical Buffers	14
SOA	Semiconductor Optical Amplifiers	50
SONET	Synchronous Optical Network	2
TDM	Time Division Multiplexing	13
TF	Tunable Filter	70
TWC	Tunable Wavelength Converter	38
WASPNET	Wavelength Switched Packet Network	13
WDM	Wavelength Division Multiplexing	3

Table of Symbols and Notations

		Page of first appearance
A	Set of edges	23
C_{ap}	Link capacity	31
$\overline{D}$	Average delay	32
$\overline{L}$	Average link load	32
L_i	Link i	31
L_{ij}	Link between node i and j	31
M	Number of links	31
N	Number of nodes	23
V	Set of nodes	23
d	Degree of a Kautz graph	23
e	Number of edges	23
$\overline{h}$	Average number of hops	31
k	Diameter of a Kautz graph	23
l	Layer	24
m	Number of layers	24
s	Node label	24
λ	Packet Arrival Rate	30
μ	Packet size	31

Chapter 1 Introduction

The evolution of the telecommunication network may have resulted in the most significant technological advances and generated the most meaningful social impacts in the development of science and technology. Over the past century, the networks that once provided only basic telephone service through a friendly local operator are now capable of transmitting the equivalent of thousands of encyclopedias per second. Today the rapid increase in Internet applications is driving unprecedented changes to the traditional communication networks. The telephone network is carrying more and more data traffic, which has altered the concept of traditional telephone network into data network, generating a new concept of their combination -- infocom.

Today's information era requires more and more flexible networks to satisfy various applications. Many new technologies have been deployed to build high-speed communication networks, among which the fiber-optic technology has played the most promising role because of its incomparable potential: huge bandwidth, low signal distortion, low power consumption, low material usage, small space requirement and low cost. Our challenge is to turn theses promises into reality in order to meet the networking demands of the future. The work in this thesis attempts to make a contribution to the solution of this problem.

1.1 Literature Review

Since the introduction of fiber optical cable in early 1980's, optical communication networking has undergone three major stages of its development [PaMa00], each of which has been driven by the invention and maturity of new technologies and techniques. The progress of optical technologies has enabled more complicated applications. Similarly, the increasing requirement of the applications has also accelerated the development of capacity, scalability and flexibility of optical communication systems.

1.1.1 The Development of Optical Communications

Optical fibers began to be used in the point-to-point optical system in early 1980's, which was normally an asynchronous network. In this early period each part of the system was designed and working independently by different vendors, without a standardized functional hierarchy. Due to the lack of a universal clock source, the variation between the internal clocks at different network elements would generate a certain amount of variation in a signal's arriving and transmitting timing, often resulting in transmission errors. Moreover, as no standards existed to mandate how network elements should format the optical signals with the increasing deployment of optical fibers, it was very difficult for network operators to interconnect equipment from different vendors. All these shortcomings limit the deployment and scalability of the first era of optical communication [Inte00].

The second generation is the Synchronous Optical Network (SONET)/Synchronous Digital Hierarchy (SDH) ring [ITU01], which is used to satisfy the need for optical standards in the expanding networks. SONET standardizes line bit-rates, coding schemes, traffic and control interface, and operation and maintenance functionality. SONET also defines the types of network elements required to implement specific functions, network architectures that vendors could implement, and the functionality each network element must perform. The operation, administration, maintenance and provisioning (OAMP) could be integrated into one centrally controlled system. Network providers could then interconnect the SONET equipment from different vendors without worrying about the compatibility. Thus, they could scale their networks along with the increasing application requirements. SONET rings are very successful and still play an important role in current backbone networks.

The third generation is the more flexible wavelength-division multiplexing (WDM) network [Ryan98], which could more efficiently exploit the capacity potential from the available waveband and fibers. Promise of scalability has enabled SONET to survive during a time of dramatic increase in network capacity needs [AlFa98]. Based on its open-ended growth plan for higher bit rates, no theoretical upper limit exists for SONET bit rates. However, as higher bit rates are used, physical limitations in the laser sources, amplifiers and optical fibers begin to restrict the endless increasing of the bit rate on the link level. Additionally, customers are demanding more services and options, and are generating more complex and different types of data traffic. To provide full end-to-end connectivity, a new paradigm is needed to satisfy the high capacity and varying QoS (Quality of Service) requirements. Consequently, the WDM/DWDM (Dense WDM) optical networks emerge to provide the required bandwidth and flexibility to enable end-to-end wavelength services.

Like SONET, WDM networks have defined network elements and architectures to provide the basis of the optical network interoperability. However, unlike SONET, rather than using a defined bit-rate and frame structure as its basic building block, the new optical network will be based on wavelengths, which implies that the components of the optical network will be defined according to how the wavelengths are transmitted, groomed or implemented in the network. There is no universal definition of the original format and property of the signal. The common opinion is that the Internet Protocol (IP) [RaPe00] will be the most viable carrier for various applications, and the Multiprotocol Label Switching (MPLS) [Li99] will be the appropriate technology for the exchange of optical signals. However, how the IP traffic is transmitted over the WDM/DWDM network is still under investigation, which generates several proposals for IP-over-WDM integration, such as IP-MPLS/SONET/WDM, IP-MPLS/ATM/SONET/WDM, or IP-MPLS/WDM hierarchy [BaDr01, GhDi00], and so on.

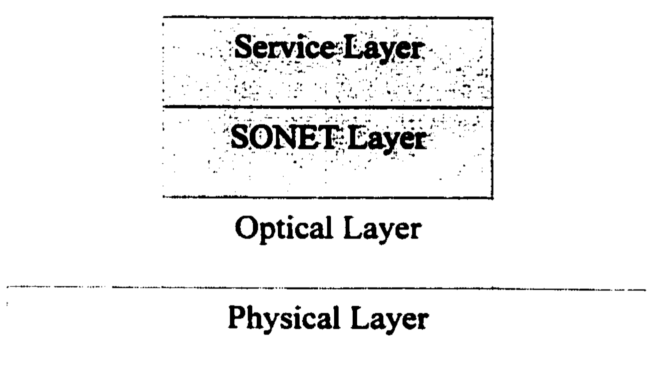


Figure 1.1 The Layered Architecture of Optical Networks

To help define current network functionality, optical networks are divided into several physical or virtual layers, as shown in Figure 1.1 [Inte01]. The first layer, the services layer, is where the services such as data traffic access the network. Although most services are IP-oriented, it is believed that the MPLS/GMPLS (Generalized MPLS) technology will be used instead of routing IP messages over the network directly. The next layer, SONET, provides restoration, performance monitoring and provisioning that is transparent to the first layer. With the emergence of the optical layer, it will eventually provide the same functionality as the SONET layer while operating entirely in the optical domain. Just as the SONET layer is transparent to the services layer, the optical layer will ideally be transparent to the SONET layer, providing restoration, performance monitoring and provisioning of individual wavelength instead of electrical SONET signals. Thus each part of the WDM signal could occupy a distinct wavelength with its own coding, modulating and signal format. The transparency is an important advantage, which is why the optical technology is becoming the first choice of modern communication networks.

1.1.2 The Optical Components

Dramatic advances in the optical component technology are accelerating the evolution of the optical networks. Remarkable progress has been achieved in the field of the optical switching components. The all-optical wavelength converters, tunable optical filters, multi-frequency lasers, amplifiers, receivers and transmitters [ElMo00a] are the many key components enabling the implementation of all-optical switching architectures. To provision a lightpath in the WDM network, a connection is established along the route using a common wavelength on all of the links along the route. To release this wavelength constraint, optical devices operating at a fixed wavelength are no longer adequate. By introducing wavelength converters, a node is able to take the data modulated in one input wavelength and transfer it to a different output wavelength.

Therefore, new components should be able to exercise selection, switching, and routing based on the wavelength [ElMo00b, StKl99].

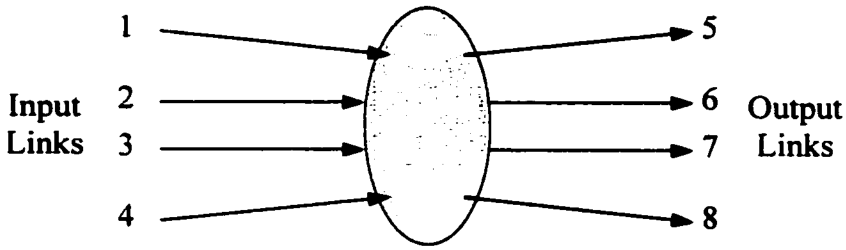


Figure 1.2 A 4x4 OPS Coupler

The optical passive star (OPS) couplers [GaMa98] are the most practical optical devices in the broadcast and select optical networks, which have the potential to be viable and cost-effective. An OPS coupler is capable of splitting the incoming signals to all output ports equally without signal amplification, which is often composed of an optical multiplexer followed by an optical fiber or a free optical space and a beam-splitter that divides the incoming light signal into several equal ports. Figure 1.2 describes a 4×4 OPS coupler. It is a passive device without the requirement of any external power supply. The properties of low-energy-loss and ease-of-use have made it to be the first choice for many applications [ChLe96, CoFe98]. In [ChLe96], a single hop Partitioned Optical Passive Start (POPS) network is modeled with groups of OPS couplers. Each coupler connects a group of nodes to another group of nodes, and each node is connected to several couplers. This network needs complicated control mechanism and the size of the network is strictly limited by the group number and the coupler's degree.

In an OPS coupler, each input or output link has the same opportunity to send or receive signals. In the real network, the packets carrying different traffic should not always be treated equally since the various applications have distinctive requirements.

Specifically, when two or more input links want to access the coupler with the same wavelength at the same time, some control mechanisms from outside of the OPS couplers are needed to allocate the couplers among these links. The hot potato routing and distance-age-priorities have been used and compared in solving the contention problems [ZhAc94]. In [Sudh93, SuMa91], some improved slotted Aloha protocols have been studied to solve the contention problems in a passive star network.

The optical switch [YaYa96, KrFo96] is the basic element for the devices with switching functions, such as Optical Add-Drop Multiplexer (OADM) [RiZe01], Optical Cross-Connect (OXC) [IaSa96], automatic protection switching and network monitoring [Inte02], which enables wavelength routing networks [HaSa00]. In the wavelength routing networks, each node routes the packets based on the availability of wavelength and the status of the links connected to it. A Routing and Wavelength Assignment (RWA) approach [KrSi01a, RaRi95] is the key to find a route along which the connection can be established and search for an appropriate wavelength to carry the signal. The switching devices provide the configurability that is necessary in such type of network.

1.1.3 The Optical Network Topology

With the evolution of optical components technology, the optical network topology has undergone necessary changes, which requires corresponding design methods. The logical topology design of lightwave networks is usually an optimization procedure with some constraints, which include the requirements of the networks, the real physical conditions and the limitation of optical devices. This procedure normally takes a long-time and high computing complexity with the solution of NP-problem. Thus, some heuristic methods have been investigated [RaSi96, KrSi01b]. By superimposing a regular topology on the physical architecture, we can calculate and analyze the logical design problem more accurately with mathematical methods. [CaBo96] shows how to embed a de Bruijn graph into an optical passive star network.

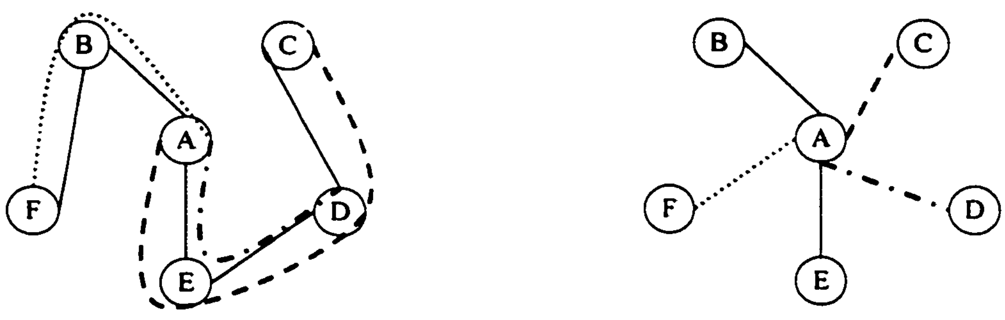


Figure 1.3 Physical vs. Logical Topology

The logical topology of a network can be independent of its physical topology. Limited by the real conditions of a network, the physical topology may not be changed at will. While at the logical level, any architecture that is best suitable for the applications and performance can be constructed. By imposing a logical topology into a real physical topology, we can apply the theoretical results in reality. As shown in Figure 1.3, the linear physical topology of a network is depicted on the left. By assigning the link capacity among not-directly-connected nodes, the network can take on a perfect logical star topology as on the right. For example, the physical bandwidth of the link between node A and E is shared by logical links A-C, A-D and A-E. Through an optimization procedure, we can find appropriate allocation of the bandwidth for each logical link. Therefore, the regular star topology is more convenient for analysis.

According to the number of intermediate nodes a message must pass, the optical network topology can be single-hop or multihop [Mukh92a, Mukh92b]. A single-hop topology normally has two major drawbacks: The need for expensive wavelength-agile transceivers and the requirement for pretransmission coordination. However, a multihop network [Siva92] does not necessitate fast-tuning transceivers or any pretransmission coordination. Traffic from a source node to a destination node has to travel via more than or equal to one node. At these nodes, the traffic undergoes optics-electronics-optics (O/E/O) conversions as well as being processed before being sent out via another link, possibly on a different wavelength. A node can be any terminal with the ability of generating, receiving or routing packets. Some researches on multihop optical networks include Rainbow [JuBo96, HaKr96], MONET [AnJa00], AON [HeSt96], HORNET [ShWh00], etc.

Multihop topologies can be regular or arbitrary. In regular topologies, the nodal connectivity pattern is very systematic and well defined. Examples of proposed regular multihop network topologies include ShuffleNet [HlKa88], Manhattan Street Network (MSN) or the Torus [KoHa01], deBruijn graph (DBG) [FeYa97], Kautz graph (KG) [PaSe95, PaSe99], Cayley graph [Tang94], Hypercube [Dowd92], as well as others [BaJa99]. Each of these networks has its specific advantages. For example, the ShuffleNet has a small average number of hops and symmetric deviation of number of hops from node to node. The de Bruijn graph can support a large number of nodes. MSN has a simple structure while the Kautz graph has the minimum average number of hops with the same degree among these regular graphs.

In an arbitrary multihop topology there is no predefined connectivity pattern. New nodes can be added to any existing node(s) in an arbitrary topology network provided hardware resources are available. Regular multihop networks are often preferred to irregular or arbitrary ones, as they tend to provide simpler routing schemes due to their well-defined connectivity patterns. However, it is difficult to add new nodes to a regular topology while still maintaining its well-defined architecture.

Table 1.1 Average Hops of Some Regular Graphs

Degree	Graph	Number of Nodes	Average Hops
3	ShuffleNet (3,3)	81	3.562500
	DBG (3,4)	81	3.516646
	KG (3,4)	108	3.261860
4	ShuffleNet (4,4)	1024	5.173021
	DBG (4,5)	1024	4.584401
	KG (4,5)	1280	4.160964

1.1.4 The Kautz Graph and Its Application in Optical Networks

The Kautz graph was first defined for multiprocessor interconnection by W.H. Kautz in 1968 [Kaut68]. Compared to other known regular graphs, such as de Bruijn graph or ShuffleNet, the Kautz graph has the minimum number of hops on the average with the same degree [Rose92], as shown in Table 1.1. This property is a major factor that motivates researchers to apply it to network research. In [PaSe95, RaRa97], the Kautz graph was proposed for multihop optical network topology. The Kautz graph has been used in optical networks in different forms, such as the single Kautz graph and the stack-Kautz graph (SKG) [ChLe96, CoFe00]. The SKG network is designed by stacking series of regular Kautz graphs and the connection between two stacks of nodes forms a small

POPS network. This topology is capable of supporting a large size network by using low-degree OPS couplers. In this network, the total device amount is still high and the control task is also heavy.

1.1.5 Mechanism of Optical Switches

The first era of optical network operated without switching features, provided only point-to-point transportation. Since the increasing application and service requirement demand a more flexible network architecture, switching becomes a necessary function of the optical networks. The switching technology includes circuit switching and packet switching. In circuit switching, an end-to-end path is setup before any data can be sent. The packet switching uses store-and-forward technique by transmitting fixed-length data blocks. Since the packet switching is capable of providing flexible bandwidth utilization and accommodating various applications, only optical packet switching [YaMu00a] is studied throughout this thesis.

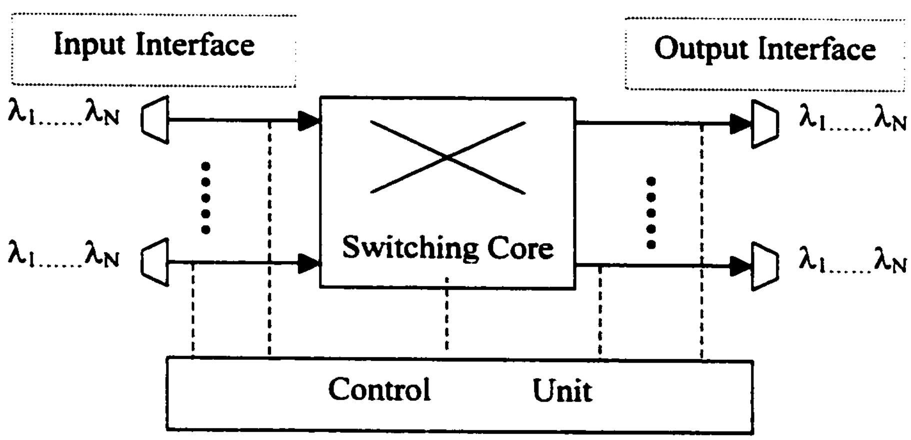


Figure 1.4 A Generic Optical Packet Switch

A generic optical packet switch usually consists of four parts: the input interface, switching fabric, output interface and control unit. As shown in Figure 1.4, the input interface mainly performs packet alignment, i.e., identifying the start and end of a packet, packet header information extraction and packet header separation from the payload. The switching fabric is the core of the switch and is responsible for routing optical packets to the appropriate output ports. The output interface is used to regenerate the optical packet header and sometimes performs wavelength conversion to accommodate the link state. The control unit controls the switching action through the information extracted from the packet headers. Currently, the control function is mainly realized electronically. For the convenience of synchronization, the optical packet switch mostly adopts a fixed-length packet. To accommodate the link availability or the port state, optical buffers and wavelength converters can be included in the input, output or switching core [XuPe01].

Typically, an optical switch operates in the following manners. When a packet arrives at an optical packet switch, it is first processed by the input interface. The header and payload of the packet are separated, and the header is converted into the electrical domain and processed by the control unit electronically. The payload remains in optical form throughout the switch. After the payload passes through the switching core, it is recombined with the header, which is converted back into the optical domain at the output interface. The packet then goes to next switch or transport link. If wavelength or link contention occurs, buffers or wavelength conventers are adopted to solve the conflict.

Depending on the stages of the switching block, the optical packet switches can be classified into two types: single stage and multistage. The single stage switch devices are capable of connecting one wavelength from the input to another one in the output ports, without passing through more than one switching core. Presently a wide variety of technologies have been developed for this purpose, including arrays of tiny tilting mirrors [OMM Inc., AD, TI], liquid crystals [Corning, Chorum Tech.], bubbles [Agilent], holograms [Trellis Photonics, Tellium], and thermo- and acousto-optics [JDSU, Lynx Photonic Networks, Alcatel] [Heyw00]. At present, however, none of these are close to being ready for widespread use in carrier networks. Each technology has pros and cons. It is unlikely that there will be a single winning technology, as all-optical switches will probably be used in a very wide range of applications. Most technologies can also be used to make other optical devices.

1.1.6 Multistage Optical Switch

Optical switches can utilize time-division multiplexing (TDM), space-division multiplexing (SDM) or WDM [ThHu96] with appropriate placement of wavelength converters and optical fiber delay line (FDL) buffers [TuZh99], while the electronic switches can only use TDM and SDM. Since the material dimension and processing techniques limit the buildup of large optical switching fabric, most proposals adopt multi-stage core, such as, "staggering switch" in [Haas93], 3-stages switch in [NaOk99], two-module stage switch in [OkOk00], layered architecture in [LiKa98], and Latin Router in [BaHu93]. To reduce the crosstalk, dilated architecture is proposed to achieve a tradeoff between the complexity and performance [PaNe87, ShCh93]. Some researchers have also considered the dominated IP traffic properties [CoRa00]. Some testbeds have been demonstrated, such as the WASPNET [HuNi99] and KEOPS [GuRe98] projects. The well-known multistage optical switch architecture also includes Banyan [VaLe98], Clos

[ChCh97], Wave-mixes, Cascaded 2*2 switches [NaUe99], SLOB (switch-with-large-optical-buffers) [HuCo98], etc. All of these switches share the same property of interconnecting small switch fabrics in certain patterns to support large switching ports, although there are different available options in the organization of switching elements. To satisfy the high-speed requirement of optical signals and to reduce the complexity of switching control, much research has taken place on the organization and simplicity of the interconnectivity. The topology of the interconnection network and the flow control technique will influence the performance significantly.

1.2 Motivations of Research

The situation and issues of the optical communication system in the previous section, along with the applicable schemes and the existing problems, have attracted much research interest.

As reviewed before, although the Kautz graph has received much attention, its limitations (such as size requirement, connectivity pattern, and routing) have prohibited its wider acceptance. A major limitation is the size of the network it can facilitate. Therefore, we would like to make this constraint less restrictive so that one topology can be applied in more situations. This scalability problem represents an important challenge to us and warrants further investigation.

When discussing the topology of optical networks, it is important to address certain implementation problems due to the limitation of optical devices which prohibit the replication of its electronic counterparts. As one can see from our literature review, many researches have not taken into account implementation problems seriously because the physical properties of optical components are very different from its electronic counterparts, thus making the implementation very difficult. Many presented architectures suffer from great complexity, huge cost, power consumption or high crosstalk. In the structure of all-optical switching, both path and wavelength contention are unavoidable. Since a lack of knowledge pertaining to lightwave devices inhibits solution, switching ports and performance are also limited. The OPS coupler is a viable choice to match the network requirement and the practical component technolgy. It is worth to exploit the potential of this simple device.

The optical technology has achieved significant success in transmission systems, but it requires matching capability in switching function of the intermediate nodes. The fast packet switching nodes have become the main bottleneck in the deployment of the fiber transmission technology in the ever-increasing network environment. Due to the particular properties of optical signal and components, the electronic architecture cannot be replicated in the optical domain when taking into account the optical properties. Thus, the multistage electronic switch architecture should be modified to work in the optical domain. This is one thing we would like to study.

There are concerns that electronic cores will not be able to cope with the explosion in the number of wavelengths in telecom networks, resulting from the deployment of DWDM. In theory, fully all-optical packet switches -- those that do not convert light pulses back into electrical signals – do not suffer from such limitations. At this time, optical technology is more suitable for circuit switching than for packet switching, because the time required to convert the wavelength or to complete O/E/O conversion is comparably longer than the light transmission time. An optical buffer is still impractical as well. Especially when the packets are very short as in Asynchronous Transfer Mode (ATM), the switching time determines that it is difficult to switch and control signals in the optical domain. However, we believe that only packet switches can match the development of future applications and we would like to initiate some work in this area.

1.3 Objectives

From above discussions, it can be seen that it is important to construct a logical topology for optical network and to integrate the available techniques into the optical packet switching networks. A primary objective of this thesis is to determine a logical network architecture with favorable properties in terms of the complexity, flexibility and implementation. The intention is that these topological results can then be embedded into the real physical networks.

Second, this thesis intends to explore the design of an optical packet switch. It would be ideal to find a dynamically flexible and reconfigurable optical core that can accommodate the dramatic increase in demand for capacity. A data-centric network requires high-speed and high-bandwidth packet exchanging and the support of diverse services. The physical potential of electronic packet switches lacks far behind the matching of the optical transportation link. Thus, switches capable of exchanging signal in the optical domain would be critical to catch up with the pace of the optical technology.

Specifically, an effort is made here to find a combination of the Kautz graph and the multilayer architecture in the optical packet switching network design. A new network architecture based on the Kautz graph is implemented with different devices. This network topology is studied under different traffic and control models. We then extend our network idea to the topological design of a multistage optical packet switch. We would like to investigate the scalability problem of this multilayer topology and determine the situations when such a design is most effective. We will also study the tradeoff between scalability, implementation and performance.

1.4 Methodology and Approaches

Since the physical conditions of real networks can be quite different, we would like to separate the logical topology from the physical topology in order to accelerate the network design procedure. The multilayer concept is very useful for network design, because it can divide the requirements into several simple sub-problems. Thus the logical topology will be analyzed first, and then the real implementation is considered. We shall also use this multilayer concept in the switch design to reduce the difficulty in large-scale fabrication, to increase the scalability of the topology and to satisfy the increasing requirement of advanced applications. The logical architecture of the switch is provided first and then the real devices used to finish the logical functions are considered.

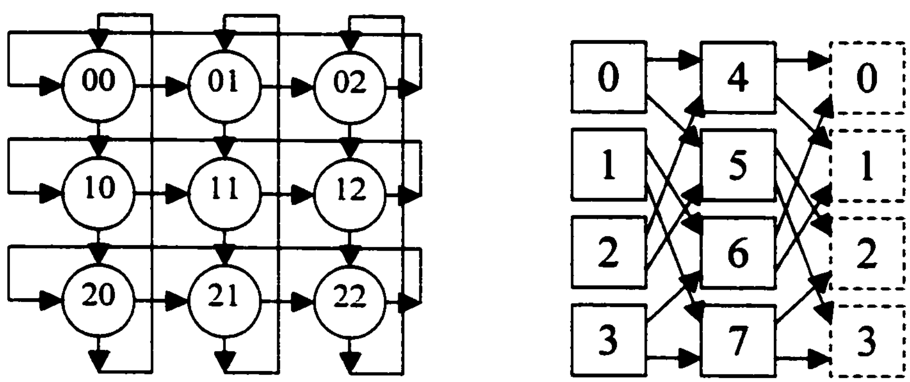


Figure 1.5 MSN and ShuffleNet

The ideas of MSN and ShuffleNet [ToLe98] have been combined to generate the new topology here. As shown in the examples of MSN and ShuffleNet in Figure 1.5, the connection between the adjacent layers is replicated as the last layer is wrapped back to the first layer. The nodes in each layer have the same connectivity pattern. By replacing these connectivity patterns from those in a Kautz graph, we can achieve a new topology, which will also be flexible and easy to implement with available devices. Therefore, the properties of these regular topologies that are easy to analyze and viable for application are exploited in this study.

In the design of an optical packet switching network, the primary goal is to achieve high performance in terms of complexity, scalability, modularity, reliability and throughput for networks. The design method should be able to provide similar usability under dynamic network situations. To do this, we start with the original Kautz graph, and duplicate several such graphs in similar ways as in the MSN and ShuffleNet. In this way, the connection inherits the regular graph and the routing operation complies with similar principle. Therefore, we can keep the simple properties of original topology while adapting to more complex network architectures. This architecture can provide more than one parameter combinations to accommodate a network. We shall analyze its performance by making some assumptions and using standard qeuing analysis methods employed in the analysis of other networks, such as those of the de Bruijn graph [e.g., FeYa97] and bilayered ShuffleNet [e.g., AyHa94]. We shall choose the best combination after comparing different variations. By combining the properties of the single-OPS and multi-OPS, the proposed topology is implemented in a novel method with OPS couplers. The topology is also modeled with the optical packet switches which provide wavelength routing function. Their variations are also proposed, analyzed and compared under some performance measures.

In the design of the multistage optical packet switch, we use a novel approach that divides the input signals into several groups and processes each group with small switches, thus achieving large switching capacity. We borrow the idea of Banyan architecture that uses the switching components with higher degrees. The connection of the adjacent switching stages is similar to the above multilayer network topology. To achieve fast switching and reduce the cost, the switch need not be strictly non-blocking. In this way, we can reach a balance between the cost/complexity and the performance.

1.5 The Contributions

The contributions of this thesis are:

1) The proposal of the multilayered Kautz graph architecture, which combines some of the merits of the Kautz graph and the multilayer topology, including the examination of the routing algorithm of the proposed topology.

2) The implementation of the architecture with single wavelength OPS couplers, multi-wavelength OPS couplers and small switch elements.

3) The analysis of three scheduling schemes for the contention resolution. The performances of these three methods for different networks.

4) The application of the MKG idea for multistage optical packet switches and the analysis of some general performance of the proposed packet switch.

5) The discussion of issues those are critical to the design and implementation of an optical packet switching network and an optical packet switch.

6) The simulation coding (in VC++) of the implementation and analysis for the above architectures.

1.6 Thesis Organization

This thesis is organized as follows.

In Chapter 2, the Kautz graph and its typical properties are reviewed. The multilayered Kautz graph is proposed and analyzed using queuing theory. This architecture inherits the good properties of the Kautz graph and is more flexible. We also give the operation model of this network.

Chapter 3 describes how to implement the proposed topology with OPS couplers and optical switches. The coupler implementation method is between the single-OPS and multi-OPS architecture. Three contention solutions are compared. The implementation with different amount of wavelength and buffer usage is compared. The node architecture suitable for the two implementations is also proposed.

The idea is then extended to the construction of a multistage optical packet switch in Chapter 4, which has a flexible architecture for different capability requirement. A control approach by grouping the signals is investigated to improve the throughput and reduce contention. The switch fabric with appropriate techniques and components is given in this part.

In Chapter 5, the design guidelines and some important issues are given based on the research and analysis of this thesis. These experiences would be helpful for the future work and further analysis.

Finally, conclusions are given in the last chapter based on the previous knowledge, while a recommendation for future research is also provided.

Part of this thesis has been published in the following conferences.

[JiYa02a] Mushi Jin and Oliver Yang, "A Multihop Optical Network Based On the Kautz Graph", CISS2002, Princeton University, March 20-22, 2002.

[JiYa02b] Mushi Jin and Oliver Yang, "Simulation Study of An Optical Network Topology", the 21st Biennial Symposium on Communications, Queen's University, June 2-5, 2002.

Chapter 2 Network Models and Operations

This chapter describes the architecture, models and operation of the network topology used in this thesis. The concept of Kautz graph is reviewed first and its optimal routing procedure is described. Then a novel architecture, multilayer Kautz graph (MKG), is introduced, which provides more choices for the optical network topology. The network operation and assumptions used for implementation analysis are given. The routing method of this network is derived from the routing properties of a single Kautz graph, and facilitated by a special numbering scheme for the links. Next, this topology is extended to arbitrary networks, and we modify the routing algorithm slightly to satisfy the new architecture. Finally, the performance of different configurations is evaluated and compared.

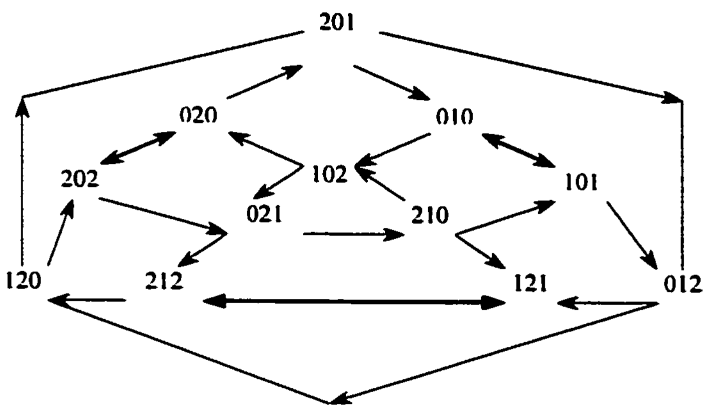


Figure 2.1 A KG(2,3) Kautz Graph

2.1 Kautz Graph [Rose92]

Let $G = (V, A)$ be a graph, where V is the set of nodes and A is the set of edges. Then a Kautz graph $KG(d,k)$ is a G with diameter k and each node has in_degree and out_degree d. Each node is labeled with a word of length k, $(x_1, x_2, ...x_k)$, from an alphabet of $d+1$ letters that do not contain two consecutive identical letters, i.e., $(x_i \neq x_{i+1}, for\, 1 \leq i \leq i-1)$. The number of nodes, N, in $KG(d,k)$ is $d^k + d^{k-1}$, and the number of edges, e, is Nd. There is an edge from vertex x to vertex y iff the last $k-1$ letters of x is the same as the first $k-1$ letters of y. In other words, the vertex $(x_1......x_k)$ is neighbor to d vertices $(x_2......x_k, z)$, where z is any letter different from x_k. Figure 2.1 is a KG(2,3) Kautz graph which has 12 nodes and whose diameter is equal to 3.

The shortest path routing algorithm of a Kautz graph is as follows. Suppose the source node is $(x_1......x_k)$, and the destination node is $(y_1......y_k)$, we need to find the maximum j $(1 \leq j \leq k)$, which satisfies $x_j = y_1, x_{j+1} = y_2, ..., x_k = y_{k-j+1}$. If such j does not exist, let $j = k+1$. Then a routing string $(x_1 ......x_k\ y_{k-j+2}......y_k)$ is formed and the routing path is a series of nodes that are defined by shifting one letter of the routing string, i.e., $(x_1x_2......x_k)$, $(x_2x_3......y_{k-j+2})$, ..., $(......y_k)$. As in KG(2,3) of Figure 2.1, suppose the source node is 121 and destination node is 101. It is easy to find that the routing string is 12101, thus the shortest path would be 121-210-101.

2.2 The Multilayer Kautz Graph

As a regular graph, a Kautz graph suffers many limitations due to its internal properties. For example, when the parameter of the network is fixed, the network size of $KG(d,k)$ can only be $d^k + d^{k-1}$. For a network size between 1000 and 2000, there are only $KG(2,10)$, $KG(4,5)$, $KG(6,4)$, $KG(10,3)$, $KG(11,3)$ and $KG(12,3)$ can be used. A more flexible architecture is given here.

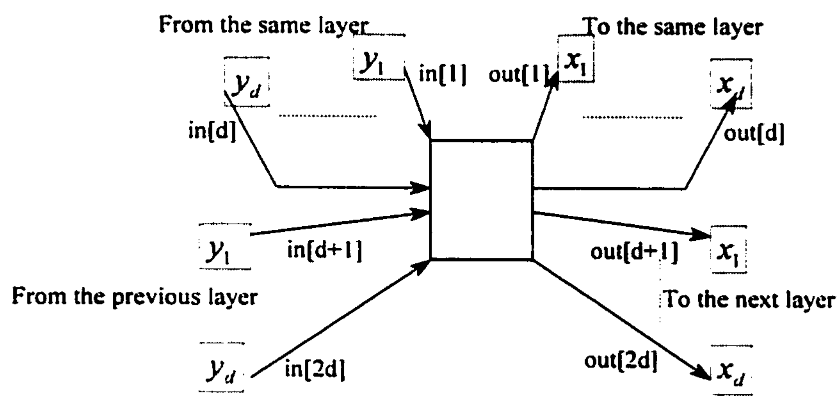


Figure 2.2 Basic Node Architecture

2.2.1 Network Architecture

By duplicating the single Kautz graph $KG(d,k)$ m times, we can construct a Multilayer Kautz Graph, $MKG(m,d,k)$, with each layer to be a Kautz graph. Let the layers be labeled from 1 to m, and node (l,s) be the node s on the lth layer. As depicted in Figure 2.2, each node has d out-links, $out[1], out[2], \ldots\ldots, out[d]$, to the nodes in the same layer, and d out-links, $out[d+1], \ldots\ldots. out[2d]$, to the nodes in the $(l+1)$st layer. Similarly, it has d in-links, $in[1], in[2], \ldots\ldots, in[d]$, from the nodes in the same layer and d in-links, $in[d+1], \ldots\ldots, in[2d]$, from the nodes in the $(l-1)$st layer. In other words, node

(l,s) on the lth layer has output links to the nodes s_1, ..., s_d on the $(l+1)$st layer, using the same connection relationship as in a single KG. At the last layer, the additional links will wrap around to the first layer. The degree of the new architecture is twice of that of the single graph.

The multilayer idea provided here is similar to the Shuffle exchange network. In a Shuffle exchange network (e.g. in Figure 1.5), the nodes in the same layer have no direct links between them, and the connection between any two adjacent layers is the same. The routing from one node to another node is a series of hops from layer to layer. A route may pass the same layer more than once to get to its destination.

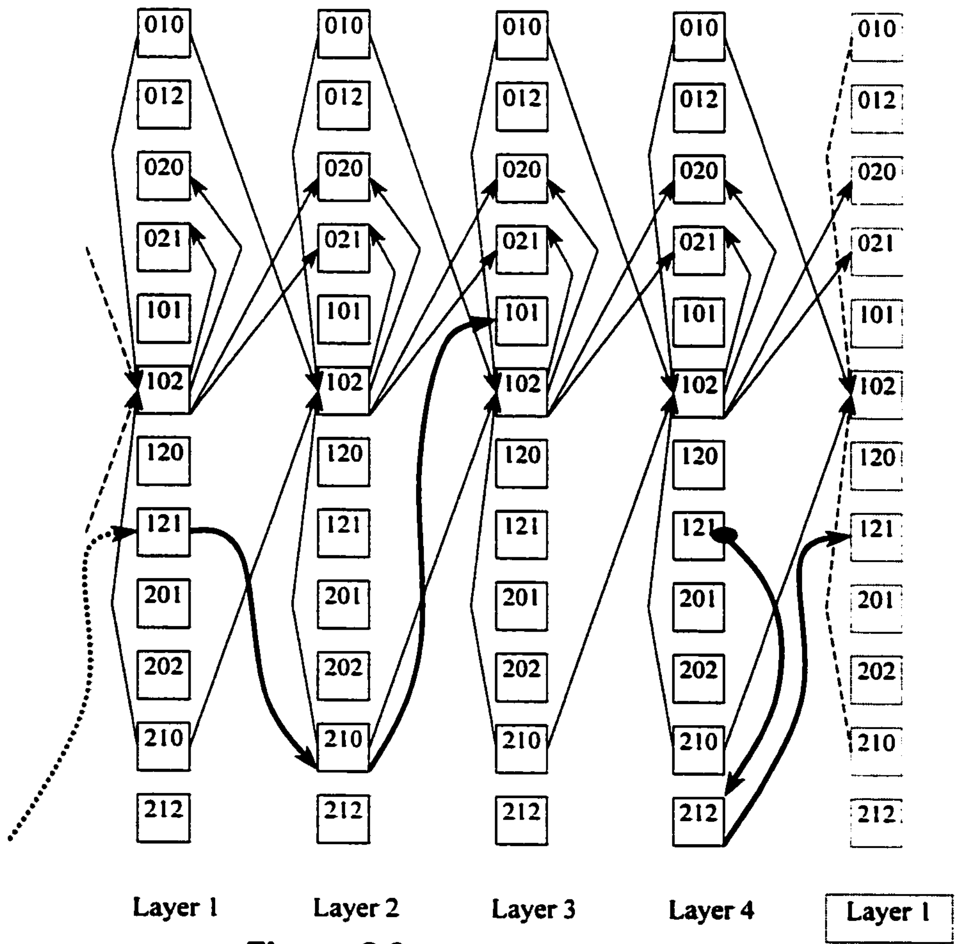


Figure 2.3 MKG(4,2,3) Architecture

In the MKG, the input and output links of each node are numbered sequentially, depending on the difference of the first or the last letter of the node label respectively. Although the links could be numbered arbitrarily, we find that it is easier to analyze if we number them in a certain order based on the definition of the network. The simple way we have come up with is to number the input links based on the first letter of the input nodes' labels and to number the output links based on the last letter of the output nodes' labels. Then we can build a routing string to represent the routing path with a series of link numbers. In case of a node failure, we can join two links together to form a new link while keeping the same routing method. The disadvantage is the increasing of node degree. However, in the implementation section it will be shown that this increase is only at the logical level. When we implement it with selected components, we can divide the links of one node into two parts, each of which is only half of the logical degree.

Figure 2.3 is $MKG(4,2,3)$, in which only the in-links and out-links of node 102 are shown. Node 102 has two out-links to node 020 and 021 in the same layer, and two out-links to node 020 and 021 in its next layer. Since the links between two adjacent layers comply with the relationship in the original Kautz graph $KG(2,3)$, the new architecture can adequately utilize the properties of the original single graph for routing with very low complexity.

2.2.2 MKG Operation and Routing

The MKG network is used for the optical packet switching network, which is operated as follows. All packets have the same length and arrive at the nodes synchronously. Each packet has a header and a payload. The header contains necessary information for routing and controlling. A viable solution is to use sub-carrier multiplexing (SCM) [BlPr94]. In this approach the header and payload data are multiplexed on the same wavelength (optical carrier). In the current that modulates the laser transmitter, payload data is encoded at the base-band, while header bits are encoded on a properly chosen subcarrier frequency at a lower bit rate in order to facilitate processing.

In a similar way as the Kautz graph, we can define the routing rule in an MKG. For two nodes in the same layer, we can use the optimal routing algorithm of single Kautz graph, which is also optimal in the new architecture. When they are in two different layers, depending on the layer locations of the source node (l_s, x) and the destination node (l_d, y), we can divide the routing algorithm into two cases. Suppose the optimal routing string from node x to node y in a single Kautz graph is $(x_1, x_2, \ldots\ldots, x_n, y_1, y_2, \ldots\ldots, y_k)$, and the number of hops is equal to n. Then we have:

Case 1: $n \geq (l_d - l_s + m) \bmod m$. The routing string will be the same as that in a single KG. The first $n-m$ nodes along the route will be in the source layer. The labels of the remaining m nodes will be the same as the last m nodes of the optimal route while the layer number of each node will increase by one in each step.

Case 2: $n \leq (l_d - l_s + m) \bmod m$. Under this case, the routing string will be exaggerated to match the layer difference. Some redundant nodes have to be inserted into the optimal (minimum) routing path to maintain the connectivity relationship, i.e. no two identical adjacent letters in the routing string. The new routing string has a number of hops that is equal to or greater than the layer difference of the source and destination node. The routing will then be similar to the first case: take some hops in the source layer, then other hops between adjacent layers until the destination node.

We can use the routing from node (4,121) to (3,101) in $MKG(4,2,3)$ as an example. In the $KG(2,3)$, the optimal route from (121) to (101) is determined by the routing string (12101) with 2 hops. While in the $MKG(4,2,3)$, the layer difference form (4,121) to (3,101) is 3 (i.e., $(3-4+4) \bmod 4$), which is greater than 2. Thus the new routing string is an extension of (12101) to increase the hops to at least 3. Since the left and right ends of the new routing string are the source node and destination node respectively, which cannot be changed, the shortest new routing string will be in the form of 121...101. Since the last letter of source node 121 and the first letter of the destination node 101 are the same, a redundant letter different from letter 1 need to be inserted between them. So the new routing string could be (*1212101*) with hops equal to 4 Thus the routing path will be (4,121), (4,212), (1,121), (2,210) and (3,101), as the bold lines in Figure 2.3.

From the above construction procedure, one can see that the routing string cannot be shortened further. Thus the routing path has the minimum number of hops, which is the optimal (shortest) one. For the implementation in a real network, the routing path can be dynamic depending on the network condition and required hops. First, as long as the number of unfinished hops is not less than the layer difference from the current node to the destination node, the hops can be in the same layer or between two adjacent layers. In the above example, without changing the node labels, we can let the layer sequence be 4, 1, 1, 2, 3 or 4, 1, 2, 2, 3 or 4, 1, 2, 2, 3. Second, we can use adaptive routing instead of the above fixed routing. Since the whole path can be represented by the routing string $(x_1, x_2, \ldots\ldots, x_n, y_1, y_2, \ldots\ldots, y_k)$, we can adjust the intermediate nodes by inserting some redundant letters to the string while keeping the first and the last k letters to be the source node and destination node respectively. This insertion will generate more hops, but will be more flexible for the changing traffic.

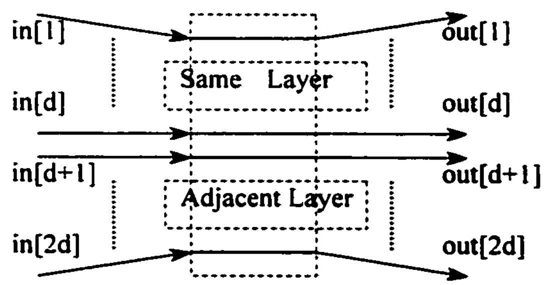


Figure 2.4 An "Absent" Node

2.2.3 Network of Arbitrary Size

If the number of nodes is not exactly equal to the multiples of the size of a single Kautz graph, the architecture in the previous section can be modified using an idea similar to that in [SeBa01]. By choosing an appropriate m and N, we can construct a multilayer architecture whose size is greater than but very close to the network size. We can then remove some nodes in the ending part of the last layer in order to achieve the proper network size, thus accommodating the real network. A removed node is considered absent during the analysis of the MKG. In an "absent" node the inputs and outputs are "transparently linked". As shown in Figure 2.4, the input links and output links of an "absent" node are directly connected according to their indices. Thus the architecture of the remaining nodes is reserved and the routing string will be kept.

The routing algorithm of the arbitrary network is similar to the fully-configured one. First we should establish a routing string to be the same as a fully-configured MKG. Then the routing path is determined by the number of hops and the layer difference as in the previous section. An important point in dealing with the routing string is that when the routing path passes through "absent" nodes in the last layer, some hops may be reduced. Therefore, in the situation when the source node is in the last layer and the destination node is in another layer, a different routing rule should be used. That is, after the routing string is generated in the same way as in a fully-configured MKG, the path is first made up of hops between adjacent layers. Then all the remaining hops will be taken in the destination layer. In this way we can avoid the hops being disrupted at the "absent" nodes. Note that in other situations we still use the conventional routing methods for a fully-configured MKG. For example, if we use $MKG(4,2,3)$ in Figure 2.3 to model a network with 46 nodes, the last two nodes 210 and 212 in layer 4 will become "absent" to match the required size. The wrap-around routing string from node (4,121) to (3,101) can still be (1212101). However, since the transit node (4,212) does not exist, the routing path mentioned in above section has to be modified. This is done by taking the only other path starting from node (4,121) to (1,212). The remaining hops can be either combination of the listed dynamic routings in Section 2.2.2.

2.3 Performance Evaluation

M/M/1 is a common and useful model to evaluate the performance of a network topology. In this section it is used to analyze the availibility of KG and MKG. The following assumptions [RaRa97] are used for our analysis.

1) All nodes have the same traffic load, i.e., the probability of a node to generate a new packet at each time slot. We suppose the Poisson arrival rate is λ packets/step.

2) Each node has an infinite buffer. At every time slot a packet is stored in the buffer if there is no available output link or wavelength. The assumption of an infinite buffer is for the convience of analysis although it can only be finite in the real implementation.

3) All packets arrive at the nodes synchronously. A small part of the signal is detected by the control unit. The routing and buffering of packets are determined by the header information, while the payload will remain unchanged through the network.

4) It takes one time slot (one sequence step) for a packet to propagate from one node to the next node. The service time is set to zero. Thus the whole network can be thought of as running in sequential time slots. All time slots consist of sequence steps. In this thesis, we will consider the performance with respect to the sequence steps.

5) The traffic is uniformly distributed, i.e., the traffic from one node to any of other nodes has the same possibility, and the distribution of the destination nodes have a normal distribution

6) Average packet size is μ bits/packet.

7) Link capacity is C_{ap} bps.

2.3.1 Comparison of KG With Other Graphs

Based on the above assumptions, we can investigate many aspects of performance as follows. Note that L_{ij} is the load of the link between nodes i and j, L_i is the load of link i, and M is the number of links.

1) Mean hops length (average number of hops): $\bar{h}$.

2) Average delay: $\overline{D} = \frac{1}{N(N-1)} \sum_{i,j} \frac{L_{ij}}{\mu C_{ap} - L_{ij} \cdot \lambda}$

and normalized delay: $\overline{D}\mu C_{ap} = \frac{1}{N(N-1)} \sum_{i,j} \frac{L_{ij}}{1 - L_{ij} \cdot (\lambda / \mu C_{ap})}$

3) Average link load: $\overline{L} = \frac{1}{M} \sum_{i=1}^{M} L_i$.

Table 2.1 Comparison of KG With Other Graphs

		Number of Nodes (N)	Average Hops ($\overline{h}$)	Link Load ($\overline{L}$)
KG	(3,3)	36	2.5810	18.4444
	(4,4)	320	3.6656	212.5781
	(4,5)	1280	4.6540	1168.3945
DBG	(3,3)	27	2.4786	21.4812
	(4,5)	1024	4.5844	2344.9206
ShuffleNet	(3,3)	18	2.7165	12.33
	(4,7)	1024	5.1730	1323.00
MSN	(4,6)	24	3.300	13.2
	(14,14)	196	7.89	15.32

Table 2.1 is the comparison of the mean hop length, and the average link load of some known regular graphs [Rose92]. Let us take $KG(4,5)$, $DBG(4,5)$ and ShuffleNet (4,7) as examples. The last two graphs have the same size of 1024 nodes, while the DBG has shorter average number of hops and lower average link load. Although $KG(4,5)$ has

more nodes, we can see that its mean hop length and average link load are smaller than other two graphs. Therefore, Kautz graph is a good choice for the network topology.

2.3.2 Performance of the MKG

Table 2.2 Performance of Some MKGs

m	d	k	Number of Nodes (N)	Average Hops ($\bar{h}$)	Link Load ($\bar{L}$)
3	3	3	108	2.614197	47.055556
4	3	3	144	2.773148	66.555557
5	3	3	180	3.018518	90.555557
6	3	3	216	3.348765	120.555557
7	3	3	252	3.727513	156.555557
3	3	4	324	3.522977	190.240738
4	3	4	432	3.579732	257.740753
5	3	4	540	3.713786	334.240753
6	3	4	648	3.928155	424.240753
7	3	4	756	4.224133	532.240723

Table 2.2 illustrates the typical performance results of some MKGs. From the table, we can see that when the number of layers m is approaching the diameter d of the single graph, the average number of hops becomes smaller. By choosing different combinations of layer and graph size, we can obtain different MKG networks to match a given size. Then we can choose the one with the best performance. This type of network design obviously has more flexibility.

Table 2.3 Average Delay of Some Graphs Under Different Load

Normalized Load	KG(4,4)	MKG(3,3,4)
0.0001	2.7246	3.6180
0.001	3.4046	4.6495
0.002	4.7244	7.1497

Table 2.3 lists the average delay of $KG(4,4)$ with 320 nodes and $MKG(3,3,4)$ with 324 nodes, under different network load. It is shown that the KG has the shorter delay and the MKG suffers longer delay under the same load. Let us compare the performance of MKG networks and KG networks in Table 2.2 and Table 2.1. Also take $MKG(3,3,4)$ and $KG(4,4)$ as an example. When the network size is almost the same, an MKG is capable of achieving similar average number of hops as a KG by properly choosing the parameters. However, this MKG needs higher node degree than the KG since the former one is not so regular as the later one.

2.4 Concluding Remarks

This chapter overviews the model of the MKG topology. The general idea is using the multilayer architecture to build scalable networks, while maintains the flexibility by setting up the routing path with the inherited advantage from the regular graphs. In the next chapter, such architecture will be modeled and different modeling methods will be compared.

Chapter 3 A Multilayer Kautz Graph Network

This chapter focuses on the implementation of the proposed MKG topology with available optical devices, the OPS couplers and optical switches. The OPS couplers are applied to the MKG architecture first. We discuss the application with both single wavelength and multi-wavelength situations. Three contention resolution methods are provided and compared. Then we consider the implementation with more general devices - optical switches, which are capable of achieving better performance with wavelength routing ability.

3.1 Implementation Using OPS Couplers

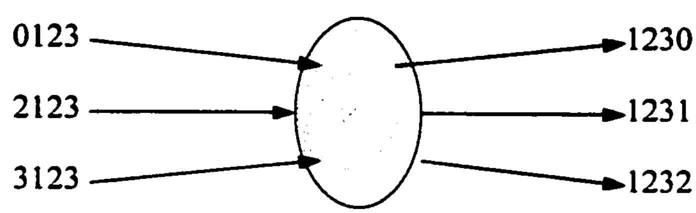


Figure 3.1 An OPS Coupler in MKG(5,3,4)

In the first implementation of MKG, $d \times d$ OPS couplers are employed. Each node in the network is connected to two OPS couplers, one connecting to the nodes in the same layer and another connecting to the nodes in the next layer. Those nodes that have links going to the same destination nodes are connected to the same OPS couplers. Although the node degree of the MKG is $2d$, as in Figure 2.2, only $d \times d$ OPS couplers are needed. Lower degree also has a benefit of easy implementation.

Figure 3.1 is one such OPS coupler used in $MKG(5,3,4)$ network. The three left links are from nodes in the same layer, while the right three may be in the same layer or the next. Observe that there are nine logical links between the three left nodes and the three right nodes, while there are only six physical links between these six nodes and the OPS coupler. The complexity of the MKG topology is therefore not high, and the cost can be reduced. Observe also that if one OPS coupler fails, the traffic could be rerouted to another one using the dynamic routing method mentioned in Section 2.3. In case of the failure of one node or one OPS coupler, the service of the network will not be interrupted by using an alternative path or by increasing the number of hops.

Consider the example used in Section 2.2.2. The routing from node (4,121) to (3,101) in $MKG(4,2,3)$ is (4,121), (4,212), (1,121), (2,210) and (3,101) (e.g. Figure 2.3). If the OPS coupler between (2,210) and (3,101) fails or node (2,210) fails, the routing path could be (4,121), (1,212), (2,121), (3,210) and (3,101) with the same number of hops. Therefore, a network with node or coupler failure can still use the same routing string. This property of not affecting the network routing performance is extremely important for high-speed networking in which each link carries very large amounts of information.

3.1.1 Control Issues

In optical networks, the traffic signals are transmitted at very high speed, up to 2.5 Gbps or 10 Gbps, while control signals are much slower and can be transmitted at a lower speed to facilitate routing and controlling. The operation of the nodes takes place under the control of the control node in the node groups. In each layer of MKG, one node is selected as the control node of the node group that shares the same two OPS couplers. Any node in the group can be the control node, who is responsible for collecting the header information of all packets passing the node group and sending the control information to instruct the operation of the nodes. When two or more packets arrive at a node with a destination to the same output link or more than one node needs to access the same OPS coupler with the same wavelength, a contention may occur. Some scheduling principles must be employed to solve the contention by giving the packets different priorities to pass an OPS coupler. The packet that cannot obtain the permission to transmit is put into the buffer as long as the buffer is not full. These buffers are equipped with each node (see assumptions in Section 2.3). The decision of which node has the highest priority to access the OPS couplers is made by the control node. Here three scheduling methods using different criteria will assign the highest priority to the packet to access the OPS couplers:

1) With the longest cumulative delay until the present sequence step, i.e., the "delay" method. The delay includes both buffeing time and transmitting time.

2) With the shortest distance left to the destination node, i.e., the "distance" method.

3) With the longest waiting time in its current buffer, i.e., the "waitingtime" method.

Note that either single wavelength or multiple wavelengths can be employed. Although the basic connection patterns are the same in both cases, the control mechanisms are different in terms of the wavelength allocation.

3.1.2 Single Wavelength Situation

The first implementation model utilizes single wavelength OPS couplers, which do not require expensive and complex tunable transceivers. Currently, tunable components are still expensive and difficult to fabricate, so a single wavelength method is considered first. Each node is equipped with fixed-wavelength transmitter and receiver. There is only one wavelength available in each coupler at one time, meaning only one node can access an OPS coupler at one time. Each input and output link can receive and transmit signals with equal priority if no controlling mechanism is used to distinguish between them. At each time slot, one node is allowed to send one packet over one OPS coupler. When contention occurs, only the packet satisfying certain criteria can pass through the coupler. Although the signal will be delivered to all output links, only one receiver can receive the packet.

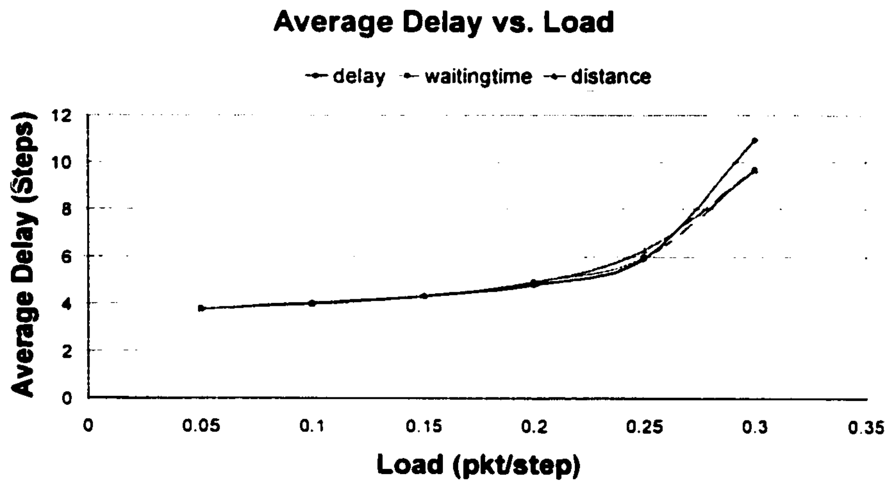
3.1.3 Multi-wavelength Situation

In the multi-wavelength implementation, the same assumptions as in the previous section are taken here except that there are d wavelengths available to each node and coupler. Therefore, each node requires tunable transceivers to utilize all possible wavelengths. At each sequence step, each input link of the star coupler is capable of sending one packet through the OPS coupler with distinguished wavelength. If more than one packet requires access to the same output link, only one of them is allowed through based on one of the three scheduling mechanisms described in Section 3.1.2. Before the packets pass through the coupler, they would pass through tunable wavelength converters (TWCs) first to ensure there is no wavelength confliction inside the coupler. The control node could allocate the wavelengths randomly while keeping the requirement of wavelength conversion to the minimum.

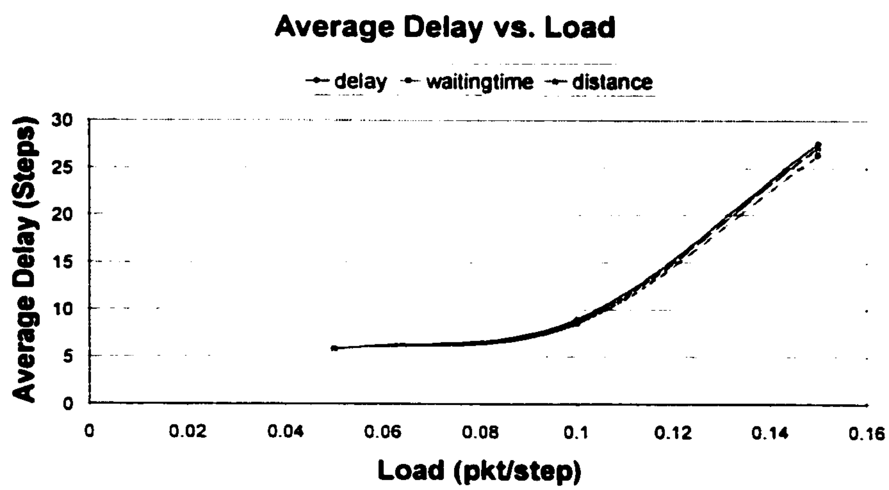
3.1.4 Performance Evaluation

Based on the above network model and assumptions, the implementation of an MKG network with OPS couplers is simulated with Microsoft Visual C++. The program runs on an Intel Pentium PC with the number of sequence steps between 200,000 and 1,000,000. To verify the results, the 95% confidence intervals of some simulations are shown in Appendix B, which has proved that the results are in acceptable level.

The main performance criterion modeled here is delay, defined as the duration (in number of sequence steps) between the time a packet is generated and the service completion time of the packet, i.e., the total time slots a packet needs for transmission and buffering. The reason why we choose delay as our analyzing goal is related to the transmission properties of optical signals. Longer delay often means more O/E/O conversions, which will increase the device cost. Even if the system works in an all-optical form, longer delay also creates longer transmission distance, which will introduce large crosstalk, power attenuation, or bit errors. In order to compensate the power loss of long fiber transmission, optical amplifiers will be employed, which will bring more non-linearity effect and degrade the signal-to-noise ratio. Therefore, in a good optical topology design, to reduce the delay is an important goal. We have chosen two networks for our analysis, $MKG(4,2,3)$ with 48 nodes and $MKG(5,3,4)$ with 540 nodes respectively. The sizes of these two networks differ ten times and the node degree is 4 and 6 respectively. Therefore, these two networks can represent networks of different order of complexity. The same two networks are modeled in different scenarios in order to compare the performance results.

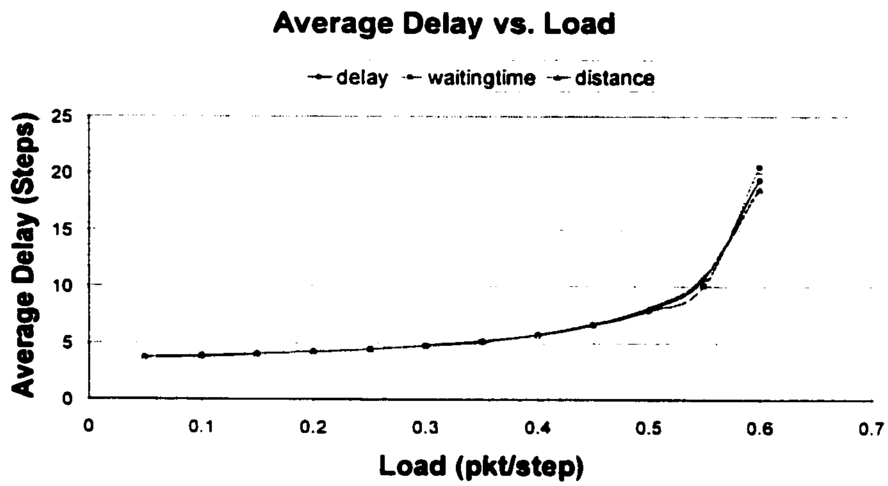


(a)

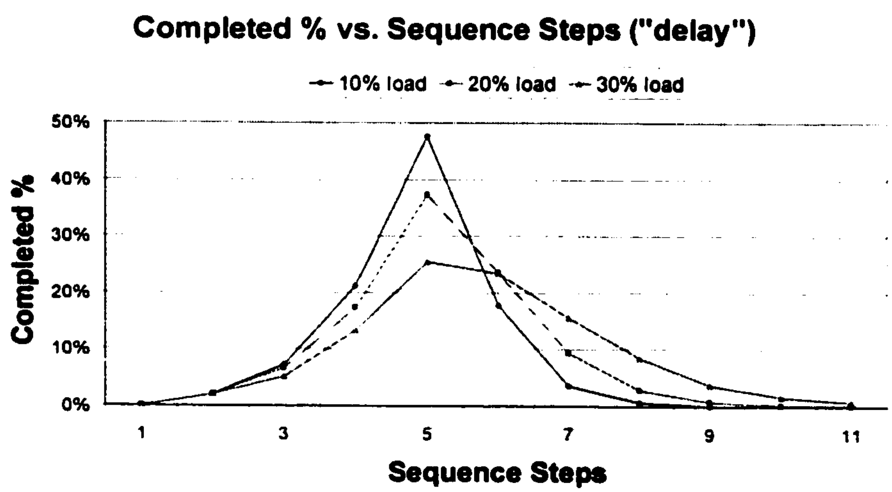


(b)

Figure 3.2 Average Delay in single wavelength situation:
(a) MKG(4,2,3) (b) MKG(5,3,4)

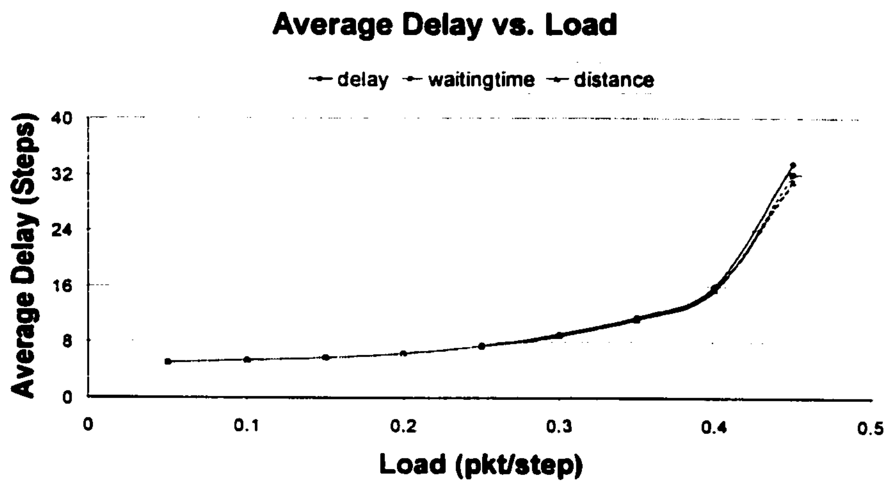


(a)

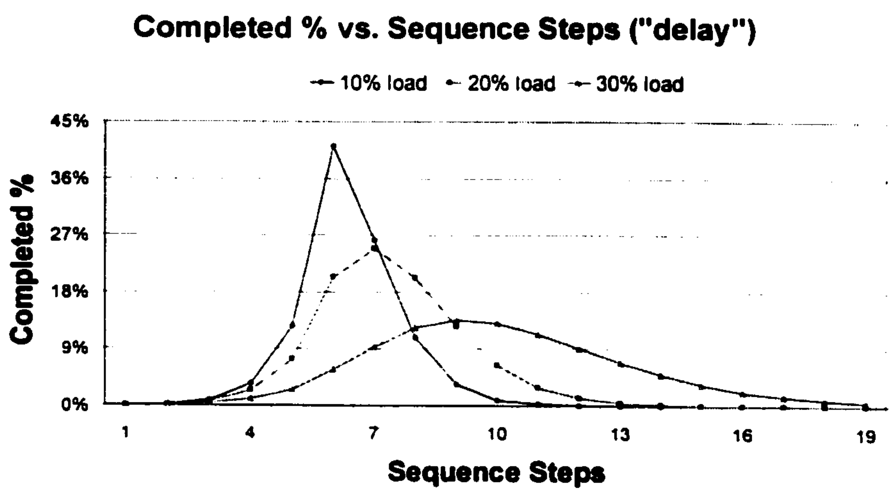


(b)

Figure 3.3 Performance of MKG(4,2,3) with multi-wavelength: (a) Average Delay (b) Percentage of delivered packets vs. needed sequence steps ("delay" method)



(a)



(b)

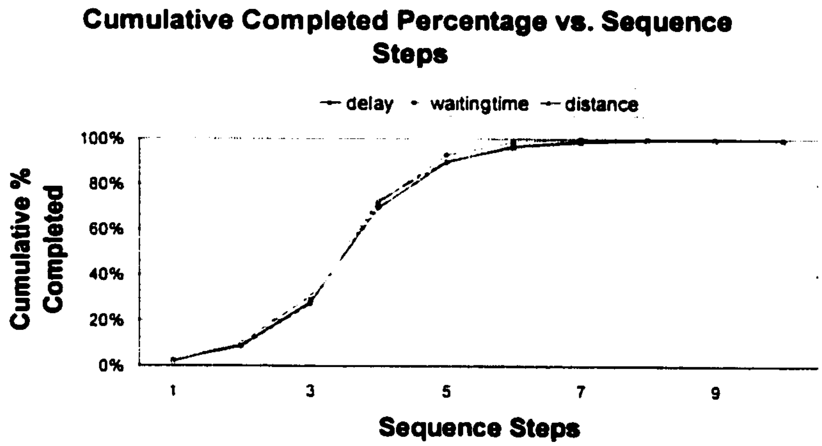
Figure 3.4 Performance of MKG(5,3,4) with multi-wavelength: (a) Average Delay (b) Percentage of delivered packets vs. needed sequence steps ("delay" method)

3.1.4.1 Average Delay

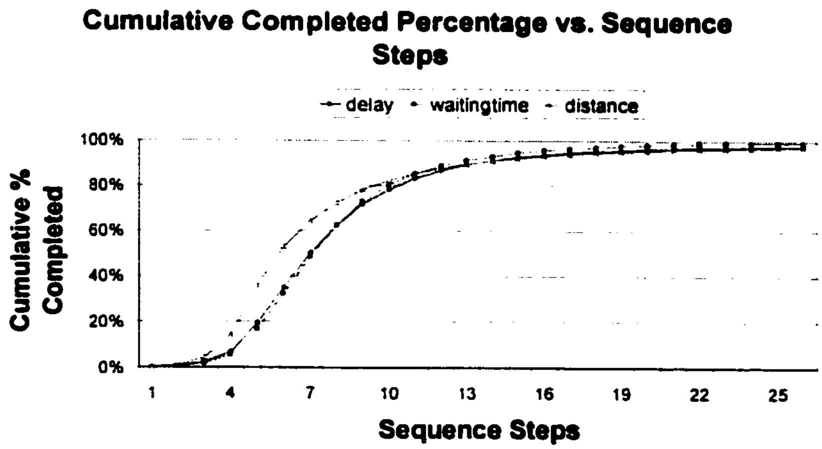
Figure 3.2 shows how the average delay varies with the packet arrival rate (packet/time slot) in single wavelength situation. Note that the horizontal axis is also the normalized load (the probability of a new packet is generated in a node at each time slot) because the service time is exactly one time unit. From the figure it is observed that in the $MKG(4,2,3)$ network the average delay increases exponentially with the increasing of the load, as expected for the general delay-throughput relationship in the queuing theory [Klei76]. When the load approaches 30 percent, the delay increases dramatically. Normally, the performance of the network will be unacceptable with very high delay. In Figure 3.2b, this point occurs earlier to $MKG(5,3,4)$, i.e., at load of 0.1 instead of 0.3 in Figure 3.2a. This is expected as the second network is ten times larger and more complex.

Figure 3.3 and Figure 3.4 are the average delay and delay distribution under multi-wavelength situation. It is shown that the average delay in Figure 3.3a increases in the same manner as the single wavelength situation for $MKG(4,2,3)$. However, the multi-wavelength network is capable of supporting almost twice the load of that in the single wavelength network for the same delay. Obviously, this improvement of performance is based on the price of two wavelengths rather than the single wavelength used. Similar conclusions can be drawn from $MKG(5,3,4)$ by comparing Figure 3.2b and Figure 3.4a.

Figure 3.3b and Figure 3.4b show the percentage of delivered packets versus the needed steps under different loads. The higher the load, the greater the steps are needed to deliver the same percentage of packets because more contention may occur. Only the results of the "delay" method is given here since the other two methods have almost identical results, which are omitted. For the small network $MKG(4,2,3)$, the peak of the step distribution does not change substantially under the load below 30 percent, which complies with the distribution of the average load in Figure 3.3a. For a larger network $MKG(5,3,4)$, the peak of the step distribution increases tremendously along with the load. Thus the saturation load of a small network is much higher than a large network. This is also the same as what is found in the previous paragraph.

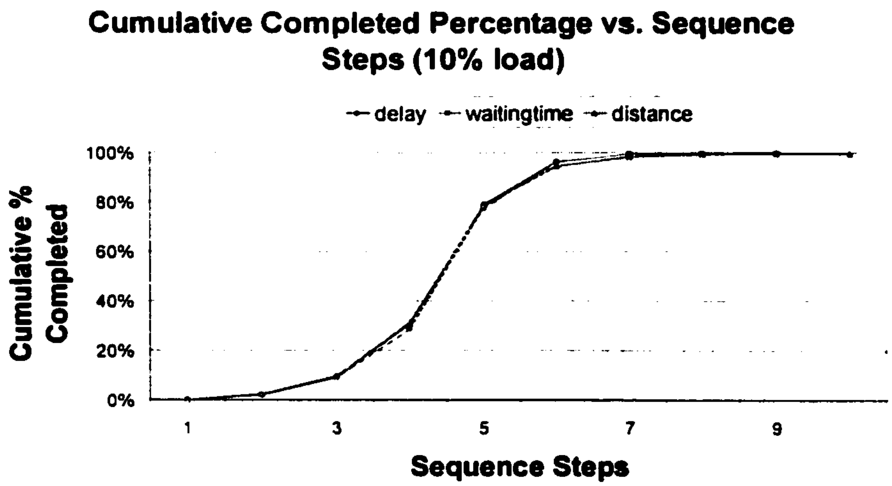


(a)

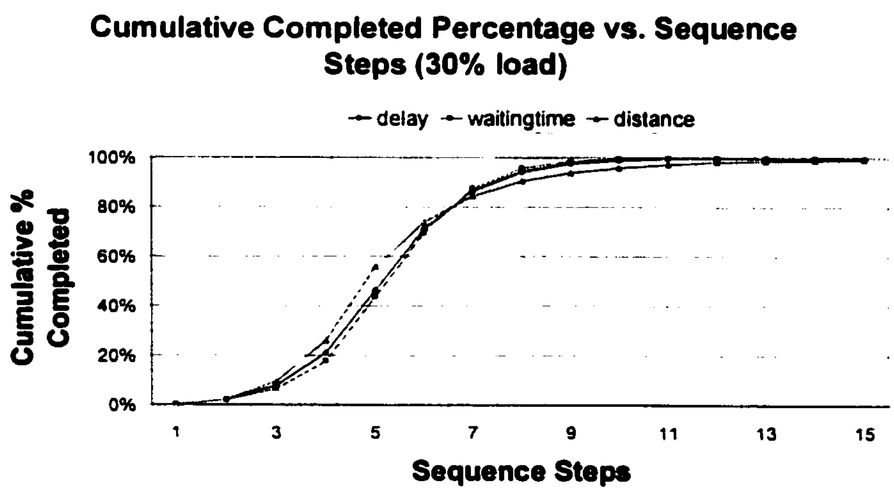


(b)

Figure 3.5 Cumulated percentage of delivered packets vs. needed sequence steps in single wavelength situation: (a) MKG(4,2,3) (b) MKG(5,3,4) (10% load)

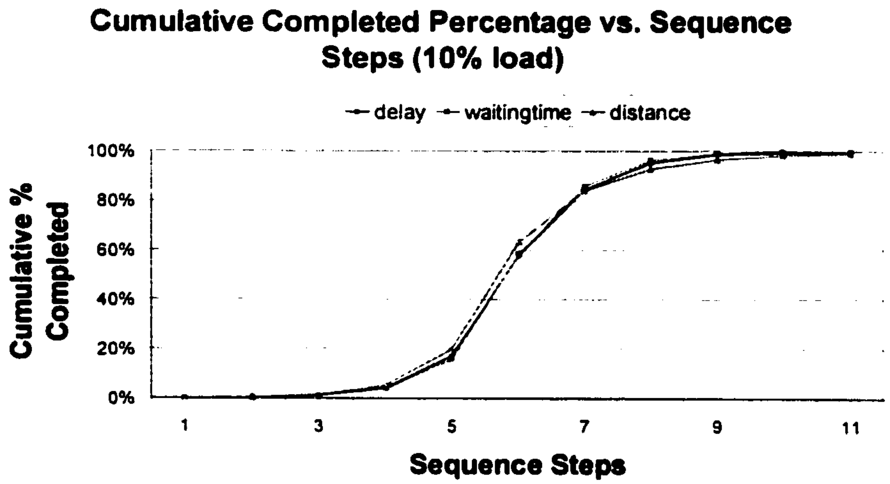


(a)

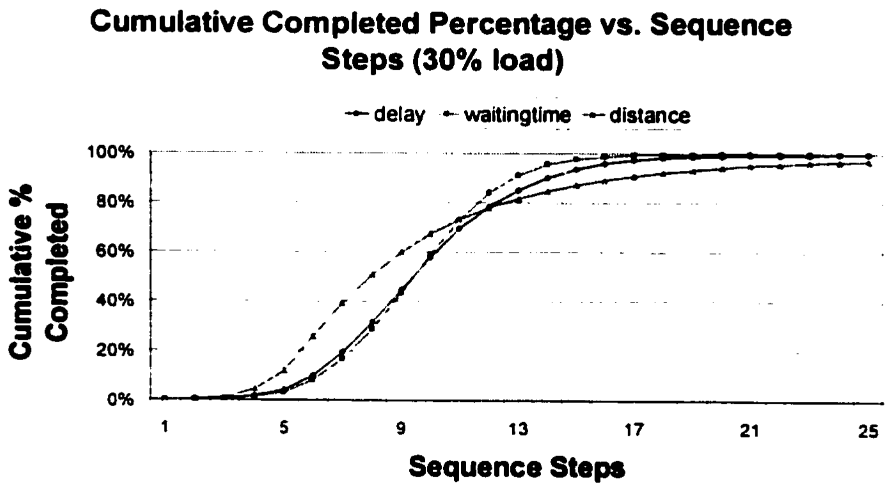


(b)

Figure 3.6 Cumulated percentage of delivered packets vs. needed sequence steps of MKG(4,2,3) with multi-wavelength situation: (a)10% load (b) 30% load



(a)



(b)

Figure 3.7 Cumulated percentage of delivered packets vs. needed sequence steps of MKG(5,3,4) with multi-wavelength situation: (a)10% load (b) 30% load

3.1.4.2 Cumulative Delay Distribution

Figure 3.5 shows the cumulative percentage of completed packets out of the total packets versus the sequence steps needed to finish the delivery when a single wavelength is used. This property tells us how much packets are delivered after certain time slots since they are generated, which will help us to choose the network parameters and operation method. To the small network $MKG(4,2,3)$, three scheduling methods do not generate significant difference. In the large network $MKG(5,3,4)$, it is observed from Figure 3.5b that in the shorter range of sequence steps, the completion speed of the "distance" algorithm is the fastest. After a certain percentage of messages have been completed, say 98 percent, the "distance" method becomes the worst, while the "waitingtime" method becomes the best. To finish the same percentage of traffic, $MKG(5,3,4)$ network needs 50 percent more or even twice steps.

From the results of the multi-wavelength situation in Figure 3.6, we can see that three scheduling methods have similar effects on the delivery speed of $MKG(4,2,3)$ as the results in Figure 3.5a, except that the speed using multiwavelengh is much faster than the single wavelength scenario. This is to be expected since the tradeoff is that more wavelengths and tunable devices are required for receiving and transmission. As the load increase from 10 to 30 percent, the needed steps for the same percentage of traffic is about one and a half times more. Similar conclusions can be drawn with $MKG(5,3,4)$ as shown in Figure 3.7.

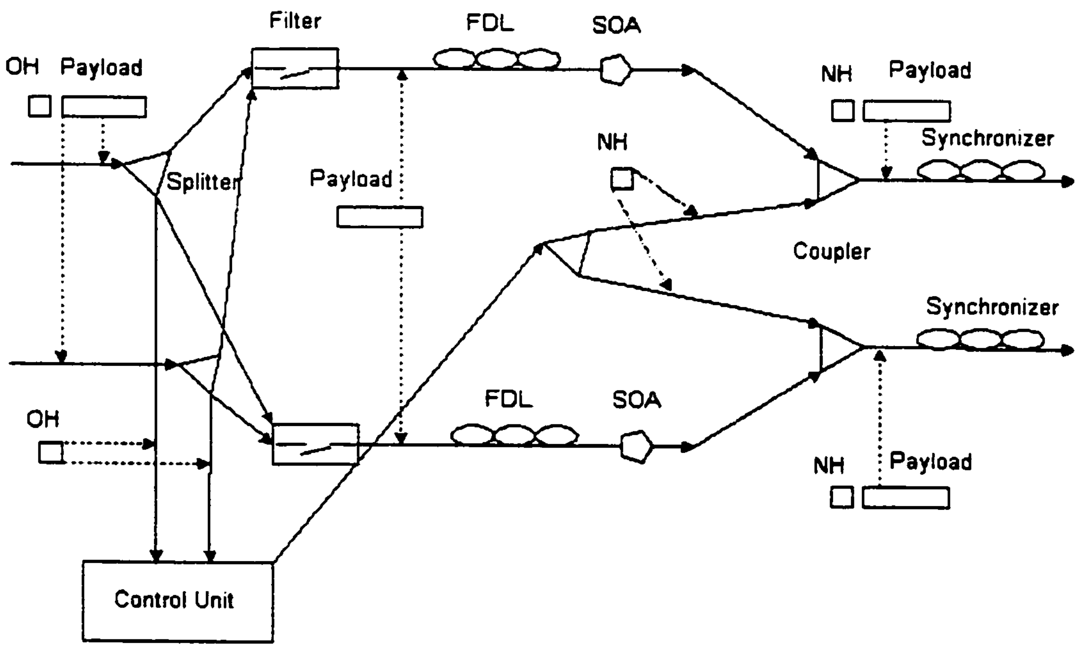
3.1.4.3 Complexity Issues

Under a single wavelength situation, each node has two states: one indicating the need to send packets to the same layer and another indicating the need to send packets to its next layer. This state word has a length of $\log(2d)$ bits in one group. At each step, each node will send this word to the control node in its group. Then the control node feeds back the availability of the couplers with a word of $2d$ bits. Therefore, the total length of the state words would be $\log(2d)+2d$ bits. If multi-wavelength is used, the control part should also include the wavelength information. Thus each node would know if wavelength converters are needed or not and which wavelength is allocated to it. Since there are d wavelengths available, each node needs another d bits to indicate the wavelength information of each OPS coupler. Then the total number of control bits would be $2d^2+2d+\log(2d)$. For $MKG(5,3,4)$ with 540 nodes, the control word would be 27 bits.

The time complexity of the control scheme in single wavelength situation can be calculated as follows. First, the d nodes in one group send the index of the OPS couplers that are needed. This step will consume time in the order of d, i.e., $O(d)$. Secondly, the control node will assign the OPS couplers to the nodes in its group that have sent requirements. This step will take twice the amount of time. Finally, the control node will take one time unit to send the control word to the nodes in its group. The total time complexity will be $(d+2d)\sim O(d)$. The multiwavelength situation will take the same amount of time as the single wavelength situation in the first and last step, i.e., $O(d)$, while the second step will cost time of $O(d^2)$ due to the increasing of the wavelength information. Therefore, the total time complexity is then $O(d^2+d)=O(d^2)$.

Modeling results in the previous section shows that the multi-wavelength implementation would provide better performance than single wavlength. It should also be pointed out that more wavelength, more complex node, and more complex control mechanism are also needed.

3.1.5 A Node Architecture Example



OH=Old Header; NH=New Header; FDL=Fiber Delay Line;
SOA=Semiconductor Optical Amplifier

Figure 3.8 Node Architecture in the MKG

Considering the above-mentioned important issues, we now present in Figure 3.8 one possible node architecture which has incorporated the schemes of synchronization, header processing, buffering, and controlling. Some techniques suitable for this architecture are the following. First, the packet format for the MKG could be sub-carrier multiplexing (SCM), in which the header is electrically modulated at a lower frequency than the payload base-band to facilitate the electrical processing in each node. The shown architecture is used in a single wavelength situation. When the input signal arrives at a node, it is split into its header and payload parts. The payload is kept in the optical form and the control unit will decide if and how long it should pass through the FDL buffer. The header will be converted into an electronic signal for controlling and new header regenerating. The semiconductor optical amplifier (SOA) is used to compensate power loss during processing and buffering. When there is an available OPS coupler, the control unit will choose the appropriate payload and generate its new header, and combine them to a new packet. In order to erase the jitter during the transmission and processing, the output packet is synchronized before being sent out.

The above architecture can also be used in the multi-wavelength situation. In this case, a tunable wavelength converter is used to avoid wavelength conflict before the signal is sent out. The transmitter and receiver must also be tunable in order to process the signal in different wavelengths.

3.1.6 Discussion

From the modeling and simulation of the multilayer Kautz graph network with OPS couplers, we find that the network implemented with OPS couplers can keep the delay within an acceptable range. When the tunable receivers and transmitters are used, the performance exhibits significant improvement. We have presented three scheduling schemes for contention resolution that can achieve our objectives. For a small network with a low load, they do not have much difference. However, for a larger network or with a higher load, the "distance" method can deliver more packets in short time steps while the "waitingtime" method can go faster after a certain step. Depending on the requirement of the network, the appropriate method can be chosen.

Table 3.1 Comparison of Some Topology Implementations

Topology	Number of Nodes	Number of Groups	Degree of Coupler	Number of Coupler	T/R Total	T/R per Coupler	Control Bits
POPS(60,30)	1800	30	60	900	54000	30	570
SKG(12,5,3)	1800	150	20	900	10800	6	108
MKG(6,3,5)	1944	486	3	1296	7776	6	27
MKG(5,4,4)	1600	320	4	800	6400	8	43

NB: The advanced control is used for SKG.
The multi-wavelength situation is considered for MKG.
T/R=Transmitter/Receiver.

Table 3.1 compares the implementation of MKG with POPS [ChLe96] and SKG [CoFe00] topologies when OPS couplers are used. As observed, the implementation of the MKG architecture in this thesis needs fewer OPS couplers and shorter control bits. When the network size is close to 1800, the MKG topology can utilize OPS couplers with much lower degree and a much smaller number of transceivers than with the other two topologies. Each node could be equipped with fewer Transmitter/Receivers. Note that the delay performance in our MKG architecture is a little longer than the POPS and SKG networks. The reason is that MKG uses fewer devices, which sacrifies the performance of delay. Simpler devices imply that the cost will be reduced.

3.2 Implementation Using Optical Switches

In this section, we investigate the implementation of MKG network with $2d \times 2d$ optical switches in the intermediate nodes. This implementation would require more devices, but is more flexible. The optical switch is capable of delivering $2d$ incoming signals to $2d$ distinct outgoing ports at the same time. Each node is equipped with an infinite buffer. The packets with contention will be put into the buffers. To access a switch, packets also obey one of the "delay", "distance" or "waitingtime" scheme as in the coupler implementation.

3.2.1 With Intermediate Buffering

Employing the similar assumptions of traffic model and replacing each node with a $2d \times 2d$ optical switch, we can analyze the MKG in the same way as we have done with the OPS couplers. Among the packets that arrive at the same switch from other nodes, those with contention will be put into the buffer while only those satisfying the scheme of "delay", "distance" or "waitingtime" method can obtain permissions to pass the switch. If there are extra output links available, some packets will be taken out from the buffer based on the scheme of "delay", "distance" or "waitingtime" method. Once a message starts to be transmitting through the network, it has priority over the buffered packets to access the switches. Therefore, the switch ports will be allocated among the packets from the input links first, then the free ports will be assigned to the buffered packets. This rule can reduce the O/E/O conversion during the packet transmission. The same two networks $MKG(4,2,3)$ and $MKG(5,3,4)$ will be modeled.

3.2.2 Without Intermediate Buffering

To store the signals in the optical form is still impractical at present. The O/E/O transform cannot match the high speed of the optical transmission. Thus, the network could be running in another more practical pattern. Suppose the buffer of a node only stores the messages generated by this local node based on first-in-first-out (FIFO). Once the message enters the switches, it cannot be turned back to the buffer of intermediate nodes until it reaches the destination node. When two or more messages arrive at the same switch and generate a contention of output port, only the one satisfying with the adopted contention resolution criterion will pass through and others must be dropped. A transmitting message has priority over the buffered messages to access a switch. If there are extra output links available, some packets will be taken out from the buffer based on the scheme of "delay" or "distance" method. Here the delay is actually the packet's waiting time in the buffer. Therefore, in the real setting, the buffer is equal to an access point and the buffered messages can stay in the electrical form. When there is a switch port available, it will be transferred to optical form for transmission. Along with the routing path, the signal will not take any O/E/O conversion which represents a bottleneck in today's design procedure.

3.2.3 Performance Evaluation

The main performance criterion modeled here is still the delay performance. When intermediate buffering is allowed, there is no packet loss. When there is no intermediate buffering, the packet loss ratio is considered. The simulation environment is the same as that for the coupler implementation described in the beginning of Section 3.1.5.

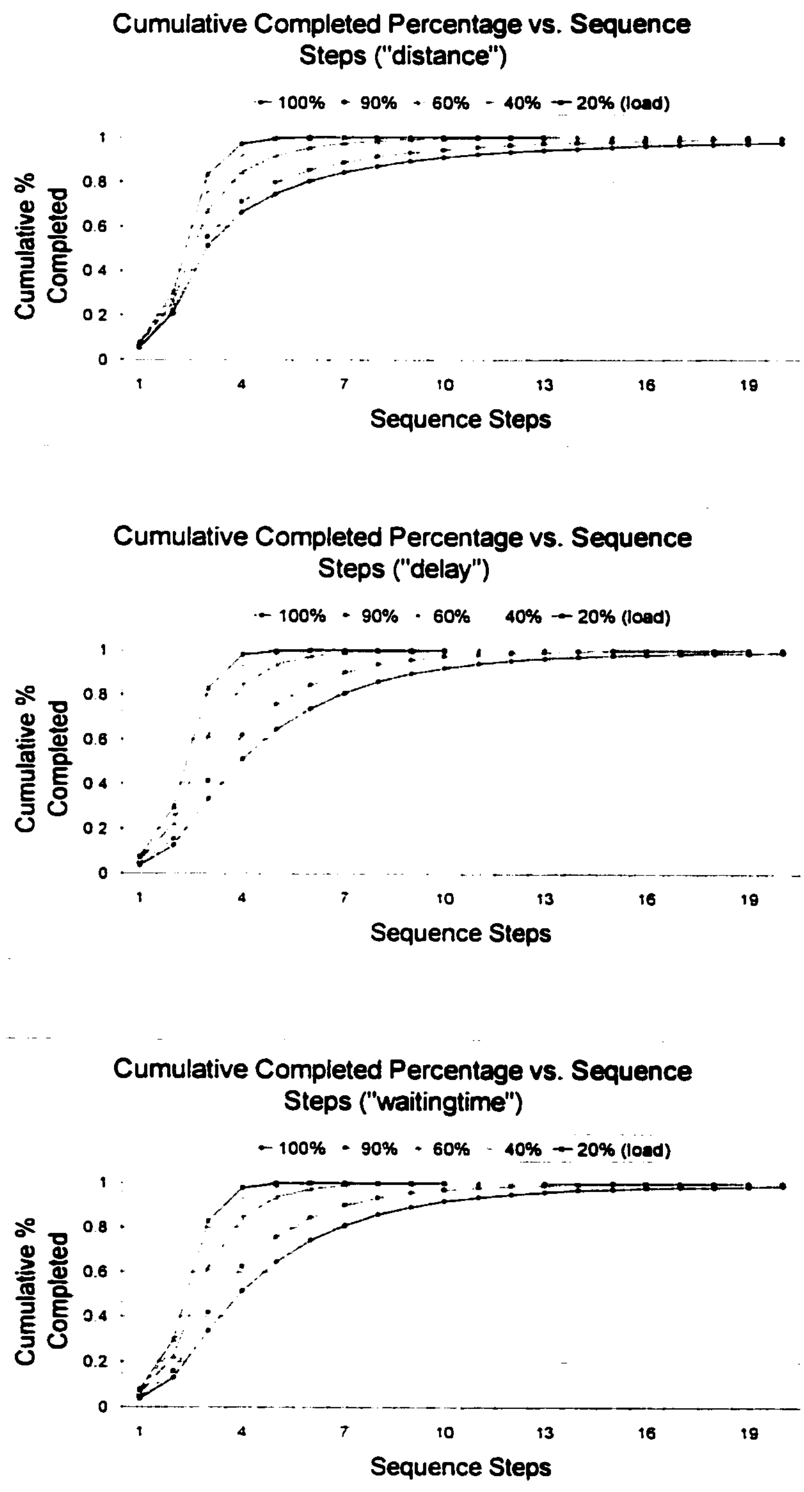


Figure 3.9 MKG(4,2,3) with switches:
Cumulated percentage of delivered packets vs. needed sequence steps

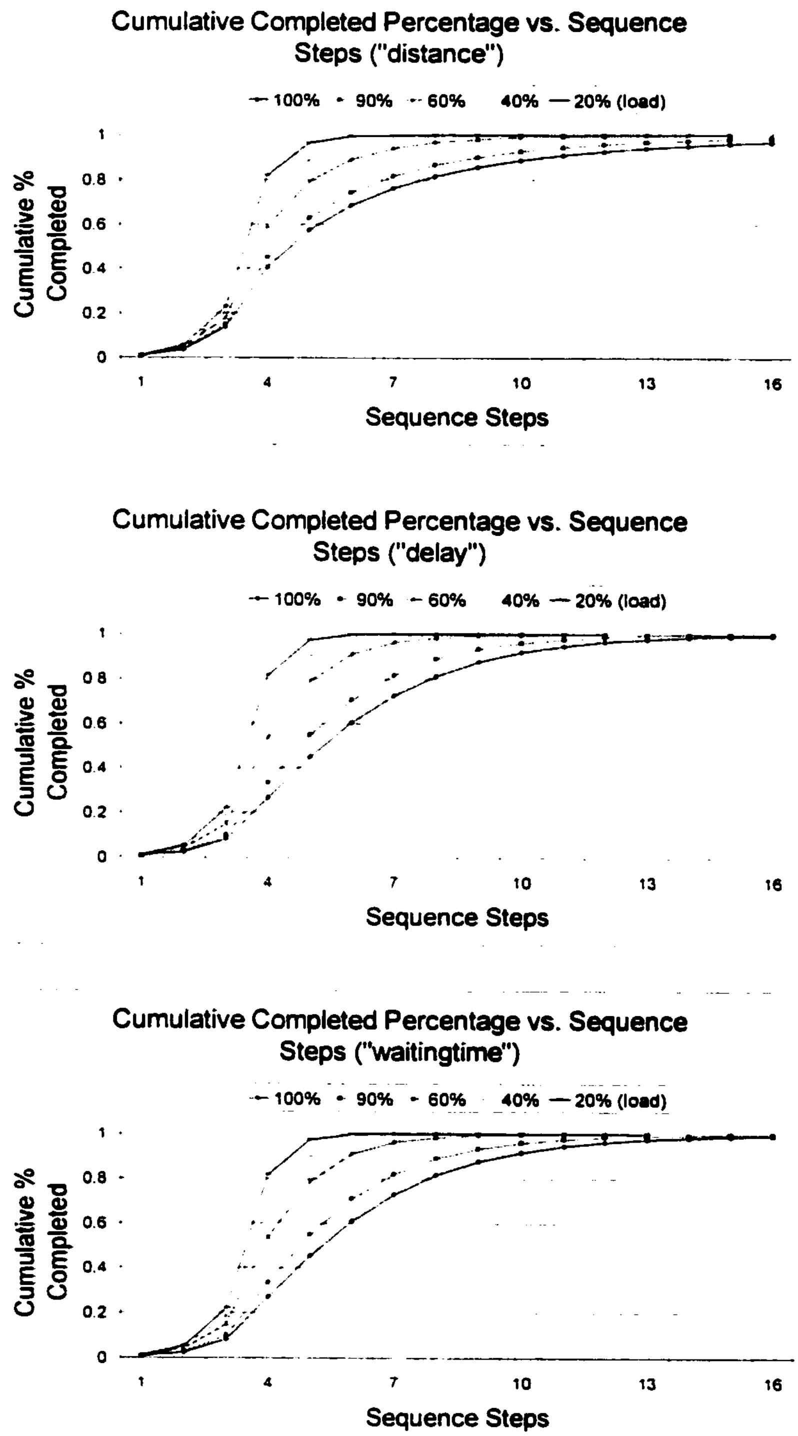
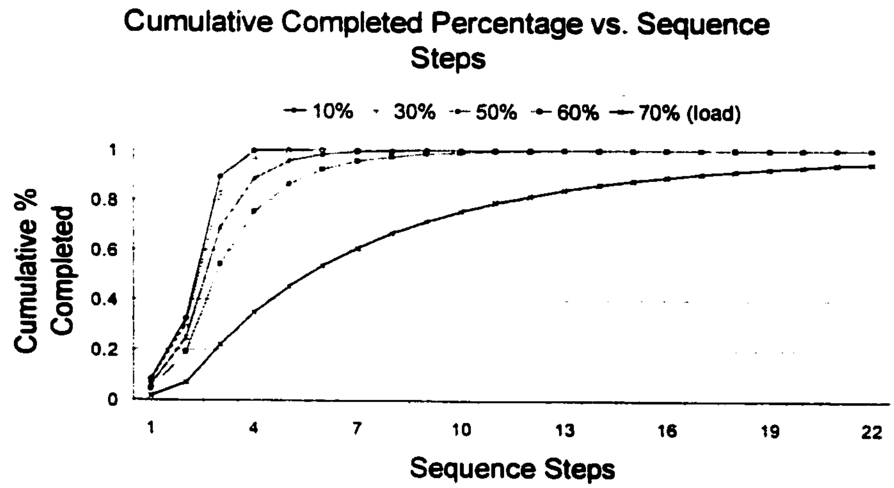
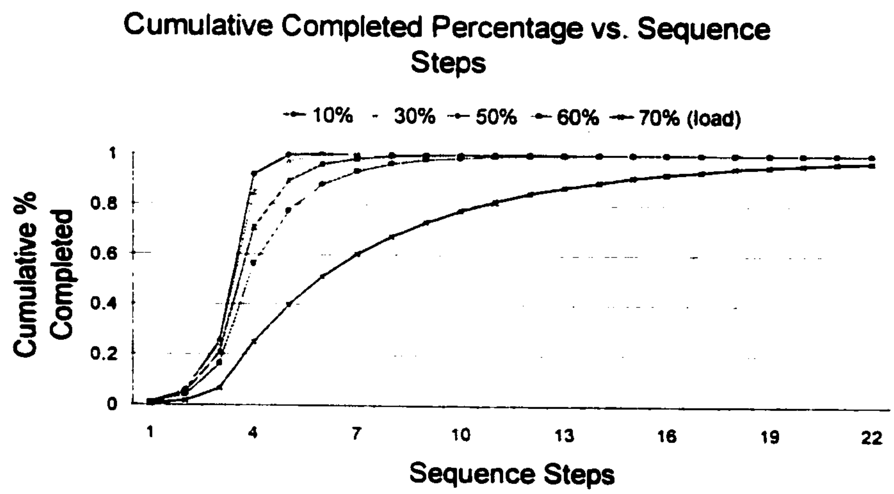


Figure 3.10 MKG(5,3,4) with switches:
Cumulated percentage of delivered packets vs. needed sequence steps



(a)



(b)

Figure 3.11 Cumulated percentage of delivered packets vs. needed sequence steps without intermediate buffering: (a) MKG(4,2,3) (b)MKG(5,3,4) ("delay" method)

3.2.3.1 Delay Performance

Figure 3.9 and Figure 3.10 show the cumulative completion percentage versus the needed sequence steps when intermediate buffering is allowed. The larger network *MKG*(5,3,4) suffers longer delay than the small network *MKG*(4,2,3) when the load and the scheduling scheme are the same since the former network is more complex. Compared to the OPS couplers, (eg. Figure 3.9 vs. Figure 3.6), the optical switches can support much higher load with very low delay and the completion speed is also very fast. Three scheduling schemes do not generate a significant difference in the results.

Figure 3.11 is the cumulative completion percentage when no intermediate buffering is allowed ("delay"method is adopted). For the two sizes of networks being modeled, the results are very close. Observe this type of network can only support a lower load with the same delay performance as in the first scenario. This is because the first simulation uses flexible buffering, i.e., all packets with contention will be put into buffers and any buffered packet satisfying the chosen criterion can access the switch. While the second uses strictly FIFO buffers for the locally generated packets only. Thus a local packet must wait for the turn to become the header of the buffer and for the time slot when there is an available output link. Therefore, the packet has more difficulty in obtaining the permission to pass through the switch under the second situation, and results in an increase of the delay overall. The packet from other nodes cannot use the local buffer. The only result of a packet with contention during transmission is to be dropped.

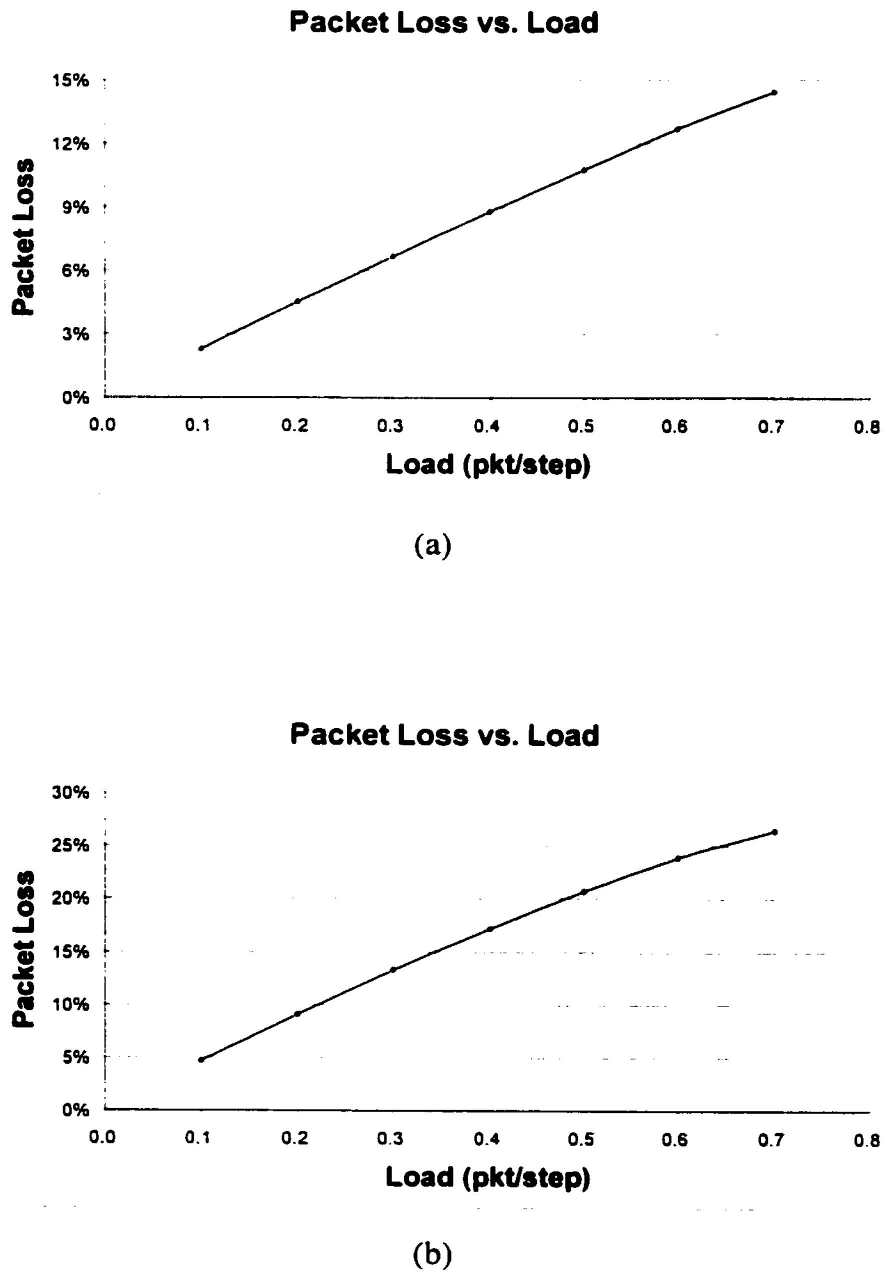


Figure 3.12 Packet Loss vs. Load without intermediate buffering: (a) MKG(4,2,3) (b) MKG(5,3,4) ("delay" method)

3.2.3.2 Packet Loss Ratio

Since the buffers are assumed to be infinite in our modeling, no packets would be discarded due to a shortage of buffering room in the first implementation. However, in the second modeling situation without intermediate buffering, the packets could be dropped due to a contention during the transmitting. From the simulation results shown in Figure 3.12 ("delay" method is adopted), we can see that the packet loss ratio versus load is pretty high. It is also found that the larger the network is, the higher loss rate the packet would have. This could be explained that the longer average routing hops in the large network causes more contention, resulting in more loss.

3.2.4 A Node Architecture Example

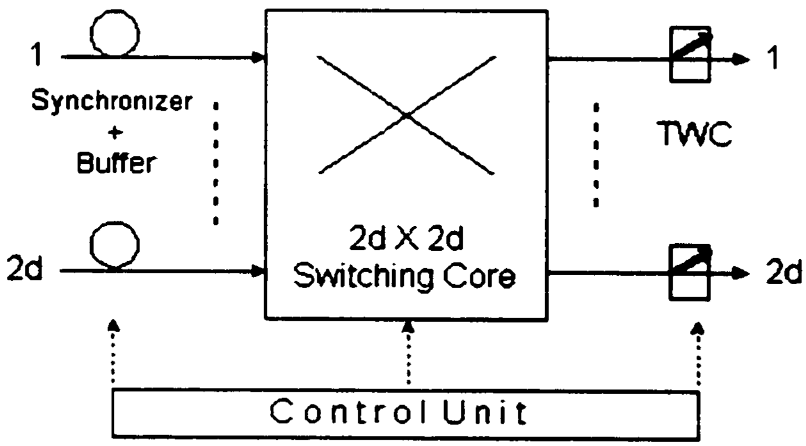


Figure 3.13 Node Architecture With Switches

A node architecture example for the switch implementation is shown in Figure 3.13. The node consists of a $2d \times 2d$ switch core with wavelength converters and buffers. When the packets arrive at the input links, they are synchronized first. A small part of the signal will be split into the control unit. If a contention takes place, the packet will be buffered or discarded depending if the intermediate buffering is adopted or not. The TWCs could be put before or after the switching fabric to remove the wavelength conflict. The input messages may come from other nodes or be generated by the local node. The most important part here is the switching core. As mentioned in Chapter 1, this switching fabric could utilize several technologies, among which the MEMS (Micro-Electro-Mechanical System) is mostly preferred at present [ChLe02, DoFa02].

3.2.5 Discussion

From the simulation results shown in Figure 3.9 to Figure 3.12, we can see that when the optical switches are applied to the MKG, the delay performance of the packets being delivered successfully under with and without intermediate-buffering situations does not change significantly. If no buffering is employed during the routing path, a large portion of the packets will be discarded which seems impractical for any application. However, what we are considering is constrained in the optical layer, and the upper layers of the network can take more responsibility for the restoration and protection. In this situation, outside mechanisms are needed to improve the throughput. Thus the loss ratio still can be reduced to some reasonable level.

3.3 OPS Couplers vs. Optical Switches

In the previous analysis of the MKG implementation, the OPS couplers are low-cost passive device with degree d, and the optical switches have degree $2d$. The coupler can be shared by several nodes, while the switch is built into a specific node. Therefore, the required number of OPS couplers is $m*2*(d^k + d^{k-1})/d$, which is less than that of the optical switches. The coupler is a simple passive device, and its control mechanism is completed by the control node. However, the optical switch in each node should be capable of dealing with its own routing and controlling operation without the help of other nodes except the information abstracted from the header part of the packets. The network with switches is more flexible to the dynamic traffic.

Although the delay performance of these two devices varies in the same manner, they differ significantly with the supported load. When the network load is higher than 30 percent, the OPS couplers network generates poor delay data, while the switches network can support a network with much higher load with acceptable delay. Even in the low load range, the switches network's delay performance is almost twice better than that of the coupler network. This is of course at the expense of loss due to finite buffers. However, the loss level is still acceptable as analyzed in Section 3.2.3.2. Thus, in the choice of the network implementation, a comprehensive consideration of the balance between the performance and the device cost is needed.

3.4 Concluding Remarks

This chapter presents the results of the implementation of MKG with two types of devices – OPS couplers and optical switches. It has been shown that this topology can achieve pretty good performance with low cost and complexity. In order to obtain more precise results, some of the assumptions which have simplified the constraints can be relaxed. The modeled results are not as good as POPS and SKG network. A possible solution is to adopt the diluted architecture to improve the performance by employing more components, which is a very common technique in the optical network design.

Chapter 4 A Multistage Optical Switch Based on Kautz Graph

As mentioned in Chapter 1, switches capable of dealing with the signals in optical domain are essential components in today's high-speed network. Due to the physical limits of optical devices, a large optical switch comprising small switching elements is a viable approach to accommodate today's material technology. This chapter studies a multistage optical switch architecture based on the MKG. First, a multistage optical switch model that divides the inputs into groups and uses the routing string as in the MKG for controlling is proposed to achieve large capacity. The switching scheme of this switch is then given, and illustrated by a 3-stage architecture. Performance is obtained to demonstrate that this architecture is applicable. Finally, a practical solution by implementing the switch with optical star couplers and other components is discussed.

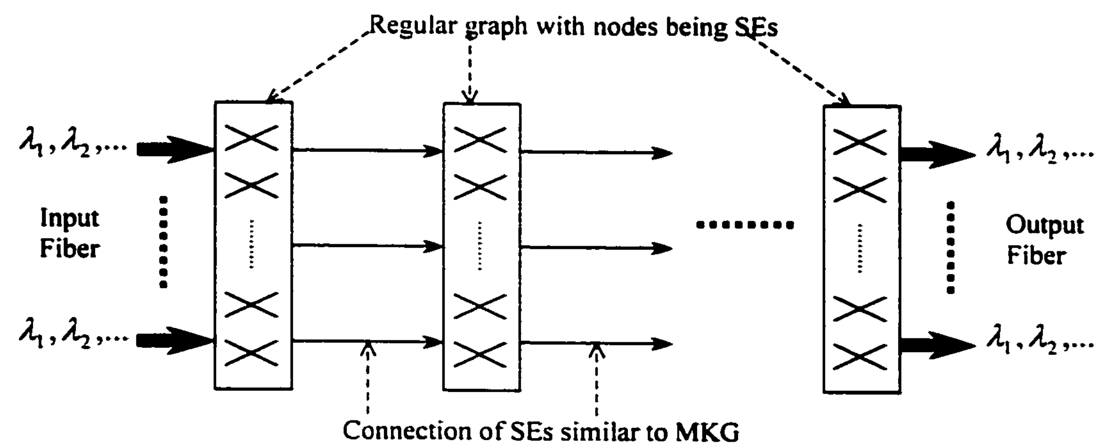


Figure 4.1 General Layout of A Multistage Optical Switch

4.1 A Multistage Switch Architecture

4.1.1 Principle

Figure 4.1 shows the general layout of a multistage optical switch we propose. Here each input fiber contains mutiple wavelength, $\lambda_1, \lambda_2, \ldots$, which is splitted into and passes through a series of switching blocks. Each swiching block is equal to a switching stage. In a normal multistage switch, non-blocking requirement at each stage usually generates high complexity in large-scale optical implementation. In this thesis, we modify the MKG topology to bypass this requirement and to switch messages by properly organizing the pathes to keep the performance at an acceptable level. Let each stage be a KG and each node of the KG be a small switching element so that several such graphs can be interconnected to form the MKG switch. Note that the switching path would be the same as the routing path from the first layer to the last layer in the MKG network. However, there can be hundreds of wavelengths in the WDM networks. The degree of the KG is relatively low. How to distribute the input signals into the first stage would affect the operation and performance. Therefore, the switching path should be constructed carefully to optimize the performance.

4.1.2 The MKG Switch

Here we present a $(k+1)$-stage switch architecture based on the Kautz graph $KG(d,k)$. Each stage is a $KG(d,k)$, the nodes of which are $d \times d$ switching elements (SEs). The connection relationship of two adjacent stages complies with the connection of the MKG graph, i.e., there is no connection between two nodes/SEs in the same stage and each node/SE has connections to its output nodes in the next stage. If there is a link from node x to node y in the $KG(d,k)$, for example, then there is a link from node x in Stage i to node y in Stage $i+1$ in the switch, where $i = 1, ..., k$.

The input ports of the switch consist of d^{k-1} fibers. Each fiber has $d*(d+1)$ wavelengths, $\lambda_1, \lambda_2, ..., \lambda_{d(d+1)}$, which are divided into $(d+1)$ groups with d wavelengths in a group. Each group corresponds to a node/SE in the Kautz graph. Thus total nodes/SEs add up to $d^{k-1} * (d+1)$, the size of $KG(d,k)$. The input buffers are used in the switch, and each input link has a distinct buffer for itself. The buffer could be infinite or finite, but strictly FIFO.

The incoming optical signals of each fiber are demultiplexed into the SEs in the first switching stage sequentially from top to bottom, starting from λ_1 to $\lambda_{d(d+1)}$. So the wavelengths distributed to each SE in the first stage are fixed. For each individual input signal, it can be identified by the fiber number, the wavelength or the SE label in the $KG(d,k)$. In the following illustration, these three identifiers will be used alternatively depending on the purpose and expression of the clarification. No matter which one is used, the signal is discussed in the link/wavelength level.

At the $(k+1)$ st stage, the output signals are sent into tunable wavelength converters (TWCs). d of these TWCs form a node and these nodes form another $KG(d,k)$. The connectivity pattern from the $(k+1)$st switching stage to the TWC stage is the same as that between two switching stages, and the signals from $(d+1)$ TWC nodes are multiplexed into one output fiber. Thus each output fiber still has $d*(d+1)$ distinctive wavelengths and the number of output fibers is d^{k-1}, the same as the input fibers.

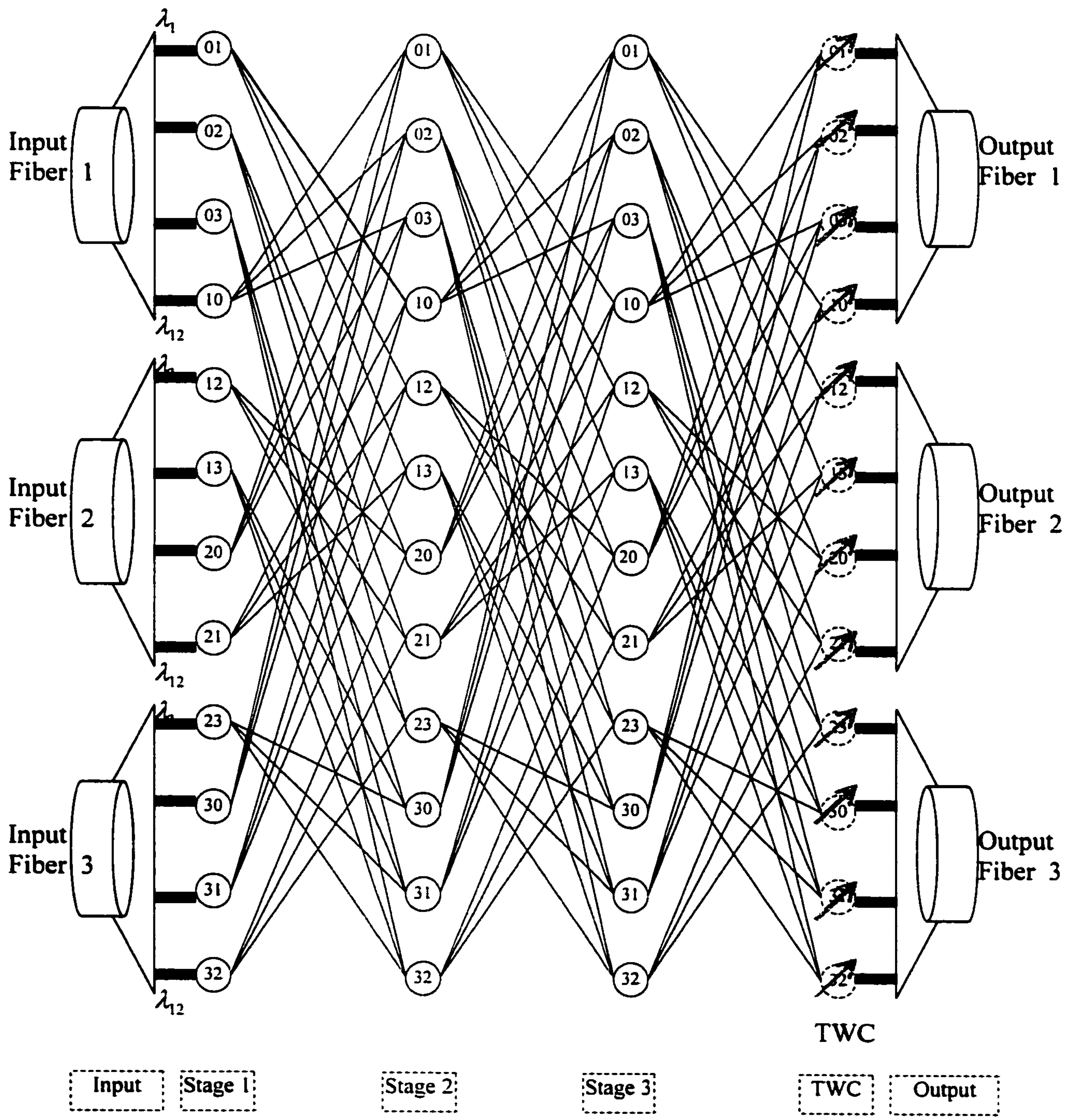


Figure 4.2 A 3-stage Optical Switch Based on KG(3,2)

Figure 4.2 is a 3×3 port (fiber) switch based on the $KG(3,2)$, with 12 wavelengths in each fiber port. This switch has 3 stages of 3×3 switching elements. Each stage has tweleve 3×3 SEs. The connection pattern between adjacent layers is the same as the $MKG(3,3,2)$. After the 3rd stage, the signals are passed through TWCs before they are multiplexed into the output fibers. The circle in each switching stage is a 3×3 switching element, and each circle in the TWC stage is just 3 TWCs in parallel. The bold lines in the figure are actually 3 links in parallel.

4.1.3 Switching Scheme

The task of switching is to route a set of packets from the source ports (in the first stage) to its destination ports (in the last stage), so that at most one packet crosses a link of the network at any time step, and as many packets as possible will reach the destination. In order to find the best paths for the packets, the control unit will decide the link assignment among the input packets. The purposes of the decision procedure are:

1) Wavelength conversion to reduce contention.

2) Buffering the packets with the same destination when no wavelength is available.

3) Discarding packets with contention when no link, buffer nor wavelength is available.

In the above $(k+1)$-stage switch architecture, the switching path can be described as a string of $(2*k+1)$ letters (e.g. Section 2.2.2). The switching path is the same as the routing path in $MKG(k+1,d,k)$ from 1st layer to (k+1)st layer with hops exactly equal to k+1 (Ref. Section 2.2). The first k letters is the label of the input SE node. The last k letter is the label of the output TWC node. Although the input node is determined by the input wavelength and fiber, the output node can be any of the $d+1$ TWC nodes inside the output fiber. Thus we have some flexibility in building the switching path. Consider the 3-stage switch example shown in Figure 4.2, if a packet has wavelength λ_1 in Fiber 1, it will be demultiplexed into SE 01 in the first stage. If the output is Fiber 2, then it may arrive at node 12, 13, 20 or 21 in the TWC stage of Fiber 2, so the switching path could be 01x12, 01x13, 01x20 or 01x21, with x being any letter different from its left and right letters. The control unit of the switch is able to select any available path to switch the packet to any TWC of the same fiber.

Some situations with high contention probability should be considered here. When the nodes ending with the same letter have packets destined to the same output fiber, the probability of contention is higher. For example, consider that each of input node 01 in input fiber 1, node 21 in input fiber 2 and node 31 in input fiber 3 has a packet to be sent to output fiber 3. Since the output fiber 3 has TWC nodes 23, 30, 31 and 32, the switching path could only be x1023, x1323, x103x, or x123x (x could be any possible letter). Therefore, the nodes in the second stage could only be 10, 12 or 13. These three nodes have 9 input links and 9 output links. Only if there is no other packet occupying any of these 9 input/output links, could every packet find its switching path. Therefore, these packets should be assigned highest priority to choose the switching pathes during the path setup procedure in order to reduce contention and packet loss. The following steps illustrate how to switch the packets.

1) Starting form input fiber 1, find all current input packets, i.e., the packets in the header of buffers.

2) Divide the input packets into d^{k-1} groups, numbered from 1 to d^{k-1}, and each group consists of the packets destinating to one output fiber.

3) Starting from Group 1 (input with destination being output fiber 1), divide the incoming packets in this group into $d+1$ arrays, from 0 to d. Each array contains the input nodes ending with the same letter, i.e., the nodes in Array 0 having the label x0, the nodes in Array 1 having the label x1, and so on. The nodes in each array are linked according to the packet's wavelength, the smallest wavelength in the header and then increasing the wavelength sequentially.

4) Repeat Step 1 to 3 to make the division to all input fibers.

5) In Group 1, starting from the largest array, find an available path for the packet in the header of the array. Choose one node in the output fiber if there is an idle path and there is no wavelength conflict in the intermediate nodes. Then assign this path and wavelength to the packet and take the packet out of the array. Otherwise, keep the packet in the buffer of the input link. Then find the path for the second packet, and so on until the array is empty.

6) Repeat Step 5 to do the switching for all output fibers.

7) If the buffer size is finite and the buffer has reached its capacity, the packet in the buffer header without available switching path will be dropped when a new packet arrives at the same input link.

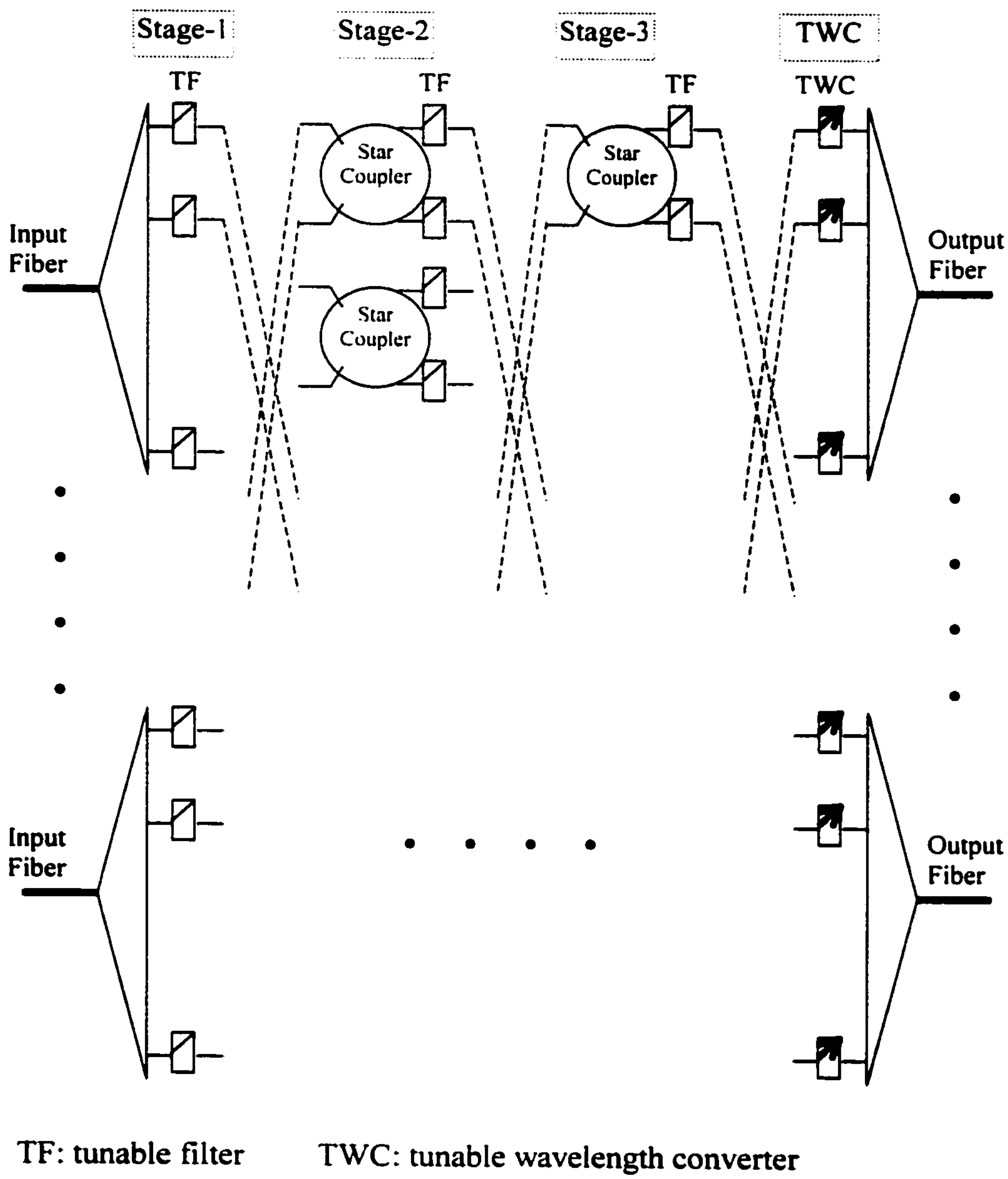


Figure 4.3 Modular Architecture of the 3-stage Switch

4.1.4 An Implementation Example

Here we propose a physical modular implementation corresponding to the logical topology of the 3-stage optical switch in Figure 4.2. As shown in Figure 4.3, the first stage of the switch consists of optical splitters and tunable filters (TFs). Each fiber is connected to a $1\times d(d+1)$ splitter. Each output link of the splitter is followed by a tunable filter. Each TF will select one signal only. The splitters and TFs also finish the switching funtion of SEs in the first stage of the Figure 4.2 architecture. The second and third stage are composed of $d\times d$ OPS couplers followed by TFs, and each "Star Coupler + TFs" module corresponds to a SE in the logical architecture in Figure 4.2. The connection relationship between two stages is as described in Section 4.1.2. The output of the third stage then goes to d^{k-1} optical combiners. Before the signals are combined to each fiber, they must pass through a TWC array to be converted to appropriate wavelengths. The wavelength conversion can remove the possible wavelength conflict in the combiner. All of the actions of these modules are controlled by an electronic controlling unit, which takes the necessary information from a small part of the original signals before they enter the core switching fabric.

4.2 Performance Evaluation

Using the above switch model and assumptions, two switches based on $KG(4,2)$ and $KG(9,2)$ are simulated with Microsoft Visual C++. The program runs on an Intel Pentium PC with simulation time around 1,000,000 time steps. Detailed numerical values of the modeling data are also tabulated in Appendix A. To check the valididility of the results, the 95% confidence intervals of some simulations are shown in Appendix B, showing that the results are acceptable.

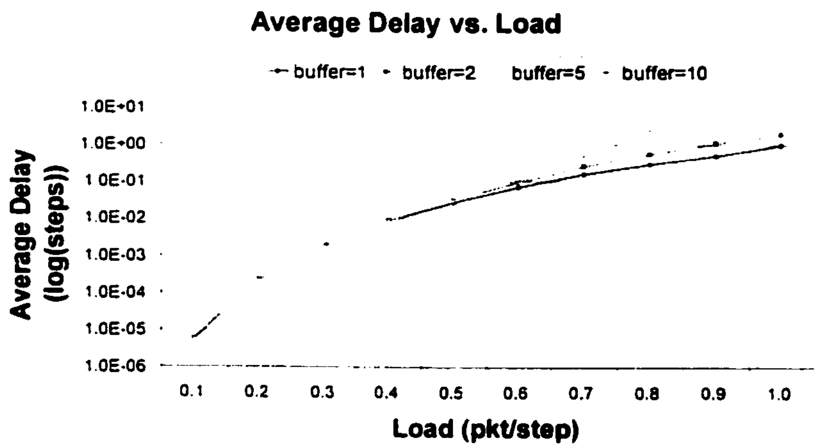


Figure 4.4 Average Delay of (4,2) Switch

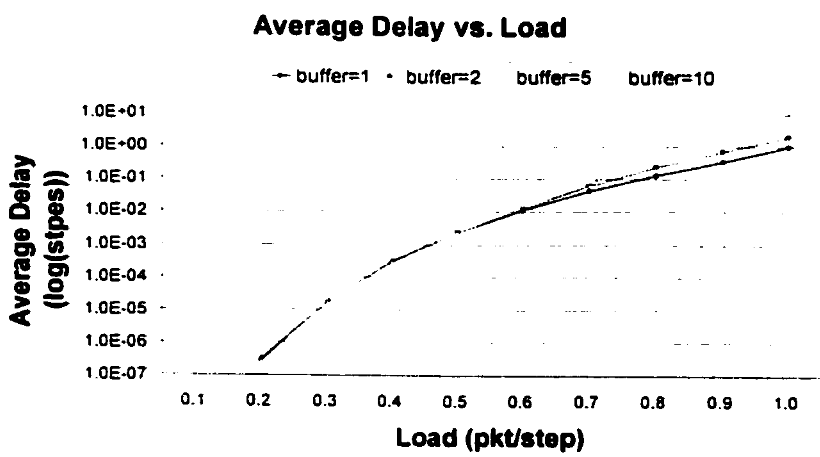


Figure 4.5 Average Delay of (9,2) Switch

Table 4.1 Maximum Delay of (4,2) Switch

Load	Buffer=1	2	5	10
0.1	1	1	1	1
0.2	2	1	2	2
0.3	3	3	3	3
0.4	3	4	4	4
0.5	6	6	7	7
0.6	6	9	14	13
0.7	7	9	15	26
0.8	6	10	15	24
0.9	6	8	12	19
1.0	1	2	5	10

Table 4.2 Maximum Delay of (9,2) Switch

Load	Buffer=1	2	5	10
0.1	0	0	0	0
0.2	1	1	1	1
0.3	1	1	1	1
0.4	2	2	2	2
0.5	4	3	3	4
0.6	5	5	6	6
0.7	6	6	11	15
0.8	5	8	14	20
0.9	5	7	12	19
1.0	1	2	5	10

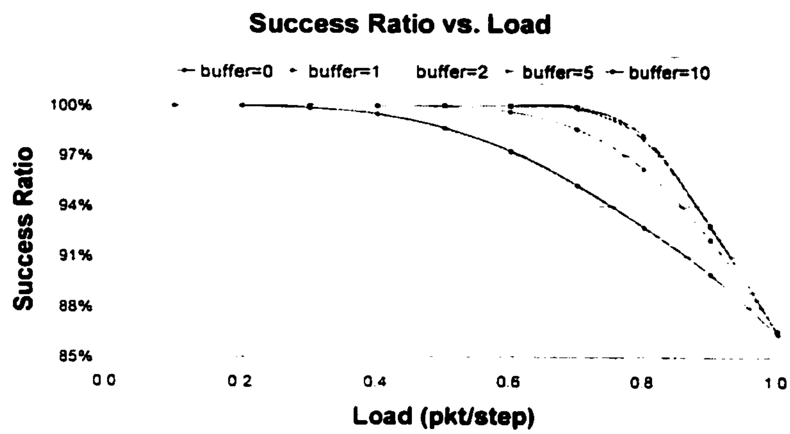
For a switch, the most important performance criteria are high capacity, high throughput, low latency, low packet loss and low material cost. The simulation results of the proposed 3-stage switches based on $KG(4,2)$ and $KG(9,2)$ are given here. The packet arrival complies with the Bernoulli model and each input link has the same load, i.e., the packet arrival rate. Packet arrival of each input is independent.

4.2.1 Delay Performance

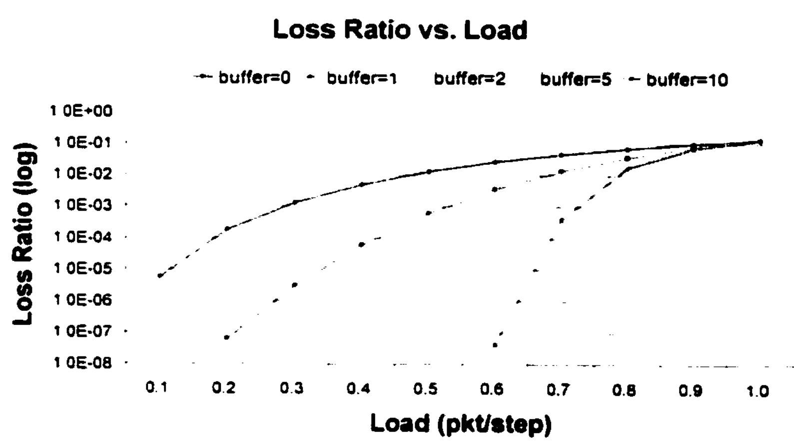
When buffers are used in the input links, the packet may spend some time slots in buffering to wait for available path. This buffering time becomes the delay. Figure 4.4 is the average delay versus load of the 3-stage switch based on $KG(4,2)$. Here the load is defined as the probability that a new packet will arrive at an input link at one time slot. With an increasing load, the delay will increase dramatically, and with an increasing buffer size, the delay time will also be longer. This is to be expected since a large buffer provides more opportunity for the packets to stay there, which will result in longer delay. If a small buffer is used, the buffered packet can more easily be dropped when a new packet arrives. The shorter delay also implies a higher packet loss ratio which will be shown below. For the larger switch based on $KG(9,2)$, the delay is shorter than the $KG(4,2)$ switch as in Figure 4.5. Since there are more output links to the SEs, the packets will have more chances to find a route, reducing delay. Tables 4.1 and 4.2 show the maximum delay data. In the low delay range, the packets can be sent out earlier. When the load is around 70 and 80 percent, the delay may be maximum. When the load is very high, say 90 or 100 percent, the packets can easily be dropped, which also reduces the maximum delay. All of the improvement of delay is based on the price of packet loss.

4.2.2 Packet Loss Performance

Since packets will be dropped when the buffer is full, the packet success ratio, defined as the number of packets arriving output ports successfully out of total input packets, and the packet loss ratio, defined as the number of packets being dropped out of total input packets, are also important performance criteria to evaluate a switch. Figure 4.6a shows

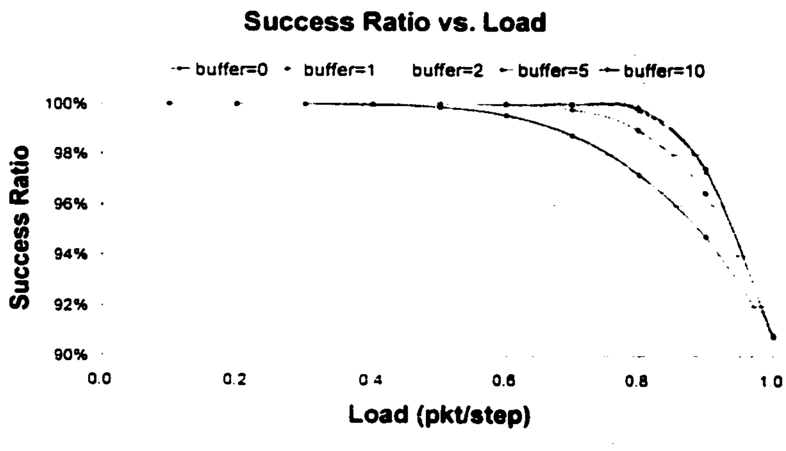


(a)

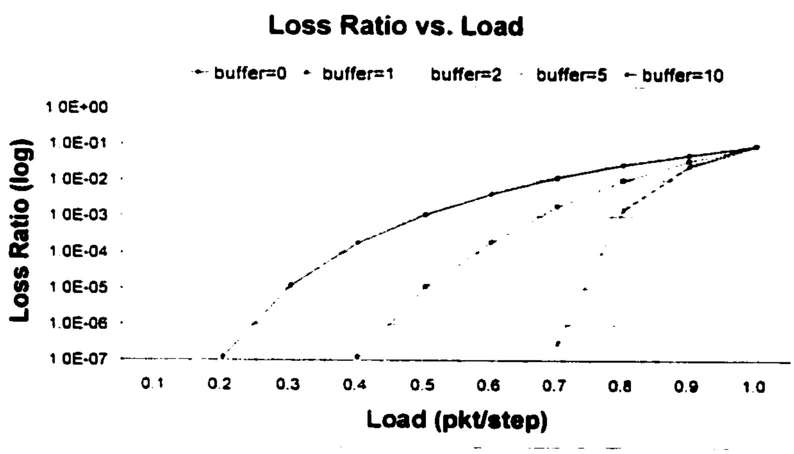


(b)

Figure 4.6 Performance of (4,2) Switch
(a) Success Ratio (b) Loss Ratio



(a)



(b)

Figure 4.7 Performance of (9,2) Switch
(a) Success Ratio (b) Loss Ratio

the packet success probability versus load, and Figure 4.6b shows packet loss probability versus load (packet arrival rate), each with different buffer sizes changing from 10 to 0. From these figures, it can be seen that when larger buffers are used, the packet loss ratio can be limited to a very low level since the packets have more chances to find a path. As for the switch based on $KG(9,2)$, its performance in Figure 4.7 is much better than the (4,2) switch. The explanation can be the same as that of the delay performance in the above section.

4.2.3 Complexity Performance

The switching path may be not possible due to no available link, thus this $(k+1)$-stage switch is not strictly non-blocking. The advantage of the proposed switch is a fewer number of cross-points when compared to some known multi-stage switches. In the switch there are $(k+1)$ stages, each of which has $d^{k-1}*(d+1)$ SEs. Each SE has $d*d$ cross-points. Thus, the total number of the cross-points is $(k+1)*d^{k-1}*(d+1)*d*d = (k+1)*d^{k+1}*(d+1)$. In a 3-stage $N \times N(n,k)$ Clos network, the cross-point is determined by $2kN + k(N/n)^2$. In a $N \times N (N = 2^n)$ Banyan network, the cross-point is equal to $n \times 2^{n+1} = 2N \log_2^N$.

To the (3,2) switch with 36 ports shown in Figure 4.2, the total cross-point is 324. If we build a single stage 36×36 switch, the cross-points will be 1296. A 3-stage Clos network with $n=3, k=3$ has 648 cross-points. A similar Banyan network will have around 372 cross-points. It is obvious that our proposed switch is superior to the other architectures in terms of the number of cross-points, which is important in the optical fabrication.

4.3 Concluding Remarks

In this chapter, a $(k+1)$ -stage optical packet switch based on the Kautz graph $KG(d,k)$ has been analyzed. This switch utilizes small $d \times d$ switching elements as the basic components to achieve large-scale capacity. Although it is not strictly non-blocking architecture, the packet loss ratio and average delay are at an acceptable level with very low cost. The advantage of the proposed switch topology lies in the simplicity of the switching path and the reduction of component cost. It is capable of sufficiently utilizing the high-speed properties of optical components and the possibility of employing small switching elements to solve the difficulty of constructing large all-optical switch fabric.

As one final remark, our MKG switch can also be used in the MKG network. However, this may not be appropriate because the intention of the MKG network architecture is to utilize simple and small-size components. While MKG switch intends to achieve large port amount switching which can be applied to more irregular networks.

Chapter 5 Design Guideline and Issues

This chapter presents the experience drawn from the design and analysis of our study. Some important issues related to the design and implementation are analyzed. The recommended choice of architecture design is discussed. Although these considerations may not answer all practical questions, they would provide an abstract solution.

5.1 Network Design

5.1.1 Network Design Issues

Since the condition of each individual network is distinct, it is most efficient to abstract the requirements of real applications, and then construct a general logical topology with the appropriate mathematical methods. The topology should be feasible under normal traffic. When the size of the network has changed, the topology should have the enough flexibility to accommodate the changing.

A nice network topology also needs to be easy to implement with available components. On the other hand, the implementation also includes the cost consideration, which determines if the architecture is really cost-effective. Therefore, we need to combine currently feasible components with long-term development of devices. The KG in our work is one such example of extending to a more general architecture and implemented with both OPS couplers and optical switches in this thesis.

5.1.2 MKG Network Design Guideline

In our MKG network design, the first step is to choose appropriate graph parameters to match the network size requirement. By making some simple queueing analysis as in Section 2.3, the best parameter combination can be chosen. For implementation using OPS couplers, we would recommend a traffic load less than 30% in order to have a satisfactory delay performance. For the implementation with optical switches, the network is capable of supporting higher load with good delay performance while the device cost is higher. Therefore, the implementation component is decided by the traffic and cost consideraton. Using the simulation results in Seciton 3.1.5, we can determine which contention resolusion method would be adopted. For example, if we hope more packets would be delivered as soon as possible, the "distance" method may be used. If we want more packets would be sent to the destination nodes after a certain sequence steps, say 13 time slots, the "waitingtime" method would be the first choice.

5.2 Switch Design

5.2.1 Switch Design Issues

The complexity, scalability, reliability and throughput are the major aspects in the switch design. With the fast-increasing application, the switch needs to support more and more ports. When the environment changes, the switch needs to scale with the requirement. Due to the limits of buffering, amplification, wavelength conversion and filtering, the implementation cost and performance are directly related to the components adopted. Furthermore, once part of the architecture fails, the effect to the traffic should be controlled in a small range. Based on these considerations, we choose the MKG as the basic topology and the passive star coupler as the basic switching element in this thesis.

5.2.2 MKG Switch Design Guideline

Depending on the traffic and size requirement, we shall choose appropriate parameters of the KG to achieve necessary ports first. To reduce the complexity of the device and controlling, we recommend 3 stages switch based on $KG(d,2)$. Thus the total input/output ports can reach to $(d+1)*d^2$. Since the switching string is 5 letters, the central one could have more than one choice, which gives some flexibility to choose the path. Thus the switch is more robust under the changing traffic or the failure of some parts. The second step is to choose the buffer. Referring to the delay and packet loss analysis in Section 4.2, we can determine if the input buffer is needed and how large the buffer should be. For example, if the (9,2) switch is used and the input load is 0.7. Then if we want to keep the packet loss ratio in the order of 1.0E-3, the buffer size should be at least 5. As for the choice of the switching elements, we choose the viable passive star couplers, which increase the availability of this type of switch. By increasing the degree of SEs, the throughput and the switch size would increase accodingly.

Chapter 6 Conclusions

In this thesis, the multilayer Kautz graph has been proposed to build an optical network topology and an optical packet switch. This multilayer architecture has been shown to have some advantage in an optical network. We have compared MKG with POPS and SKG architecture. The MKG can support very large sized networks. When all these topologies are implemented with OPS couplers, the MKG needs fewer devices. We have also discussed three methods for contention resolution. In a small size of network, these methods do not have significant difference. In a large network, they have different advantages in the short delay range and long delay range respectively. When the MKG is implemented with optical switches in each node, the performance has very significant improvement while the device cost is also higher.

This multilayer concept has been extended to the design of an optical packet switch. As for the proposed multistage switch architecture, it is not a strictly non-blocking switching fabric. By properly assigning the ports and wavelengths to the input and output fibers, we show that this switch can provide superior advantages in terms of the average delay, the loss ratio and the number of cross-points. Even when no buffer is used, this switch can still perform well within a wide range of traffic load.

Our attemps to implement our MKG network and switch in this thesis expose the difficulties and dilemma we face nowadays. When considering the properties of the optical transmission and processing, some critical issues related to the implementation of the proposed network and switch topology have also been examined thoroughly. These issues are necessary for overcoming the difficulty in real operation and implementation. With available technologies and components, some possible building blocks for our proposed topologies are given.

For further work, some improvements can be taken into account for our topologies. For example, the three scheduling schemes can be combined to obtain better throughput of the MKG. More general dynamic routing can be modeled to produce a more flexible network. Since the IP-centric traffic will dominate the future network application, its traffic pattern could also be considered for the network analysis. As for the switch design, a different partition of the switch ports may be done. Furthermore, we can simplify the control scheme in deciding the switching pathes to achieve faster packet switching. The traffic properties may also be combined with the switch fabric analysis. Finally, the properties of optical devices have more interesting impacts on the whole system, which motivates more research on the physical blocks of each part of the network.

References:

[AlFa98] D. Y. Al-Salameh, M. T. Fatehi, et al., "Optical Networking", Bell Labs Technical Journal, Jan-Mar 1998, pp.39-61.

[AnJa00] W. T. Anderson, J. Jackel, et al., "The MONET project-a final report", IEEE/OSA J. of Lightwave Tech., vol.18, no.12, Dec. 2000, pp.1988-2009.

[AyHa94] F. Ayadi, J. F. Hayes and M, Kavehrad, "WDM Cross-Connected Star Topology for the Bilayered ShuffleNet", IEEE/OSA J. of Lightwave Tech., vol.12, no.9, Sep. 1994, pp.1668-1678.

[BaDr01] A. Banerjee, J. Drake, J. P. Lang, et al., "Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements", IEEE Commun. Mag., vol.39, no.1, Jan. 2001, pp.144-150.

[BaHu93] R. A. Barry and P. A. Humblet, "Latin Routers, Design and Implementation", IEEE/OSA J. of Lightwave Tech., vol.11, no.5/6, May/Jun. 1993, pp.891-899.

[BaJa99] S. Banerjee, V. Jain and S. Shan, "Regular Multihop Logical Topologies for Lightwave Networks", IEEE Commun. Surveys, First Quat. 1999, pp.2-18.

[BlPr94] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer, "Photonic Packet Swtiches: Architectures and Experimental Implementations", Proceedings of IEEE, vol.82, no.11, Nov. 1994, pp.1650-1667.

[CaBo96] Feng Cao and Al Borchers, "Optimal Transmission Schedules for Embeddings of the de Buuijn Graphs in an Optical Passive Network", 5th IC3N, Oct. 1996, pp.220-225.

[ChCh97] K. S. Chan, S.Chan and K. L. Yeung, "Wide-sense Nonblocking Multicast ATM Switches", Proceedings, Globecom 97, Phoenix, November 1997, pp. 526-530.

[ChFu93] Imrich Chlamtac and Andrea Fumagalli, "An Optical Switch Architecture for Manhattan Networks", IEEE JSAC, vol.11, no.4, May 1993, pp.550-559.

[ChLe02] P. B. Chu, Shi-Sheng Lee and Sangtae Park, "MEMS: the path to large optical crossconnects", IEEE Commun. Mag., vol.40, no.3, Mar 2002, pp.80-87.

[ChLe96] D. M. Chiarulli, S. P. Levitan, R. P. Melhem, et al.., "Partitioned Optical Passive Star (POPS) Multiprocessor Interconnection Networks with Distributed Control", IEEE/OSA J. of Lightwave Tech., vol.14, no.7, Jul. 1996, pp.1601-1612.

[CoFe00] D. Coudert, A. Ferreira and X. Munoz, "A Multihop Multi-OPS Optical Interconnection Network", IEEE/OSA J. of Lightwave Tech., vol.18, no.12, Dec. 2000, pp.2076-2085.

[CoFe98] D. Coudert, A. Ferreira and S. Perennes, "Multiprocessor architectures using multi-hop multi-OPS lightwave networks and distributed control", ", IEEE IPPS 1998, IEEE Press, 1998, pp.151-155.

[CoRa00] F. C. G. Corazza and C. Raffaelli, "Design of a WDM Optical Packet Switch for IP Traffic", GLOBECOM'2000.

[DoFa02] De Dobbelaere, P. Falta, K. Gloeckner, S. Patra, S., "Digital MEMS for optical switching", IEEE Commun. Mag., vol.40, no.3, Mar 2002, pp.88-95.

[Dowd92] P. W. Dowd, "Wavelength division multiple access channel hypercube processor interconnection", IEEE Transactions on Computers, vol.41, no.10, Oct 1992, pp.1223-1241.

[ElMo00a] Jaafar M.H. Elmirghani and Hussein T. Mouftah, "Technologies and Architectures for Scalable Dynamic Dense WDM Networks", IEEE Commun. Mag., vol.38, no.2, Feb. 2000, pp.58-66.

[ElMo00b] Jaafar M. H. Elmirghani and Hussein T. Mouftah, "All-Optical Wavelength Conversion: technologies and Applications in DWDM Networks", IEEE Commun. Mag., vol.38, no.3, Mar. 2000, pp.86-92.

[FeYa94] Zhi Feng and O. Yang, "Routing Algorithms in the Bidirectional de Bruijn Graph Metropolitan Area Networks", IEEE Military Commun. Conf., New Jersey, Oct 1994.

[FeYa97] Zhi Feng and O. Yang, "Designing an Optical BdBG Network", J. of Optical Communications, 18.5, 1997, pp.181-186.

[GaMa98] D. Gardner, P. Marchand, et al., "Photorefractive beamsplitter for free space optical interconnection systems", OSA Appl. Opt., vol.37, no.26, Sept 1998, pp.6178-6181.

[GhDi00] Nasir Ghani, Sudhir Dixit and Ti-Shiang Wang, "On IP-over-WDM Integration", IEEE Commun. Mag., vol.38, no.3, Mar. 2000, pp.72-84.

[Guiz97] M. Guizani, "ATM Architectures Using Optical Technology: An Overview of Switching, Buffering and Multiplexing", Intl. J. of Network Management, vol.7, 1997, pp.198-220.

[GuRe98] C. Guillemot, M. Renaud, P. Gambini, et al., "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach", IEEE/OSA J. of Lightwave Tech., vol.16, no.12, Dec. 1998, pp.2117-2134.

[Haas93] Zygmunt Haas, "The 'Staggering Switch": An Electronically Controlled Optical Packet Switch", IEEE/OSA J. of Lightwave Tech., vol.11, no.5/6, May/Jun. 1993, pp.925-936.

[HaKr96] E. Hall, J. Kravitz, et al., "The Rainbow-II gigabit optical network", IEEE JSAC, vol.14, no.6, June 1996, pp.814-823.

[HaSa00] K. Habara, H. Sanjo, H. Nishizawa, et al., "Large-Capacity Photonic Packet Switch Prototype Using Wavelength Routing Techniques", IEICE Trans. Commun., vol.E83-B, no.10, Oct. 2000, pp.2304-2311.

[HeSt96] B. R. Hemenway, M. L. Stevens, et al., " Performance of a wavelength-routed all-optical network at over 100 Gigabit/s", Advanced Applications of Lasers in Materials Processing/Broadband Optical Networks/Smart Pixels/Optical MEMs and Their Applications, IEEE/LEOS 1996 Summer Topical Meetings, 1996, pp. 10-13.

[Heyw00] Peter Heywood, "Optical Switching Fabric", www.Lightreading.com.

[HlKa88] M. G. Hluchyj and M. J. Karol, "ShuffleNet: An application on generalized perfect shuffles to multihop lightwave networks", IEEE INFOCOM'88, Mar 1988, pp.379-390.

[HsWe97] D. F. Hsu and D. S. L. Wei, "Efficient Routing and Sorting Schemes for de Bruijn Networks", IEEE Trans. on Parallel and Distributed Systems, vol.8, no.11, Nov. 1997, pp.1157-1170.

[HuCh98] D. K. Hunter et al., "Buffering in Optical Packet Switching", IEEE/OSA J. of Lightwave Tech., vol.16, no.12, Dec. 1998, pp.2087-2094.

[HuCo98] D. K. Hunter, W. D. Cornwell, et al., "SLOB: a Switch with Large Optical Buffers for Packet Switching", IEEE/OSA J. Lightwave Tech., vol.16, no10, Oct. 1998, pp.1725-36.

[HuNi99] D. K. Hunter, M. H. M. Nizam, M. C. Chia, et al.., "WASPNET: A Wavelength Switched Packet Network", IEEE Commun. Mag., vol.37, no.3, Mar. 2001, pp.120-129.

[IaSa96] E. Iannone, R. Sabella, "Optical Path Technologies: a Comparison among Different Cross-Connect Architectures", IEEE/OSA J. of Lightwave Tech., vol. 14, no. 10, pp. 2184 - 2196, October 1996.

[ITU01] ITU G.774, "SDH information model for the network element view", Feb 2001.

[Inte00] Internet, http://optcom.korea.ac.kr/optinet/content01.htm

[Inte01] Internet, http://www.alcatel.com/doctypes/articlepaperlibrary/pdf/ATR2001 Q3 /gb/06ErmaGB. pdf

[Inte02] Internet, http://www.lightreading.com/document.asp?site=lightreading&doc_id=2254&page_number=1

[JuBo96] J. P. Jue, M. S. Borella, and B. Mukherjee, "Performance analysis of the Rainbow WDM optical network prototype", IEEE JSAC, vol.14, no.6, June 1996, pp.945-951.

[Kaut68] W.H. Kautz, "Bounds on directed (d,k) graphs." Theory of cellular logic networks and machines. AFCRL-68-0668, SRI Project 7258, Final report, pp.20-28, 1968.

[Klei76] Leonard Kleinrock, "Queueing Systems. Volume 1/2". 1976.

[KoHa01] O. Komolafe, D. Harle, and D. Cotter, "Ultrafast optical packet switching over arbitrary physical topologies using the Manhattan Street Network", ICC'2001, vol.2, pp.489-493.

[KrSi01a] R. M. Krishnaswamy, and K. N. Sivarajan, "Algorithms for routing and wavelength assignment based on solutions of LP-relaxations", IEEE Communications Letters, vol.5, no.10, Oct. 2001, pp.435-437.

[KrSi01a] R. M. Kishnaswamy and K. N. Sivarajan, "Design of Logical Topologies: A Linear Formulation for Wavelength-Routed Optical Networks with No Wavelength Changers", IEEE Trans. Networking, vol.9, no.2, Apr.2001, pp.186-198.

[Li99] T. Li, "MPLS and the evolving Internet architecture", IEEE Commun. Mag., vol.37, no.12, Dec 1999, pp.38-41.

[MoMu98] G. Mohan, C. S. R. Murthy and R. Vijayashankar, "Permutation routing in wavelength-routed wrapped-around shuffle networks using fewer wavelengths", Computer Networks and ISDN Systems 30, 1998.

[Mukh92a] Biswanath Mukherjee, "WDM-Based Local Lightwave Networks Part I: Single-Hop Systems", IEEE Network, May. 1992, pp.12-26.

[Mukh92b] Biswanath Mukherjee, "WDM-Based Local Lightwave Networks Part I: Multi-Hop Systems", IEEE Network, Jun. 1992, pp.

[NaOk99] K. Nakai, E. Oki and N. Yamanaka, "Scalable 3-Stage ATM Switch Architecture Using Optical WDM Grouped Links Based on Dynamic Bandwidth Sharing", IEICE Trans. Electron., vol.E82-C, no.2, Feb. 1999, pp.213-218.

[NaUe99] S. Nakamura, Y. Ueno, and K. Tajima, "Femtosecond Operation of a Polarization-Discriminating Symmetric Mach-Zehnder All-Optical Switch and Improvement in Its High-Repetition Operation", IEICE Trans. Electron., vol.E-82-C, no.2, Feb 1999, pp.327-334.

[OkOk00] H. Okayama, Y. Okabe, T. Arai, et al., " Two-Module Stage Optical Switch Network", IEEE/OSA J. of Lightwave Tech., vol.18, no.4, Apr. 2000, pp.469-476.

[PaMa00] A. Patttavina, M. Martinelli, G. Maier and P. Boffi, "Techniques and Technologies Towards All-Optical Switching", Optical Networks, vol.1, no.2, Apr. 2000, pp.75-93.

[PaNe87] Krishnan Padmanabhan and Arun N. Netravali, "Dilated Networks for Photonic Switching", IEEE Trans. Commun., vol.35, no.12, Dec. 1987, pp.1357-1365.

[PaSe95] Geetha Panchapakesan and Abhijit Sengupta, "On Multihop Optical Network Topology Using Kautz Digraphs", IEEE INFOCOM'1995, pp.675-682.

[PaSe99] Geetha Panchapakesan and Abhijit Sengupta, "On a lightwave network topology using Kautz digraphs", IEEE Transactions on Computers, vol.48, no.10, Oct. 1999, pp.1131-1137.

[Qiao00] Chunming Qiao, "Labeled Optical Burst Switching for IP-over-WDM Integration", IEEE Commun. Mag., vol.38, no.9, Sep. 2000, pp.104-114.

[RaPe00] B. Rajagophlan, D. Pendarakis, et al., "IP over optical networks: architectural aspects", IEEE Commun. Mag., vol.38, no.9, Sep. 2000, pp.82-92

[RaRa97] C. P. Ravikumar, T. Rai, and V. Verma, "Kautz graphs as attractive logical topologies in Multihop Lightwave networks", Computer Communications, 20, 1997, pp.1259-1270.

[RaRi95] R. Ramaswami. and K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks". IEEE/ACM Transactions on Networking, vol.3, no.5, Oct. 1995, pp.489-500.

[RaSi96] R. Ramaswami and K. N. Sivarajan, "Design of Logical Topologies for Wavelength-Routed Optical Networks", IEEE JSAC, vol.14, no.5, Jun, 1996, pp.840-851.

[RiZe01] C. Riziotis. and M. N. Zervas, "Design considerations in optical add/drop multiplexers based on grating-assisted null couplers", IEEE/OSA J. of Lightwave Tech., vol.19, no.1, Jan 2001, pp.92-104.

[Rose92] Christopher Rose, "Mean Internodal Distance in Regular and Random Multihop Networks". IEEE Trans. Commun., vol.40, no.8, Aug. 1992, pp.1310-1318.

[Ryan98] J.P. Ryan, "WDM: North American deployment trends", IEEE Commun. Mag., vol.36, no.2, Feb. 1998, pp.40-44.

[SeBa01] A. Sen, S. Bandyopadhyay and B. P. Sinha, "A New Architecture and A New Metric for Lightwave Networks", IEEE/OSA J. of Lightwave Tech., vol.19, no.7, Jul. 2001, pp.913-925.

[ShCh93] J. Sharony, K. cheung, and T. E. Stern, "The Wavelength Dilation Concept in Lightwave Networks-Implementation and System Considerations", IEEE/OSA J. of Lightwave Tech., vol.11, no.5/6, May/Jun. 1993, pp.900-907.

[ShWh00] K. V. Shrikhande, I. M. White, et al., "HORNET: a packet-over-WDM multiple access metropolitan area ring network", IEEE JSAC, vol.18, no.10, Oct. 2000, pp.2004-2016.

[SiRa94] K. N. Sivarajan and R. Ramaswami, "Lightwave Networks Based on de Bruijn Graphs". IEEE Trans. Networking, vol.2 no.1, Feb. 1994, pp.70-79.

[Siva92] Kumar N. Sivarajan, "Multihop Logical Topologies For Gigabit Lightwave Networks", IEEE LTS, Aug. 1992.

[StKl99] Kristian E. Stubkjaer, "Wavelength Converter Technology", IEICE Trans. Commun., vol.E82-B, no.2, Feb. 1999, pp.390-400.

[Sudh93] G.N.M. Sudhakar, "Access protocols and network architectures for very high-speed optical fiber local area networks", Ph.D thesis, University of Ottawa, 1993.

[SuMa91] G.N.M. Sudhakar, N.D. Georganas, and M. Kavehrad, "Slotted Aloha and reservation Aloha protocols for very high-speed optical fiber local area networks using passive star topology", IEEE/OSA J. of Lightwave Tech., vol.9, no.10, Oct 1991, pp.4144-1422.

[Tang94] K. Wendy Tang, "CayleyNet: A Multihop WDM-Based Lightwave Network", Proc. INFOCOM'94, Toronto, Jun, 1994, pp.1260-1267.

[ThHu96] R. A. Thompson and D. K. Hunter, "Elementary Photonic Switching Modules in Three Divisions", IEEE JSAC, vol.14, no.2, Feb. 1996, pp362-373.

[ToLe98] Philip P. To and Tony T. Lee, "The Multi-dimensional Shuffle-exchange Network: A Novel Topology for Regular Network Architectures", 1998, IEEE.

[TuZh99] R. S. Tucker and W. D. Zhong, "Photonic Packet Switching: An Overview", IEICE Trans. Commun., vol.E82-B, no.2, Feb. 1999, pp.254-264.

[VaLe98] M. Mehdi Vaez, and Chin-Tau Lea, " Wide-Sense Nonblocking Banyan-Type Switching Systems Based on Directional Couplers", IEEE JSAC, vol.16, no.7, Sep. 1998, pp.1327-1332.

[XuPe01] L. Xu, Harry G. Perros and G. Rouskas, "Techniques for Optical packet switching and optical burst switching", IEEE Commun. Mag., vol.39, no.1, Jan. 2001, pp.136-142.

[YaMu00a] S. Yao, B. Mukherjee and S. Dixit, "Advances in Photonic Packet Switching: An Overview", IEEE Commun. Mag., vol.38, no.2, Feb. 2000, pp.84-94.

[YaMu00b] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit, "All-Optical Packet-Switched Networks: A Study of Contention-Resolution Schemens in an Irregular Mesh Network with Variable-Sized Packets", Proc. SPIE vol.4233, OptiComm 2000: Optical Networking and Communications, Richardson, Texas, USA, Oct. 2000, pp.235-246.

[YaYa96] M. Yamaguchi, T. Yamamoto, K. Hirabayashi, et al., "High-density digital free-space photonic-switching fabrics using excitation absorption reflection-switch (EARS) arrays and microbeam optical interconnections", IEEE J. Selected Topics in Quantum Electronics, vol.2, no.1, Apr. 1996, pp.4454.

[YaYo01] S. Yao, S. J. Yoo, B. Mukherjee and S. Dixit, "All-Optical Packet Switching for Metropolitan Area Networks: Opportunities and Challenges", IEEE Commun. Mag., vol.39, no.3, Mar. 2001, pp.142-148.

[ZhAc94] Zhensheng Zhang and Anthony S. Acampora, "Performance Analysis of Multihop Lightwave Networks with Hot Patato Routing and Distance-Age-Priorities", IEEE Trans. Commun., vol.42, no.8, Aug. 1994, pp.2571-2581.

Appendix A: Simulation Results of 3-Stage Switches

The data in this section provide more numerical details to Figure 4.4 to 4.7 in Section 4.2.

Table A.1 Average Delay of (4,2) Switch

Load	Buffer=1	2	5	10
0.1	6.11868017E-06	5.87033675E-06	5.49481770E-06	7.24387440E-06
0.2	2.34076522E-04	2.34804848E-04	2.34895871E-04	2.34586754E-04
0.3	1.95538123E-03	1.93068343E-03	1.95463993E-03	1.94266234E-03
0.4	8.72390844E-03	9.04930943E-03	9.02976491E-03	8.97965930E-03
0.5	2.85295939E-02	3.15640384E-02	3.19291563E-02	3.20090304E-02
0.6	7.42443965E-02	9.66210009E-02	1.06603079E-01	1.07151680E-01
0.7	1.60870063E-01	2.61384335E-01	4.06009678E-01	5.09239497E-01
0.8	2.99300071E-01	5.79705685E-01	1.32571454E+00	2.44913002E+00
0.9	5.13102605E-01	1.09224819E+00	2.93510763E+00	6.06522831E+00
1.0	9.99987602E-01	1.99996107E+00	4.99980785E+00	9.99932796E+00

Table A.2 Average Delay of (9,2) Switch

Load	Buffer=1	2	5	10
0.1	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.2	3.20680102E-07	2.71285458E-07	2.46667351E-07	1.48017586E-07
0.3	1.84034472E-05	1.81534395E-05	1.82304513E-05	1.82035378E-05
0.4	3.01418965E-04	2.93952764E-04	2.95919303E-04	2.94844371E-04
0.5	2.30538832E-03	2.30308550E-03	2.31232419E-03	2.34012777E-03
0.6	1.15572846E-02	1.28046816E-02	1.30449753E-02	1.30293696E-02
0.7	4.52301425E-02	6.12222830E-02	7.03528146E-02	7.12836890E-02
0.8	1.36558614E-01	2.36826743E-01	4.44025117E-01	7.01788140E-01
0.9	3.48617972E-01	7.14909069E-01	1.82407529E+00	3.66623922E+00
1.0	9.99582249E-01	1.99876804E+00	4.99407909E+00	9.97867240E+00

Table A.3 Success Ratio of (4,2) Switch

Load	Buffer=0	1	2	5	10
0.1	9.99994383E-01	1.00000000E+00	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.2	9.99812122E-01	9.99999938E-01	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.3	9.98654357E-01	9.99996878E-01	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.4	9.94984614E-01	9.99939372E-01	9.99998689E-01	1.00000000E+00	1.00000000E+00
0.5	9.86856896E-01	9.99391160E-01	9.99959811E-01	9.99999950E-01	9.99999975E-01
0.6	9.72688369E-01	9.96442076E-01	9.99316060E-01	9.99984891E-01	9.99999917E-01
0.7	9.52422518E-01	9.86153750E-01	9.93950704E-01	9.98621763E-01	9.99598420E-01
0.8	9.27362077E-01	9.62027088E-01	9.72363416E-01	9.80147913E-01	9.82292522E-01
0.9	8.99283400E-01	9.19911458E-01	9.25251315E-01	9.28075563E-01	9.28411144E-01
1.0	8.63509000E-01	8.65028263E-01	8.63810413E-01	8.64354175E-01	8.64454913E-01

Table A.4 Loss Ratio of (4,2) Switch

Load	Buffer=0	1	2	5	10
0.1	5.61651275E-06	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.2	1.87878255E-04	6.24370517E-08	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.3	1.34564338E-03	3.12200652E-06	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.4	5.01538562E-03	6.06284555E-05	1.31117117E-06	0.00000000E+00	0.00000000E+00
0.5	1.31431035E-02	6.08839706E-04	4.01645217E-05	2.49807180E-08	0.00000000E+00
0.6	2.73116307E-02	3.55790269E-03	6.83836394E-04	1.49837903E-05	4.16240835E-08
0.7	4.75774815E-02	1.38460533E-02	6.04899318E-03	1.37773738E-03	4.00902013E-04
0.8	7.26379229E-02	3.79725996E-02	2.76359593E-02	1.98507595E-02	1.77049486E-02
0.9	1.00716600E-01	8.00880282E-02	7.47472404E-02	7.19213818E-02	7.15822597E-02
1.0	1.36491000E-01	1.34970738E-01	1.36187588E-01	1.35640825E-01	1.35535088E-01

Table A.5 Success Ratio of (9,2) Switch

Load	Buffer=0	1	2	5	10
0.1	1.00000000E+00	1.00000000E+00	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.2	9.99999877E-01	1.00000000E+00	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.3	9.99987936E-01	1.00000000E+00	1.00000000E+00	1.00000000E+00	1.00000000E+00
0.4	9.99820371E-01	9.99999877E-01	9.99999975E-01	1.00000000E+00	1.00000000E+00
0.5	9.98888079E-01	9.99989121E-01	9.99999877E-01	1.00000000E+00	1.00000000E+00
0.6	9.95655894E-01	9.99808571E-01	9.99985718E-01	9.99999803E-01	9.99999877E-01
0.7	9.87624601E-01	9.98067930E-01	9.99522848E-01	9.99980288E-01	9.99998840E-01
0.8	9.72163736E-01	9.89757154E-01	9.94243045E-01	9.97492139E-01	9.98398235E-01
0.9	9.47412056E-01	9.64646456E-01	9.69682960E-01	9.73534060E-01	9.74507590E-01
1.0	9.07943136E-01	9.07918049E-01	9.07975728E-01	9.07940000E-01	9.07829506E-01

Table A.6 Loss Ratio of (9,2) Switch

Load	Buffer=0	1	2	5	10
0.1	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.2	1.23339383E-07	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.3	1.20643758E-05	0.00000000E+00	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.4	1.79629098E-04	1.23320075E-07	0.00000000E+00	0.00000000E+00	0.00000000E+00
0.5	1.11192104E-03	1.08296715E-05	1.23358858E-07	0.00000000E+00	0.00000000E+00
0.6	4.34410562E-03	1.91330704E-04	1.41088522E-05	0.00000000E+00	0.00000000E+00
0.7	1.23753995E-02	1.93125575E-03	4.76559605E-04	1.87740908E-05	2.96062999E-07
0.8	2.78362639E-02	1.02402052E-02	5.75293180E-03	2.50090124E-03	1.59127761E-03
0.9	5.25879435E-02	3.53471007E-02	3.03035382E-02	2.64324922E-02	2.54215563E-02
1.0	9.20568642E-02	9.20619506E-02	9.19842716E-02	9.19600000E-02	9.19704938E-02

Appendix B: Confidence Interval Verification

Each simulation has been running for 10 times with different random seeds and $\alpha = 0.05$ confidence interval is shown. They correspond to results in Section 3.1.4, 3.2.3, and 4.2.

Table B.1 Average Delay of MKG(4,2,3): With Single Wavelength OPS Couplers ("delay" method)

	Load=0.1	Load=0.2	Load=0.3
1	4.00854089	4.92347846	11.6984453
2	4.00873720	4.92338977	11.3943564
3	4.00557214	4.91788294	12.2097564
4	4.00786421	4.92698152	11.4670132
5	4.00901272	4.91795351	12.1970722
6	4.00895930	4.92869697	12.0263944
7	4.00667760	4.91937768	11.5389089
8	4.00677898	4.92321704	11.4747442
9	4.00964393	4.92289159	11.6422426
10	4.00845794	4.92133635	11.4445061
Average	4.00802449	4.92252058	11.70934397
Upbound	4.00882038	4.92472635	11.90594883
Lowbound	4.00722860	4.92031482	11.51273910

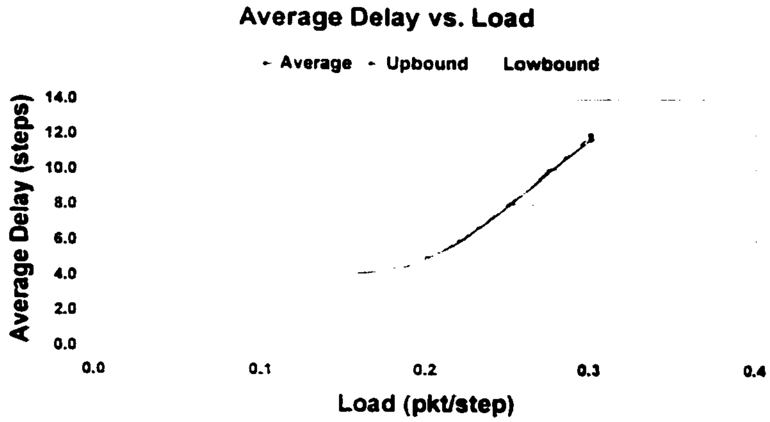


Figure B.1 Average Delay of MKG(4,2,3): With Single Wavelength OPS Couplers ("delay" method)

Table B.2 MKG(4,2,3) With Single Wavelength OPS Couplers(20% load): Cumulative Completed % vs. Needed Sequence Steps

Step											Average	Upbound	Lowbound
1	0. 02087	0. 02077	0. 02077	0. 02072	0. 02082	0. 02073	0. 02093	0. 02081	0. 02075	0. 02084	0. 02080	0. 02084	0. 02076
2	0. 06798	0. 06790	0. 06793	0. 06769	0. 06796	0. 06778	0. 06801	0. 06794	0. 06785	0. 06783	0. 06789	0. 06795	0. 06782
3	0. 18287	0. 18270	0. 18285	0. 18243	0. 18311	0. 18245	0. 18291	0. 18269	0. 18243	0. 18276	0. 18272	0. 18286	0. 18258
4	0. 44390	0. 44415	0. 44437	0. 44319	0. 44483	0. 44280	0. 44444	0. 44391	0. 44376	0. 44404	0. 44394	0. 44431	0. 44357
5	0. 67148	0. 67129	0. 67258	0. 67062	0. 67240	0. 67005	0. 67214	0. 67130	0. 67175	0. 67180	0. 67154	0. 67202	0. 67106
6	0. 82826	0. 82796	0. 82941	0. 82765	0. 82911	0. 82724	0. 82910	0. 82825	0. 82882	0. 82868	0. 82845	0. 82888	0. 82801
7	0. 91937	0. 91946	0. 92051	0. 91909	0. 92036	0. 91886	0. 92018	0. 91963	0. 91981	0. 91994	0. 91972	0. 92006	0. 91939
8	0. 96516	0. 96529	0. 96595	0. 96509	0. 96593	0. 96490	0. 96576	0. 96543	0. 96541	0. 96565	0. 96546	0. 96568	0. 96524
9	0. 98575	0. 98596	0. 98629	0. 98577	0. 98623	0. 98567	0. 98602	0. 98587	0. 98582	0. 98604	0. 98594	0. 98607	0. 98581
10	0. 99441	0. 99450	0. 99467	0. 99432	0. 99466	0. 99436	0. 99416	0. 99443	0. 99434	0. 99448	0. 99446	0. 99454	0. 99439
11	0. 99782	0. 99789	0. 99801	0. 99781	0. 99793	0. 99781	0. 99789	0. 99786	0. 99778	0. 99789	0. 99787	0. 99791	0. 99783
12	0. 99916	0. 99921	0. 99924	0. 99916	0. 99920	0. 99916	0. 99921	0. 99918	0. 99913	0. 99918	0. 99919	0. 99921	0. 99917
13	0. 99968	0. 99971	0. 99972	0. 99967	0. 99969	0. 99968	0. 99971	0. 99969	0. 99967	0. 99970	0. 99969	0. 99970	0. 99968
14	0. 99988	0. 99989	0. 99989	0. 99988	0. 99989	0. 99988	0. 99990	0. 99988	0. 99988	0. 99989	0. 99989	0. 99989	0. 99988
15	0. 99995	0. 99996	0. 99996	0. 99995	0. 99996	0. 99995	0. 99997	0. 99995	0. 99995	0. 99996	0. 99996	0. 99996	0. 99995
16	0. 99998	0. 99998	0. 99999	0. 99998	0. 99999	0. 99998	0. 99999	0. 99998	0. 99998	0. 99998	0. 99998	0. 99998	0. 99998
17	0. 99999	0. 99999	0. 99999	0. 99999	1. 00000	0. 99999	1. 00000	0. 99999	0. 99999	0. 99999	0. 99999	0. 99999	0. 99999
18	0. 99999	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000
19	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000
20	1. 00000		1. 00000	1. 00000		1. 00000		1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000

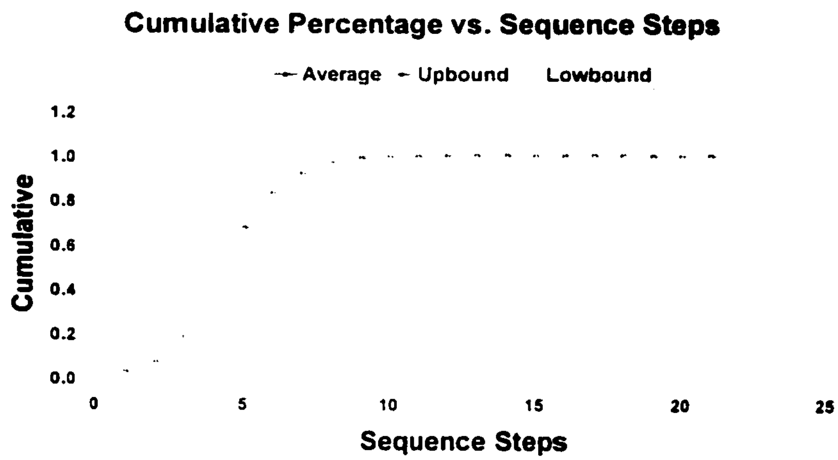


Figure B.2 Cumulative Completed Percentage vs. Needed Sequence Steps (MKG(4,2,3), Single Wavelength OPS Couplers, 20% load, "delay" method)

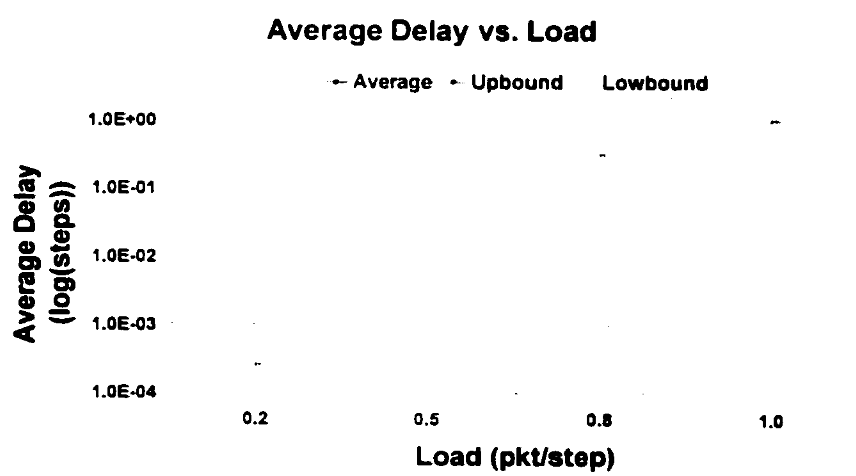


Figure B.3 Average Delay of (4,2) Switch (Buffer=2)