Sliding Banyan Network

Michael W. Haney, *Member, IEEE,* and Marc P. Christensen, *Member, IEEE*

***Abstract*—The *Sliding Banyan* network is described and evaluated. The novel three-dimensional (3-D) multistage network topology employs a *macro*-lenslet array in a retroreflective configuration to effect the required shuffle link patterns across a *single* two-dimensional (2-D) multichip array of "smart pixels." An electronic deflection routing scheme, based on simple destination-tag self-routing, is employed within the smart pixels. Internal packet blocking is efficiently avoided because deflected packets are routed through individualized banyan networks that have "slid" in the time dimension to accommodate each packet's routing needs. Simulations show that this self-routing approach reduces the number of stages, and hence the number of switching and interconnection resources necessary to achieve a specified blocking probability. Experimental focusing and registration results, using arrays of vertical cavity surface emitting lasers, show that conventional optical imaging technology is suitable for this architecture. The results indicate that the sliding banyan approach will overcome the current performance constraints of conventional metallic interconnections and be scalable to ATM switching applications with aggregate throughputs in the Tb/s regime.**

I. Introduction

THE EXPLOSIVE growth in the asynchronous transfer mode (ATM) equipment industry is just one indication of the ever-increasing demand for high throughput, cost effective, broadband data switching networks. The high throughput demands of future systems will be driven by the growing number of nodes on any given network, the increasing bandwidth of data communications between nodes, and the desire to transmit video data over the same networks. Aggregate capacities in the Tb/s regime will be required to meet the demand [1].

A common measure of interconnection difficulty in networks is the bisection width, defined as the minimum number of "wires" that must be removed to partition the network into two halves with identical numbers of processors [2]. Multistage interconnection networks (MIN's) suffer from bisection widths that grow nearly linearly with the number of nodes. A banyan network is defined as a MIN with a unique path from any input to any output. Banyan-based MIN's offer the potential for using simple self-routing algorithms that will be critical to the effective operation of high throughput switching circuits, where global control is impractical. Though theoretically powerful, banyan-based switching architectures have been limited to fairly small network sizes owing to the high bisection widths of large networks. The link interconnections in a banyan network consist of perfect shuffles [3], or interconnections isomorphic to the perfect shuffle. These interconnection patterns do not lend themselves to implementation with traditional metallic interconnection techniques. The global interconnection topology of banyans leads, in VLSI approaches, to internode communications performance limits in speed, crosstalk, and power consumption. Very often the limitations of electronic banyans have caused designers to give up on exploiting the banyan network structure altogether, and adopt wholly different topologies, such as mesh networks, which have simpler interconnection requirements, but much higher switching complexity. These approaches are not suitable for high throughput packet switching applications due to the large amount of buffering and contention control required. There is a need to use MIN approaches with new technologies that overcome the metallic interconnection bottleneck.

In this paper, the sliding banyan (SB) [4], [5], a new MIN switching architecture that uses optoelectronic "smart pixels" and free-space optical interconnections to overcome the limitations of electronic interconnections, is described and evaluated. The SB provides a significant reduction in the number of switch, control, and interconnection resources that would be required in an equivalent all-electronic approach. The resource reduction stems from a novel partitioning of the resources—achieved by spatially interleaving the stages of the switching network in a way that is possible only with 3-D optical interconnections. The interleaving is possible because each stage in the shuffle-based SB multistage interconnection network requires the same shuffle link pattern. With interleaving, each stage's identical I/O pattern is slightly shifted from those of the other stages. A *single* optical system, suitably configured, can thus be used to interconnect all of the MIN's stages *simultaneously.* The SB uses a pipelined destination-tag self-routing approach within a deflection routing strategy. Since each node is physically colocated with all of its sister nodes at each stage, successfully routed packets may exit the network immediately, at whatever stage they finally arrive. This is the key feature of the SB. Deflected packets are effectively routed to a new banyan that has "slid" in time to accommodate those packets' needs. The result is that packets are removed from the network as rapidly as possible, leading to an overall low blocking probability and high resource utilization.

As a preface to the SB architecture description, Section II provides background on banyan networks and the optical shuffle-connected approach that are the key elements of the SB concept. The SB architecture is then detailed in Section III. In addition to the SB architecture description, Section III contains the results of simulations, analysis, and experiments that demonstrate the SB's important features. The simulations

Manuscript received August 21, 1995. This research is supported by the Advanced Research Projects Agency under a contract from the Air Force Office of Scientific Research.

The authors are with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA 22030.

Publisher Item Identifier S 0733-8724(96)03894-7.

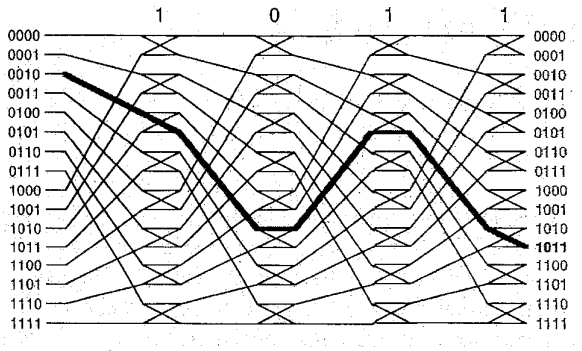


Fig. 1. Shuffle-based banyan depicting destination tag self-routing approach.

and analysis show that the blocking performance of the SB compares favorably with other approaches under fully loaded permutation traffic. In fact, the SB is shown to be within a factor of three of the theoretical minimum number of switching resources, despite using a simple self-routing control strategy. Furthermore, the results of experiments, in which vertical cavity surface emitting laser (VCSEL) and detector arrays were used to simulate future smart pixel I/O, indicate that an optical shuffle interconnection system based on conventional refractive elements is feasible. Section IV contains a discussion of the smart pixel technology needed to implement the SB network and the anticipated performance parameters of a future implementation. The conclusion, contained in Section V, summarizes the key features of the SB network.

II. Background

A. Shuffle-Based Banyan Networks

Fig. 1 depicts a shuffle-based banyan network for $N = 16$ nodes. A banyan network connects any given input/output node pair with a unique path. There are typically $\log_k N$ stages, with each stage consisting of a permutation link pattern, such as a shuffle, and a set of $k \times k$ crossbar switches. In many banyans, the unique path between input and output is determined by following a simple self-routing algorithm. This algorithm is called a destination-tag algorithm because it has the advantage of requiring only the destination address—the routing algorithm is independent of source address. Self-routing eliminates the need for external control of the switching elements and offers the means to the high throughput desired in future switching circuits. Fig. 1 illustrates this approach for a banyan comprised of perfect shuffle permutations and 2×2 switches. The self-routing algorithm is performed on the output address (located in the packet header in ATM switching) as follows: beginning with the most significant bit (MSB) of the output address, inspect one bit at each stage—if the bit is a one, exit the stage on the lower node; if it is a zero, exit on the upper node. Since the perfect shuffle performs a bit rotation on the address, this algorithm will work for any destination and is independent of the source address. Fig. 1 shows this algorithm for a 16- node banyan.

In a single banyan architecture, a blocking error occurs when two (or more if $k > 2$) packets wish to continue to the next stage using the same connection. The connection can only transmit one of the packets, so the other(s) must be lost, or blocked. A blocked packet has no method for recovery in this single banyan architecture. The blocking probability of a banyan increases as the load (number of active inputs) increases.

Redundant banyan architectures were proposed to overcome the internal blocking problem. Examples include replicated [6], dilated [6], and tandem [7] banyan networks, which all use auxiliary banyan networks to reduce the blocking probability to an acceptable level. Fig. 2 depicts a tandem banyan (TB) architecture. In the TB, packets are not blocked, but rather tagged as having been routed incorrectly, and then passed through the remainder of the banyan. If two packets contend for the same output pin and one of them has been tagged as having been routed incorrectly, then the untagged packet would "win" the contention. If both packets have not been tagged, the winner could be determined by a random selection. At the end of one banyan, the packets which made it to their destination are removed from the network; the remaining packets then begin anew in a second banyan. Since some of the packets have been removed, the chances of blocking are diminished. Additional banyans are appended, in tandem, until an acceptable overall blocking performance is achieved.

B. Optical Shuffle Interconnections

To overcome the limitations of metallic interconnection, banyan based topologies can be implemented using free-space optics. An important step was the proposal of a free-space optical implementation of the perfect shuffle (PS) [8]. Several implementations of 1-D and 2-D optical shuffles were investigated [9]–[16]. All of these approaches use optics to effect the magnification and interleaving needed to perform the PS link pattern. Examples of 2-D shuffles are the Folded Perfect Shuffle [13], and the Separable Perfect Shuffle [9], [11].

A useful generalization of the PS pattern comes from partitioning the nodes into k equal sized groups and interleaving them. This is referred to as a k-shuffle; by this definition the PS is a 2-shuffle. The higher order shuffle does not avoid the high bisection width problem of the PS. However, the number of stages, and hence the switch latency, in k-shuffle based banyans is reduced to $\log_k N$, which is a strong function of k when $k \ll N$. The price paid for fewer stages in a k-shuffle banyan (with $k > 2$) is a need for more complicated $k \times k$ active self-routing switching elements.

One implementation of the separable k-shuffle uses two $k \times k$ lens arrays [16], a side view of which is depicted in Fig. 3, for $k = 4$ and $N = 16 \times 16$ nodes. In this arrangement each pair of lenses, from the two lenslet arrays, perform a unity magnification operation that achieves the desire k-shuffle pattern as shown. Fig. 3 also depicts a grouping of the 16 smart pixel nodes into four subgroups in which the nodes are in close proximity to each other. This subgrouping concept places the nodes on a self-similar grid [17] rather than a regular square grid. As shown in the figure, the self-similar grid concept groups the smart pixels in a manner more amenable to packaging on separate OEIC's and increases the

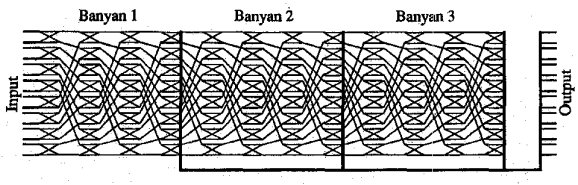


Fig. 2. Tandem banyan architecture.

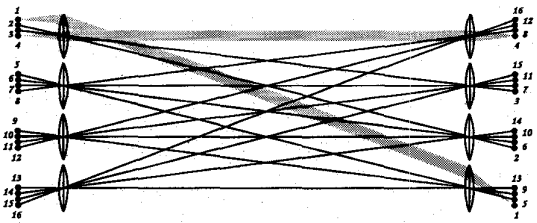


Fig. 3. Side view of 4 × 4 shuffle system showing a 1-D four-shuffle pattern arranged on self-similar grid layout.

optical efficiency of the shuffling optics [17]. The symmetry of the optical k-shuffle depicted in Fig. 3 is a key aspect that is exploited in the SB implementation described below.

III. Sliding Banyan Architecture

A. Optically Interleaved Interconnection Topology

Interleaving of multiple shuffle stages was previously proposed to make better use of the shuffle optics' space bandwidth product (SBWP) and simplify the optical complexity by simultaneously using a single optical system for all stages in the MIN [18]. Fig. 4 depicts the central notion of the interleaved topology used in the SB. Previously proposed MIN's were comprised of physically separated stages—essentially emulating the traditional VLSI approach by replacing interchip and interboard metallic interconnections with interchip free-space optical interconnections. Such a scheme, implemented with 2-D optical PS's [9], [11], [13], [14], is depicted in the top half of the figure. This approach shows promise for overcoming the massive interconnection requirements between MIN stages, but has some implementation difficulties that stem from the physically separated multistage topology and the lack of compatibility with broad area multichip packaging conventions. The multistage implementation shown in the top of the figure requires one array for each stage.

Inspection of the network depicted in the top of Fig. 4 reveals that the interstage interconnections are all identical and are shift invariant within the field of view (FOV) of the imaging optics. Therefore, with slight physical offsets of the I/O of each stage, multiple parallel interconnections can be implemented in an *interleaved* fashion, with a *single* optical system. This is schematically shown in the bottom of Fig. 4, where a single optical system effects all of the required interconnections *simultaneously*. With this topology, the switching resources are distributed *laterally* across a single physical plane, rather than longitudinally across several planes. Since all of the stages have been collapsed onto a single plane, the bisection width implemented by the optics has been increased by the number of stages being implemented. To

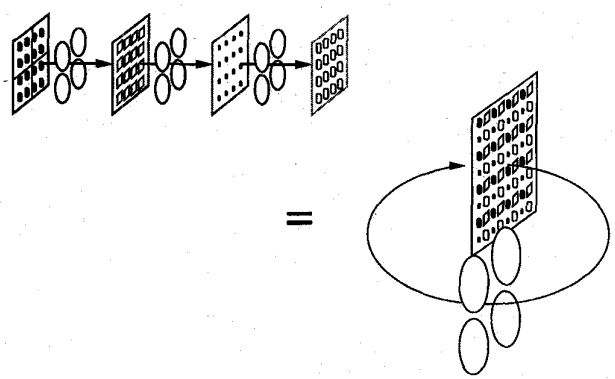


Fig. 4. Optically interleaved shuffle-based MIN topology.

handle the increase in I/O resources, the plane on which the smart pixel array resides will be a PC board or multichip module (MCM) package that can accommodate an array of pixel optoelectronic integrated circuits (OEIC's) that is large enough to contain all of the smart pixel resources. As discussed below, for a network with ~1024 nodes, such a system will likely be ~10–20 cm across. The associated optics, therefore, will consist of one or two *macro*-lenslet arrays, in which the size of individual elements correspond roughly to the size of an OEIC chip (e.g., 1–3 cm^2).

As discussed in the next section, the SB will require the equivalent of several banyans in stages to achieve the desired low blocking probability—e.g., for $N = 1024$, approximately 30–50 stages will be needed. The number of stages that can be interleaved in this fashion is theoretically bounded by the SBWP. For example, consider a 1024 node system in which the PE's are arrayed in a 32×32 array. A typical high quality imaging system will have a SBWP $> 10^6$, meaning that many more stages than required could theoretically be interleaved in this manner and maintain good isolation. A more practical limitation on the number of stages, however, is obtained from the real-estate constraints of the smart pixels and the related heat dissipation issues that determine the closest separation of emitter elements on the array. For example, typical air cooled

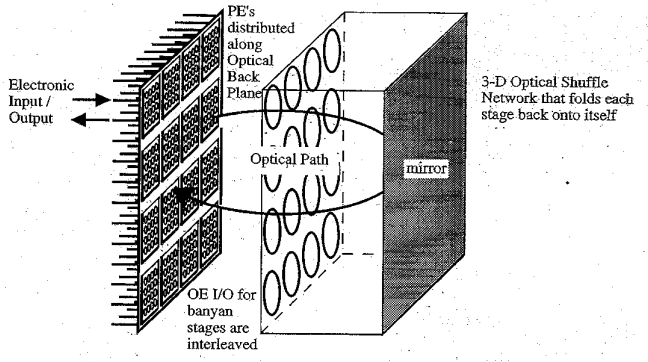


Fig. 5. Single lenslet array k-shuffle interconnection approach. An array of processing elements (PE's) is interconnected through a shuffling retro-reflective optical system in an interleaved fashion.

IC's are limited to a few watts/cm^2 of power dissipation. If the optoelectronic elements of the smart pixel are assumed to dominate the power dissipation, and the power needed for a smart pixel link is of the order of a few mW, then the number of emitters/detectors will be limited to several hundred/cm^2. This density limitation will determine the number of smart OEIC's used, the number of nodes/OEIC, and the number of stages/node located at each smart pixel site.

In the schematic depiction of Fig. 4, the shuffling optics interconnect the front of the plane to the back of the plane with the smart pixel logic connecting the two sides through the substrate. It is desirable to place emitters and detectors on the same side of the substrate, leaving the backside of this plane to function as a "backplane" and interface with electronic boards behind the optoelectronic backplane. A number of implementations of this approach can be considered; this paper focuses on an implementation based on the reversible k-shuffle interconnections, of the type depicted in Fig. 3.

In the SB, each node in Fig. 3 should be considered to actually be an identical cluster of optoelectronic elements, corresponding to all of the stages of the network. For example, in a 256 node SB switch, with 25 interleaved stages, there will be an array of 25 light sources interleaved with an array of 25 detectors located at each smart pixel location in the figure. Since the optical system is shift invariant within the field of view of each pair of elements, each of the rays in Fig. 3 should be considered to be interconnecting the cluster of sources at the input with the cluster of detector elements at the output, in a $1:1$ manner. The second plane (e.g., the one on the right in Fig. 3) could be used as an auxiliary active plane for routing purposes. However, a preferred and more compact variation of this concept is shown in Fig. 5. Here a mirror is used to retro-reflect the shuffled image of the interleaved array back onto itself. A single lenslet array performs the k-shuffle interconnection in each dimension. When coupled with the self-similar grid array concept, the optical architecture depicted in Fig. 5 uses a single lens for each OEIC. Thus, each lens is simultaneously the input and output optical element for an OEIC.

The high resolution interleaved shuffle interconnection scheme depicted in Fig. 5 demands lenses that provide wide field imaging with very low aberrations and good lens-to-lens (and hence OEIC-to-OEIC) registration. It is therefore likely that the required lenses will be multielement and/or aspherical in design, with the possible inclusion of diffractive optical techniques. Fortunately, there is a wide body of applications that demand similar requirements, most notably high performance video camera and projection lenses. In the SB, the basic imaging system consists of a lenslet, a mirror, and another lenslet that is laterally offset from the first lens. The amount of lateral displacement is determined by the relative positions of the OEIC's on the back plane. The SB, therefore, has several optical design issues to be resolved. These include: alignment, interleaved registration, distortion, focal length variation tolerances, VCSEL image resolution, and the retro-reflective folding of the optical system. These issues combine to determine the ability of the optical system to image the interleaved VCSEL array onto the interleaved detector array with good efficiency and low crosstalk. The alignment tolerances are determined by the size and spacing of the detectors and VCSEL's. One compensating feature of the smart pixel design for the SB will be the use of monochromatic sources, such as VCSEL's. Chromatic aberrations will therefore not be an issue.

B. Routing Control Approach

The physical collocation of stages afforded by the 3-D optics offers an important feature not practical with a VLSI implementation. Because the interleaving scheme allows all stages of a single node of the network to be in close proximity to each other, they may all reside on a single OEIC. This means that packets may be removed from the network *at any stage*, not just at the end of a fixed banyan, as in the TB. This is the essential feature of the SB architecture.

Fig. 6 depicts the *unfolded* shuffle-based SB architecture. The SB takes advantage of the constant availability of output paths to reduce the network traffic quickly and thereby increase the networks performance. When a packet's route is blocked it is misrouted *once*, then it begins routing immediately. If the packet is not misrouted again, it will reach its destination in $\log_k N$ stages from the point of its deflection. This will be the end of this packet's banyan, which has slid to align with the misrouting incident, and the packet will be removed. With the SB, resources are not wasted by simply routing the misrouted packets to the end of the banyan; the rerouting begins immediately. After the first banyan, packets can physically leave the network at any stage.

The sliding banyan routing strategy is possible only because the interleaved topology afforded by the 3-D shuffle optics provides colocation of the outputs for any stage of a given output node. Thus, only a single IC output driver is required for each node. If the unfolded network depicted in Fig. 6 were to be implemented in VLSI technology, then an output driver for each node *and* for each stage would be required. In other words, each of the thick vertical lines in Fig. 6, comprising N links for N nodes, would require a physical link and link driver of some sort. This would be totally impractical for a large network due to the large power consumption required for each of the output drivers. For example, consider a 1024

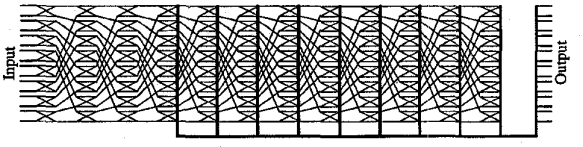


Fig. 6. Shuffle based sliding banyan architecture (shown unfolded).

node SB with 40 stages, in a perfect shuffle ($k = 2$) based banyan network. The first banyan requires $\log_2(1024) = 10$ stages. The optically folded and interleaved topology of the SB requires just 1024 output drivers from the smart pixel plane. A VLSI implementation, in which the outputs from each stage reside on different IC's or boards, however, would require (40-10)×1024 output drivers—a factor of 30 more than the optical SB.

This advantage is not without some added packet coding complexity. If packets are to be removed from the network at any stage, the packet must contain information about the number of stages correctly routed. When this number reaches $\log_k N$, the packet exits. Self-routing algorithms use a header with the destination address. The TB also requires a conflict bit to determine if the packet has been misrouted. Often the destination address will be rotated by each stage as it is routed; this way, the next stage need only inspect the first bit (if $k = 2$) to determine the switching. For the SB, the conflict bit would be replaced with a header containing the number of successfully routed stages, and the destination address would not be rotated. The number of successful stages is used to determine on which bit of the destination address the packet is to be routed. It is a simple inspection—if the number successful stages is m, then this bit of the destination is the determining one. Furthermore, the SB can give priority to those packets which had the highest number of successful stages in their history, i.e., are closest to their destination. This would prevent a packet, which had just begun rerouting, from interfering with a packet that is close to completing its routing through the network. The conflict bit is replaced by this priority number. Misrouted packets simply set this number to zero, then begin routing again.

C. Sliding Banyan Performance

The type of folded optical shuffle approach employed will determine the order (k) of the local cross-bar switches on the OEIC's. The routing algorithm must then accommodate the local switching scheme within the smart pixel associated with the k adjacent nodes that must pass through the local switch. A digital simulation and analytical model have been developed for estimating the blocking performance of the SB and other similar networks, under various operating configurations and traffic conditions. Following are results which validate the SB in terms of blocking performance, latency, and switching resource requirements.

First, the TB was simulated on a 1024-node network with a 2-D (32×32) separable shuffle interconnection ($k = 4$). The number of stages per banyan is $n = \log_4(1024) = 5$. The simulation tagged misrouted packets so that they would not interfere with any correctly routed packets. Randomly

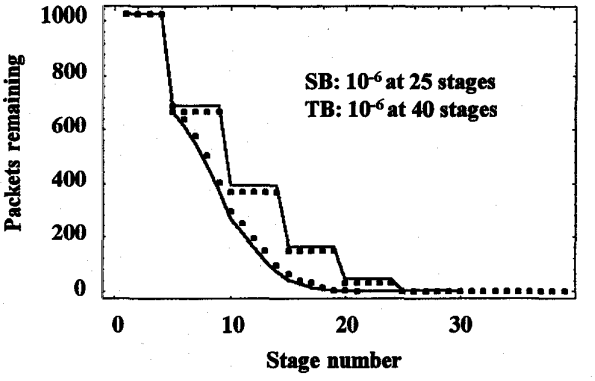


Fig. 7. Performance of sliding banyan and tandem banyan networks. The solid lines show the simulated performance results, while the dotted lines show the analytical performance prediction results.

generated unity permutation traffic was used in the simulations. In this traffic pattern every input and every output is used exactly once; there is no output conflict. This is a standard type of traffic used in evaluating such networks. A typical plot of the number of packets remaining on the network versus the stage number is shown in Fig. 7. Packets are removed only at the end of banyans; this is why the number changes only at integral numbers of five stages. Notice that the final banyan in vastly underutilized; it only routed one packet for this run. A total of seven banyans were required to route the packets, so 35 stages were needed in all.

Next, the SB was simulated, using the identical unity permutation traffic pattern (originally generated in a random fashion) that was used to evaluate the TB. The packets were tagged with the number of consecutive successfully routed stages, and this number was used to prioritize the routing of any conflicts that arose. These results are also shown in Fig. 7. Note that no packets are removed before the fifth stage, but then packets are removed at every stage thereafter. The packets removed in the sixth stage are those which were misrouted in stage 1, then had five successful stages. The network required 21 stages to route all 1024 packets, resulting in 14 fewer stages then the standard TB.

As a check on the simulated blocking performance results, a statistical model of the probability of blocking in the SB and TB was developed. This model is based on a modification to a banyan performance expression derived in [6]. The results are plotted in Fig. 7 alongside of the simulation results, and show close agreement to them. Using the analytical approximation, the number of stages required for an arbitrary packet blocking probability (P_B), was determined; the results are plotted in Fig. 8. These show that the SB maintains an advantage in number of stages over a wide range of operational P_B.

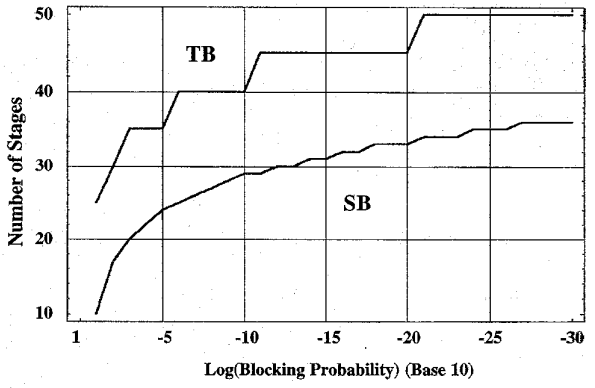


Fig. 8. Number of network stages as a function of required packet blocking probability for the sliding banyan and tandem banyan networks.

The analytical model used to generate the data in Fig. 8 is based on the approximation that the probability that a packet survives the first m stages of a network composed of $k \times k$ switch elements, when the probability of a packet entering the first switch is p, is given by [6]:

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

The probability that a packet entered this series of m stages and was blocked is the difference between p and p_m:

$$p_b(k,m,p) = p - \frac{2k}{m(k-1) + \frac{2k}{p}}. \tag{2}$$

Using this expression for packet survival, the analysis of a TB network is straightforward. Since unity permutation traffic is assumed, the initial probability (p) is 1, and one banyan's worth or $\log_k N$ stages is considered at a time. Using this iterative expression the probability of blocking in the ith banyan can be expressed in terms of the $(i-1)$th:

$$p_i = p_b(k, \log_k N, p_{i-1}). \tag{3}$$

In this manner the probability of blocking in one banyan is used as the input probability in the next. The load of the network is reduced until the probability of blocking of the final banyan is below the threshold required by an application, in this case 10^{-6}. i banyans or $i \times \log_k N$ stages are required to perform the routing with the requisite blocking probability.

The analysis of the SB architecture relies on (1) as well, only it is a more complicated process. For the SB to be successful, packets which have completed the greatest number of consecutive successful routings must have priority over packets with fewer consecutive successful routings. When implementing the SB, a counter placed in the header indicated which bit of the destination address on which to decide—the higher the counter, the greater the priority. To simulate this mechanism all packets must be grouped in the network into $\log_k N$ groups. The designation of each group is the number of consecutive successful routings it has made $(P_\#)$. All packets begin with zero successful routings. Again, unity permutation traffic is used, so the initial input probability is 1. The number of successful routings after 1 stage is determined as $p_m(k,1,1)$. This successful packet probability is placed in group 1 (i.e., one successful routing), and the probability of zero successes (group 0) is $1 - P_1$. At the next stage priority is given to those packets in group 1, $P_2 = p_m(k,1,P_1)$, but the packets with lower probabilities must contend with others like themselves, as well as the successful packets of group 1, so $P_1 = p_m(k,1,P_0 + P_1)$. In this fashion every one of the probabilities of the $\log_k N$ sized groups are calculated and the zero group is set to the remainder of all the packets (exclusive of those which successfully completed $\log_k N$ stages and were removed). Thus, the manner of computing the probability of packet existence at every stage is:

$$P_i = p_m\left(k, 1, \sum_{j=i}^{\log_k N} P_j\right)$$
$$P_0 = 1 - \sum_{j=1}^{\log_k N} P_j. \tag{4}$$

When all of these probabilities have dropped below the threshold (i.e., 10^{-6}) then the network has achieved this probability of blocking and the number of stages required to achieve this is determined. These analytical expressions are a very close approximation for the simulations run, as shown in Fig. 7. In order to achieve a probability of blocking of 10^{-6} for a 4 shuffle $(k = 4)$ the TB required 40 stages, whereas the sliding banyan required only 25. This is a savings of approximately 30%.

As a measure of resource utilization efficiency, the optical SB self-routing network may be compared with the Benes network, which is known to be the smallest network for which all permutations are realizable, but for which no pipelineable self-routing algorithm exists [19]. The equivalent Benes network requires $2(\log_k N) - 1$ stages, or nine in our example traffic above. The 25 stages for the SB needed to achieve very low blocking probability is within a factor of three to the Benes, yet has the critical advantage of self-routing.

D. Optical Module Experiments

Current chip placement technology provides the ability to align IC's to a registration accuracy of approximately 10 μm across a multichip substrate. This registration accuracy will be suitable for the multichip SB implementation. If the smart pixel OEIC's are assumed to have, on each chip, submicron optoelectronic registration (comparable to modern photolithographic IC technology capabilities), then the dominant source of optical misalignment and loss of efficiency will come from the lenslet array itself. Ultimately, custom wide angle imaging optics will be used for the SB. However, the requirements of the SB optical interconnection module are not unlike those of existing wide angle video and projection lenses. Therefore, preliminary experiments were conducted to evaluate the resolution and registration capability of commercially available lenses for use in the SB interconnection concept.

In the experiments, the smart pixel emitters were simulated with Honeywell-supplied arrays of VCSEL's, including a 4×4

array of 10 μm VCSEL's on a grid with a center-to-center spacing of 630 μm and a 1×32 array of VCSEL's with a center-to-center spacing of 140 μm. The VCSEL arrays were precisely placed at the various positions of smart pixel OEIC's in the SB backplane, as depicted in Fig. 5. The smart pixel output detector array was simulated by capturing the VCSEL array imagery on a high resolution CCD camera array, precisely positioned at other OEIC positions in the smart pixel backplane of the test set-up. Pairs of lenses under test were positioned to emulate the shuffle interconnection lens positions depicted in Fig. 5. The results were analyzed assuming 40 μm detectors spaced on the same grid as emitter arrays.

Fig. 9 shows the results of a registration and resolution experiment in which three collinear VCSEL elements, spaced by 630 μm, were imaged onto a CCD array with a lens array system consisting of $f/1.5$, 25 mm focal length miniature video camera lenses. The overlaid white outline squares indicate the size and precisely registered locations of evenly spaced 40 μm detectors that would be part of the smart pixel. As shown in the figure, the off-the-shelf video lenses perform fairly well in this off-axis imaging system. The resolution across the FOV of the system indicates that most of the light emitted by the VCSEL's would be captured by an appropriately positioned 40 μm detector element. Some blurring occurred at the widest angle position (primarily due to vignetting of the narrow VCSEL beam by the barrel of the lens mount). The inherent distortion of the imaging system leads to misregistration of the VCSEL images with respect to the correct detector positions (indicated by the square box outlines). At the widest angles the array's images are beginning to misalign with the target detector array patterns. This distortion becomes especially apparent for total fields of view greater than about 20 degrees, occurring when the VCSEL's were placed at the extreme points of the input field (approximately 4.4 mm from the axis). As shown in Fig. 9, registration errors of approximately 25 μm occurred for the widest angle VCSEL images.

The focusing and registration results show good performance, despite the fact that the inexpensive test lenses were not selected to be precisely matched in focal length or other performance criteria. These results, therefore, suggest that better matching of commercially available lenses, or custom designed lenses, will provide the performance necessary for the SB optical module.

IV. Discussion

Several promising smart pixel technologies are now emerging as candidates for use in the SB packet switching architecture, including both emitter and modulator based approaches in monolithic and hybrid technologies [20]. At this stage of the study it appears that integrated emitter based (with either VCSEL's or LED's) smart pixels will more readily be incorporated into the envisioned optical system (as shown, for example, in Fig. 5) than modulator based technologies.

In practice, it is envisioned that packets will enter the SB switching fabric on a fiber optic or coax bundle that interfaces to "line cards" stacked across the optoelectronic backplane. These boards then interface to the SB OEIC's. For a 1024 node switch, consisting of 40 stages, there will be over 40 000 VCSEL/detector pairs distributed across the backplane. The power consumption is conservatively estimated to be 10 mW/smart pixel/stage, to include all electronic and optoelectronic power dissipation sources. Estimating power dissipation on a chip at a maximum of 2 W/cm^2 results in a maximum smart pixel I/O density of 200 /cm^2. The SB architecture, consisting of 40 000 optical links, would then require $\sim$150 cm^2 of OEIC chip area. A backplane of 20 cm $\times$ 20 cm would have an OEIC fill factor of $\sim$50%, which is consistent with practical MCM packaging.

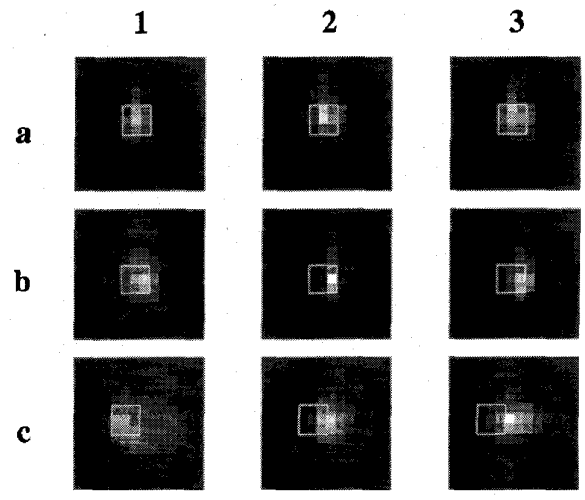


Fig. 9. Focusing and registration data for three colinear VCSEL's separated by 630 μm in the smart pixel plane. (a) Output for on-axis imaging, (b) Output for VCSEL's centered $\sim$2.5 mm from axis. (corresponding to $\sim$5° off-axis), (c) Output for VCSEL's centered $\sim$4.4 mm from axis. (corresponding to $\sim$10° off-axis). Box outlines correspond to the positions of properly registered 40μm wide detectors.

V. Conclusions

Current all-electronic control and routing technology is not cost-effectively scalable to the anticipated high throughput networks of the future owing to the fundamental limitations of metallic interconnections. The sliding banyan uses a fundamental advantage of free-space optical interconnections to reduce the switching and routing resources necessary in high throughput ATM switching applications. The novel free-space optical interconnection scheme provides the necessary high bisection width shuffle interconnection, while eliminating the need for large numbers of power hungry chip-to-chip drivers. Furthermore, the new 3-D interleaved topology, based on the rapidly maturing smart pixel technology, obviates the need for distributing the control and switching resources across numerous optical or electronic boards and instead provides a single backplane interface for the nodes of the switch. Preliminary experiments suggest that the high precision optical system needed to implement the sliding banyan can use existing high performance lens design techniques to achieve the need resolution and registration accuracy. Simulations and analysis show the sliding banyan to significantly reduce the resources required for a given blocking probability. The

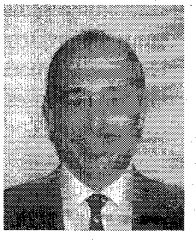
switching resources required to achieve blocking probabilities of 10^{-6} are within a factor of three of the Benes network, known to have the minimum number of stages for a non-blocking network. However, whereas the Benes network is totally impractical for high throughput ATM switching, owing to its lack of efficient routing control, the sliding banyan's fundamental advantage stems from its simple self-routing deflection control strategy, in which packets are removed from the network as soon as they find their way to their destination node. Self-routing control overhead is thus minimized in the sliding banyan. The combination of lowered switching resources and minimized control overhead of the sliding banyan topology provides an ATM switching architecture that is scaleable to aggregate bandwidths in the Tb/s regime.

ACKNOWLEDGMENT

The authors wish to thank J. J. Levy for his technical support during the experimental portions of this effort and Dr. M. Hibbs-Brenner, of Honeywell Research Center, who provided the VCSEL arrays.

REFERENCES

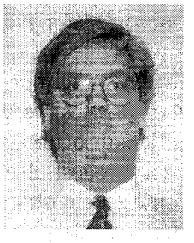
[1] J. Hui, "Switching integrated broadband services by sort-banyan networks," in *Proc. of IEEE,* vol. 79, pp. 145–154, Feb., 1991.
[2] F. T. Leighton, *Introduction to Parallel Algorithms and Architectures; Arrays, Trees, Hypercubes,* San Mateo, CA: Morgan Kaufmann, 1992.
[3] H. S. Stone, "Parallel processing with the perfect shuffle," *IEEE Trans. Comput.,* C-20, pp. 81–89, 1971.
[4] M. W. Haney and M. P. Christensen, "Optical freespace sliding tandem banyan architecture for self-routing switching networks," *Digest Int. Conf. Optic. Comput.,* pp. 249–250, Aug., 1994.
[5] M. W. Haney and M. P. Christensen, "Free-space optical sliding banyan network," *Digest OSA Topic. Meeting: Photon. Switching,* pp. 27–29, Mar., 1995.
[6] C. P. Kruskal and M. Snir, "The performance of multistage interconnection networks for multiprocessors," *IEEE Trans. Comput.* C-32, no. 12, pp. 1091–1098, Dec., 1983.
[7] F. A. Tobagi, T. Kwok, and F. M. Chiussi, "Architecture, performance, and implementation of the tandem banyan fast packet switch," *IEEE J. Select. Areas Commun.* no. 8, pp. 1173–1193, Oct., 1991.
[8] A. W. Lohmann, *et al.*, in *Digest Conf. Optic. Comput.,* Washington, D. C.: Optical Society of America, paper WA3, 1985.
[9] A. W. Lohmann, "What classical optics can do for the digital optical computer," *Appl. Optic.,* vol. 25, pp. 1543–1549, 1986.
[10] G. Eichmann, and Y. Li, "Compact optical generalized perfect shuffle," *Appl. Optic.,* vol. 26, pp. 1167–1169, Apr. 1987.
[11] S.-H. Lin, T. F. Krile and J. F. Walkup, "2-D optical multistage interconnection networks," in *Proc. SPIE,* vol. 752, pp. 209–216, 1987.
[12] K.-H. Brenner and A. Huang, "Optical implementations of the perfect shuffle interconnection," *Appl. Optic.,* vol. 27, pp. 135–137, Jan. 1988.
[13] C. W. Stirk, R. A. Athale, and M. W. Haney, "Folded perfect shuffle optical processor," *Appl. Optic.,* vol. 27, pp. 202–203, 1988.
[14] A. A. Sawchuk and I. Glaser, "Geometries for optical implementations of the perfect shuffle," in *Proc. SPIE,* vol. 963, p. 270, 1988.
[15] M. W. Haney and J. J. Levy, "Optically efficient free-space folded perfect shuffle network," *Appl. Optic.,* vol. 30, no. 20, pp. 2833–2840, July, 1991.
[16] G. C. Marsden, P. J. Marchand, P. Harvey, and S. C. Esener, "Optical transpose interconnection system architecture," *Optic. Lett.,* vol. 18. no. 13, pp. 1083–1085, July 1, 1993.
[17] M. W. Haney, "Pipelined optoelectronic free-space permutation network," *Optic. Lett.,* vol. 17, no. 4, pp. 283–285, Feb., 1992.
[18] M. W. Haney, "Self-similar grid patterns in free-space shuffle/exchange networks," *Optic. Lett.,* vol. 18. no. 23, pp. 2047–2049, Dec. 1, 1993.
[19] F. Tobagi, "Fast packet switch architectures for broadband integrated services digital networks," in *Proc. IEEE* 78, pp. 133–167, 1990.
[20] *IEEE Digest Catalog* No. 94TH0606-4, July, 1994. IEEE/LEOS Topical Meeting on Smart Pixels,



Michael W. Haney (M'80) received the B.S. degree in physics from the University of Massachusetts in 1976, the M.S. degree in electrical engineering from the University of Illinois in 1978, and the Ph.D. in electrical engineering from the California Institute of Technology in 1986.

From 1978 to 1986 he was with General Dynamics, where his work ranged from the development of electrooptic sensors to research in photonic signal processing. In 1986 he joined BDM International, Inc., where he became a senior principal staff member and the Director of Photonics Programs. In 1994 he joined George Mason University as an Associate Professor of Electrical and Computer Engineering. His current research activities center on the application of hybrid optoelectronics to free-space optical interconnection architectures for high throughput switching and signal processing applications.

Dr. Haney is a member of OSA and SPIE and has participated as the chair of several conferences and workshops. He is the IEEE Communications Society's technical committee chairman for Highspeed Interconnections in Digital Systems. He has contributed to more than 45 journal and conference papers and holds one patent in optical information processing.



Marc P. Christensen (M'95) received the B. S. degree in engineering physics from Cornell University in 1993. He is currently pursuing a Ph.D. at George Mason University in the School of Information Technology.

In 1991 he joined the optoelectronics technology group at BDM Federal. Inc., and is currently a part-time Associate Staff Member. His current research interests include free space optical interconnects for packet switching and distributed computing. He has contributed to 12 technical conference papers and journal articles and holds one patent.