Analytical Evaluation of Improved Access Techniques in Deflection Routing Networks

Alberto Bononi, *Member, IEEE*, and Paul R. Prucnal, *Fellow, IEEE*

***Abstract*— This paper presents an extension of a known analytical model for the performance evaluation of nonpriority deflection routing networks in uniform traffic. The extension allows the analysis of improved access techniques. The key features of the analytical technique are described by casting it in a very simple setting: nonpriority hot-potato in a two-connected slotted shufflenet (SN) network. Results are presented for three access techniques: transmit-no-hold (TXNH), transmit-hold (TXH), and bypass queuing (BQ).**

I. Introduction

THIS paper presents extensions to the standard model [1], [2] for the analysis of deflection routing [3] in slotted, regular networks under uniform traffic.

A common assumption of the standard model is that a node injects a packet whenever it has one ready at its transmitter (TX) and at least one of the input slots is free after reception (RX) of packets destined to the node itself. A transmission occurs even if such event causes a deflection.

The contribution of this paper is a novel formulation of the standard model that allows extensions to improved access techniques. Specifically, the transmitter could hold up its packet whenever injecting it would cause a deflection. Better yet, the transmitter could select, among the packets waiting for transmission at the node, another one that is not in conflict with the packet in transit, thus avoiding the head-of-line blocking caused by hold-ups.

The analysis is cast in a very simple setting: nonpriority deflection routing without buffers (hot-potato, [3]) in a two-connected shufflenet (SN) topology [4]. Section II introduces the novel formalism, while Section III details the key steps leading to the desired throughput and hop-delay curves. Section IV presents the numerical results for the chosen topology and concludes the paper.

II. Network Model

A network is said to be *regular* when all nodes are topologically equivalent. The traffic is said to be *uniform* when 1) all nodes are equally active, generating a new packet at each slot with the same probability, and 2) the destinations of packets generated at each node's TX are chosen uniformly among all nodes (except the source) in the network, and independently slot by slot.

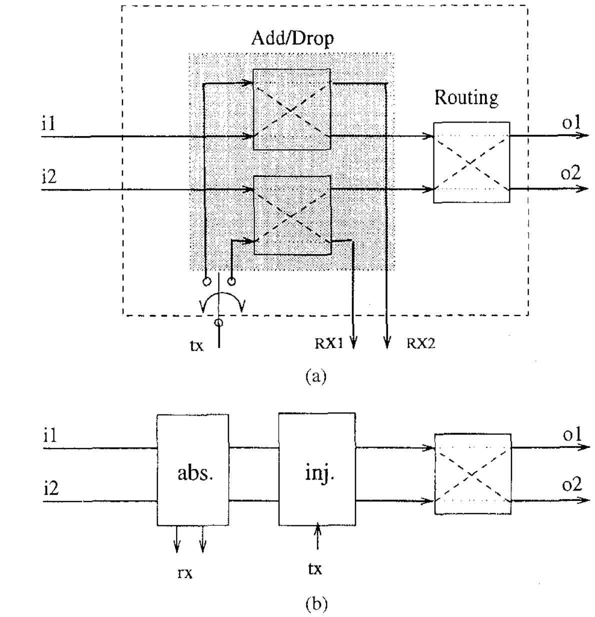


Fig. 1. (a) Physical and (b) logical node structure.

Regularity and uniform traffic ensure that the traffic flowing through a node is statistically the same for every node. Hence, performance evaluation is accomplished by focusing on a single node.

We illustrate the extensions to the known analytical model with the simplest node structure for a two-connected network, the one shown in Fig. 1(a). The node consists of two crossbar switches for injection/absorption of traffic destined to the node (Add/Drop block), followed by a crossbar routing switch. The logical flow of node operations is *absorption, injection* and *routing,* as depicted in Fig. 1(b).

Slots arrive aligned at the node's inputs i_1 and i_2. They can be empty (E), can carry a packet *f*or the *n*ode (FN), or a packet that *c*ares (C) to exit on output 1 or on output 2, or a *d*on't *c*are (DC) packet whenever both node outputs provide shortest-paths to its destination. When two care competing packets are present at the input links, the nonpriority hot-potato routing algorithm assigns one at random to the desired output, and deflects the second.

Manuscript received May 1, 1995; revised May 12, 1996; approved by IEEE/ACM Transactions on Networking Editor R. Ramaswami. This paper was presented in part at the 28th Annual Conference on Information Sciences and Systems, Princeton, NJ, March 16–18, 1994. This work was supported by The Defense Advanced Research Project Agency under Contract F19628-94-G-0045.

A. Bononi is with the Dipartimento di Ingegneria dell'Informazione, Universita di Parma, I-43100 Italy.

P. R. Prucnal is with the Department of Electrical Engineering, Princeton University, NJ 08544 USA.

Publisher Item Identifier S 1063-6692(96)07565-6.

Define u as the input slot utilization, i.e., the probability that an input slot carries a packet. Define P_{dc} and r as the probability that an incoming packet is DC and FN, respectively. We make here the usual key assumption that, at every time-slot k, the input arrivals $i_1(k), i_2(k)$ are independent random variables (rv's). We also assume that they have the same probability distribution

$$\boldsymbol{f}_i = \{\Pr[i_j = s], s \in \{E, DC, C2, C1, FN\}\} \\ j = 1, 2.$$

From the above definitions one gets

$$\begin{aligned}\boldsymbol{f}_i &= \{f_i(E), f_i(DC), f_i(C), f_i(FN)\} \\ &= \{1-u, uP_{dc}, u(1-P_{dc}-r), ur\}\end{aligned}$$

and it is assumed that, among care packets, outputs one and two are equally likely.

To keep the analysis simple, we assume the TX has no local input queue. New TX packets arrive in each slot with probability g, the generation probability. If both input links contain a flow-through packet not destined to the node, local blocking occurs and the local packet is discarded. The uniform traffic pattern assures that all destinations except the source are equally likely for TX packets. Let P_{dc0} be the fraction of DC destinations, i.e., those that can be reached from the source from either output link in the same minimal number of hops. Regularity of the network ensures that half of the remaining care destinations will be for output one and half for output two. With these definitions, the local arrival $tx(k)$ at any time-slot k is a rv, independent of $i_1(k), i_2(k)$, with distribution

$$\begin{aligned}\boldsymbol{f}_{tx} &= [f_{tx}(E), f_{tx}(DC), f_{tx}(C)] \\ &= [1-g, gP_{dc0}, g(1-P_{dc0})].\end{aligned}$$

III. Throughput and Delay Evaluation

We will now detail the throughput and delay analysis for three access techniques:

1) *transmit no-hold*, where a TX packet is injected whenever at least one input slot is empty, i.e., when an injection is possible;
2) *transmit hold*, where a TX packet is injected when an injection is possible *and* there is no conflict with a (possibly present) flow-through packet; and
3) *bypass queueing*, where, if an injection is possible and there is a conflict, the present packet is discarded and a new packet is drawn uniformly among all *care* destinations nonconflicting with the flow-through packet. At loads g before saturation, an actual input queue will not always have the "right" nonconflicting packet, and hence results for BQ provide an upper bound on the performance of an actual BQ system. However, at full load, $g = 1$, BQ is indeed equivalent to the case of a saturated infinite shared input queue at the TX.

A. Slot Utilization

Expressions for the steady-state slot utilization u will now be derived. Refer to Fig. 1(b). After the absorption block, packets FN are removed and replaced by empty slots. Hence, after absorption, the distribution of the two inputs changes to $\tilde{\boldsymbol{f}}_i$, which differs from $\boldsymbol{f}_i$ only in the entries: $\tilde{f}_i(E) = (1-u+ur)$ and $\tilde{f}_i(FN) = 0$. This is the new input distribution.

After absorption we can evaluate the quantities

$$P_{\text{bcf}} = u^2(1-r)^2 \quad \text{and} \quad P_{\text{ocoe}} = 2u(1-P_{\text{dc}}-r)(1-u+ur)$$

indicating the probability of having both input channels full (bcf), and one care and one empty (ocoe), respectively.

At steady-state, at each node, the average number of absorbed packets T_{abs} must equal the average number of injected packets per slot T_{inj}, their common value being the throughput per node T. Since on average ru FN packets reach the node from each input and are all absorbed, we have $T_{\text{abs}} = 2ru$. By Little's law, the throughput per node in two-connected networks is $T = 2u/H$, so that one immediately gets: $r = 1/H$.

The injected throughput can be expressed as

$$T_{\text{inj}} = \sum_{s \in \{DC, C\}} \Pr[tx \,\text{is injected}/tx = s] \; \Pr[tx = s].$$

For TXNH and BQ we have

$$T_{\text{inj}} = (1-P_{\text{bcf}})(f_{tx}(DC) + f_{tx}(C)) = (1-P_{\text{bcf}})g \qquad (1)$$

since for all packet types an injection is possible with probability $(1-P_{\text{bcf}})$.[1]

For TXH we have instead

$$T_{\text{inj}} = (1-P_{\text{bcf}})[f_{tx}(DC) + f_{tx}(C)(1-P_{b0})] \qquad (2)$$

where

$$P_{b0} = \left[\frac{P_{\text{ocoe}}/2}{(1-P_{bcf})}\right] \qquad (3)$$

represents the blocking probability for TX care packets, i.e., the probability of having a conflicting care flow-through packet conditioned on the event "an injection is possible".

Solving the equation $T_{\text{abs}} = T_{\text{inj}}$ gives an explicit expression for u. For TXNH and BQ we get

$$u = \frac{\sqrt{r^2 + g^2(1-r)^2} - r}{g(1-r)^2}$$

while for TXH we get $u = (\sqrt{B^2 + 4gA} - B)/2A$, where

$$\begin{aligned}A &= g(1-r)[(1-P_{dc0})P_{dc} + P_{dc0}(1-r)] \quad \text{and} \\ B &= 2r + g(1-P_{dc0})(1-P_{dc}-r).\end{aligned}$$

[1] In BQ, when a Care packet is blocked, a nonconflicting packet of the same type, i.e., a Care packet, is drawn.

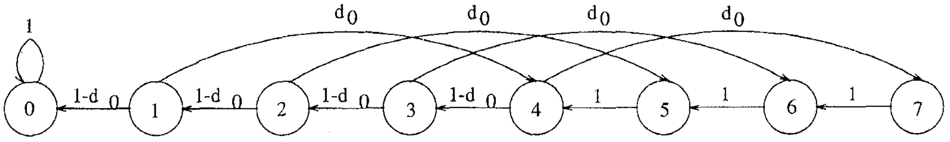


Fig. 2. Markov chain describing the random walk of the test packet in a SN(2, 4) topology.

B. Deflection Probability

Because of the regularity of the topology and the uniform traffic assumption, the global network traffic is a merger of independent, statistically identical traffic streams directed to each destination. Any packet will be a "typical" packet, whose trajectory toward destination can be modeled as a random walk in a homogeneous "gas" of interfering packets [1], [2]. We now evaluate the deflection probability d of a flow-through test packet entering a care (with respect to its destination) intermediate node, and the deflection probability d_0 of a test care packet at its injection node.

Refer again to Fig. 1(b). The flow-through care test packet is at one of the two inputs and bypasses the absorption and injection blocks, reaching the routing block. The probability P_b that another care conflicting packet reaches the routing block on the second channel is, for TXNH, $P_b = \frac{1}{2}\{f_i(C) + \tilde{f}_i(E) f_{tx}(C)\}$, where 1/2 is the probability that the care test and the other care collide.

In TXH and BQ this reduces to $P_b = \frac{1}{2} f_i(C)$, since a conflicting packet is never injected at the TX. P_b is the probability of a conflict for the test packet. Thus, in no-priority hot-potato, the flow-through deflection probability for the care test is

$$d = P_b/2. \tag{4}$$

As for the initial deflection probability of a care test, d_0, this is by definition zero for TXH and BQ, while for TXNH it is $d_0 = P_{b0}/2$.

C. Average Number of Hops and Don't Care Probability

The random walk of the test packet toward its destination is modeled as an absorbing Markov chain whose states are defined by the network nodes, the only absorbing state being the destination node [2], [6].

For some topologies, like SN, it is possible to speed up the computation by drastically reducing the number of states in the chain. This is done by combining in a single state all nodes with same distance to destination. The test packet thus performs a random walk on the integers $0, 1, \cdots d_{\max}$, where $d_{\max}$ is the maximum distance to destination [7].

The solution procedure presented next can be applied to any regular topology, whether or not a reduced state-space can be obtained.[2] However, for illustration purposes, a SN topology will be used.

A specific example of the absorbing Markov chain is given in Fig. 2 for a 64-node SN(2,4) topology.

A $\mathrm{SN}(q,k)$ topology has $N = kq^k$ nodes arranged in k columns of q^k nodes each, and there is a perfect shuffle connection among nodes in adjacent columns [4]. The maximum distance between nodes is $d_{\max} = 2k - 1$. Fix a destination node. All nodes reachable in less than $k + 1$ hops proceeding backward are care with respect to that destination. All the remaining nodes, at distance $k+1, \cdots, 2k-1$ are don't care. A deflection of the test packet flowing toward that destination at a node at distance i brings the packet back to the set of nodes at distance $i + k - 1$. Finally, the number of nodes $n(i)$ at distance i is

$$n(i) = \begin{cases} q^i, & 1 \le i \le k-1 \\ q^k - q^{i-k}, & k \le i \le 2k-1. \end{cases} \tag{5}$$

The don't care probability at the injection step P_{dc0} is easily obtained from (5).

In Fig. 2, the states represent the distance in hops of the test packet to its destination. State zero is the absorbing state of the chain.

Fig. 2 refers to the initial step of the walk, when the packet is at its injection node. For every step after the first one, the packet is flow-through at one input port, and the transition probability d_0 must be changed to d.

For all steps $t = 1, 2, \cdots$, the transition probabilities $\pi(l, m)$ from state m to state $l, l, m = 0, 1, \cdots, 7$, can be organized in a transition matrix $\Pi = \{\pi(l, m)\}$. Analogously, a matrix Π_0 can be written for the injection step $t = 0$.

Since zero is the only absorbing state, matrix Π is in its canonical form. Taking off the first row and the first column, a matrix Q is obtained. From this, the *fundamental matrix* of the absorbing chain $\mathcal{N} = (I - Q^T)^{-1}$ is obtained [8], where I is the 7×7 identity matrix. The entries of $\mathcal{N} = \{n(l,m)\}, l, m = 1, \cdots, 7$, give the expected number of times in each nonabsorbing state m for each possible nonabsorbing state l after the first hop [8].

Let $\boldsymbol{p}_0$ be the probability state (column) vector at the injection step. The state after the first hop is $\boldsymbol{p}_1 = \Pi_0 * \boldsymbol{p}_0$. Let $\hat{\boldsymbol{p}}_1$ and $\hat{\boldsymbol{p}}_0$ indicate respectively $\boldsymbol{p}_1$ and $\boldsymbol{p}_0$ with the first component removed. The ith entry of vector $\mathcal{N}^T \hat{\boldsymbol{p}}_1$ represents the average number of visits before absorption of state i after the first hop. The sum of all entries[3] is thus the average number of hops excluding the first hop

$$\hat{H} \triangleq \|\mathcal{N}^T \hat{\boldsymbol{p}}_1\|_1.$$

Since $\hat{\boldsymbol{p}}_0$ represents the probability of visits at the first hop, the average number of hops before absorption is

$$H = \|\hat{\boldsymbol{p}}_0 + \mathcal{N}^T \hat{\boldsymbol{p}}_1\|_1 = \hat{H} + 1. \tag{6}$$

[2] Once the states of the absorbing chain are defined as the network nodes, as shown for instance in [6, Fig. 4], the destination node being the absorbing state, the procedure outlined here applies verbatim.

[3] Indicated as vector norm 1, $\|\cdot\|_1$.

The sum of the entries of $\mathcal{N}^T\hat{\boldsymbol{p}}_1$ relative to don't care states $i = 5, 6, 7$ represents the expected number of visits before absorption at don't care nodes at which the test packet is flow-through, V_{dc}. The don't care probability P_{dc} is estimated as the fraction of time the test packet is don't care flow-through

$$P_{dc} = \frac{V_{dc}}{H}. \qquad (7)$$

This procedure, making use of the fundamental matrix of the absorbing chain, allows obtaining closed-form expressions for both H and P_{dc} as functions of d, d_0. Such expressions have been explicitly obtained for SN and TXNH in [9] as functions of d only, by neglecting the initial injection step and thus, the dependency on d_0. Similar formulas for the case of two transmitters per node appeared in [10].

A very delicate point is to properly establish the value of $\boldsymbol{p}_0$, which must sum to one. Each entry $\boldsymbol{p}_0(i)$ represents the probability that a packet for 0 has been *injected at distance* i, conditioned on the event "one packet destined to zero has been injected in the network". Now, from (5) and the uniform traffic assumption, the conditional probability that a packet for zero has been *generated* at distance i is $n(i)/N - 1$. This also represents $\boldsymbol{p}_0(i)$ when TXNH and BQ are adopted, since care and don't care TX packets have the same injection probability. However, in TXH, the injection probabilities for TX don't care packets and TX care packets are in ratio one to $(1 - P_{b0})$, as can be seen from (2). Therefore $\boldsymbol{p}_0$ must first be changed in

$$\boldsymbol{p}_0(i) = \frac{n(i)}{N-1}(1 - P_{b0}) \qquad (8)$$

for all distances i corresponding to care nodes, namely $i = 1, 2, 3, 4$ in our example, and then renormalized to one.

IV. Conclusion

The previous results can be put together to get the desired expressions of the throughput $T(g)$ and the hop delay $D(g)$ as functions of the parameter g, the generation probability. The procedure involves the solution of a 2×2 system of nonlinear equations. We start with an initial guess of the quantities $[d, d_0]$. Using the results of Section III-C, the average number of hops H and the don't care probability P_{dc} are expressed as functions of d, d_0, P_{dc0}. Then $r = 1/H$ is obtained. Next $u = u(g, P_{dc0}, r, P_{dc})$ is evaluated as outlined in Section A. Finally, new values for $[d, d_0]$ are obtained from (3) and (4). The process is repeated up to convergence of $[d, d_0]$.

Fig. 3 shows the Hop-Delay/Throughput analytical curves for both a 24-node SN (SN24) and a 64-node SN (SN64). Simulation results are marked with circles. For both SN24 and SN64, simulation statistics were collected for 10 000 clock cycles, after discarding 1000 initial cycles to allow for transients to die out.

It is confirmed that TXH gives improved performance with respect to TXNH. The main reason is that traffic destined to nodes one hop away is never deflected.

An interesting feature of TXH in Shufflenet is that transmission to nodes far away from the source (don't care nodes) occurs more often than those at smaller distance (care nodes), which compensates for the longer propagation delay.

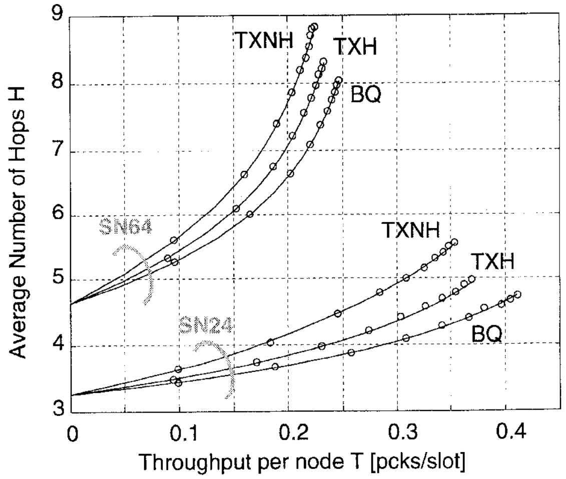


Fig. 3. Hop-delay versus throughput in a 24-node SN (SN24) and a 64-node SN (SN64) for TXNH, TXH, and BQ.

The improvement of the hold-up technique is larger for SN24, and decreases with increasing network size for a fixed node in/out degree of two. This is due to the fact that avoiding one deflection doesn't do much good when the total packet path is already very long.

However, the hold-up technique is quite valuable when the average number of hops is intrinsically small, as is the case with compact, low diameter networks, such as SN24 or in general large networks with large node in/out degree. In fact, this is the most meaningful case, since multihop networks have decent throughput/delay only when the average number of hops is small.

The gain in performance of the hold-up technique is achieved at essentially no added complexity to the node control, since in any case the node has to find out the desired port for the two input slots, and the decision whether or not to inject the TX packet is just based on observing whether or not the TX packet is in conflict with those destinations. Therefore our recommendation is to always use the hold-up technique instead of the standard TXNH.

The best performance for pure hot-potato is achieved by BQ. There is indeed an added complexity in managing the TX input queue according to BQ, and the performance curves suggest that in the complexity/gain tradeoff the hold-up technique TXH is preferable.

References

[1] A. G. Greenberg and J. B. Goodman, "Sharp approximate models of deflection routing in mesh networks," *IEEE Trans. Commun.*, vol. 41, pp. 210–223, Jan. 1993.

[2] A. S. Acampora and A. Shah, "Multihop lightwave networks: a comparison of store-and-forward and hot-potato routing," *IEEE Trans. Commun.*, vol. 40, pp. 1082–1090, June 1992.

[3] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.*, vol. 12, pp. 1–9, Mar. 1964.

[4] A. S. Acampora, M. J. Karol, and M. G. Hluchyj, "Terabit lightwave networks: the multihop approach," *AT&T Tech. J.*, vol. 66, pp. 21–34, Nov./Dec. 1987.

[5] A. K. Choudhury and V. O. K. Li, "An approximate analysis of the performance of deflection routing in regular networks," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 1302–1316, Oct. 1993.

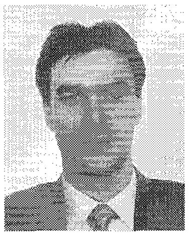
[6] F. Forghieri, A. Bononi, and P. R. Prucnal, "Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks," *IEEE Trans. Commun.,* vol. 43, no. 1, pp. 88–98, Jan. 1995.

[7] A. Krishna and B. Hajek, "Performance of Shuffle-like switching networks with deflection," in *Proc. INFOCOM '90,* San Francisco, CA, June 1990, vol. 2, pp. 473–480.

[8] J. G. Kemeny, H. Mirkil, J. L. Snell, and G. L. Thompson, *Finite Mathematical Structures.* Englewood, NJ: Prentice-Hall, 1959, pp. 404–409.

[9] S.-H. Chan and H. Kobayashi, "Performance analysis of shufflenet with deflection routing," in *Proc. GLOBECOM'93,* Houston, TX, Dec. 1993, vol. 2, pp. 854–859.

[10] A. V. Ramanan, H. F. Jordan, J. R. Sauer, and D. B. Blumenthal, "An extended fiber optic backplane for multiprocessors," in *Proc. 27th Hawaii Int. Conf. Syst. Sci.,* Maui, Hawaii, Jan. 1994, vol. I, pp. 462–470.



Alberto Bononi (S'92–M'95) received the "Laurea in Ingegneria Elettronica" degree (cum laude) from the University of Pisa, Italy, in 1988, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, in 1992 and 1994, respectively.

In 1989, he was a visiting researcher at the University of Parma, Italy, working on coherent optical communications. In 1990, he worked at GEC-Marconi Hirst Research Centre in Wembley, UK, on a Marconi S.p.A. Project on coherent FSK systems. Since 1994, he has been an Assistant Professor of Electrical Engineering at SUNY at Buffalo, NY, teaching courses on circuits and optical networks. His research interests include system design and performance issues in fast packet switching and high-speed all-optical networks.

Paul R. Prucnal (S'75–M'78–SM'90–F'92) received the A.B. degree (summa cum laude) from Bowdoin College, Brunswick, MN, in 1974, and the M.S., M.Phil., and Ph.D. degrees from Columbia University, NY, in 1976, 1978, and 1979, respectively.

He was an Assistant Professor of Electrical Engineering at Columbia University from 1979 to 1984, and was an Associate Professor from 1984 to 1987. He was a member of the Executive Board of Columbia's NSF Engineering Research Center in Telecommunications Research from 1985 to 1987. He is now a Full Professor at Princeton University, where he is Director of the Lightwave Communications Research Laboratory, and served as Acting Director of the newly-established New Jersey Advanced Technology Center in Photonics and Optoelectronic Materials. He has taught courses in the areas of fiber-optic communications systems, quantum electronics, and digital signal processing. He has been a Technical Consultant for Philips Labs, Optical Information Systems Inc., GTE Labs, IBM, Dove Electronics, and AT&T Bell Labs. He has published more than 50 journal papers in the areas of optical networks, photonic switching, optical techniques for advanced VLSI/VHSIC interconnections, and optical signal processing, and holds three patents. He is an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE CIRCUITS AND DEVICES MAGAZINE, and the IEEE LIGHTWAVE COMMUNICATIONS MAGAZINE

Dr. Prucnal is a member of OSA and SPIE.