Feedforward SDL Constructions of Output-buffered Multiplexers and Switches with Variable Length Bursts

Yi-Ting Chen, Cheng-Shang Chang, Jay Cheng, and Duan-Shin Lee
Institute of Communications Engineering
National Tsing Hua University
Hsinchu 30013, Taiwan, R.O.C.
Email: ytchen@gibbs.ee.nthu.edu.tw
cschang@ee.nthu.edu.tw
jcheng@ee.nthu.edu.tw
lds@cs.nthu.edu.tw

***Abstract*— In this paper, we study the problem of exact emulation of two types of optical queues: (i) N-to-1 output-buffered multiplexers with variable length bursts, and (ii) $N \times N$ output-buffered switches with variable length bursts. For both queues, the delay of a packet is known upon its arrival. As such, one can emulate such queues by finding a delay path that yields the exact delay for each packet. For emulating the delay of a packet in such queues, in this paper we consider a mutlistage feedforward network with optical crossbar Switches and fiber Delay Lines (SDL). For any fixed delay d, there exist multiple delay paths in such a network. A delay path is feasible if it satisfies the following three constraints: (i) conflict constraint: no more than one packet can be scheduled at the same input/output ports of each crossbar switch at the same time, (ii) causality constraint: no packet can be scheduled before its arrival, and (iii) strong contiguity constraint: packets in the same burst should be routed through any fiber delay lines contiguously. By the worst case analysis, we find sufficient conditions for the numbers of delay lines needed in each stage of such a feedforward network to achieve exact emulation of both queues. For N-to-1 output-buffered multiplexers, our sufficient conditions are also necessary when each burst contains exactly one packet. By computer simulation, we also show that the number of delay lines in each stage can be greatly reduced due to statistical multiplexing gain.**

I. Introduction

One of the key challenges for optical packet switching is to construct optical buffers needed for conflict resolution. Unlike electronic packets, optical packets cannot be easily stopped, stored, and forwarded. A typical way to construct an optical buffer is to route optical packets through a set of optical Switches and fiber Delay Lines (SDL) so that optical packets appear at the right time and at the right place. Recent advances in SDL constructions have shown several promising results for multiplexers in [1]–[5], FIFO queues in [6], linear compressors, non-overtaking delay lines, and flexible delay lines in [7], and priority queues in [8][9].

In this paper, we address the problem of exact emulation of two types of optical queues: (i) N-to-1 output-buffered multiplexers with variable length bursts, and (ii) $N \times N$ output-buffered switches with variable length bursts. For both queues, the delay of a packet (in a burst) is known upon its arrival. As such, one can emulate such queues by finding a delay path that yields the exact delay for each packet. For emulating the delay of a packet in such queues, in this paper we propose a generalization of the mutlistage feedforward network in [3]. Instead of using a single routing path for each delay in [3], we allow multiple delay paths in such a network. Moreover, routing of each delay path can be found by the r-ary representation of the desired delay. By so doing, we not only allow these multiple delay paths to be shared by packets destined for different outputs, but also enable self-routing for packets. As such, both the construction complexity and the routing complexity can be greatly reduced.

To achieve exact emulation, we need to find a *non-conflicting* delay path when a burst of packets arrives. One of the main contributions in this paper is to find sufficient conditions for the numbers of delay paths needed for exact emulation of both N-to-1 output-buffered multiplexers and $N \times N$ output-buffered switches. This is done by the worst case analysis that identifies the minimum distance between two packets routed to the same set of delay lines. For N-to-1 output-buffered multiplexers, our sufficient conditions are also necessary when each burst contains exactly one packet.

Even though our results are derived for feedforward networks, they can be easily extended to feedback systems constructed by a single switch. Such a feedback system, originally proposed by Karol [10], was intended for finding a good approximation of an output-buffered switch. Here we strengthen the result by giving specific conditions on selecting the lengths of the delay lines that can be used for exact emulation of an output-buffered switch.

As our sufficient conditions are based on the worst case analysis, its complexity is still high. For the engineering purpose, one might only need to consider the average case. For this, we preform various computer simulations and show that the number of delay paths can be greatly reduced due to statistical multiplexing gain.

The paper is organized as follows: in Section II, we first introduce basic assumptions and the architecture of the feedforward network. By the worst case analysis, we then provide a sufficient condition for exact emulation of an N-to-1 multiplexer with variable length bursts. In Section III, we consider the constructions of $N \times N$ output-buffered switches. There are three types of constructions: direct construction, feedforward construction, and feedback construction. In Section IV, we perform various simulations for the average case by exploiting statistical multiplexing gain. The paper is concluded in Section V, where we address possible extensions of our work.

II. Constructions of N-to-1 Output-buffered Multiplexers with Variable Length Bursts

A. Basic Assumptions

In this paper, we partition time into slots and assume that packets are of the same size such that each packet can be transmitted within a time slot. We further assume that a burst consists of an integer number of fixed size packets. Also, each burst length is known when a burst arrives. To do this, we can add the burst length information in the header of each burst or transmit the information in another channel (see e.g., [11] [12]).

A packet entering a fiber delay line with d units of delay can be accessed again after d time slots. Therefore, we can use the fiber delay lines to store optical packets. An $M \times M$ crossbar switch can realize all $M!$ permutations between its inputs and outputs. A network element constructed by crossbar Switches and fiber Delay Lines is called an *SDL* element.

B. Lindley's Recursion

An N-to-1 FIFO multiplexer subject to arrivals of variable length bursts can be regarded as a discrete-time G/G/1 queue. There are at most N arriving bursts in any time slot, and an arriving burst is attached to the tail of the queue if the buffer is not full. Otherwise, an arriving burst is lost if there is not enough space in the buffer to accommodate the whole burst of packets. At each time slot, a packet departs if there are still packets in the buffer. Let ℓ_k (resp. τ_k, x_k) be the burst length (resp. arrival time, burst delay) for the k^{th} burst. Then it is well-known that the burst delay in a G/G/1 queue is governed by the following Lindley recursion:

$$x_k = (x_{k-1} + \ell_{k-1} - (\tau_k - \tau_{k-1}))^+ \tag{1}$$

where $(a)^+$ denotes $\max(0, a)$. Define a busy period of an N-to-1 multiplexer as the period of time that there are packets stored in the N-to-1 multiplexer. If the first k bursts are in the same busy period, then the k^{th} burst must arrive before the departure of the $(k-1)^{th}$ burst, i.e., $x_{k-1} + \ell_{k-1} + \tau_{k-1}$. Therefore, we have

$$x_k = x_{k-1} + \ell_{k-1} - (\tau_k - \tau_{k-1}). \tag{2}$$

By recursively expanding (2), we further have

$$x_k = x_1 + \sum_{s=1}^{k-1} \ell_s + \tau_1 - \tau_k \tag{3}$$

if the first k bursts are in the same busy period. Note that the first burst arrives when the multiplexer is empty. The delay of the first burst is thus 0. In view of (3), the delay for each burst is known upon its arrival.

C. A Feedforward SDL Network

As pointed out in [4], there are three constraints that need to be satisfied when we schedule the bursts.

(i) Conflict constraint: no more than one packet can be scheduled at the same input/output ports of each crossbar switch at the same time.

(ii) Causality constraint: no packet can be scheduled before its arrival.

(iii) Contiguity constraint: packets from the same burst should be scheduled so that they leave the system contiguously.

To avoid segmenting or reassembling bursts, here we use a stronger constraint than (iii).

(iiiA) Strong contiguity constraint: packets in the same burst should be routed through any fiber delay lines contiguously.

One of our main results in this section is to construct a self-routing discrete-time N-to-1 multiplexer with variable length bursts satisfying the above three constraints, *i.e.*, (i) (ii) and (iiiA). As shown in Figure 1, we propose an architecture consisting of M stages of SDL units and a bufferless multiplexer (i.e., a crossbar switch). In each stage, there is a crossbar switch with fiber delay lines connected to the next stage. The fiber delay lines in stage i consists of r bundles, indexed from $j = 0, 1, 2, , \ldots, r-1$. The length of the delay lines of the j^{th} bundles in stage i is jr^{i-1}. Let D_{ij} be the j^{th} bundle of fiber delay lines in stage i and $|D_{ij}|$ be the number of delay lines in D_{ij}. Then $D_i = \sum_{j=0}^{r-1} |D_{ij}|$ represents the total number of delay lines in stage i. Therefore, the first crossbar switch consists of N input ports and $N - 1 + D_1$ output ports (with the additional $N-1$ output ports for routing loss packets due to buffer overflow). For $i = 2, \ldots, M$, the i^{th} crossbar switch has D_{i-1} input ports and D_i output ports. The last crossbar switch, as a bufferless multiplexer, has D_M input ports and a single output port.

The delay of a path in such a feedforward network is the sum of the delays of the fiber delay lines along the path. Note that the maximum delay among all the paths in such a feedforward network is $r^M - 1$, which is the delay of the path by taking a delay line in the r^{th} bundle in each stage. Since the delay of a burst of packets is known when it arrives, we can route a burst of packets using the r-ary representation of its delay, as long as the delay does not exceed $r^M - 1$. Specifically, suppose the delay of a burst of packets is x for some $x \leq r^M - 1$. Using the r-ary representation, we can write

$$x = \sum_{i=1}^{M} I_i(x) r^{i-1}, \tag{4}$$

where $I_i(x) = 0, 1, 2, \ldots$ or $r-1$ for each $1 \leq i \leq M$. We route the burst to a delay line of the $I_i(x)^{th}$ bundle in stage i,

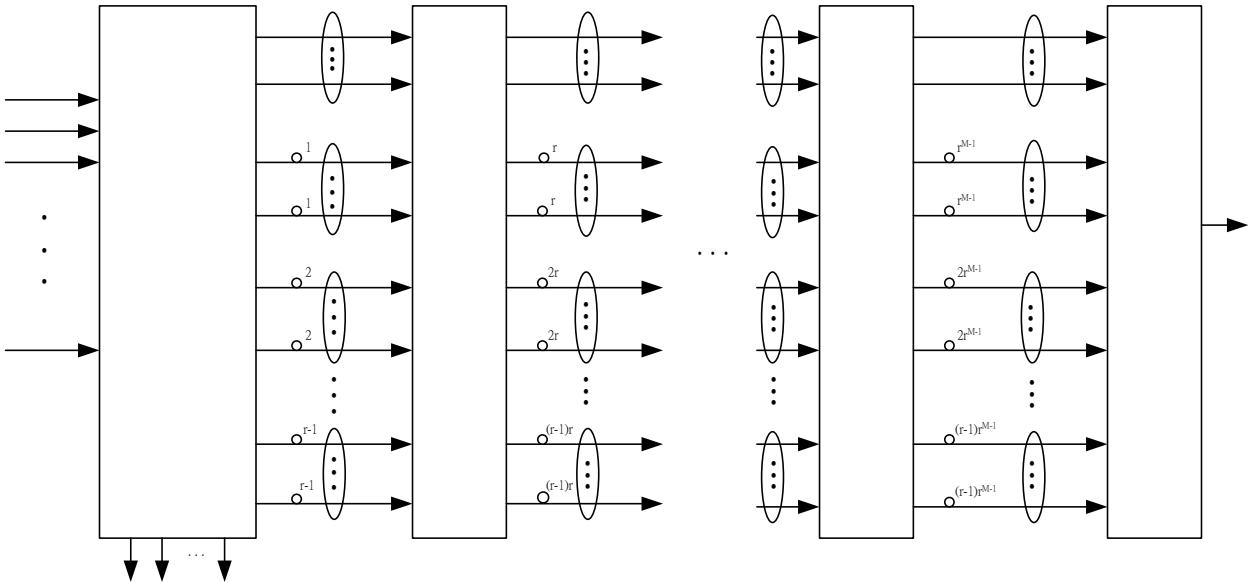

Fig. 1. A feedforward construction of a discrete-time N-to-1 multiplexer with variable length bursts

for $i = 1, 2, \ldots, M$. To be precise, suppose the burst arrives at time t with burst length ℓ. The burst that contains ℓ packets is first routed to a delay line in $D_{1I_1(x)}$ during $[t, t+\ell-1]$. The burst is then routed to a delay line in $D_{iI_i(x)}$ in stage i at time $[t + \sum_{s=1}^{i-1} I_s(x)r^{s-1}, t + \sum_{s=1}^{i-1} I_s(x)r^{s-1} + \ell - 1]$ for $2 \leq i \leq M$. Thus, the burst departs the feedforward network at the right time and at the right place. Otherwise, if $x \geq r^M$, then an arriving burst is routed to one of the lost ports immediately. Clearly, under the self-routing rule, every packet in a burst has the same delay and is routed through the same path. As such, the causality constraint and the strong contiguity constraint are satisfied. The problem left is whether there are enough fiber delay lines in each bundle so that the conflict constraint can still be satisfied.

D. A Sufficient Condition

In this section, we provide a sufficient condition for the number of fiber delay lines needed to construct a discrete-time N-to-1 multiplexer with variable length bursts. For this, we need to compute the maximum number of packets that can be routed to D_{ij} in a time slot. This is shown in the following theorem.

Theorem 1 *Suppose that the burst lengths are bounded by $\ell_{\max}$ and that the feedforward network in Figure 1 is started from an empty system. If*

$$|D_{ij}| \geq \min\left(r^{M-i}, Nr^{i-1}, \left\lceil \frac{N}{r} + \frac{2(N-1)(\ell_{\max}-1)}{r^i} \right\rceil\right), \quad (5)$$

then under the self-routing rule it is a discrete-time N-to-1 FIFO multiplexer with buffer $r^M - 1$ for variable length bursts.

For the proof of Theorem 1, we need the following lemmas. Define the k^{th} packet of a busy period as the k^{th} *departing* packet in its busy period. Suppose that the ${k_1}^{th}$ (resp. ${k_2}^{th}$) packet arrives at time t_1 (resp. t_2) with delay x_1 (resp. x_2). Then, we have

$$k_2 - k_1 = (t_2 + x_2) - (t_1 + x_1) \quad (6)$$

as both sides of (6) are simply the interdeparture time of these two packets.

Lemma 2 *Suppose the ${k_1}^{th}$ packet of a busy period is routed to a delay line in D_{ij} at time t. The ${k_2}^{th}$ packet of the same busy period is routed to a delay line in D_{ij} at time t if and only if the ${k_2}^{th}$ packet arrives in the time interval $[t - r^{i-1} + 1, t]$ and $k_2 - k_1 = cr^i$ for some integer c.*

Proof. Without loss of generality, we assume the ${k_1}^{th}$ (resp. ${k_2}^{th}$) packet arrives at time t_1 (resp. t_2) with delay x_1 (resp. x_2).

($\Rightarrow$) Since the two packets are routed to D_{ij} at time t, we have

$$t_1 + \sum_{s=1}^{i-1} I_s(x_1)r^{s-1} = t_2 + \sum_{s=1}^{i-1} I_s(x_2)r^{s-1} = t. \quad (7)$$

Thus,

$$t_2 = t - \sum_{s=1}^{i-1} I_s(x_2)r^{s-1} \geq t - r^{i-1} + 1.$$

This shows that the arrival time of packet k_2 is in $[t - r^{i-1} + 1, t]$.

Using (6), (4), and (7), we have

$$\begin{aligned}
k_2 - k_1 &= (t_2 + x_2) - (t_1 + x_1) \\
&= \left(t_2 + \sum_{s=1}^{M} I_s(x_2) r^{s-1}\right) - \left(t_1 + \sum_{s=1}^{M} I_s(x_1) r^{s-1}\right) \\
&= \left(\sum_{s=i}^{M} I_s(x_2) r^{s-1} - \sum_{s=i}^{M} I_s(x_1) r^{s-1}\right) \\
&\quad + \left(t_2 + \sum_{s=1}^{i-1} I_s(x_2) r^{s-1}\right) \\
&\quad - \left(t_1 + \sum_{s=1}^{i-1} I_s(x_1) r^{s-1}\right) \\
&= \sum_{s=i}^{M} I_s(x_2) r^{s-1} - \sum_{s=i}^{M} I_s(x_1) r^{s-1}. \qquad (8)
\end{aligned}$$

Since the two packets are both routed to D_{ij}, *i.e.*, $I_i(x_1) = I_i(x_2) = j$, we further have

$$k_2 - k_1 = \sum_{s=i+1}^{M} I_s(x_2) r^{s-1} - \sum_{s=i+1}^{M} I_s(x_1) r^{s-1}. \qquad (9)$$

Therefore, $k_2 - k_1 = cr^i$ for some integer c.

($\Leftarrow$) First, we show that

$$r^{i-1} > \sum_{s=1}^{i-1} I_s(x_1) r^{s-1} + t_1 - t_2 \geq 0. \qquad (10)$$

Since the $k_1{}^{th}$ packet is routed to D_{ij} at time t, we have

$$t = t_1 + \sum_{s=1}^{i-1} I_s(x_1) r^{s-1}. \qquad (11)$$

On the other hand, since the $k_2{}^{th}$ packet arrives in the time interval $[t - r^{i-1} + 1, t]$, we have $t - r^{i-1} < t_2 \leq t$. Therefore, we have

$$r^{i-1} > t - t_2 \geq 0. \qquad (12)$$

Hence, replacing the t in (12) by (11) yields (10).

Now, we show that the $k_2{}^{th}$ packet is routed to D_{ij} at time t. That is, we need to show $I_i(x_2) = j$ and $t_2 + \sum_{s=1}^{i-1} I_s(x_2) r^{s-1} = t$. Using (6), (4), and the assumption that $k_2 = k_1 + cr^i$, we have

$$\begin{aligned}
x_2 &= x_1 + (k_2 - k_1) - (t_2 - t_1) \\
&= \sum_{s=1}^{M} I_s(x_1) r^{s-1} + cr^i + t_1 - t_2 \\
&= \left(\sum_{s=i+1}^{M} I_s(x_1) r^{s-1} + cr^i\right) + I_i(x_1) r^{i-1} + \\
&\qquad \left(\sum_{s=1}^{i-1} I_s(x_1) r^{s-1} + t_1 - t_2\right). \qquad (13)
\end{aligned}$$

In (13), the delay x_2 consists of three parts: $\sum_{s=i+1}^{M} I_s(x_1) r^{s-1} + cr^i$, $I_i(x_1) r^{i-1}$, and $\sum_{s=1}^{i-1} I_s(x_1) r^{s-1} + t_1 - t_2$. The first part is equal to $\tilde{c} r^i$ for some integer $\tilde{c}$. Using (10) and the fact $0 \leq I_i(x_1) \leq r - 1$, we have

$$\begin{aligned}
&I_i(x_1) r^{i-1} + \left(\sum_{s=1}^{i-1} I_s(x_1) r^{s-1} + t_1 - t_2\right) \\
&< (r-1) r^{i-1} + r^{i-1} = r^i. \qquad (14)
\end{aligned}$$

As the sum of the second part and the third part is less than r^i, we have from the fact $x_2 \geq 0$ and (13) that $\tilde{c}$ must be a nonnegative integer. Hence, the first part is equal to a nonnegative integer times r^i, the second part is equal to $I_i(x_1)$ times r^{i-1}, and the third part is smaller than r^{i-1}. Since the r-ary representation is unique, we must have

$$\sum_{s=i+1}^{M} I_s(x_2) r^{s-1} = \sum_{s=i+1}^{M} I_s(x_1) r^{s-1} + cr^i \qquad (15)$$

$$I_i(x_2) r^{i-1} = I_i(x_1) r^{i-1} \qquad (16)$$

$$\sum_{s=1}^{i-1} I_s(x_2) r^{s-1} = \sum_{s=1}^{i-1} I_s(x_1) r^{s-1} + t_1 - t_2. \qquad (17)$$

From (16), we have $I_i(x_2) = I_i(x_1) = j$. From (17) and (11), we also have $t_2 + \sum_{s=1}^{i-1} I_s(x_2) = t$. ■

Lemma 3 *Let t_1^d (resp. t_2^d) be the earliest (resp. largest) departure time of the packets that arrive in the time interval $[t_1, t_2]$. Then the number of packets that depart in the time interval $[t_1^d, t_2^d]$ is bounded by*

$$2(N-1)(\ell_{\max} - 1) + N(t_2 - t_1 + 1).$$

Proof. Without loss of generality, let us assume that the packet departs at t_1^d (resp. t_2^d) arrives at input i_1 (resp. i_2). Because of the FIFO property, a packet that departs later than t_1^d must arrive either at input i_1 after t_1 or at the other $N-1$ inputs after $t_1 - \ell_{\max}$. Similarly, a packet that departs earlier than t_2^d must arrive either at input i_2 before t_2 or at the other $N-1$ inputs before $t_2 + \ell_{\max}$. Thus, the number of packets that depart in the time interval $[t_1^d + 1, t_2^d - 1]$ is bounded by

$$2(N-1)(\ell_{\max} - 1) + N(t_2 - t_1 + 1) - 2.$$

Adding the two packets that depart at t_1^d and t_2^d completes the argument. ■

Proof. (Theorem 1)

To prove this theorem, we just need to show that the number of packets that are routed to D_{ij} in a time slot is bounded by

$$\min\left(r^{M-i}, Nr^{i-1}, \left\lceil \frac{N}{r} + \frac{2(N-1)(\ell_{\max} - 1)}{r^i} \right\rceil\right).$$

The first two bounds are related to the number of input/output ports. Note that the number of paths that have different delays from $D_{i,j}$ to the output port is r^{M-i}. Since there is at most one packet that can be routed to the output port in each time

slot, we conclude that there are at most r^{M-i} packets in D_{ij} in each time slot. Similarly, the number of paths that have different delays from $D_{i,j}$ to an input port is r^{i-1}. Since there are N input ports, we also know that there are at most Nr^{i-1} packets that can be routed to D_{ij} in each time slot.

Now we use Lemma 2 and Lemma 3 to prove that the number of packets routed to D_{ij} at time t is bounded by $\left\lceil \frac{N}{r} + \frac{2(N-1)(\ell_{\max}-1)}{r^i} \right\rceil$. As the packets from different busy periods can not be routed to D_{ij} at the same time, we only need to consider the busy period with packets routed to D_{ij} at time t. Let t_1^d (resp. t_2^d) be the earliest (resp. largest) departure time among the packets that arrive in $[t - r^{i-1} + 1, t]$. From Lemma 3, we know that the number of packets that depart in $[t_1^d, t_2^d]$ is bounded by

$$2(N-1)(\ell_{\max}-1) + Nr^{i-1}.$$

From Lemma 2, we also know that there is at most one packet routed to D_{ij} at time t for every consecutive r^i packets departing the system in the time interval $[t_1^d, t_2^d]$. Thus, there are at most

$$\left\lceil \frac{2(N-1)(\ell_{\max}-1) + Nr^{i-1}}{r^i} \right\rceil = \left\lceil \frac{N}{r} + \frac{2(N-1)(\ell_{\max}-1)}{r^i} \right\rceil \quad (18)$$

packets routed to D_{ij} at time t. ■

Corollary 4 *In the case that all the bursts contain exactly one packet, i.e.,*$\ell_{\max} = 1$*, the sufficient condition in (5) can be simplified as follows:*

$$|D_{ij}| \geq \min\left(r^{M-i}, \left\lceil \frac{N}{r} \right\rceil\right). \quad (19)$$

When M is large and i is small, the sufficient condition in (19) is then lower bounded by $\lceil N/r \rceil$. In this case, we argue that it is also a necessary condition and the lower bound cannot be improved further. To see this, suppose that there are $(j+1)r^{i-1} - 1$ packets in the N-to-1 multiplexer at time $t - r^{i-1}$. Then at time $t - r^{i-1} + 1, t - r^{i-1} + 2, \ldots, t$, there is exactly one packet arrival at each input port at each time slot. Clearly, the delay of the first arrival at $t - r^{i-1} + 1$ is simply $(j+1)r^{i-1} - 1$ and that packet will be routed to one of the delay lines in $D_{i,j}$ at time t (according to the r-ary representation of $(j+1)r^{i-1} - 1$). Since there are exactly $r^{i-1}N$ packets that arrive in the time interval $[t - r^{i-1} + 1, t]$ (and these packets are admitted to the multiplexer when M is large), it follows from Lemma 2 that there are exactly $\lceil N/r \rceil$ packets routed to $D_{i,j}$ at time t.

For the special case that $r = N$ and $\ell_{\max} = 1$, we have from Corollary 4 that there is at most one packet routed to D_{ij} at each time. Therefore, we can simply choose $|D_{ij}| = 1$ to implement a self-routing N-to-1 multiplexer with buffer $N^M - 1$ as previously shown in [3].

Finally, we address the issue of choosing the optimal r-ary representation in the construction of an N-to-1 multiplexer. We argue that the best choice is to use the N-ary representation when $\ell_{\max} = 1$ and M is large. To see this, note that we need to have at least one fiber delay line in $D_{i,j}$ for all i, j. When $\ell_{\max} = 1$, we have already known that we can choose $r = N$ so that only one fiber delay line is needed for $D_{i,j}$. As such, we only need N fiber delay lines at each stage by using the N-ary representation. On the other hand, the number of fiber delay lines at each stage is lower bounded by $r\lceil N/r \rceil \geq N$. Thus, using the N-representation not only requires the minimum number of fiber delay lines, but also requires the minimum construction complexity of the crossbar switches used in the feedforward construction in Figure 1.

III. Exact Emulation of N×N Output-buffered Switches

A. Direct Construction

An $N \times N$ output-buffered switch can be viewed as a network element with N parallel G/G/1 queues. An arriving burst of packets destined for output port i is added to the tail of the i^{th} G/G/1 queue. Therefore, the delay of a burst of packets is characterized by the Lindley recursion in (1) and it is known when the burst arrives at the switch. As such, each G/G/1 queue can be implemented by an N-to-1 multiplexer, and one can then construct an $N \times N$ output-buffered switch using N 1-to-N demultiplexers in the first stage and N N-to-1 (buffered) multiplexers in the second stage. A 4×4 output-buffered switch by such a construction is shown in Figure 2. Arrivals destined for output i are routed to the i^{th} multiplexer using the demultiplexer in the first stage. By so doing, each packet departs at the right time and at the right output port.

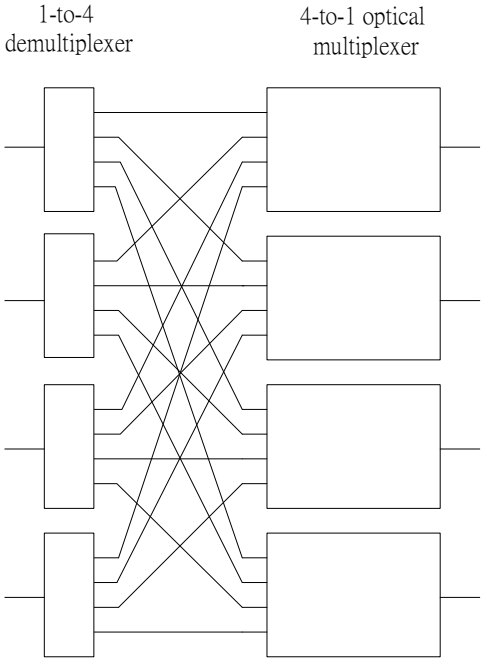

Fig. 2. A direct construction of a 4×4 output-buffered switch

Though the direct construction of an $N \times N$ output-buffered switch is straightforward, it may not be efficient. To see this, suppose that each of the N N-to-1 (buffered) multiplexers is constructed by using the feedforward construction in Figure 1. Then the buffers, constructed by fiber delay lines, in these N N-to-1 (buffered) multiplexers are not shared by packets with different outputs. As such, the construction complexity (by such a construction for an $N \times N$ output-buffered switch) is N times of that for an N-to-1 (buffered) multiplexer, both in the number of fiber delay lines and in the number of crossbar switches.

B. Feedforward Construction

The key insight is then to allow the buffers to be shared to reduce the construction complexity. For this, we propose the feedforward construction for an $N \times N$ output-buffered switch in Figure 3. The architecture in Figure 3 is the same as that for an N-to-1 multiplexer in Figure 1 except that there are N outputs at the last stage. The selection of a routing path for each packet (in a burst) is basically the same as that in Figure 1. Specifically, we use the r-ary representation of the delay of a packet to route it through the feedforward network until it reaches the last stage. At the last stage, a packet is then routed to its destined output.

In Theorem 5 below, we provide a sufficient condition on the number of fiber delay lines needed to construct an $N \times N$ output-buffered switch. Let E_{ij} be the set of fiber delay lines with jr^{i-1} units of delay in stage i and $|E_{ij}|$ be the number of delay lines in that set.

Theorem 5 *Suppose that the burst lengths are bounded by $\ell_{\max}$ and that the feedforward network in Figure 3 is started from an empty system. If*

$$|E_{ij}| \geq \min(Nr^{M-i}, Nr^{i-1}, N + \left\lfloor \frac{N}{r} + \frac{2(N-1)(\ell_{\max}-1)-N}{r^i} \right\rfloor), \quad (20)$$

then under the routing rule it is a discrete-time $N \times N$ output-buffered switch with buffer $r^M - 1$ for variable length bursts.

Proof. As there are N input ports and N output ports, the first two bounds can be derived as in Theorem 1. Now, we use Lemma 2 and Lemma 3 to prove the third bound. Similar to the proof in Theorem 1, we only consider the busy period of each output port with packets routed to E_{ij} at time t. Let $t^d_{n,1}$ (resp. $t^d_{n,2}$) be the earliest (resp. largest) departure time of packets destined to output port n that arrive in $[t-r^{i-1}+1, t]$. Also, let K_n be the number of packets that depart from output port n in the time interval $[t^d_{n,1}, t^d_{n,2}]$.

Using Lemma 2, we have at most

$$\sum_{n=1}^{N} \left\lceil \frac{K_n}{r^i} \right\rceil \quad (21)$$

packets routed to E_{ij} at time t. Using the property $\lceil \frac{a}{b} \rceil = \lfloor \frac{a-1}{b} \rfloor + 1$ when a and b are integers, we have

$$\sum_{n=1}^{N} \left\lceil \frac{K_n}{r^i} \right\rceil = \sum_{n=1}^{N} \left\lfloor \frac{K_n - 1}{r^i} \right\rfloor + N \leq \left\lfloor \sum_{n=1}^{N} (\frac{K_n - 1}{r^i}) \right\rfloor + N. \quad (22)$$

Now let $t^d_1 = \min_{1 \leq n \leq N} t^d_{n,1}$ and $t^d_2 = \max_{1 \leq n \leq N} t^d_{n,2}$. Then $\sum_{n=1}^{N} K_n$ is the number of packets that depart in $[t^d_1, t^d_2]$ and arrive in $[t-r^{i-1}+1, t]$. From Lemma 3, it follows that

$$\sum_{n=1}^{N} K_n \leq 2(N-1)(\ell_{\max}-1) + Nr^{i-1}. \quad (23)$$

Using (23) in (22) yields the desired bound.

■

Corollary 6 *In the case that all the bursts contain exactly one packet, i.e., $\ell_{\max} = 1$, the sufficient condition in (20) can be simplified as follows:*

$$|E_{ij}| \geq \min(Nr^{M-i}, N + \left\lfloor \frac{N(r^{i-1}-1)}{r^i} \right\rfloor). \quad (24)$$

In view of the second lower bound in (24), the number of fiber delay lines in each stage is roughly $(r+1)N$. Thus, the best choice that minimizes the number of fiber delay lines in each stage is $r = 2$. For such a choice, one only needs $3N$ delay lines at each stage, which is comparable to that of N-to-1 multiplexer in Figure 1. As such, its construction complexity is considerably lower than that using the direct construction.

C. Feedback Construction

In this section, we show that we can also use a feedback construction (see Figure 4) to replace the feedforward construction in Figure 3. The key difference between the feedback construction in Figure 4 and the feedforward construction in Figure 3 is that there is no need to use fiber delay lines with delay 0 in the feedback construction as a packet can be routed directly to the next fiber delay line with nonzero delay. Specifically, suppose that each burst contains exactly one packet, i.e., $\ell_{\max} = 1$ and that the binary representation is used, i.e., $r = 2$. As discussed before, we only need $3N$ delay lines at each stage and only half of them are with nonzero delay. Thus, to construct an $N \times N$ output-buffered switch with buffer $2^M - 1$, we can use a single $(N + \frac{3}{2}NM) \times (N + \frac{3}{2}NM)$ crossbar switch as shown in Figure 4. For $i = 1, 2, \ldots, M$, there are $\frac{3}{2}N$ delay lines with delay 2^{i-1}. These $\frac{3}{2}NM$ delay lines are connected from $\frac{3}{2}NM$ outputs of the crossbar switch back to $\frac{3}{2}NM$ inputs of the crossbar switch, leaving N inputs and N outputs of the crossbar switch as the N inputs and the N outputs of the $N \times N$ output-buffered switch. We note that the feedback construction was originally proposed in [10] for an approximation of an output-buffered switch. The sufficient

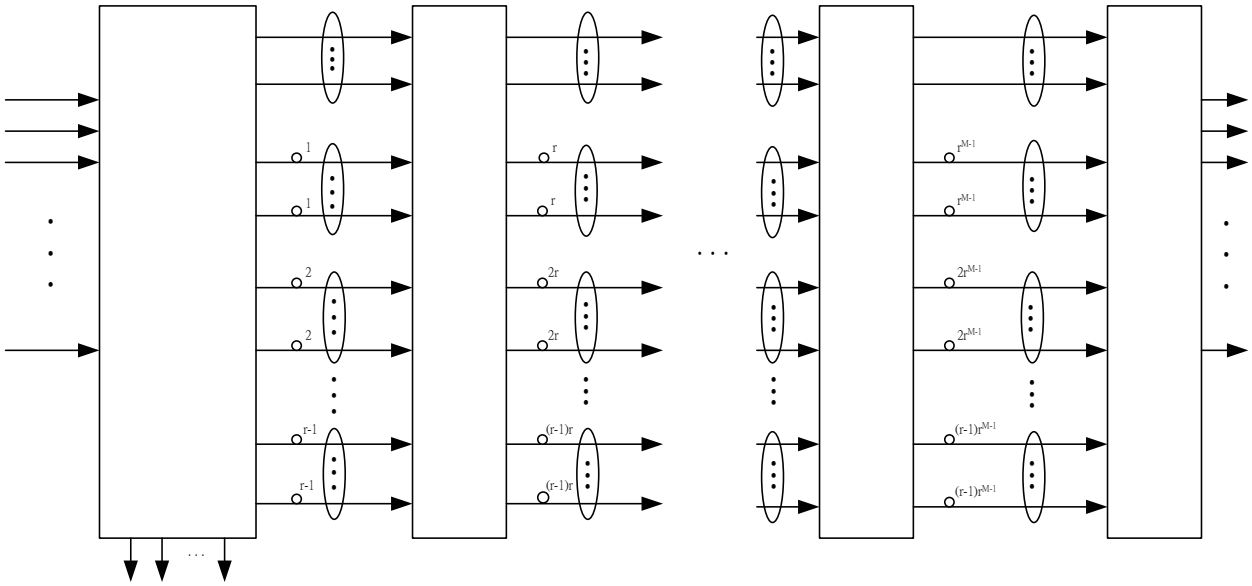

Fig. 3. A feedforward construction of an $N \times N$ output-buffered switch

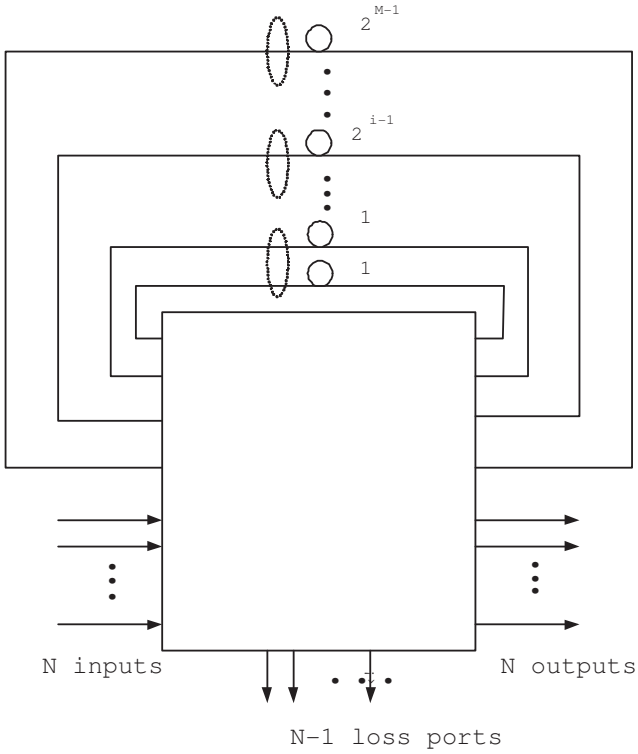

Fig. 4. A feedback construction of the $N \times N$ output-buffered switch

condition in Theorem 5 (and Corollary 6) gives the specific number of fiber delay lines (with specific delay) needed for *exact emulation* of an output-buffered switch.

IV. Simulation Results

Since in practice the buffer size is finite in an $N \times N$ output-buffered switch, bursts are lost due to buffer overflows. For the engineering purpose, it may not be efficient to use the worst case to design an $N \times N$ output-buffered switch as we did in Theorem 5. In stead of achieving exact emulation, in this section we consider good approximations in the sense that their packet loss probabilities are comparable to that of the exact emulation.

Consider the feedforward construction in Figure 3 with the binary representation, i.e., $r = 2$. Even for the case that all the bursts contain exactly one packet, i.e., $\ell_{\max} = 1$, we know that in order to achieve exact emulation one needs roughly $3N/2$ delay lines for $E_{i,j}$. Here we argue that one can achieve a good approximation if $|E_{i,j}| > N/2$. Our intuition is that with equal probability a packet is routed to a delay line with zero delay or a delay line with nonzero delay in each stage. Thus, even with N packet arrivals in each time slot, there are (on average) roughly $N/2$ packets that are routed to $E_{i,j}$ in each time slot. From the law of large numbers, the probability that a burst of packets cannot find a feasible path can be made arbitrarily small if $|E_{i,j}| > N/2$. We note that our intuition may not be correct in light traffic as most packets are routed to delay lines with zero delay. However, we expect that this is the case in heavy traffic.

To verify our intuition, we perform various computer simulations. Each run of our simulations contains 10^5 time slots. We use the bursty traffic model as described in [14]. However, the burst lengths are chosen independently according to the following distribution: with probability 0.35 (resp. 0.45, 0.2), a burst is chosen to be 1 (resp. 16, 38) unit of time slot. The reason for choosing such a distribution, instead of using the (truncated) Pareto distribution in [14], is that there are three typical packet lengths in Ethernet traffic [13]. With probability 0.35 (resp. 0.45, 0.2), an Ethernet packet is found to be of 40 (resp. 572, 1500) bytes. Our choice for the distribution of burst lengths is then corresponding to the three typical

N	Exact emulation	$\|E_{ij}\| = 0.6N$
32	0.044872	0.055889
64	0.045428	0.047908
128	0.045614	0.046329
256	0.045808	0.045857

TABLE I

COMPARISON OF PACKET LOSS PROBABILITIES BETWEEN EXACT EMULATION AND APPROXIMATION

Ethernet packet lengths when we set one unit of time slot as the time interval needed to transmit 40 bytes of data. The buffer size for $N \times N$ output-buffered switches is set to be 256 packets (time slots). With $N = 64$, we plot in Figure 5 the packet loss probabilities for $|E_{ij}| = cN$ for $c = 0.5, 0.55,$ and 0.6. In Table I, we report the packet loss probabilities for $N = 32, 64, 128,$ and 256 when $c = 0.6$.

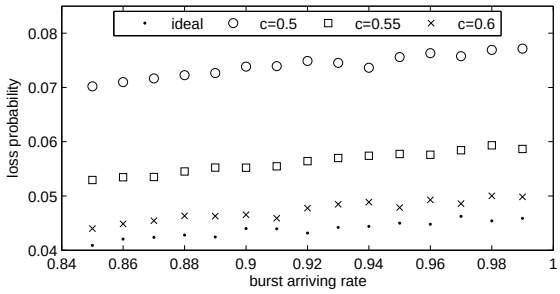

Fig. 5. Packet loss probability as a function of arrival rate

Two interesting observations can be found from our simulations. First, as expected, we can make the packet loss probability very close to that in exact emulation by choosing $c > 0.5$. This is verified in Figure 5 for $c = 0.6$. Second, statistical multiplexing gain can be achieved by increasing N. This is shown in Table I that the loss probability of the approximation using $|E_{i,j}| = 0.6N$ tends to that of the exact emulation of an output-buffered switch as N increases.

V. CONCLUSIONS

In this paper, we proposed feedforward SDL constructions of output-buffered multiplexers and switches with variable length bursts. By the worst case analysis, we gave specific sufficient conditions for the number of fiber delay lines needed in each stage for achieving exact emulation. By computer simulation, we also showed that the number of delay lines in each stage can be greatly reduced due to statistical multiplexing gain.

There are several possible extensions of our work.

(i) Statistical analysis: here our analysis is based on the deterministic worst case. As commented in Section IV, it might be of interest to have a statistical analysis for packet loss probability. Note that the feedforward networks are in fact some form of *loss networks* in circuit switching [15]. Here queueing is coupled with routing in finding "circuits".

(ii) Continuous time: we used a discrete-time setting in our analysis. It is plausible that our feedforward construction can also be used for the continuous-time setting, where burst arrivals are asynchronous. In the continuous-time setting, there is a granularity problem for choosing the *right* unit for delay lines (see e.g., [16], [17]). Moreover, it is expected that the worst case for the continuous-time setting is much worse than that for the discrete-time setting. However, we still expect that statistical multiplexing gain can be used for reducing the number of delay lines.

ACKNOWLEDGMENT

This research was supported in part by the National Science Council, Taiwan, R.O.C., under Contract NSC-93-2213-E-007-040, Contract NSC-93-2213-E-007-095, Contract NSC-94-2213-E-007-046, and the Program for Promoting Academic Excellence of Universities NSC 94-2752-E-007-002-PAE.

REFERENCES

[1] R. L. Cruz and J. T. Tsai, "COD: alternative architectures for high speed packet switching," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 11–20, February 1996.

[2] D. K. Hunter, D. Cotter, R. B. Ahmad, D. Cornwell, T. H. Gilfedder, P. J. Legg and I. Andonovic, "2×2 buffered switch fabrics for traffic routing, merging and shaping in photonic cell networks," *IEEE Journal of Lightwave Technology*, vol. 15, pp. 86–101, 1997.

[3] C.-S. Chang, D.-S. Lee, and C.-K. Tu, "Recursive construction of FIFO optical multiplexers with switched delay lines," *IEEE Transactions on Information Theory*, vol. 50, pp. 3221-3233, 2004.

[4] C.-S. Chang, D.-S. Lee and C.-K. Tu, "Using switched delay lines for exact emulation of FIFO multiplexers with variable length bursts," *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 4, pp. 108-117, 2006.

[5] C.-C. Chou, C.-S. Chang, D.-S. Lee, and J. Cheng, "A necessary and sufficient condition for the construction of 2-to-1 optical FIFO multiplexers by a single crossbar switch and fiber delay lines," to appear in *IEEE Transactions on Information Theory*, Oct. 2006.

[6] C.-S. Chang, Y.-T. Chen, and D.-S. Lee, "Construction of optical FIFO queues," *IEEE Transactions on Information Theory*, Vol. 52, No. 6, pp.2838-2843, 2006.

[7] C.-S. Chang, Y.-T. Chen, J. Cheng, and D.-S. Lee, "Multistage constructions of linear compressors, non-overtaking delay lines, and flexible delay lines," *Proceedings of IEEE INFOCOM 2006*

[8] A. D. Sarwate and V. Anantharam, "Exact emulation of a priority queue with a switch and delay lines," to appear in *Queueing Systems: Theory and Applications*, Vol. 53, pp. 115–125, July 2006.

[9] H.-C. Chiu, C.-S. Chang, J. Cheng, and D.-S. Lee, "A simple proof for the constructions of optical priority queues," submitted to *Queueing Systems: Theory and Applications*, 2005.

[10] M. J. Karol, "Shared-memory optical packet (ATM) switch," SPIE vol. 2024: Multigigabit Fiber Communication Systems(1993), pp. 212–222, October 1993.

[11] M. Yoo, C. Qiao, and S. Dixit, "QoS performance of optical burst switching in IP-over-WDM networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 2062-2071, October 2000.

[12] E. A. Varvarigos and V. Sharma, "An efficient reservation connection control protocol for gigabit networks, " *Computer Networks and ISDN Systems* ,vol. 30, (no. 12), 13 July 1998, pp. 1135-1156.

[13] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the Sprint IP backbone,"*Network, IEEE*, vol. 17, pp. 6-16, November-December 2003.

[14] C.-S. Chang, D.-S. Lee and Y.-S. Jou, “Load balanced Birkhoff-von Neumann switches, part I: one-stage buffering,” *Computer Communications*, Vol. 25, pp. 611-622, 2002.
[15] F. P. Kelly, “Loss networks,” *Ann. Appl. Probab.*, Vol. 1, pp. 319–378, 1991.
[16] L. Tancevski, S. Yegnanarayanan, G. Castanon, *et al.* “Optical routing of asynchronous, variable length packets” *Journal on Selected Areas in Communications*, Vol. 18, pp. 2084-2093, 2000.
[17] F. Callegati, “Approximate modeling of optical buffers for variable length packets,” *Photonic Network Communications*, Vol. 3, pp. 383-390, 2001.