SCALABLE DESIGN OF OPTICAL BURST SWITCH BASED ON DEFLECTION ROUTING

Deng Yun

A thesis submitted in Partial

Fulfillment of the Requirements for

the Degree of Master of Philosophy

in

Information Engineering

July 2003

Acknowledgments

Many people support my effort during my M.Phil study in CUHK. I would like to thank them for making this two-year study period a fruitful and rewarding experience.

First of all, I would like to express my deepest gratitude towards my supervisor Prof. Tony T. Lee for his continuous support, for his invaluable advice, and most importantly, for being a true teacher to me. I am grateful to him. Many thanks to Dr. Soung Y. Liew for his sincere encouragement, for his constructive criticism, and for the overall directions he provided.

I would also like to thank all the members in broadband lab in department of Information Engineering: Mr. Choy Manting and Mr. Mui Szewai for their comments and continuous discussion on my research work. Mr. Zhang liang, Mr. Man Waihong , Mr. Lam Luifuk, Mr. Wong Tszchung and Mr. Lo Yinghang for their various kinds of help during these two years without them these two years life will lack of many beautiful memory.

Finally, I am grateful to my family for morally supporting during all these years.

摘要

最近提出的光突發交換是粗密度的光電路交換和細密度的光分組交換的一種折衷。在光突發交換中，在發出請求連接信號一段偏移時間後，數據以突發數據流的形式進入光通信網絡中。在這種混合型的交換技術中，交換機以電信號控制路由，而突發數據則保持光信號的形式通過交換系統。因此，光突發交換可以在不需要光緩存器的情況下達到較好的統計複用性能。

如何降低在每個交換節點中的由于輸出競爭而造成的光突發數據丟失仍然是一個有興趣的問題。我們提出使用一種偏轉路由交換技術，以雙重洗牌-交換網絡實現光突發交換機。然而，以目前的光交換技術較難以實現雙重洗牌-交換網絡中的 4×4 交換模塊，我們亦提出用只需要 2×2 交換模塊的洗牌-交換網絡來實現光突發交換節點，但是這種交換網絡與雙重洗牌-交換網絡相比在同樣的數據丟失率下需要更多的交換級數（也就是更大的複雜度）。

采用偏移路由的洗牌-交換網絡和雙重洗牌-交換網絡的可以運用在無緩存及异步交換的環境下，并且具有自路由的特性。然而，

考慮單個洗牌-交換網絡或者雙重洗牌-交換網絡平面，我們必須延伸交換級數的數目以滿足一定突發數據丟失率的要求。這會導致額外的複雜度和延遲。

我們進一步提出一種新的多平面交換結構來實現光突發交換，這個結構由并行的 k 個洗牌-交換網絡或者 k 個雙重洗牌-交換網絡組成。假定到達突發可以平均分配在各個交換平面，每個平面中的突發偏移將會急劇降低，這樣能够有效減少用于錯誤糾正的交換級數。通過分析，我們得到一個在一定數據丟失率下交換平面數目 k 的優化值以達到最小的交換複雜度。

Abstract

Optical burst switching (OBS) has been proposed to compromise between the coarse-grained optical circuit switching and the fine-grained optical packet switching. In OBS, data enter the network as a burst stream, following their connection request signals with an offset time. As a hybrid switching technology, each OBS switch controls routing decisions electronically, but keeps data bursts in optical form when they pass through the switch. Therefore, OBS can achieve excellent statistical sharing performance without buffering requirement.

Reducing the optical burst loss rate caused by output contention at each switching node is still a challenging issue, nevertheless. We propose using the deflection routing scheme, with the underlying dual shuffle-exchange network (DSN), to implement OBS switches. However, the basic building block of DSN is 4×4 switch module which is difficult to implement with the current optical switching technology. We also propose using an alternative switching architecture shuffle-exchange network (SN) with 2×2 switch module but having larger required number of stages (i.e. the complexity) than DSN under the same loss rate bound. We propose two routing schemes based on routing by wavelength or routing by fiber, and we compare the performance of them.

The attractive features of the deflection routed SN and DSN include that they can be adapted to the bufferless and asynchronous environment easily, and they are self-routed in nature. However, considering a single SN plane, it is necessary to extend the number of stages of the plane in order to correct as many deflection errors as possible to reduce the burst loss rate as required. This incurs some additional complexity and delay.

We further propose a novel multi-plane architecture, which consists of k SNs in parallel, for the OBS switches. Assuming that the arriving bursts at the switch can be evenly distributed among the planes, the burst deflection in each plane can then be greatly alleviated and thus the required number of stages for the error correction can be significantly reduced. We derive the optimal value of k, which depends on the loss rate bound and loading of switch, such that the overall complexity of the switch is minimized.

TABLE OF CONTENTS

LIST OF FIGURES

Number *Page*

Chapter 1

Introduction

1.1 Optical Switching

Owing to the explosive growth of the Internet traffic, there is increasing demand for the transmission capacity in telecommunications networks. It has also triggered a great deal of research on high-speed transmission and switching technologies, such as dense wavelength division multiplexing (DWDM) in which a single fiber is pushed to support hundreds of wavelengths. However, the mismatch of the switching capacity of electronic routers and the transmission capacity of fibers leads to the bottleneck of network.

Many switching schemes have been proposed for all-optical networks, aiming at efficiently and flexibly utilizing the huge potential capacity of fiber, such as wavelength routing (optical circuit switching) [10], optical packet switching [15], and optical burst switching (OBS) [1]. Wavelength-routed optical networks have already been deployed and currently, they represent the most promising technology for optical networks. However wavelength-routed optical networks are variants of the circuit switching networking paradigm and, as such, will suffer inevitably from various shortcomings. The granularity of optical circuit switches is so coarse that their bandwidth utilization is considered poor and so it may not be the most appropriate technology for the different applications that will use the emerging optical Internet. On the other hand, although optical packet switches can utilize the bandwidth more efficiently with sharing the resources statistically, they are hard to implement because of the absence of cost-effective optical buffering devices, and the difficulty of optical packet synchronization.

Optical burst switching is a switching technique that occupies the middle of the spectrum between the well-known circuit switching and packet switching paradigms, borrowing ideas from both to deliver a completely new functionality.

1.1.1 Optical Circuit Switching

Even though wavelength-routed optical networks have already been deployed, however, they are variants of the circuit-switching networking paradigm and, as such, will suffer inevitably from various shortcomings. For example, because of the bursty nature of the data traffic also leads to a poor wavelength utilization, therefore, in order to fully utilize the wavelength, at the network edge, sophisticated traffic aggregation mechanisms are needed to support applications requiring only a sub-wavelength of bandwidth of bandwidth cost-efficiently using optical pipes of granularity of at least one wavelength of bandwidth. In addition, as the number of client nodes connected to an optical network increase, the number of wavelengths required to provide a full mesh, as well as the corresponding size of the wavelength switches, may exceed technological limits.

1.1.2 Optical Packet Switching

In order to achieve efficient utilisation of bandwidth (i.e. wavelength or sub-wavelength) and simplify the process of resource provisioning, the bandwidth provisioning timescales maybe reduced. Optical packet switching [15] is an alternative technology that appears to be the optimum choice. On the other hand, although optical packet switches can utilize the bandwidth more efficiently with sharing the resources statistically, they are hard to implement because of the absence of cost-effective optical buffering devices, and the difficulty of optical packet synchronization.

One major challenge is current lack of optical random access memory. The basic idea of packet switching is the "Store and Forward", which means that packets are burst stored in the packet switches, and then forwarded to the next switch. This is necessary due to output port contention. At present, it seems that the most feasible method of implementing optical buffering is to use fiber-delay lines (FDLs), which, however, can only provide limited delays for optical signals.

Another is the stringent requirement for synchronisation, both between multiple packets arriving at different input ports of an optical switching fabric and between a packet's header and its payload. However, it is difficult and expensive to implement the synchronization component.

A disadvantage of optical packet switching, which uses small, fixed-length packets, is that the percentage of control overhead is high and the bandwidth utilization is low.

1.1.3 Optical Burst Switching

Optical burst switching (OBS) [1], [2] has been proposed to compromise between the coarse-grained circuit switching and the fine-grained packet switching. In OBS, data enter the network as a burst stream, following their connection request signals with an offset time. The offset time is for OBS switches to reconfigure their switching states so as to accommodate the burst. As a hybrid switching technology, each OBS switch controls routing decisions electronically, but keeps data bursts in optical form when they pass through the switch. Therefore, OBS can achieve excellent statistical sharing performance without buffering requirement.

In Optical burst switching mechanisms, the transmission unit size of optical burst switching called granularity is between the optical circuit switching and optical packet switching and burst header is transmitted on a separate wavelength (or channel).

To make a wavelength reservation for burst data, the source nodes of OBS network setup an offset time between the transmission of the first bit of the control header and the transmission of the first bit of the data burst. There are three classes of OBS based on different offset time length.

- No Reservation: In this kind of OBS, the offset time length is very short, and immediately after the control packet, the burst will be sent out. That is, the offset is only the transmission time of the control packet. This scheme is usable only when configuration time of switch and processing time of a control packet are very short. The optical burst switching scheme called Tell And Go (TAG) [11] belongs to this class.

- One-Way Reservation: After sending the control packet a short time, the data burst will be sent out in this scheme. This short offset is between transmission time of the control packet and the round-trip delay of the control packet. And the source node does not wait for the acknowledgement back from the destination node. Different optical burst switching mechanisms may choose different offset values in this range.

In just-enough-time (JET) proposed by Qiao and Yoo [1], the offset is selected in a manner that takes into account the processing delays of the control packet at the intermediate switches. Yoo et al. [3] also described a priority scheme for JET to support multi-class traffic. Dolzer and Gauger [13] proposed an iteration algorithm to analyze approximately this priority scheme considering the class interference, Verma et al. [14]

and Chaskar et al. [16] presented a traffic shaping scheme which randomizes the offset to reduce the burst loss probability.

- Two-Way Reservation: This scheme requires receiving an acknowledgement from the destination before sending the data burst. This class is very close to optical circuit switching. The offset time will be a round-trip delay to set up the transmission, and since the control packet reserves resources, delivery of the burst is guaranteed. The optical burst switching scheme called Tell And Wait (TAW) [11] belongs to this class. The major drawback of this class is the long offset time, which causes the long data delay. DÄuser and Bayvel [17] also proposed a wavelength-routed optical burst switching (WR-OBS), which uses two-way reservation.

Most of OBS schemes use one-way reservation, that is, a source node does not need to wait for the acknowledgement back from the destination node, before it starts transmitting of the burst.

The intermediate node in the optical network does not require optical buffers. Bursts in optical domain go through the intermediate node without any delay. The routing decision of each burst is preceded by its control header, which is transmitted by a separate signalling channel. Unlike circuit switching, a source node does not wait for confirmation that a path with available resources has been set up; instead, it starts transmitting the data burst soon after the transmission of the control packet.

We will refer to the interval of time between the transmission by the source node of the first bit of the control packet and the transmission of the burst bit of the data burst as the offset. The control packet carries information about the burst, including the offset value, the length of the burst, its priority, etc. The purpose of the control packet is to inform intermediate nodes of the upcoming data burst, so that they can make a routing decision and configure their fabric to switch the burst to the appropriate output port. However, in case of congestion or output port contention, an intermediate node may drop bursts. Thus, as in connectionless packet switching, there is no guarantee of delivery. Also, consecutive bursts between a given source-destination pair may be routed independently of each other.

1.2 Design of Optical Burst Switching Node

1.2.1 Burst Switched Network Architecture

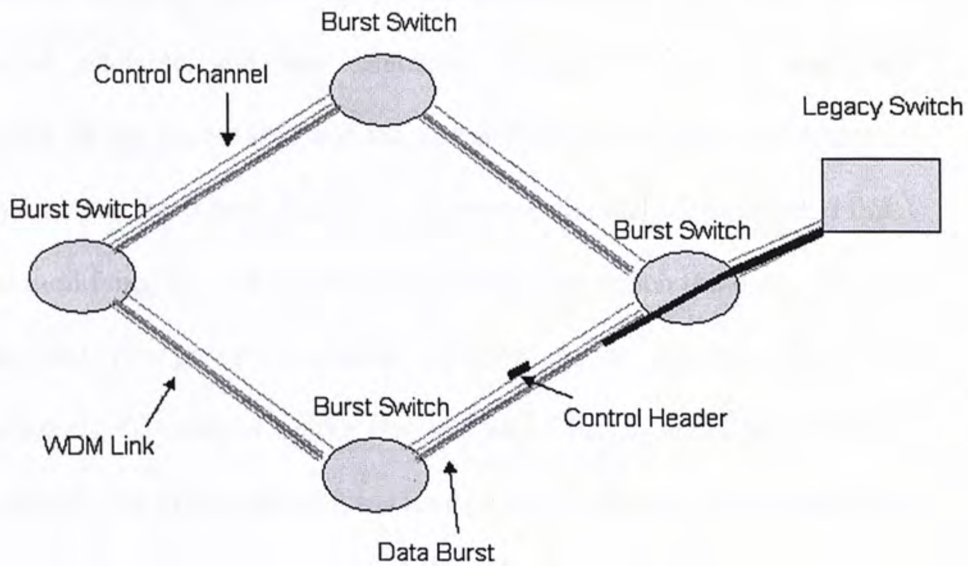

Figure 1-1Optical Burst Switched Network Architecture

Figure.1-1 [2] shows the network architecture for a burst switched network. The transmission links in the system can carry multiple channels in with multiplexing technique DWDM. One channel on each link is designated a control channel, and is used to transmit control headers to assign the remaining channels to user data bursts.

When a border switch has a burst of data to send, it will select an idle channel on the access link to send the data burst. Shortly before the burst transmission begins, a control header is sent on the control channel, specifying the channel on which the burst is being transmitted and the destination of the burst. A burst switch, on receiving the control header, selects an outgoing link leading toward the desired destination with an idle channel available, and then establishes a path between the specified channel on the access link and the channel selected to carry the burst. It also forwards the control header on the control channel of the selected link, after modifying the cell to specify the channel on which the burst is being forwarded. This process is repeated at every switch along the path to the destination. The control header also includes a length field specifying the amount of data in the burst. This is used to release the path at the end of the burst.

1.2.2 Design of Optical Burst Switching Node

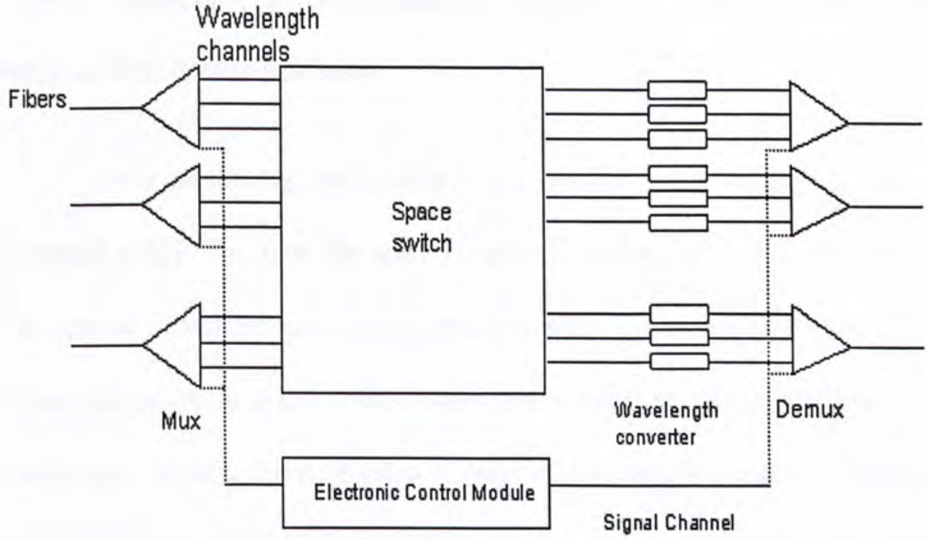

Figure 1-2 Structure of an Optical Switching Node

Fig.1-2 illustrates a possible structure of an Optical Switching Node (OSN). There are incoming and outgoing fiber links, each of which has a number of wavelengths for carrying data bursts (solid lines) and one additional wavelength for carrying control packets (dotted lines). Every control packet is processed by the electronic control module inside an OSN, which generates appropriate control signals to set up the wavelength converters, FDL buffers, and switching fabric. The optical switching fabric switches each burst on an incoming wavelength as it arrives (i.e., without having to synchronize it with other incoming bursts). Output contention occurs when there are too many input bursts destined to the same output fiber at the same time. If buffering is not provided, these bursts have to be dropped. Signaling protocol is needed to setup an all-optical path from source node to destination node.

Two switching architectures for optical burst switching were discussed in [2]. The first one uses Tunable Wavelength Converters along with optical crossbars and passive multiplexors and demultiplexors. The second design replaces the optical crossbars of the first design with passive Wavelength Grating Routers (also known as Arrayed Waveguide Grating Multiplexors or AWGMs). Unfortunately, these architectures may introduce extra blocking and require complex scheduling algorithms that make them less appealing especially in the fast optical network.

Reducing the optical burst loss rate caused by burst contention at each switching node is still a challenging issue, nevertheless. Some papers [3]-[5] proposed using channel scheduling and signaling protocol to reduce the burst loss rate but with the price of high architectural and scheduling complexities. J. Ramamirtham [6] proposed to associate the tunable wavelength converters with the optical crossbars or passive wavelength grating routers to construct the OBS switch, but unfortunately this approach introduces extra burst loss rate and requires complicated scheduling algorithms.

We propose using the deflection routing scheme, with the underlying shuffle-exchange network (SN) or dual shuffle-exchange network (DSN) [9], to implement OBS switches. In the above two kinds of deflection routed OBS switches, the bursts that lost contention for the desired link are deflected to other links in the hope that the deflection errors can be corrected later. In other words, the deflected bursts can detour the contended link such that they can continue their journey for the destination without internal blocking. The attractive features of the deflection routed SN and DSN include that they can be adapted to the bufferless and asynchronous environment easily, and they are self-routed in nature. However, considering a single SN or DSN plane, it is necessary to extend the number of stages of the plane in order to correct as many deflection errors as possible to reduce the burst loss rate as required. This incurs some additional complexity and delay.

1.2.3 Scalable Architecture With Multi-plane Fabric

We further propose a novel multi-plane architecture, which consists of k SNs or k DSNs in parallel, for the OBS switches. Assuming that the arriving bursts at the switch can be evenly distributed among the planes, the burst deflection in each plane can then be greatly alleviated and thus the required number of stages for the error correction can be significantly reduced. We derive the optimal value of k, as a function of the loss rate bound, such that the overall complexity of the switch is minimized.

1.3 Organization

The rest of my thesis is organized as follows. In chapter 2 we introduce the switch architecture based on deflection routing. In chapter 3 we propose design based on SN and Multi-plane. Chapter 4 is design based on DSN and comparison with SN. Conclusion is at last given in Chapter 5.

Chapter 2

Proposed OBS Node and Blocking probability due to Output Contention

2.1 OBS Node Architecture

One way to solve the burst contention problem without having to buffer the losing bursts is to use deflection routing. The basic idea is to route (deflect) the losing bursts to "wrong" outgoing link rather than drop them. Redundancy is built into the switch design so that deflection can be routed in later switching stages in a way that corrects for the earlier mistakes.

In our following switch architecture based on deflection routing, the bursts failing to get selected for intended link are sent along different links, in the willing that they later return, or detour the contended link and continue their journey to the destination. The attractive feature of deflection routing based switch is that it works in distributed self-routing and asynchronized way and can solve burst contention without buffer.

Figure.2-1 shows our proposed switching node architecture. There are d incoming and d outgoing optical fibers, each contains h wavelength channels. The central switching fabric is a dh-by-dh deflection routed switch. n is the number of stages one burst has to pass without any deflection, and with the deflection routing property, burst can exit from the network beginning from the n^{th} stage with $n = \log_2(d \times h)$. L is the length of the switch—number of switch stages and there are L-n+1 stages can output bursts.

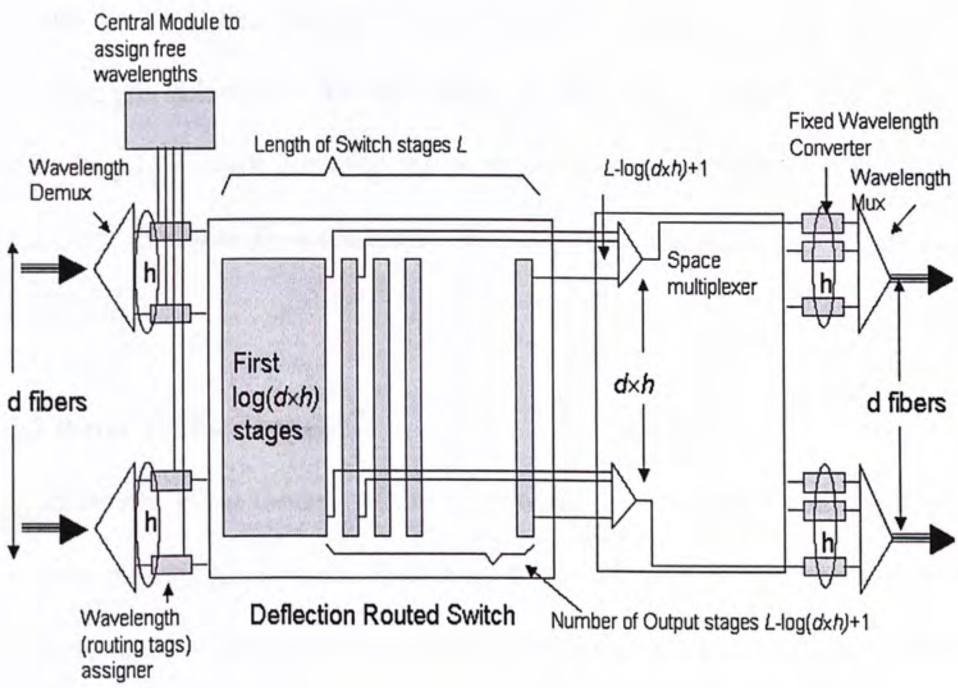

Figure 2-1 Proposed OBS Node Architecture

With deflection routed switch architecture such as SN or DSN that will be introduced later, we can solve internally blocking by deflection routing, but there are still a lot of output contentions existing in output space multiplexers.

Output contention happens when two coming bursts are routed to same output port (wavelength channel), and a simple way to avoid it is to find a free wavelength of the destination fiber before sending the burst into switch and then route the burst to that wavelength. To do that, it need a central module to record which wavelength channel of destination fiber is being occupied so that it could easily find a free wavelength channel for a newly arrived data burst. We can set the output port that the free wavelength belongs as the destination of burst. If it cannot find a free wavelength, we will just drop the burst. As showed in figure 2.1, to route the burst to a specify wavelength, we have to set the route tags length be $\mathrm{Log}(d \times h)$.

2.2 Burst Traffic Model

In most analytical models of OBS networks, it is assumed that the burst arrival process is Poisson. However, it is well known that the Poisson process is not a good model for wide area traffic, and it is unlikely that the burst arrival processes in future optical networks will be accurately characterized by the Poisson model. Therefore, more sophisticated models are required in order to advance our understanding of the performance and the potential of OBS networks.

We use the two-state Markov process shown in Figure.2-2 to model arrivals on a given burst wavelength. The arrival process may be in one of two states: burst, or idle. We assume burst time-length is exponential distribution with mean $1/\mu$ and idle time-length is exponential distribution with mean $1/\lambda$. We define the loading of each input

$$\rho = \lambda/(\mu + \lambda)$$

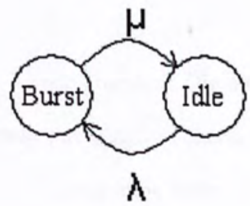

Figure 2-2 Markov process of burst traffic

2.3 Blocking Probability due to Output Contention

With our scheme, the bursts those who cannot be assigned a free wavelength channel of the destination fiber will be blocked instead of input to switch. As illustrated by the state-transition diagram in Fig 2-3, in which the state is the number of busy wavelength channels of destination fiber, and the blocking probability is limiting probability of state *h,* where *h* is the number of wavelength channels of each fiber.

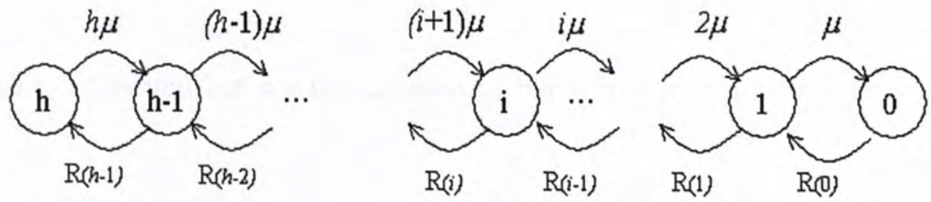

Figure 2-3 State Transition Digram--Number of busy wavelength channels in the destination fiber

In this state-transition diagram, state i means that there are i busy wavelength channels, and h-i free wavelength channels in the destination fiber. It also means that there are i input ports in burst state, h-i input ports in idle state and other N-h input ports those are unknown be burst or idle, where N=$d\times h$.

From above, we can get the rate from state i to i-1 is $i\mu$. As for the rate from state i to i+1:

$$R_{(i)} = [(h-i)+(N-h)(1-\rho)]\cdot\lambda/d, i = 0,1,2,....,h-1.$$

By solve this Markov chain, we can get

$$P_{Blocking} = P_{state(h)}$$
$$= 1/(\sum_{i=1}^{h}(\prod_{j=1}^{i}(\frac{(h+1-j)u}{[j+(N-h)(1-\rho)]\cdot\lambda/d}))+1)$$

In above equation, $\mu / \lambda = \frac{1}{\rho} - 1$, so $P_{Blocking}$ just depends on loading of switch but not the duration of burst time length or idle time length.

When $d = 8, h = 64, \rho \geq 0.5$, we have numerical results of blocking probability due to output contention, as illustrated in Figure 2-4.

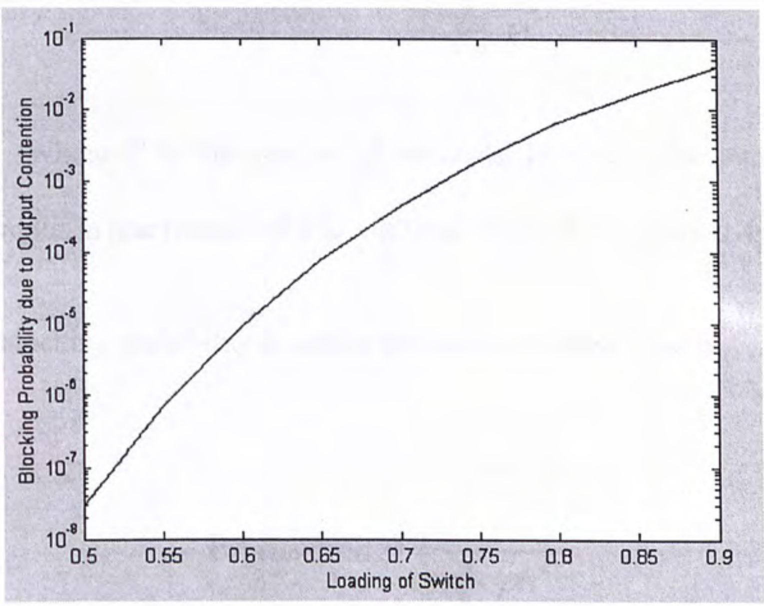

Figure 2-4 Numerical Results of Blocking Probability due to Output Contention

2.4 Poisson Approximation of Burst Traffic

If we consider the burst interarrival time --- burst time + idle time be exponentially distributed with mean $1/\mu+1/\lambda$ for approximation. The traffic will be Poisson arrival with mean service time $1/\mu$.

Because the burst blocking happens when there is no any free wavelength in the destination fiber, we can get this blocking probability by Erlang B formula with above approximation.

$$\Pr[Blocking] = \frac{\left(\frac{A^C}{C!}\right)}{\sum_{k=0}^{C}\frac{A^k}{k!}}$$

Where C is the number of channels, here C is the number of wavelengths in one fiber, $C=h$ and A is total overload to the fiber, $A=\rho h$.

So the blocking probability in each wavelength of output fiber is

$$\Pr[Blocking] = \frac{\left(\frac{\rho h^h}{h!}\right)}{\sum_{k=0}^{h}\frac{\rho h^k}{k!}}$$

Fig 2-5 is the compare of two results with On/Off burst model and Poisson arrival model, and we can see that the approximation with Poisson arrival model will introduce more blocking probability. That is because the variance of interarrival time of Poisson arrival traffic $(1/\mu+1/\lambda)^2$ is bigger than the variance of interarrival time of On/Off traffic model $(1/\mu)^2+(1/\lambda)^2$.

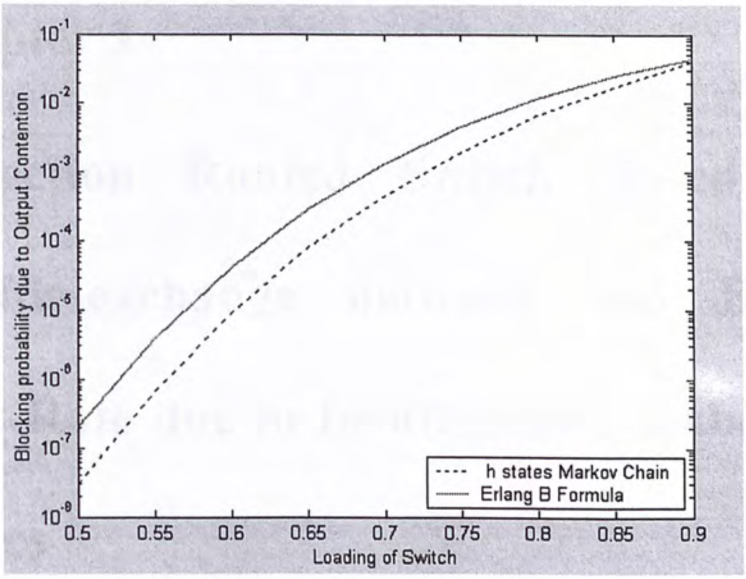

Figure 2-5 Compare of Two Traffic Models

2.5 Simulation Results

Fig 2-6 is the simulation results of blocking probability due to output contention with same *h* and *d*, and we can see it is very close to our analysis results.

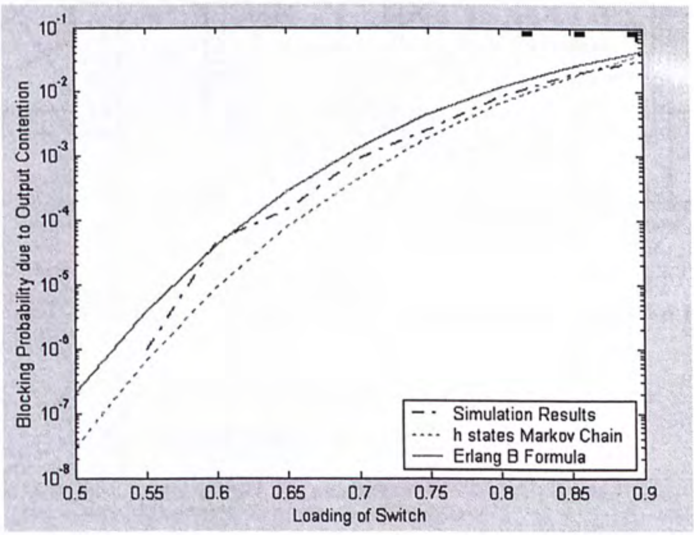

Figure 2-6 Simulation Results of Blocking

Probability due to Output Contention

Chapter 3

Deflection Routed Switch Based on Shuffle-exchange network and Burst Loss Rate due to Insufficient Number of Stages

3.1 Architecture of Shuffle-exchange Network

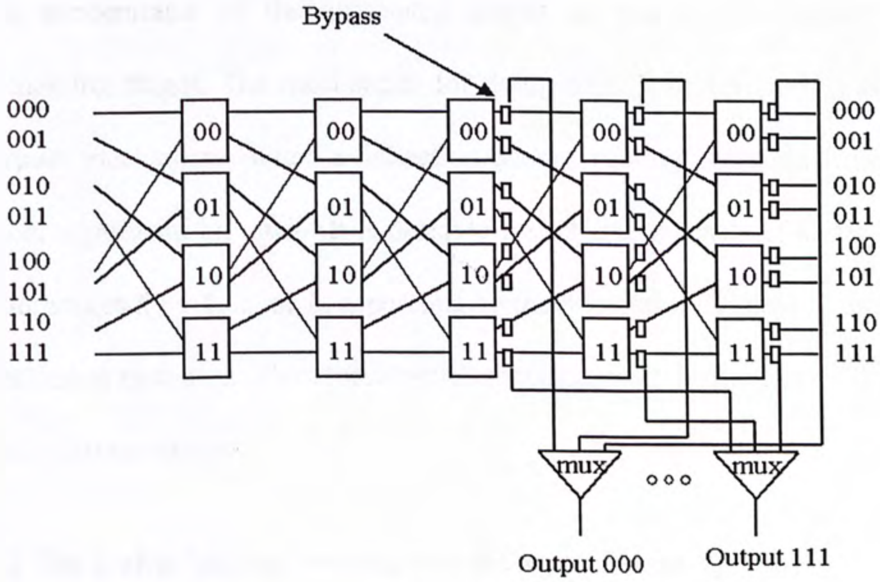

Figure 3-1 Structure of Shuffle-exchange network

Shuffle-exchange network [8] [9] with n stages is isomorphic to the Banyan network. When deflection routing is used to tackle the contention problem, instead of a regular shuffle-exchange network with n stages, we have an elongated shuffle-exchange network with $L \geq n$ stages, as depicted in Fig. 3-1.

The output address of a burst is represented in binary form for routing, and the most significant bit is used in the first stage, the second most significant bit is used in the second state, and so on. When all the bits are used, the burst has reached its desired destination, and it will be removed from the elongated shuffle-exchange network and forwarded to the concentrator of the associated output so that it will bypass the remaining stages. The mechanism for doing so will be referred to as the bypass mechanism. When a deflection occurs, routing is started from the most significant bit again. It is possible for a burst to reach its destination after stages i, $i + 1$, ..., or L, depending on the number of deflections and the deflection distances. Therefore, bypass mechanisms must be installed after each of these stages.

3.2 The traffic loading entering into the second stage

When bursts with loading ρ go through our proposed OBS, the burst loss first happen in the first stage—the burst blocking due to output contention, and we can get $P_{Blocking}$ from chapter 2. After first stage, the bursts with loading ρ' will enter the second stage—the shuffle-exchange network of switch and the bursts with loading $\rho_{Blocking}$ will be dropped.

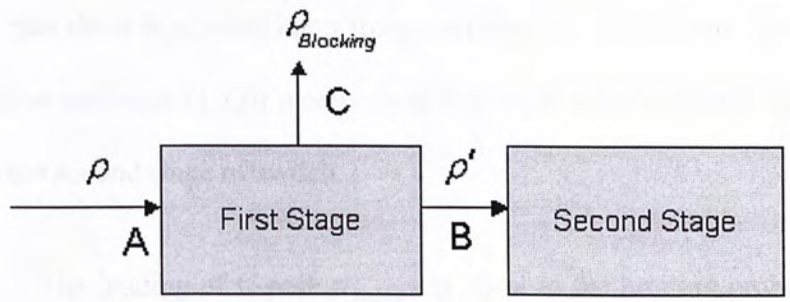

Figure 3-2 Loading of Two stages

When a burst coming into first stage, it will be blocked with probability $P_{Blocking}$, or be sent into second stage with probability 1- $P_{Blocking}$. To get the loading ρ' and $\rho_{Blocking}$, we use a Markov chain with three states as illustrated in Fig 3-3.

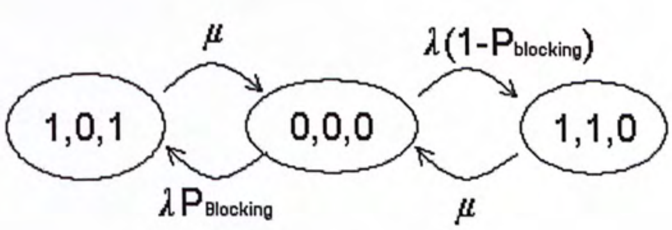

Figure 3-3 Transition-Diagram of states of

three ports

The bits of this Markov chain's states are the state of A, B, C ports, respectively. 0: Idle state, 1: Burst State.

State (0,0,0) means no burst coming into system, state (1,0,1) means that there is a burst input to system but be blocked due to output contention and state (1,1,0) means there is a burst input to system and be sent to the second stage of switch.

The loading of C port $\rho_{Blocking}$ is equal to the limiting probability of state (1,0,1) and the loading of B port ρ' is equal to the limiting probability of state (1,1,0).

We have

$$P_{0,0,0} + P_{1,1,0} + P_{1,0,1} = 1$$

$$P_{0,0,0} \times \lambda(1 - P_{blocking}) = P_{1,1,0} \times \mu$$

$$P_{0,0,0} \times \lambda P_{blocking} = P_{1,0,1} \times \mu$$

We can get

$$\rho_{Blocking} = P_{1,0,1} = \frac{\lambda P_{Blocking}}{\lambda + \mu} = \rho \cdot P_{Blocking}$$

$$\rho' = P_{1,1,0} = \frac{\lambda(1 - P_{Blocking})}{\lambda + \mu} = \rho \cdot (1 - P_{Blocking})$$

3.3 The Deflection Probability in a 2×2 Switching Module of SN

Let us first derive the deflection probability of a 2×2 switching module. When a new burst arrives at a 2×2 switching module, it will see at least one output port is idle, or equivalently 0 or 1 output ports are being occupied. Assuming each burst is equally likely to destine to one of the two output ports, and as described in last section the burst length is exponential with mean $1/\mu$ and idle time is exponential with mean $1/\lambda$. Figure 3-4 depicts a Markov chain for this 2×2 non-blocking switching module, and states of the Markov chain are the number of busy output ports seen by a new arriving burst. After obtaining the limiting probabilities

$$p_0 = \frac{\mu}{\mu+\lambda}, \quad p_1 = \frac{\lambda}{\mu+\lambda},$$

the deflection probability q for non-blocking switching module can be calculated as follows

$$q = 1-(P_0+0.5P_1) = \frac{\rho_{2\times 2}}{2}, \text{ where } \rho_{2\times 2} = \frac{\lambda}{\lambda+\mu}$$

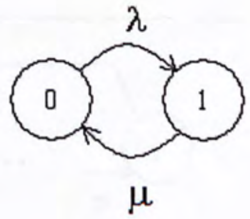

Figure 3-4 The Markov chain of 2x2 switching module: state means the number of busy output ports

3.4 Analysis of Burst Loss Rate due to Insufficient Number of Stages in SN

With the deflection routing algorithm of SN, when a burst loses contention in one switching module, its routing tag will be reset to its original destination address and start anew from the deflection point. As illustrated by the state-transition diagram in Fig 3-5, in which the state is the distance or the number of stages away from destination, once the burst is deflected, it have to start from initial state *n*.

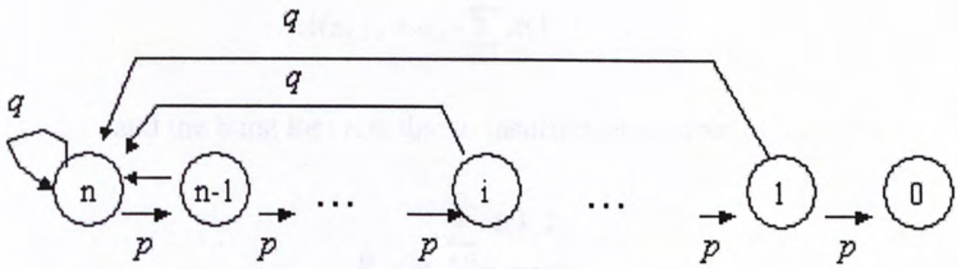

Figure 3-5 The state-transition diagram of a packet in the SN with deflection probability q and P=1-q

In a shuffle-exchange network with size $N \times N$, $n = \log N$, number of stages = L, and loading of switch = ρ'.

$B(j)$ is the loading of stage j, $A(i,j)$ is the loading of stage j with distance i steps to destination.

We have

$$A(n,0) = \rho'$$

$$A(i,0) = 0,\ i = 0,1,,\ldots,n-1$$

and the deflection probability of stage j is q_j

$$q_j = \frac{\sum_{k=1}^{n} A(k, j-1)}{2}, p_j = 1 - q_j, \quad j = 1,2,\ldots,L$$

From Markov chain of fig 3-3, we have

$$A(i,j) = p_j \cdot A(i+1, j-1),\ i = 0,1,\ldots,n-1$$

$$A(n,j) = q_j \cdot \sum_{k=1}^{n} A(k, j-1)$$

and the burst loss rate due to insufficient number of stages is

$$P_{Loss} = \frac{\sum_{k=1}^{n} A(k,L)}{\rho'}$$

We can get numerical results from above equation, as illustrated in Figure 3-6, with $h = 64, d = 8$, $n = LogN = Log(h*d) = 9$, loading of switch $\rho' = 0.5$.

We do the simulation with $\lambda = 1, \mu = 1$, d=8, h=64, and loading of switch $\rho' = \lambda/(\lambda + \mu) = 0.5$. The result is presented in Fig 3-4.

Compare simulation with analysis, simulation has better performance, because when we did simulation, the switch is empty initially without any traffic in it.

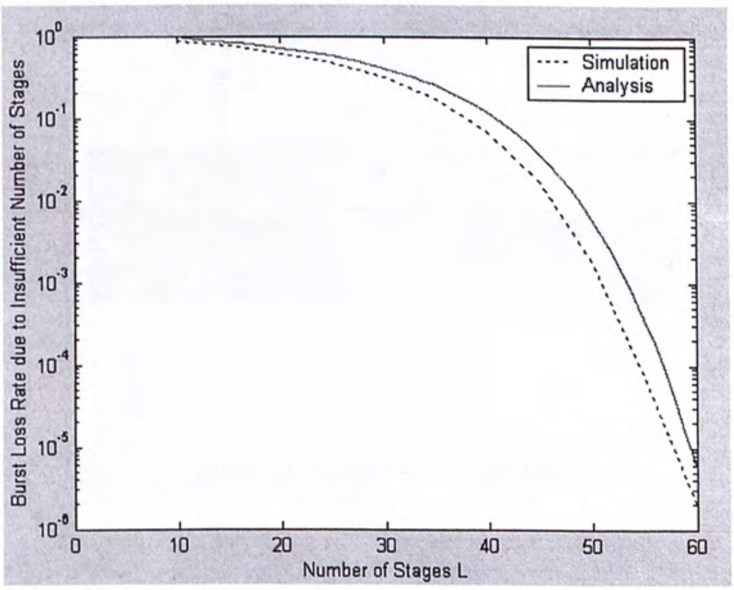

Figure 3-6 The analysis and simulation

results with SN

3.5 Total Burst Loss Probability

When bursts with loading ρ go through our proposed OBS, the burst loss first happen in the first stage—the burst blocking due to output contention, and we can get $P_{Blocking}$ (with On/Off traffic model) from chapter 2. After first stage, the bursts with loading $\rho' = \rho \cdot (1 - P_{Blocking})$ will enter the second stage—the shuffle-exchange network of switch.

With limited number of stages of SN, in Fig 3-7, the bursts with loading ρ'' cannot be successfully routed to its destination port and will be dropped due to insufficient number of stages. Thus, the total burst loss probability of our proposed OBS architecture will be

$$P_{Total\_loss} = \frac{\rho_{Blocking} + \rho''}{\rho}$$

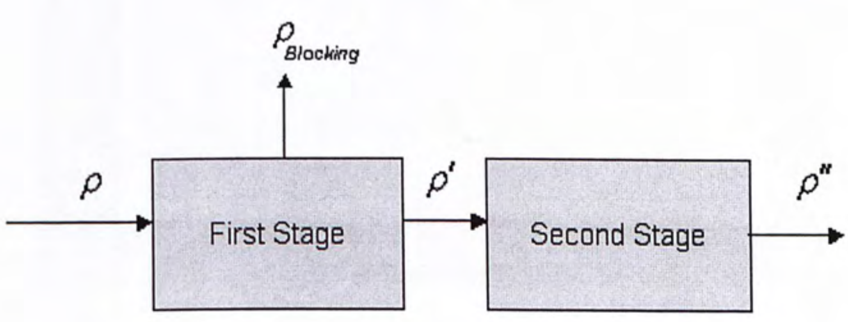

Figure 3-7 Loading of switching in two stages

We can get ρ'' from the numerical results of last section with known ρ'.

When $h = 64, d = 8$, $n = LogN = Log(h*d) = 9$, we can get the total burst loss probability as showed in Fig 3-8 with different loading of switch from 0.6 to 0.8. From the results, for example, loading = 0.6, we can say when number of stages is less than 70, P_{Loss} dominates over the total burst loss probability and if larger than 70, $P_{Blocking}$ dominates.

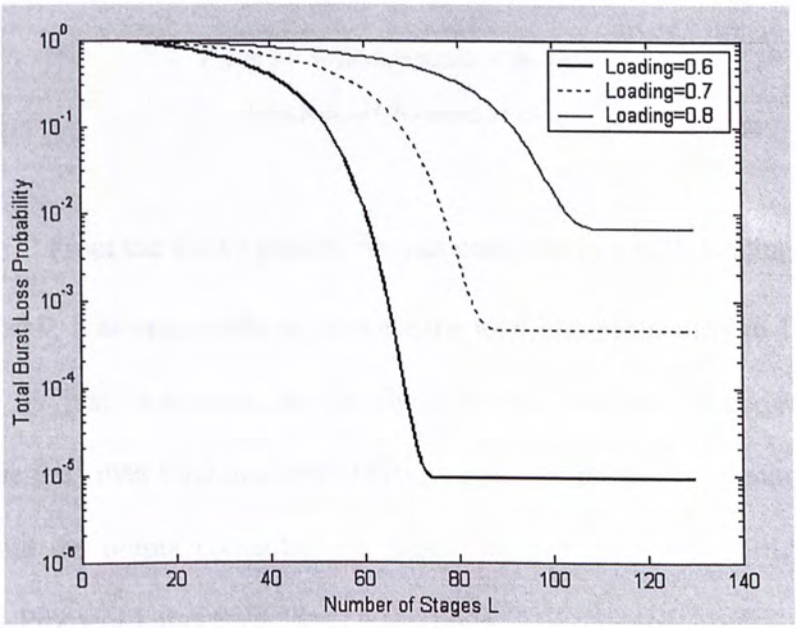

Figure 3-8 Numerical results of the Total Loss Rate of OBS based on SN

Fig 3-9 is the simulation results of total burst loss probability with $h = 64, d = 8$, $n = LogN = Log(h*d) = 9$.

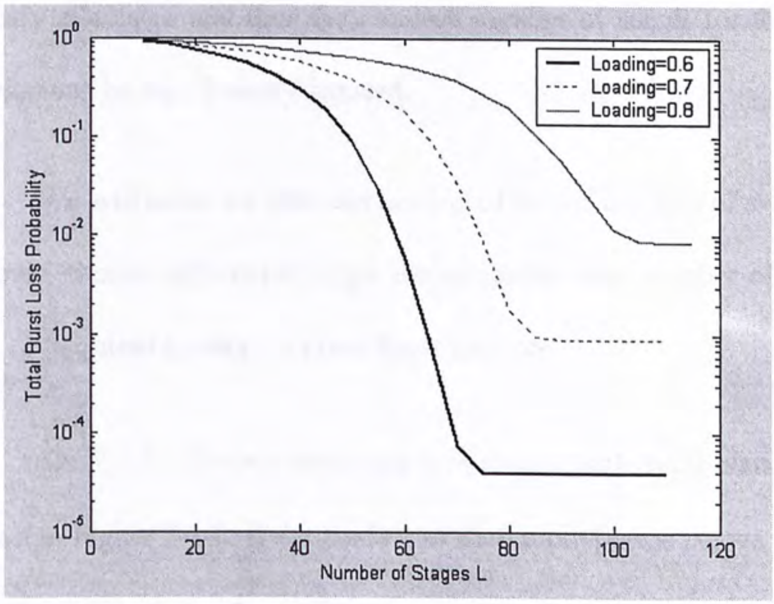

Figure 3-9 Simulation results of the Total

Loss Rate of OBS based on SN

From the above results, we can conclude that with loading > 0.6 when n=9, it is impossible to decrease the total loss probability to 10^{-5} or lower by just increasing the length of switch (number of stages). To achieve the lower total loss probability, we have to adopt some technology to avoid the output contention of bursts, for example, with MPLS [18] (Multi-Protocol Label Switching) technology.

3.6 Multi-plane Architecture

We further propose a novel multi-plane architecture, which consists of k planes in parallel, for the OBS switches. Assuming that the arriving bursts at the switch can be evenly distributed among the planes, due to the reduced offered load per plane, the burst deflection in each plane can then be greatly alleviated and thus the required number of stages for the error correction can be significantly reduced.

We will show for different loading of switch and size of switch *N*, we should choose different *k* to get the minimum total number of stages $L' = L \cdot k$ required to obtain a given burst loss rate.

An $N \times N$, *N=d×h*, switching architecture with multi-plane fabric is shown in Figure 3-10. If the loading of each input port is ρ, we assume it is evenly divided by demultiplexer and then the input loading of each $N \times N$ switching plane will be $\rho' = \rho / k$.

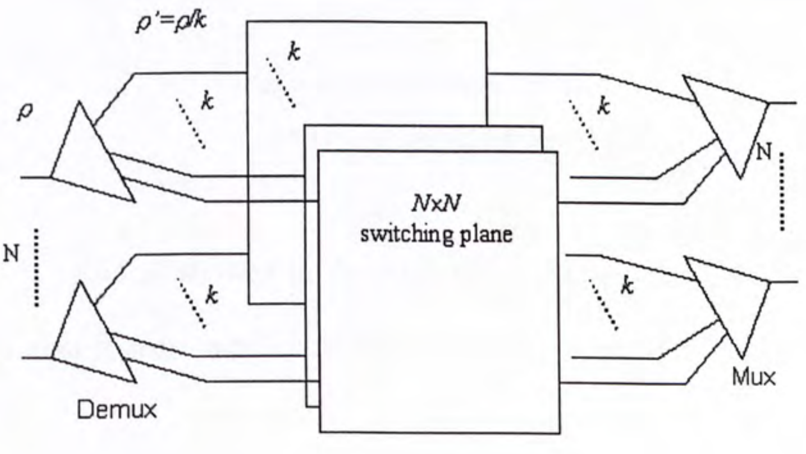

Figure 3-10 Multi-plane Fabric

3.6.1 Relationship between *k* and loading of SN

With the numerical results of burst loss rate due to insufficient number of stages, we can find the relationship between *k* and loading of SN.

For example, in fig 3-11, with n=9, when loading = 0.9, if for required burst loss rate 10^{-5}, with k=2 we can get best performance with the least total number of stages.

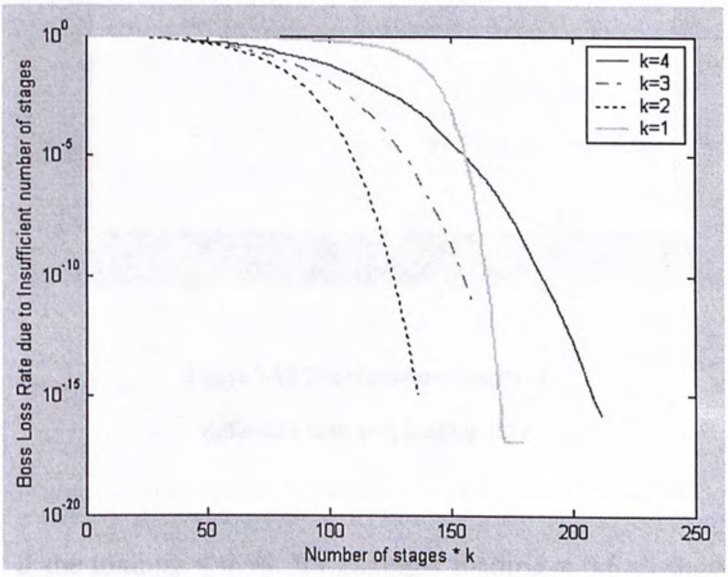

Figure 3-11 The Numerical Results of

different k with n=9, loading=0.9

And as showed in fig 3-12, with loading = 0.74, k=2 and k=1 need almost same number of stages to achieve required burst loss rate 10^{-5}.

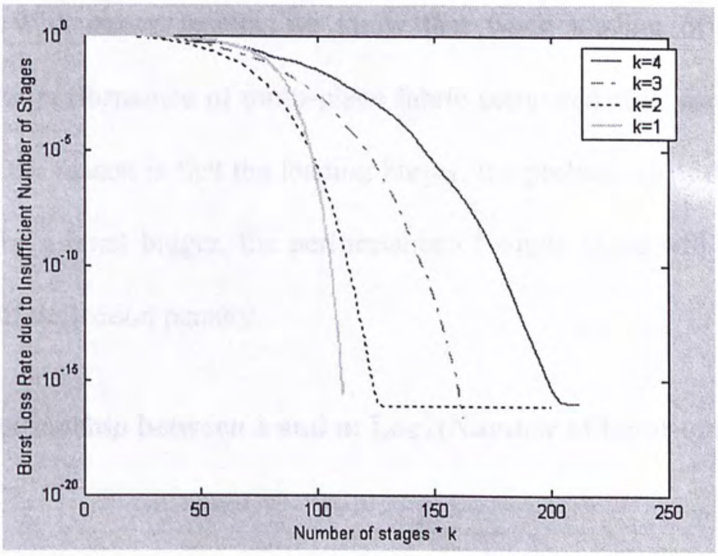

Figure 3-12 The Numerical Results of

different *k* with n=9, loading=0.74

If the loading < 0.74, for example loading = 0.6 as showed in fig 3-13, *k*=1 need the least total number of stages.

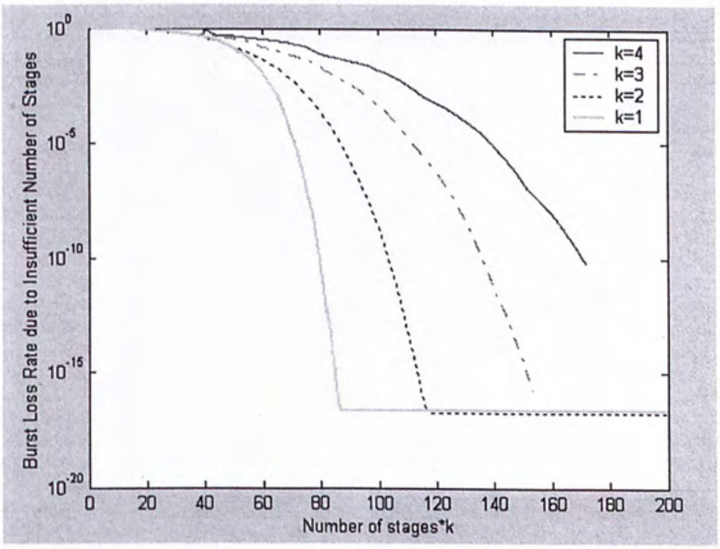

Figure 3-13 The Numerical Results of

different *k* with n=9, loading=0.6

With above results, we know that when loading of SN gets bigger, the performance of multi-plane fabric compared with single plane is better, the reason is that the loading bigger, the probability of deflection routing for a burst bigger, the performance of single plane will be worst because of deflection penalty.

3.6.2 Relationship between *k* and n: Log2(Number of input-output ports)

With the numerical results of burst loss rate due to insufficient number of stages, we also can find the relationship between *k* and *n*.

For example, in fig 3-11, with loading = 0.7, when n=7, if for required burst loss rate 10^{-5}, with k=1 we can get best performance with the least total number of stages.

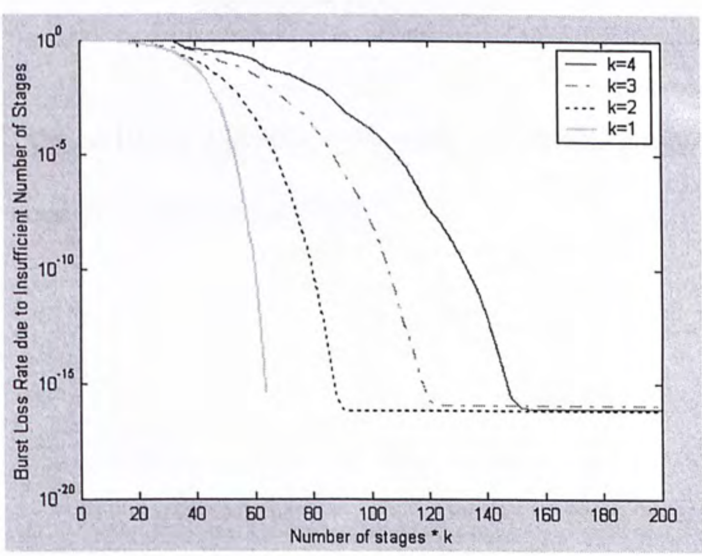

Figure 3-14 The Numerical Results of

different *k* with n=7, loading=0.7

And as showed in fig 3-15, with loading = 0.7, but n=10, k=2 and k=1 need almost same number of stages to achieve required burst loss rate 10^{-5}.

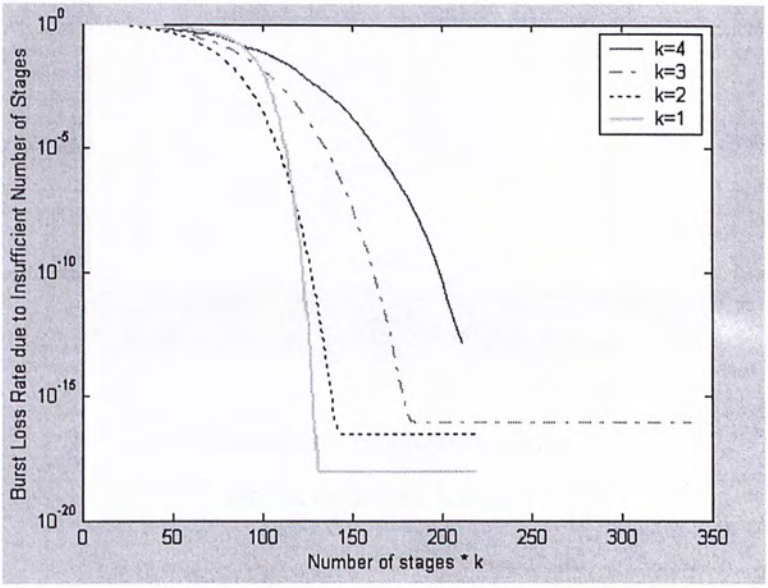

Figure 3-15 The Numerical Results of different k with n=10, loading=0.7

If the n>10, for example n=14, with loading=0.7 as showed in fig 3-16, k=2 need the least total number of stages.

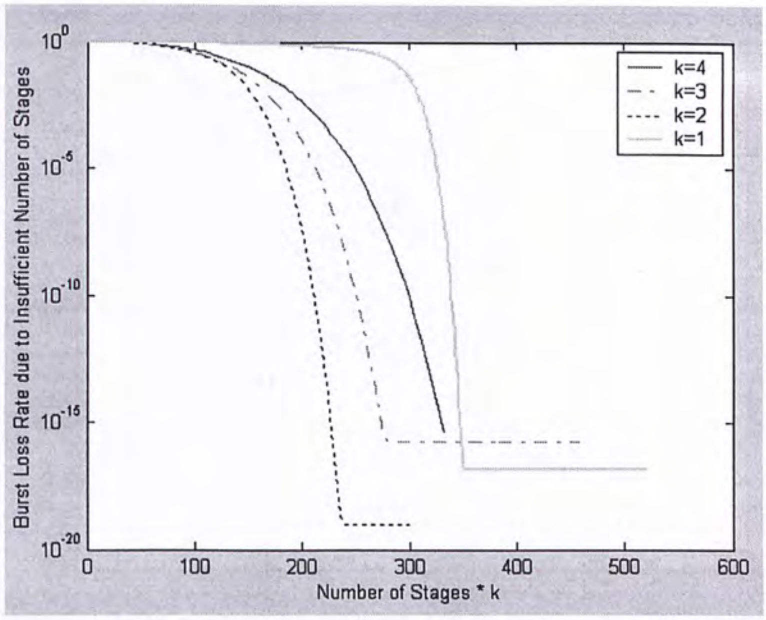

Figure 3-16 The Numerical Results of

different *k* with n=14, loading=0.7

And from above results, we know that when *n* gets bigger, the performance of multi-plane fabric compared with single plane is better. The reason is that the *n* bigger, the penalty of deflection routing for a burst bigger, the performance of single plane will be worst.

3.6.3The result of appropriate number of planes *k*

From last two sections, the performance of *k* multi-plane fabric depends on loading of SN and *n*: Log2(Number of input-output ports). Based on the numerical results of this chapter, as illustrated in the Fig 3-17, we can choose a appropriate number of planes *k* according to different loading of switch and *n*, to achieve the required burst loss rate 10^{-5} with minimal total number of stages.

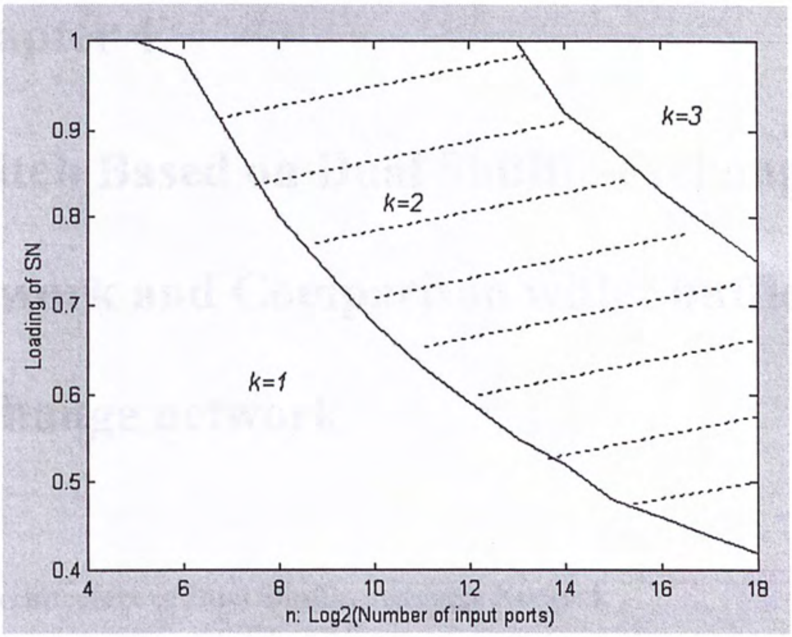

Figure 3-17 Appropriate k depending on loading and n

Chapter 4

Switch Based on Dual Shuffle-exchange network and Comparison with Shuffle-exchange network

4.1 Architecture of Dual Shuffle-exchange Network

In this section, we introduce a dual shuffle-exchange network (DSN) [9] that reduces the penalty of deflection. The dual shuffle-exchange network consists of a shuffle exchange network (SN) and an unshuffle-exchange network (USN) -the mirror image of the shuffle network, as illustrated in Fig 4-1.

Both SN and USN have the self-routed property and dual Shuffle-exchange Network is designed so that when a burst is deflected from node *i* to node *j*, we must have a link in the reverse direction connecting node *j* to node *i*, the burst can travel back to node *j* from node *i*, correcting the deflection in one step.

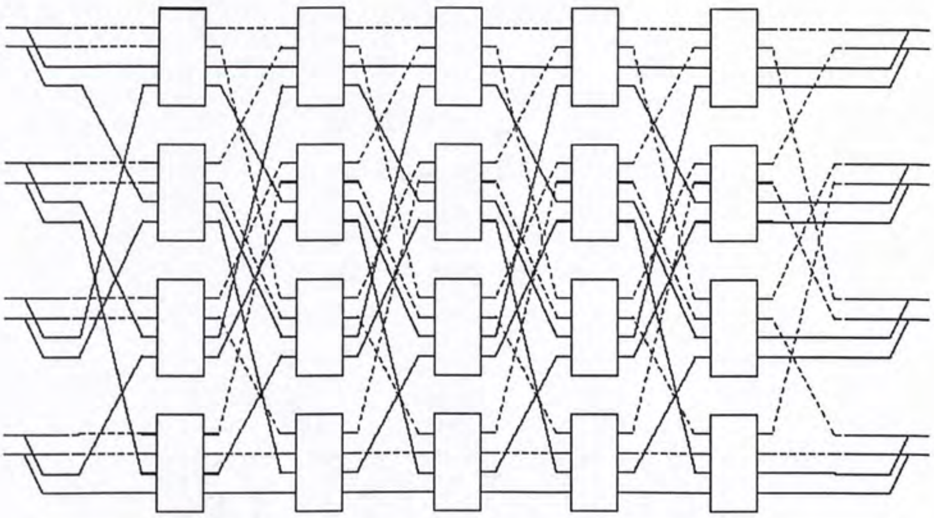

Figure 4-1 Structure of dual shuffle-exchange network

4.2 The deflection Probability in a 4×4 Switching Module of DSN

To derive the deflection probability of an internally non-blocking switching module, when a new burst arrives at this non-blocking switching module, it will see at least one output port is idle, or equivalently 0 to 3 output ports are being occupied. Assuming each burst is equally likely to destine to one of the four output ports, the burst length is exponential with mean μ and idle time is exponential with mean λ, Figure 4-2 depicts a Markov chain for this 4×4 non-blocking switching module. The states of the Markov chain are the number of busy output ports seen by a new arriving burst. After obtaining the limiting probabilities p_0, p_1, p_2, p_3

With,

$$\begin{aligned} p_0 \cdot 3\lambda &= p_1 \cdot \mu \\ p_1 \cdot 2\lambda &= p_2 \cdot 2\mu \\ p_2 \cdot 1\lambda &= p_3 \cdot 3\mu \\ p_0 + p_1 + p_2 + p_3 &= 1 \end{aligned}$$

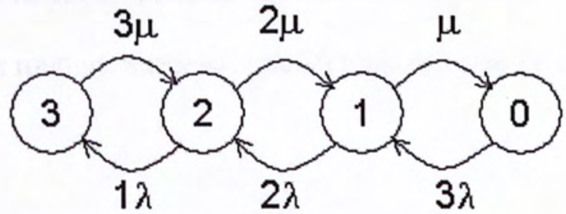

Figure 4-2 The Markov chain of switching element: state means the number of busy output ports.

The deflection probability for one 4×4 switching module q can be calculated as follows

$$q = 1 - (P_0 + 0.75P_1 + 0.5P_2 + 0.25P_3) = \frac{3}{4\rho}$$

$$\begin{aligned} &\textit{with} \\ &P_0 = (1-\rho)^3 \\ &P_1 = 3(1-\rho)^3 \frac{\rho}{1-\rho} \\ &P_2 = 3(1-\rho)^3 (\frac{\rho}{1-\rho})^2 \\ &P_3 = (1-\rho)^3 (\frac{\rho}{1-\rho})^3 \end{aligned}$$

4.3 Burst Loss Rate due to Insufficient Number of Stages of DSN

When contention occurs inside a 4×4 switching module of DSN, the loser burst will be deflected to one of the idle output ports available. A one-stage routing instruction will be added to this burst based on which output port this it is deflected. By successfully following this routing instruction in the next stage, the deflected burst can return to the state where it was deflected and resume its routing. Successive deflections can also be corrected by this algorithm.

The deflection penalty distance is reduced to one and figure 4-3 shows the state-transition diagram of this error-correcting algorithm. Each state represents the number of remaining stages that the bursts still have to go through until it reaches the destination.

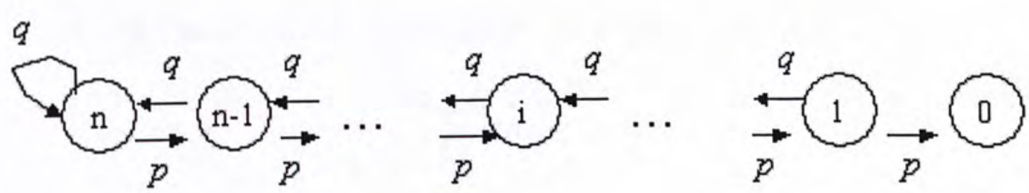

Figure 4-3 The state-transition diagram of Shuffle-exchange network

In a dual shuffle-exchange network with size $N \times N$, $n=\log N$, number of stages = L, and loading of switch = ρ.

A (i,j) is the loading of stage *j* with distance *i* steps to destination.

We have

$$A(n,0) = \rho$$

$$A(i,0) = 0,\ i = 0,1,\ldots,n-1$$

and the deflection probability of stage *j* is q_j

q_j can be got by last part, $p_j = 1 - q_j,\quad j = 1,2,\ldots,L$

From Markov chain above, we have

$$A(i,j) = p_j \cdot A(i+1, j-1) + q_j \cdot A(i-1, j-1),\ i = 2,3\ldots,n-1$$

$$A(n,j) = q_j \cdot [A(n,j-1) + A(n-1,j-1)]$$

$$A(1,j) = q_j \cdot A(2,j-1)$$

$$A(0,j) = q_j \cdot A(1,j-1)$$

and the burst loss rate due to insufficient number of stages is

$$P_{Loss} = \frac{\sum_{k=1}^{n} A(k,L)}{\rho}$$

Fig 4-4 showed the comparison of simulation results and analysis results based on above equations, with $\lambda = 4, \mu = 1$, *d*=8, *h*=64, and loading of switch $\rho = \lambda/(\lambda + \mu) = 0.8$. We can see that two curve are

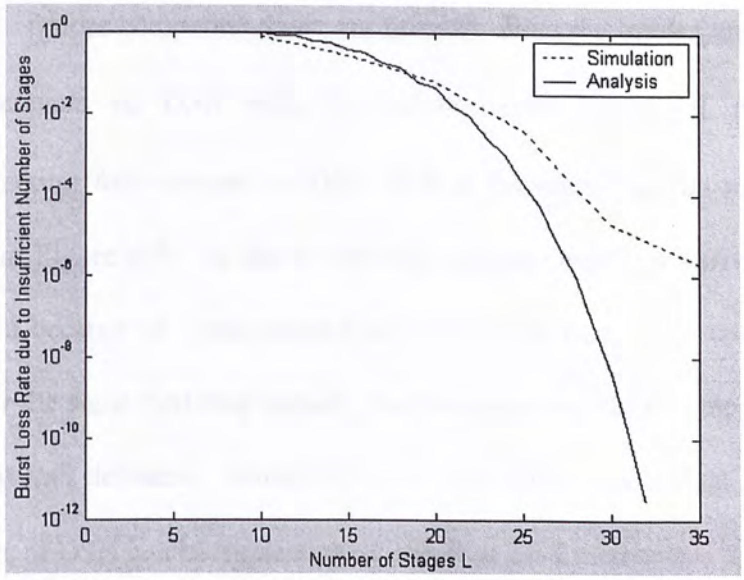

Figure 4-4 Analysis and Simulation Results

of DSN

4.4 Comparison of SN and DSN

According to our analysis results, we can compare the performance of SN and DSN. The advantage of DSN is that its deflection penalty is just one step, much less than SN. We guess that when loading is very high so that the deflection probability is big, the DSN should have better performance than SN. And we also can guess when size of switch is huge so that the deflection penalty of SN that depends on n gets very large, the DSN will have better performance.

We will compare the performance of SN and DSN in different *n* and different loading of switch.

Before comparing them, we first talk about the implementation of 4x4 elements of DSN with 2 × 2 elements. Paper [9] proposed implementing 4x4 element of DSN with a two-stage banyan switch as shown in Figure 4-5. In this a two-stage banyan switch, a burst may be deflected because of "internal conflict" even when there is no contending burst for the same "external output", but this paper proved that the increase in the overall deflection probability is actually quite small. Thus, one 4x4 elements of DSN can be implemented with four 2×2 elements.

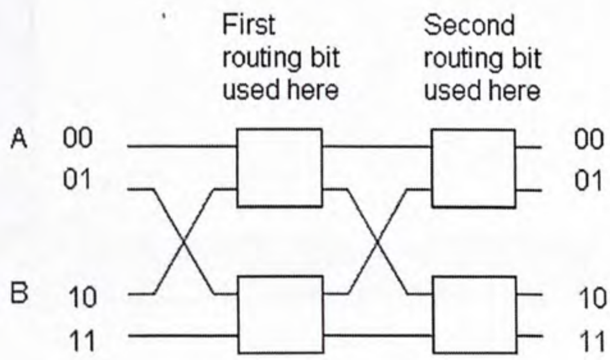

Figure 4-5 Implementation of 4x4 element of DSN with a two-stage banyan switch

Assuming the required burst loss probability is 10^{-5}, we will compare total needed 2×2 elements of switch based on DSN or SN.

4.4.1 Comparison with different *n*

With loading = 0.8, we will compare the performance of DSN and SN with n=6, 11 and 15 to achieve the required burst loss rate if be 10^{-5} as showed in Fig 4-6.

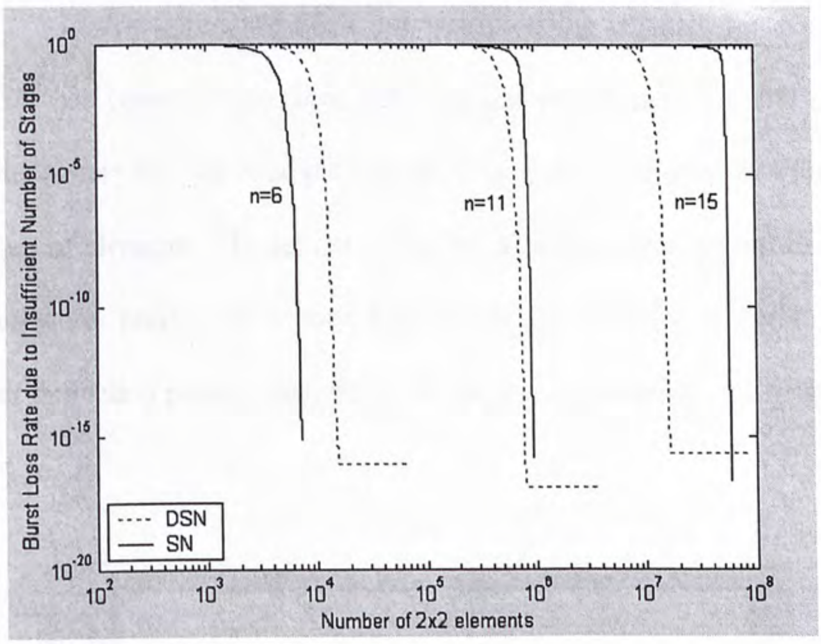

Figure 4-6 Comparison of SN and DSN with

n=6, 11 and 15, loading=0.8

In this figure, we can know when *n*<11, for example *n*=6, SN has better performance with less number of elements; but when *n*>11, for example *n*=15, DSN has better performance with less number of elements. That is because when n gets bigger, the penalty of deflection routing of switch based on SN is bigger, but the penalty of deflection routing of switch based on DSN is just one which is independent of *n*.

4.4.2 Comparison with different loading

With n=13, we compared the performance of DSN and SN with loading = 0.4 and 0.9 to achieve the required burst loss rate if be 10^{-5} as showed in Fig 4-7.

In this figure, we can know when loading is small, for example 0.4, SN has better performance with less number of elements; but when loading is very big, for example 0.9, DSN has better performance with less number of elements. The reason is that the loading bigger, the probability of deflection routing for a burst bigger, and also because of SN's much bigger deflection penalty than that of DSN, the performance of SN will be worst

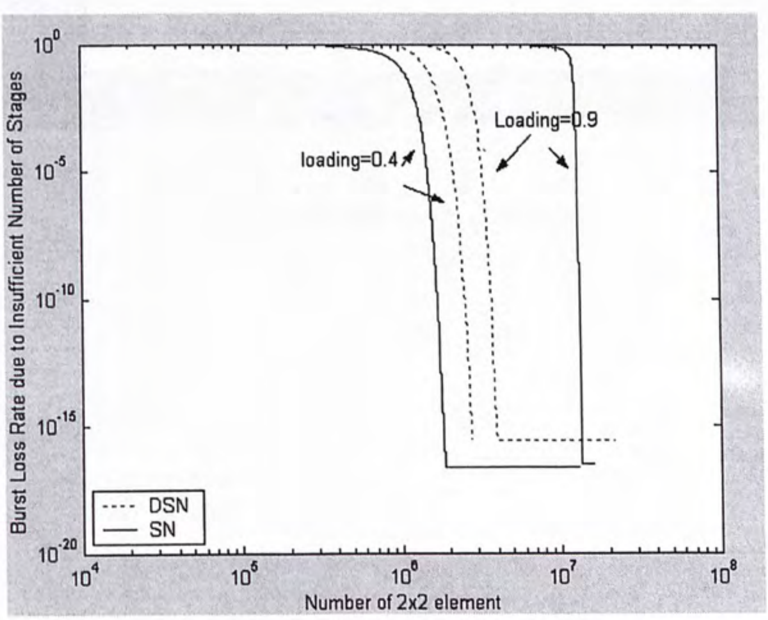

Figure 4-7 Comparison of SN and DSN with

n=13, loading=0.4 and 0.9

4.4.3The result of comparison

Based on the numerical results, as illustrated in the Fig 4-8, we can choose to use SN or DSN according to different loading of switch and n, to achieve the required burst loss rate 10^{-5} with minimal total number of 2×2 elements.

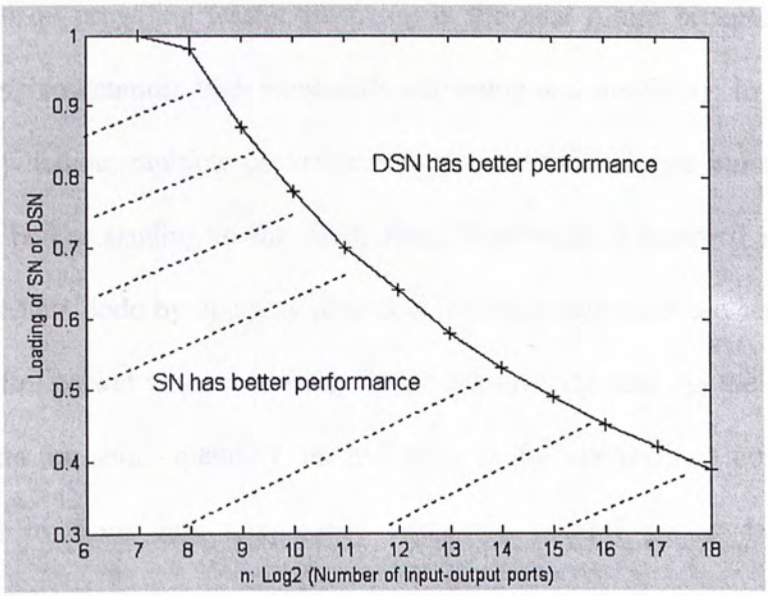

Figure 4-8 Choice for SN or DSN

Chapter 5

Conclusions

The optical burst switch (OBS) has been highly regarded as a viable solution on providing terabit switching in the near future because of its easy implementation, high bandwidth utilization and flexibility. In optical burst switching, multiple packets are aggregated into a larger burst at the source before sending to the destination. Bandwidth is reserved in each intermediate node by one-way protocols in which data burst are sent after its control packet without waiting for the acknowledgment. As the control and data are sent separately, no buffering in the intermediate nodes are needed to store data temporarily while the control packet is being processed.

A challenging issue on OBS is to reduce the blocking caused by contention at each switching node. Output contention occurs when there are too many input bursts destined for the same output fiber simultaneously. If no buffering is provided, the bursts that have lost contention must be dropped.

5.1 The Burst Loss Probability of Proposed OBS Based on SN

Firstly we describe a deflection-based network including shuffle-exchange network (SN) that can be employed in optical burst switching. This switching architecture provides a significantly reduce in complexity.

The blocking probability due to output contention was derived by two different traffic models and they are very close and both useful. The simulation results also prove the analysis is right. The burst loss rate due to insufficient number of stages of SN switching architecture was derived with numerical results and the simulation also proves it.

Total burst loss probability is given out when $h = 64, d = 8$, $n = LogN = Log(h * d) = 9$, as showed in Fig 3-8 with different loading of switch from 0.6 to 0.8. From the results, we can conclude that with loading > 0.6 when n=9, it is impossible to decrease the total loss probability to 10^{-5} or lower by just increasing the length of switch (number of stages). To achieve the lower total loss probability, we have to adopt some technology to avoid the output contention of bursts, for example, with MPLS [18] (Multi-Protocol Label Switching) technology.

5.2 The multi-plane Fabric with appropriate number of planes k

We also propose a novel multi-plane architecture based on SN. We find that with k which depends on loading of switch and size of switch, we can significantly reduce the complexity of switch to reach some loss rate bound and make the switch with great scalability.

Based on the numerical results of this chapter, as illustrated in the Fig 5-1, we can choose a appropriate number of planes *k* according to different loading of switch and *n*, to achieve the required burst loss rate 10^{-5} with minimal total number of stages.

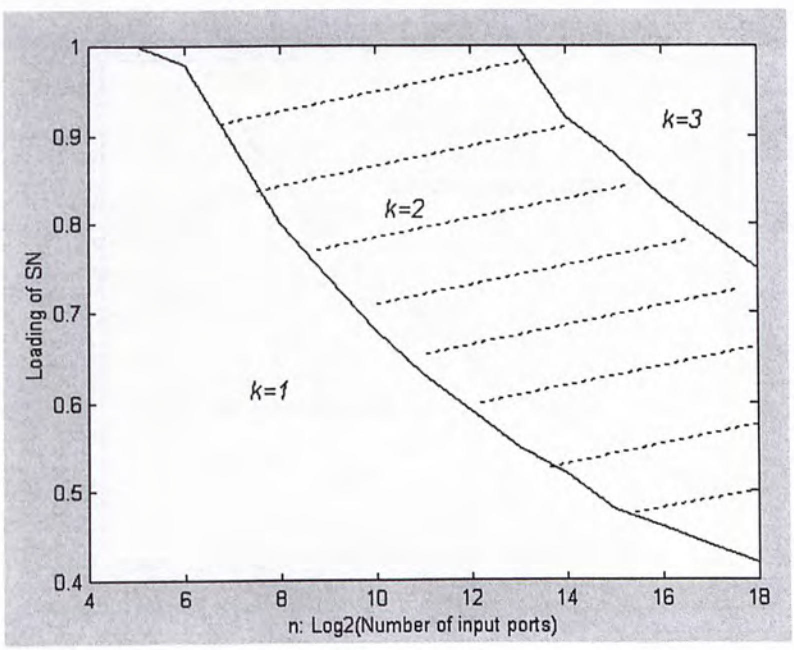

Figure 5-1 Appropriate *k* depending on loading and *n*

5.3 Performance of OBS Design Based on DSN and Comparison of SN and DSN

We also propose to use a dual shuffle-exchange network (DSN) for OBS design whose deflection penalty is just one step. But to implement the 4x4 element of DSN with 2×2 elements, we use a two-stage banyan switch as shown in Figure 4-5. Thus, one 4x4 elements of DSN can be implemented with four 2×2 elements.

Based on the numerical results, as illustrated in the Fig 5-2, we give out either SN or DSN has better performance when in different loading of switch and *n*, to achieve the required burst loss rate 10^{-5} with minimal total number of 2×2 elements.

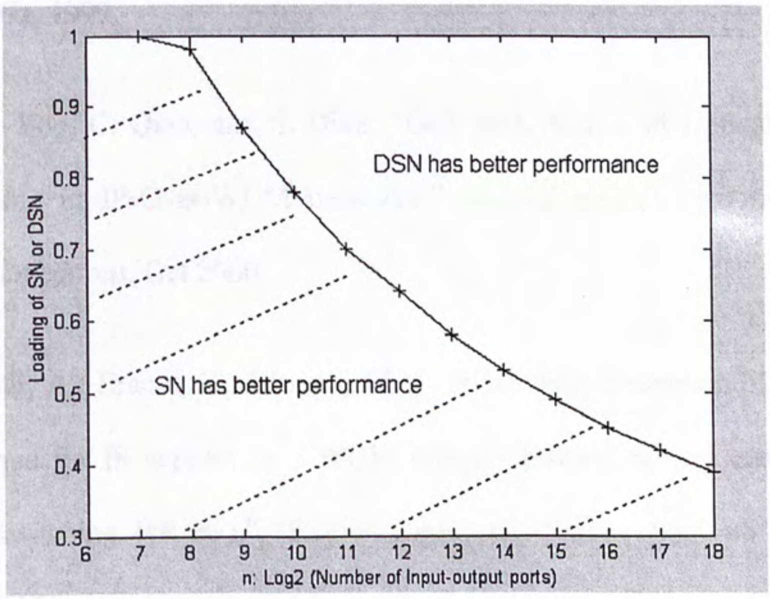

Figure 5-2 Choice for SN or DSN

Bibliography

[1] C. Qiao and M. Yoo. "Optical burst switching (OBS)-A new paradigm for an optical Internet." *Journal of High Speed Networks*, 8(1):69-84, January -999.

[2] Jonathan S. Turner, "Terabit Burst Switching," *Journal of High Speed Networks*, 1999.

[3] M. Yoo, C. Qiao, and S. Dixit, "QoS performance of Optical Burst Switching in IP-Over-WDM networks," *Journal on Selected Areas in Communications*, Oct 2000.

[4] Detti, A.; Eramo, V.; Listanti, M., "Performance evaluation of a new technique for IP support in a WDM optical network: optical composite burst switching (OCBS)," *IEEE J. Lightwave Technology*, Feb 2002 , pp.154-165

[5] Jin-Bong Chang, Chang-Soo Park, "Efficient channel-scheduling algorithm in optical burst switching architecture," *High Performance Switching and Routing, 2002. Merging Optical and IP Technologies. Workshop* on , 2002, pp. 194 -198

[6] J. Ramamirtham and J. Turner. "Design of wavelength converting switches for optical burst switching". *In proceedings of INFOCOMM*, volume 1, pages 362-370, 200

[7] F. A. Tobagi, and T. kwok, "The Tandem Banyan Switching Fabric: A Simple high-Performance Fast Packet Switch," *IEEE INFOCOM'1991*

[8] A. Krishna and B. Hajek, "Performance of Shuffle-Like Switching Networks with Deflection," *Proceeding of IEEE Infocom'90*, San Francisco, CA, pp. 473-480.

[9] Liew, S.C., Lee, T.T., "NlogN Dual Shuffle-Exchange Network with Error-correcting Routing," *IEEE Communications*, vol. 42, pp. 754-766, Feb-Apr 1994.

[10] B. Mukherjee. Optical Communication Networking. McGraw-Hill, 1997.

[11] I. Widjaja. Performance analysis of burst admission control protocols. IEEE Proceeding Of Communications, 142:7-14, February 1995

[12] J.Y. Wei and R.I. McFarland, "Just-in-time Signaling for WDM Optical Burst Switching Networks," J.Lightwave Tech., vol.18, no. 12, Dec. 2000, pp. 2019-37.

[13] K. Dolzer and C. Gauger. "On burst assembly in optical burst switching networks- a performance evaluation of Just-Enough-Time". *In*

Proceedings of the 17th International Teletraffic Congress, pages 149-161, September 2001

[14] S. Verma, H. Chaskar, and R. Ravikanth. “Optical burst switching: a viable solution for terabit IP backbone” *IEEE Network*, pages 48-53, November/December 2000.

[15] S. Yao, S. Dixit, and B. Mukherjee. “Advances in photonic packet switching: An overview.” *IEEE Communications*, 38(2):84{94, February 2000.

[16] H. M. Chaskar, S. Verma, and R. Ravikanth. “A framework to support IP over WDM using optical burst switching”. *In IEEE/ACM/SPIE Optical Network Workshop*, January 2000.

[17] M. DÄuser and P. Bayvel. “Analysis of a dynamically wavelength-routed optical burst switched network architecture”. Journal of Lightwave Technology, 20(4):574-585, April 2002.

[18] Rosen, Eric C., et al., “Multiprotocol Label Switching Architecture,”

RFC 3031, Jan 2001