Performance Analysis of Output Buffers in Multistage Interconnection Networks with Multiple Paths *

Boseob Kwon, Byungho Kim, Jaehyung Park, H. Yoon, and Jungwan Cho
Department of Computer Science
Korea Advanced Institute of Science and Technology
373-1 Kusung-Dong, Yusung-Gu, Taejon 305-701, Korea

Abstract

Multistage Interconnection Networks with multiple paths can support higher bandwidth than those of nonblocking networks by passing multiple packets to the same destination simultaneously. In the multiple path networks, the performance of the output buffer affects the whole system performance and is closely coupled with the output traffic distribution that is the packet arrival rate at each output link destined to a given output module. Many multiple path networks produce the nonuniform output traffic distribution even if the input traffic pattern is random and uniform. In this paper, performances of the output buffers for several multiple path networks are investigated by our proposed analysis model. It is shown that the output traffic distributions are different with the various multiple path networks and the output buffer performance such as packet loss probability and delay gets better as nonuniformity of the output traffic distribution becomes higher.

1 Introduction

Multistage Interconnection Networks (MINs) have been well researched as an interconnection mechanism between processors and memory modules of parallel machines [1]. Most of them are blocking networks as like banyan networks [2, 3]. In the blocking networks, packets may collide with each other in the network even if their destinations are different. Resource contentions arise if more than one cell needs to access the same link or the same internal buffer location. Although these networks have simple structure, the internal packet conflicts degrade the performance in the network as the dimension of the network becomes larger. In the nonblocking networks as like the crossbar [4] or Clos [5], packets with distinct destination addresses can be forwarded to the destinations without any contention in the network. Although internal blocking dose not occur in nonblocking networks, blocking can occur if more than one packet is destined to the same destination.

Nonblocking networks are divided into two categories according to the position of the buffers: input queueing networks with buffers at the sources of the network and output queueing networks with buffers at the destinations of the network. In the nonblocking network with input queueing, the throughput performance is approximately 58 percent under uniform traffic [5]. This degradation is due to packet blocking in the head of line buffers at the inputs. To cope with this problem, networks with larger bandwidth are required. In recent years, there have been many proposals for switching networks which take advantage of multiple paths available in the networks to improve the performance. Multiple path networks can support higher bandwidth than those of nonblocking networks by passing multiple packets to the same destination simultaneously, and these packets are buffered at the output buffer. Examples of multiple path networks based on MINs are the replicated or dilated banyan network [6], the tandem banyan network [7], the multistage shuffle network [8], and the Fly network [9].

In the multiple path networks, the performance of the output buffer such as the mean delay and the packet loss probability affects the whole system performance, and is determined by two factors: the number of output links destined to an output and the *output traffic distribution* which means the packet arrival rate at each output link. Generally, the output traffic distribution is independent of the input traffic pattern offered to the switching network, but is dependent upon the particular network architecture. For example, even if the input traffic pattern is random and uniform, many switching networks produce the nonuniform output traffic distribution and the degree of the nonuniformity varies also with the different network architectures.

While there have been many isolated studies of the network performance and the output buffer performance for individual network architecture, studies of the output traffic distribution which is tightly coupled with the performance of the output buffer have been relatively few. In this paper, we investigate the performance of the output buffers according to our proposed analysis model for several multiple path networks as like the tandem banyan network [7], the multistage shuffle network [10], the Fly network [9] and the *multiple path crossbar*.

The remainder of this paper is organized as follows. Section 2 illustrates several multiple path networks. Section 3 analyzes the output traffic distributions of

*This work was partially supported by *Center for Artificial Intelligence Research*of KAIST and *Electronics and Telecommunications Research Institute*.

each network. Section 4 proposes an analytical model for the output buffer analysis in regard to the nonuniform output traffic distribution and analyzes the output buffer performance of the networks using the proposed model. Conclusions are presented in Section 5.

2 Multiple path networks

We consider, in this section, four kinds of multiple path networks which employ different concepts to provide multiple paths: the tandem banyan network [7], the multistage shuffle network [10], the multiple path crossbar, and the Fly network [9].

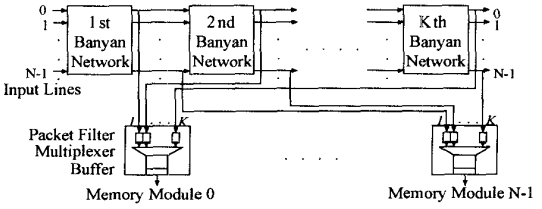

Figure 1: The tandem banyan network.

A. Tandem Banyan Network

The tandem banyan network consists of multiple banyan networks placed in series (see Figure 1), such that each output of every banyan network is connected to both corresponding input of the following network and the corresponding memory module [7]. Upon a conflict between two packets at a switching element, one of the two packets is properly routed, while the other is routed to the wrong way. The packet which gets misrouted is marked and at the end of a banyan network it is fed into the following banyan network for further processing.

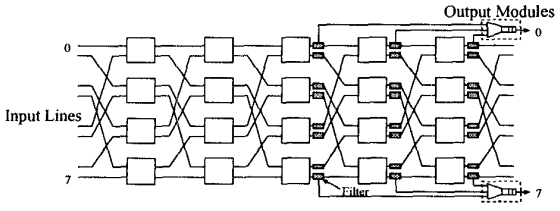

Figure 2: An 8 × 8 multistage shuffle network with 5 stages.

B. Multistage Shuffle Network

The multistage shuffle network [10] [8] is constructed using $H(\geq n = log_2 N)$ stages, each with $N/2$ unbuffered switching elements. Successive stages are interconnected through the perfect shuffle pattern. At the stages after the $log_2 N$th stage, all packets which have succeeded in reaching their desired destination proceed to the memory modules. Figure 2 shows an 8 × 8 multistage shuffle network with $H = 5$.

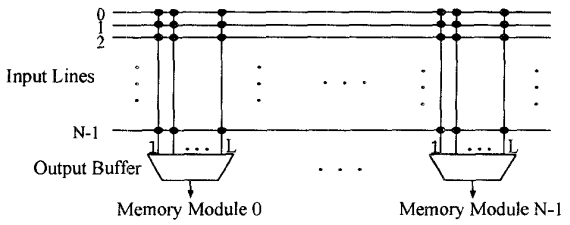

Figure 3: The multiple path crossbar.

C. Multiple Path Crossbar

The multiple path crossbar (Figure 3) provides L disjoint paths between each input line and memory module. Hence, at most L packets for the same output module can reach their destinations simultaneously. Assuming that the left most output link for a given memory module has higher priority to contain the packet when more than one packet is destined to the same memory module, the multiple path crossbar produces the most nonuniform output traffic distribution such that a multiple path network can achieve.

The multiple path crossbar requires very high network complexity of N^2L crosspoints. We consider, in this paper, the multiple path crossbar as an ideal multiple path network.

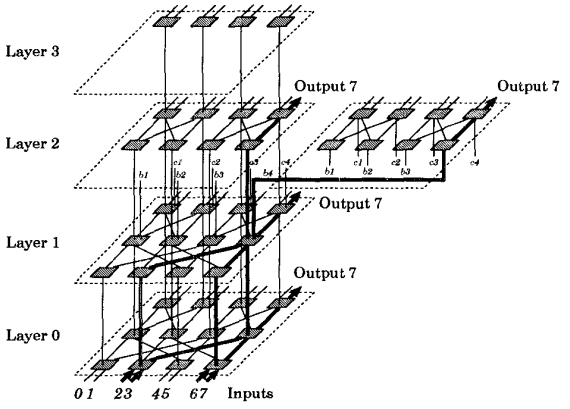

Figure 4: An 8 × 8 4-layered Fly network.

D. Fly Network

The Fly network [9] is a self-routing and nonblocking network architecture based on the fundamental property that a k by $k+1$ switching element is nonblocking. The Fly network consists of multiple banyan networks connected in three dimensional direction. Each switching element consists of three inlets and four outlets, in which two inlets and two outlets are used for the horizontal connections within a banyan switching plane and remaining one inlet and two outlets are used for the vertical inter-layer connections.

Figure 4 shows an 8 × 8 4-layered Fly network and

the routing example. The thick links in that figure illustrate four disjoint routing paths for the four packets arrived in the 2nd, the 3rd, the 6th, and the 7th input line respectively, all of which are destined to the 7th memory module simultaneously.

3 Analysis of Output Traffic Distributions

We investigate output traffic distributions of each networks illustrated in the last section. Let us define several parameters related to the the output traffic distribution :

$\mathcal{L}$: The number of output links destined to a memory module.

$\mathcal{T}(i)$: The packet arrival rate on the ith output link, for $1 \leq i \leq \mathcal{L}$.

$\alpha(i)$: The probability of having i packet arrivals at a time slot all destined for a given memory module, for $0 \leq i \leq \mathcal{L}$.

Then, the total throughput on the switching network is represented by $\sum_{i=1}^{\mathcal{L}} \mathcal{T}(i)$.

3.1 Output Traffic Distributions

At first, we compute $\mathcal{T}(i)$ of each network. Our analysis assumes that the offered input traffic to the switching network is uniform.

A. Tandem Banyan Network

For an n-stage banyan network built of k by k unbuffered switching elements, the probability p_m that there is a packet on any particular input at the mth stage of the network is approximated by [11]

$$p_m = \frac{2k}{(k-1)m + \frac{2k}{p}}, \quad (1)$$

where p is the input load. Assuming that the traffic of each banyan network in the tandem banyan switch is uniform, we can apply the above equation to a banyan network of the tandem banyan switch. The input load for the ith banyan network is defined as the remained packets which get misrouted in the preceding banyan networks. Thus, the throughput at the ith banyan network in the tandem banyan network with $m = log_2 N$ stages and $k = 2$ is obtained by

$$\mathcal{T}(i) = \frac{4}{log_2 N + \frac{4}{p - \sum_{j=1}^{i-1} \mathcal{T}(j)}}, \; for\ 1 \leq i \leq K. \quad (2)$$

B. Multistage Shuffle Network

The analysis of the multistage shuffle network is based on Awdeh's work [8]. In the multistage shuffle network (see Figure 2), packets are delivered to the addressed output out of the stages n $(= log_2 N)$ to H. A packet in the switching network has own counter field C, which represents the number of remained stages to the correct output. Thus, a packet can be in any of $n+1$ states, ranging from $C = n$ (which corresponds to a packet n stages far from its destination), to $C = 0$ (which corresponds to a packet that has reached its destination, *or there is no packet*). The corresponding state transition diagram is shown in Figure 5. A packet in state i $(1 \leq i \leq n)$ can move to state $i-1$ in case of correct routing, or to state n in case of deflection.

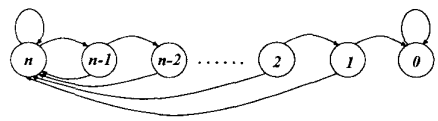

Figure 5: The state transition diagram of a packet in the multistage shuffle network.

From $q(i,k)$ which is the probability that an output link of stage k $(0 \leq k \leq H)$ carries a packet with $C = i$ $(0 \leq i \leq n)$ [8], $\mathcal{T}(i)$ for the multistage shuffle network is obtained by

$$\mathcal{T}(i) = q(0, n+i-1) - q(0, n+i-2) \quad for\ 1 \leq i \leq H-n+1. \quad (3)$$

C. Multiple Path Crossbar

The output traffic distribution of the multiple path crossbar can be obtained by the probabilistic model used in the analysis of the Knockout switch [12]. For instance, considering the first output link between $\mathcal{L}$ output links, we can assume that if there is at least one packet destined to the memory module, the first output link always delivers the packet. Accordingly, the second output link can deliver a packet when there are at least two packets destined to the memory module. Then, assuming that the input pattern is uniform, each output traffic on the ith output link is obtained by

$$\mathcal{T}(i) = 1 - \sum_{j=0}^{i-1} \left[\binom{N}{j} (\frac{p}{N})^j (1 - \frac{p}{N})^{N-j} \right] \quad for\ 1 \leq i < \mathcal{L}, \quad (4)$$

where p is the offered input load.

D. Fly Network

Considering a 3 by 4 switching element of the Fly network, the probability that there is a packet at each four outlets in a switching element is obtained by [9]

$$\rho_{o1} = \rho_{o2} = 1 - (1 - \frac{\rho_{i1}}{2})(1 - \frac{\rho_{i2}}{2})(1 - \frac{\rho_{i3}}{2}), \quad (5)$$

$$\rho_{o3} = \frac{1}{2}(\rho_{i1} + \rho_{i2})\rho_{i3} + \frac{1}{2}\rho_{i1}\rho_{i2} - \frac{3}{8}\rho_{i1}\rho_{i2}\rho_{i3}, \quad (6)$$

$$\rho_{o4} = \frac{1}{8}\rho_{i1}\rho_{i2}\rho_{i3}. \quad (7)$$

Applying the above equations to the whole switching network recurrently, we can obtain $T(i)$.

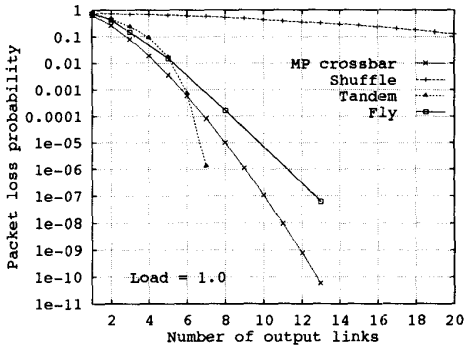

Figure 6: Number of output links versus packet loss probabilities in 1024 × 1024 multiple path networks.

3.2 Analysis Results

In the first place, we might decide the appropriate number of output links for each network, because the number of output links provided by a multiple path network is tightly coupled with the throughput of the switching network which is then identical to the traffic load at the output buffer. As shown in Figure 6 representing the number of output links versus the probability of packet loss, if we consider the probability of packet loss less than 10^{-7} with 100 percent load or 10^{-9} with 80 percent load, the corresponding number of output links for each network is 8 for the tandem banyan network, 45 for the multistage shuffle network, 11 for the multiple path crossbar, and 13 for the Fly network. These are used for the fair comparison of the performance of the output buffer in the following part.

Figure 7 shows the output traffic distributions of four networks. The *output link* is just an index which is extraneous to the particular position, and the *throughput* is obtained by $T(i)$ for each network. In the figures, if we regard the degree of the nonuniformity as the concentration or deviation of the curve, it can be respected that the curve of the multiple path crossbar is most nonuniform, and that of the Fly network follows. We also notice that the patterns of the output traffic distribution are consistent for the various loads and the network dimensions as shown in the figures.

4 Performance of Output Buffers

The performance of the output buffer can be regarded as a function of the number of output links and the output traffic distribution at the output links. In this section, we propose an analytical model for the output buffer analysis, which is not dependent upon a particular network architecture but applicable to the general multiple path networks.

4.1 Output Buffer Analysis

Input arguments for the analysis of the output buffer are $\mathcal{L}$ and $T(i)$. We model the output buffer

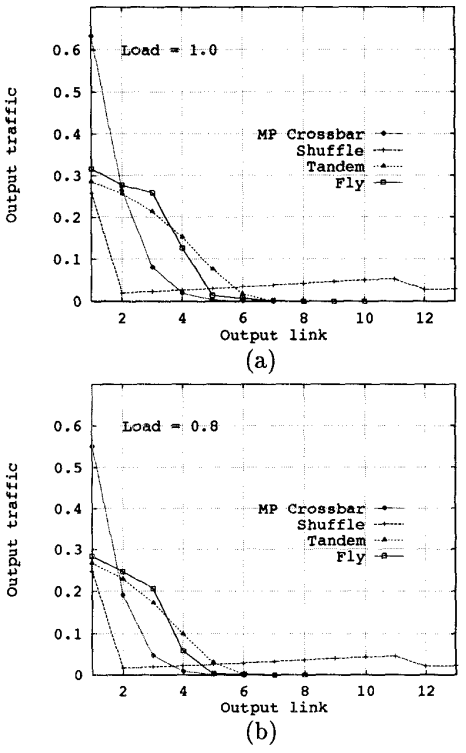

Figure 7: Output traffic distributions of 1024 × 1024 networks.

as a discrete time Markov chain. With the buffer size of m, let Q_t denote the number of packets in a given output buffer just after the tth time slot, and A_t denote the number of packet arrivals during tth time slot, then,

$$Q_t = \min(\max(0,\ Q_{t-1} + A_t - 1),\ m). \quad (8)$$

Let Q denote the steady-state buffer size obtained from Q_t. Figure 8 shows the Markov chain state transition diagram. The transition probabilities are in terms of $\alpha(i)$. Then, $\alpha(i)$ for the nonuniform output traffic distribution is obtained by

$$\alpha(i) = \sum_{\Omega(\mathcal{L},i)} \left(\prod_{j=1}^{\mathcal{L}} T(j)^{\omega_j} \cdot (1 - T(j))^{1-\omega_j} \right) \quad for\ \ 0 \le i \le \mathcal{L}, \quad (9)$$

where

$$\Omega(\mathcal{L},i) = \{(\omega_1, \omega_2, \ldots, \omega_{\mathcal{L}}) \mid \sum_{j=1}^{\mathcal{L}} \omega_j = i,\ \omega_j = \{0,1\}\}.$$

However, in case of the uniform output traffic distri-

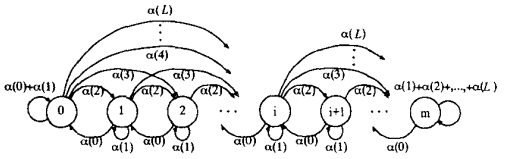

Figure 8: Markov chain state transition diagram of the output buffer.

bution, i.e., if all of $\mathcal{T}(i)$ are identical, $\alpha(i)$ has the binomial probabilities and is obtained by [12]

$$\alpha(i) = \binom{N}{i} (\frac{p}{N})^i (1 - \frac{p}{N})^{N-i} \; for \; 0 \leq i \leq \mathcal{L}, \quad (10)$$

where p is the offered input load. We obtain the steady-state buffer size probabilities directly by the transition matrix for the Markov chain from the q_i. The normalized throughput ρ is then [13]

$$\rho = 1 - q_0 \alpha(0), \quad (11)$$

and the probability of the packet loss in the output buffer is

$$Loss = 1 - \frac{\rho}{\lambda}, \; where \; \lambda = \sum_{i=1}^{\mathcal{L}} \mathcal{T}(i). \quad (12)$$

The mean waiting time in the output buffer is obtained by

$$\overline{W} = \frac{1}{\rho} \sum_{i=1}^{m} i q_i. \quad (13)$$

4.2 Performance Results

Figure 9 shows the packet loss probability versus the output buffer size curves. These curves are obtained by Equation (12) with the following input parameters; $\mathcal{L}$ is 11, 45, 8, and 13 for the multiple path crossbar, the multistage shuffle network, the tandem banyan network, and the Fly network, respectively. In that figure, to achieve packet loss probability less than 10^{-9} for 1024 × 1024 networks with the 80 percent load, the Fly network requires 32 packet buffers, while the tandem banyan network and the multistage shuffle network require 33 and 39 packet buffers respectively. However, those of the multiple path crossbar are 23, which is the ideal performance as a multiple path network can achieve. Such results are analogous to the analysis results of the output traffic distribution consistently, therefore, we can assert that the performance of the output buffer increases as the nonuniformity of the output traffic distribution becomes higher.

Figure 10 shows the load versus delay curve (obtained by Equation (13)) of the networks with the buffer size 40. The figure shows that all of the networks have small delay up to load 0.9 and the delay variance among the networks is very small.

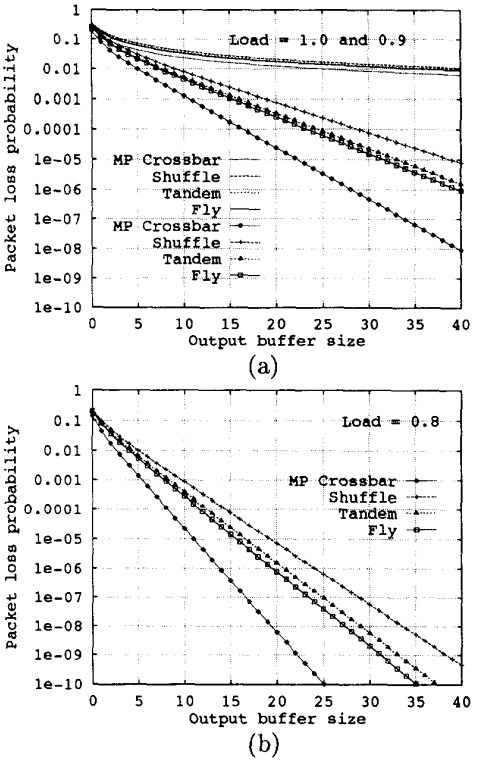

Figure 9: Packet loss probability versus output buffer size.

5 Conclusions

This paper has proposed an analytical model for the output buffer analysis in multiple path networks with respect to the nonuniform output traffic distribution, and has shown that the performance of the output buffer increases as the nonuniformity of the output traffic distribution becomes higher. To demonstrate our proposed performance model, we investigated the nonuniformity of the output traffic distribution for the representative multiple path networks: the multistage shuffle network, the tandem banyan network, the multiple path network, and the Fly network. The Fly network has the best throughput and delay performance of the output buffer by providing a high degree of nonuniform output traffic distribution, except the multiple path crossbar which shows an ideal output buffer performance such that a multiple path network can achieve. Although not considered in this paper, it is also applicable under nonuniform input loads such as a hot spot traffic or a bursty traffic. We may expect better output buffer performance under the nonuniform input traffic pattern rather than under the uniform one.

References

[1] G. Broomell and J. R. Health, "Classification Categories and Historical Development of Circuit

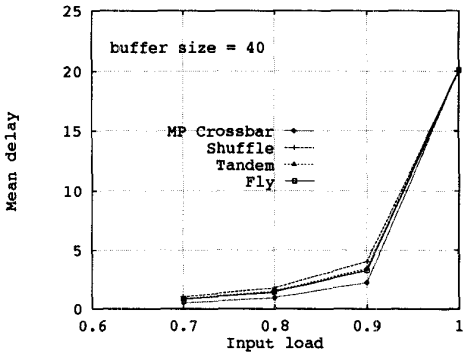

Figure 10: Load versus delay in the output buffer.

Switching Topologies," *ACM Computing Surveys*, Vol. 15, No. 2, Jun., 1983.

[2] L. R. Goke and G. J. Lipovski, "Banyan Networks for Partitioning Multiprocessor System," *Proc. 1st Annu. Symp. Comput. Arch.*, 1973, pp. 21-28.

[3] H. Yoon, K. Y. Lee and M. T. Liu, "Performance Analysis of Multibuffered Packet-Switching Networks in Multiprocessor Systems," *IEEE Transaction on Computers*, Vol. 39, No. 3, Mar., 1990, pp. 319-327.

[4] S. Nojima, et. al., "Integrated Services Packet Networking Using Bus Matrix Switch," *IEEE Journal of Selected Areas in Communications*, Vol. SAC-5, No. 8, Oct., 1987, pp. 1284-1292.

[5] Mark J. Karol, Michel G. Hluchjy, and Samuel P. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Transaction on Communications*, Vol. COM-35, No. 12, Dec., 1987, pp. 1347-1356.

[6] Ted H. Szymanski, and V. Carl Hamacher, "On the Permutation Capability of Multistage Interconnection Networks," *IEEE Transaction on Computers*, Vol. C-36, No. 7, Jul., 1987, pp. 810-822.

[7] Fouad A. Tobagi, Timothy Kwok, and Fabio M. Chiussi, "Architecture, Performance, and Implementation of the Tandem Banyan Fast Packet Switch," *IEEE Journal of Selected Areas in Communications*, Vol. 9, No. 8, Oct., 1991, pp. 1173-1193.

[8] Ra'ed Y. Awdeh and H. T. Mouftah, "Design and Performance Analysis of an Output Buffering ATM Switch withh Complexity of $O(Nlog_2N)$," *INFOCOM '94*, 1994, pp. 420-424.

[9] Byungho Kim, Boseob Kwon, Jinchun Kim, Hyunsoo Yoon, and Jungwan Cho, "Performance Analysis of an ATM Switch with Multiple Paths," *Proc. International Conference on Network Protocols*, Nov., 1995.

[10] S. Bassi, M. Dècina, P. Giacomazzi, and A. Pattavina, "Multistage Shuffle Networks with Shortest Path and Deflection Routing for High-Performance ATM Switching: The Open-Loop Shuffleout," *IEEE Transaction on Communications*, Vol. 42, No. 11, Oct., 1994, pp. 2881-2889.

[11] Clyde P. Kruskal and Marc Snir, "The Performance of Multistage Interconnection Networks for Multiprocessors," *IEEE Transaction on Computers*, Vol. C-32, No. 12, Dec., 1983, pp. 1091-1098.

[12] Yu-Shuan Yeh, Michael G. Hluchyj, and Anthony S. Acampora, "The Knockout Switch: A Simple, Modular Architecture for High-Performance Packet Switching," *IEEE Journal of Selected Areas in Communications*, Vol. SAC-5, No. 8, Oct., 1987, pp. 1274-1283.

[13] Hyoung S. Kim and Alberto Leon-Garcia, "A Self-Routing Multistage Switching Network for Broadband ISDN," *IEEE Journal of Selected Areas in Communications*, Vol. 18, No. 3, Apr., 1990.