fied in the switched state of a very first device using carrier injection.

REFERENCES

[1] Y. Silberberg, P. Perlmutter, and J. E. Baran, "Digital optical switch," *Appl. Phys. Lett.*, vol. 51, pp. 1230-1232, 1987.

[2] L. Thylen, P. Svensson, B. Lagestrom, B. Stolz, P. Granestrand, and W. K. Burns, "Theoretical and experimental investigations of 1 × 2 digital optical switches," *ECOC* '89, Gothenburg, Sweden, Sept. 1989, paper WeA12-4.

[3] P. Granestrand, B. Lagerstrom, P. Svensson, L. Thylen, B. Stoltz, K. Bergvall, J. E. Falk, and H. Olofsson, "Integrated optics 4 × 4 switch matrix with digital optical switches," *Electron Lett.*, vol. 26, pp. 4-5, 1990.

[4] J. A. Cavailles, M. Renaud, J. F. Vinchant, M. Erman, P. Svensson, and L. Thylen, "First digital optical switch based on InP/GaInAsP double heterostructure waveguides," *Electron Lett.*, vol. 27, pp. 699-700, 1991.

[5] R. R. A. Syms and R. G. Peall, "The digital optical switch: Analogous directional coupler devices," *Opt. Commun.*, vol. 69, pp. 235-238, 1989.

[6] M. G. F. Wilson and G. A. Teh, "Tapered optical directional coupler," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 85-92, 1975.

[7] A. F. Milton and W. K. Burns, "Tapered velocity couplers for integrated optics: Design," *Appl. Opt.*, vol. 14, pp. 1207-1212, 1975.

[8] Y. Cai, T. Mizumoto, and Y. Naito, "Analysis of the coupling characteristics of a tapered coupled waveguide system," *J. Lightwave Technol.*, vol. 8, pp. 90-98, 1990.

[9] D. R. Rowland, Y. Chen, and A. W. Snyder, "Tapered mismatched couplers," *J. Lightwave Technol.*, vol. 9, pp. 567-570, 1991.

Demonstration of a Deflection Routing 2 × 2 Photonic Switch for Computer Interconnects

D. J. Blumenthal, K. Y. Chen, J. Ma, R. J. Feuerstein, and J. R. Sauer

Abstract—**We present the first reported demonstration of a 2 × 2 self-routing photonic switch. Output port contention is handled using a deflection routing protocol implemented by a pipelined electrooptic processor. Routing and data bits are encoded in separated multiwavelength channels occupying only a portion of a full packet period. This encoding technique accommodates timing uncertainty, reduces the effective electronic processing rate at the switching node, and maintains a high-link throughput. In our single-node demonstration, the packet header contains two address bits plus one priority bit for contention resolution, and the packet payload is represented by a single bit. Control information is encoded at 830 nm with 3.6 nm channel spacing and data at 1300 nm. The switch, implemented with longer wavelength transmission bands, is an efficient building block for large scale, wide area, high capacity networks.**

I. INTRODUCTION

INTERCONNECTION networks are a critical component of multiprocessor computing systems, and their performance, to a large degree, determines the ultimate performance of the system itself. Multiprocessor computers which utilize shared memory and frequent interprocessor communications are efficiently implemented as *tightly coupled* architecture [1]. Tightly coupled systems can be defined as architectures which require high efficiency for single word transfers. A multistage architecture for this network can be illustrated as node hosts (processors and memories) interconnected by a series of switching nodes and links as shown in Fig. 1.

The wide bandwidth of fiber-optic links and photonic switches allow implementation of architectures which can better meet the performance requirements of a tightly coupled interconnect. Wavelength division multiplexing has been used to implement shared access protocols for both circuit switched computer networks and packet switched metropolitan area networks because it efficiently uses the abundant bandwidth of the optical fiber [2], [3]. However, the various costs of these access protocols and their centralized passive optical star topology do not scale well with network size. Multistage networks with multipath characteristics provide efficient use of bandwidth as the number of nodes increases [4].

The 2 × 2 switch reported in this letter is the building block for a multistage, multipath computer interconnect [5]. In this interconnect, a single network node is composed of three 2 × 2 switching elements and a control processor as shown in Fig. 2. A fully functional node has the capability to route information between the node host and network, and act as a 2 × 2 switching element for traffic which remains on the network. A more detailed description of this node and network may be found in [5]. The 2 × 2 switch demonstrated in this letter is extendable to this fully functional node without any significant change in technology.

Manuscript received October 10, 1991; revised November 19, 1991. This work was supported by NSF Grant CDR-8622236.

The authors are with the Optoelectronic Computing Systems Center, Department of Electrical and Computer Engineering, University of Colorado at Boulder, Boulder, CO 80309.

IEEE Log Number 9105974.

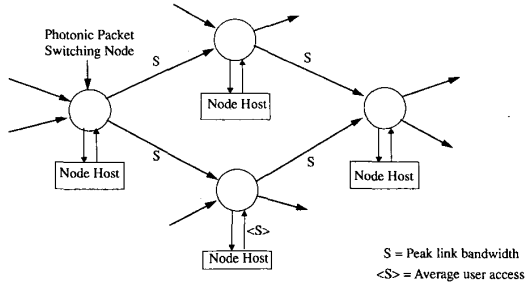

Fig. 1. A high-performance multistage interconnect for tightly coupled multiprocessors.

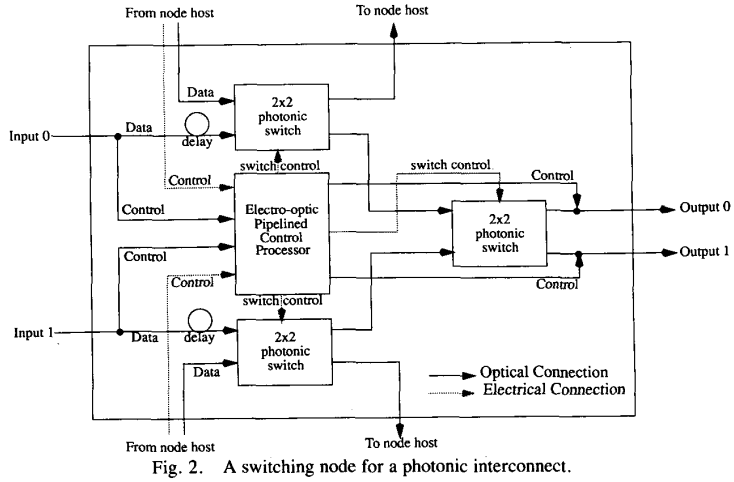

Fig. 2. A switching node for a photonic interconnect.

Two fundamental characteristics of a 2×2 switching node are the ability to route information and resolve contention for an output port. Conventionally, static electronic storage at the switch point is used to accommodate contention. However, in photonic flow-through architectures (i.e., no static memory at the switch point), memory is at a premium and bandwidth is an abundant resource. Self-routing control, without contention resolution, of 1×2 photonic switches has been demonstrated [6], [7]. Practical implementation of a full 2×2 photonic switch requires higher level processing than that demonstrated to date in order to resolve output port contention. This work demonstrates self-routing in a flow-through photonic switch with processing of a contention-resolving protocol on the fly. A primary choice for a flow-through contention-resolving protocol is deflection routing [8], [9]. Deflection routing provides a minimally complex procedure for handling conflicts in multi-path networks, and is amenable to flow-through photonic switch architectures.

II. Encoding Technique

The technique used to encode both routing information and data is bit per wavelength (BPW) encoding. A packet is transmitted in parallel wavelength channels in one time slot, at low duty cycle, such that the network is sparsely populated in the time domain as illustrated in Fig. 3. The number of wavelengths required to encode a packet is $O(\log_2 N) + M$

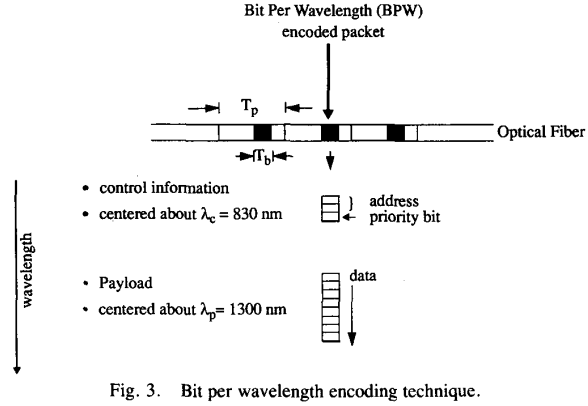

Fig. 3. Bit per wavelength encoding technique.

where N is the number of nodes and M is the number of data bits in a packet. The network operates synchronously with a packet period T_p and actual optical signals occupying a single bit width T_b. The $O(\log_2 N)$ control wavelengths $\{\lambda_c\}$ and M data wavelengths $\{\lambda_p\}$ are transmitted at different wavelength bands to simplify demultiplexing. The practical number of wavelength channels on a single fiber over large distances is $O(100)$ due to fiber nonlinearities [13]. For N nodes, wavelength division multiplexed shared medium protocols require $O(N)$ wavelengths, whereas BPW encoding only requires $O(\log_2 N)$ wavelengths, allowing

scalability to significantly larger networks. Due to the abundant bandwidth of optical fiber, this technique is efficient for transmission of single-word packets. Also with BPW encoding, parallel pipelined architectures can operate on all bits simultaneously with maximum processor throughput.

III. Node Architecture

A flow-through switching node is shown in Fig. 4. Individual packets are synchronously incident at both input ports. Data and routing information are coded in the optical frequency domain as described in Section II. The control information is separated from the data prior to the switch input, and is directed to a parallel pipelined control processor (CP) for switch state control and contention resolution. The fiber delays in the data path compensate for processor latency. Contention for an output port is handled by a deflection routing protocol. The CP is programmed to correctly route or deflect a packet based on routing priority. The penalty for deflecting packets in a multipath topology is low, and the algorithm provides for extremely simple and rapid contention resolution.

In this demonstration, a two-bit address field, corresponding to four addresses, is mapped onto one of the two output ports by the CP, which functions as a lookup table. The packet rate is limited by the slowest pipeline stage in the CP. The unused portion of the packet time slot T_p is used to increase the timing tolerance of the system as discussed in Section VI. The dispersion between control and data over long fiber lengths can be accommodated by the data path delays since the internode distances are fixed. Dispersion between data bits will limit node throughput as discussed in Section VI. The priority bit is used by the CP to arbitrate contention. A routing decision is made based on the simultaneous input of all address and priority bits from both incoming packets. Contention resolution is handled on the fly by including all possible contention configurations in the CP. If two packets desire to route to the same output port, the higher priority packet will be switched to the correct port while the other packet is deflected to the remaining port. In the case where priorities are equal, a fair algorithm must be used. We maintain the switch in its prior state under this condition. This method promotes fairness for statistically independent packets. Other measures of priority can include the time a packet has been in the network, number of times a packet has been misrouted, or the distance to the final destination node [10], [11].

IV. Experiment

The experimental setup is shown in Fig. 5. Four 4-bit packets are serially generated and converted to 4 parallel bits with a period of 400 ns, at 25% duty cycle (low speed for TTL implementation). The control bits are transmitted at wavelengths of 831.6, 828, and 824.4 nm. The data bit (representing many payload bits) is transmitted at 1300 nm and subsequently combined with the routing bits into 1300 nm single mode optical fiber. Two copies of each packet are created using the two fiber splitters and two fiber couplers. One copy is delayed by a single packet period in order to

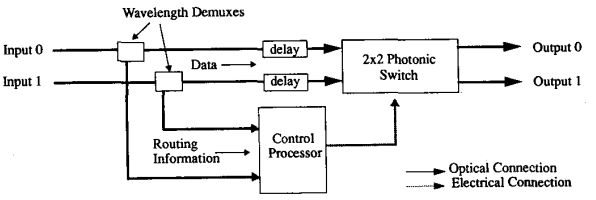

Fig. 4. 2 × 2 switch node architecture.

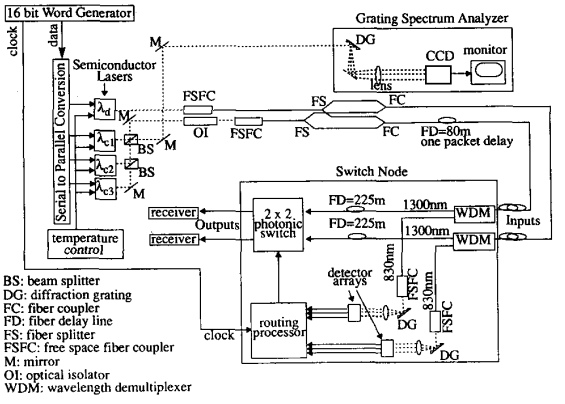

Fig. 5. Experimental setup.

present the switch with two different packets in any time slot. Packets 1 and 3 are transmitted with data bits "1" while 2 and 4 are transmitted with a data bit "0" to facilitate verification of deflection routing. The four packets [4/3/2/1] are output from a word generator as a serial bit stream [0010/1110/0101/1010]. The left most bit in each packet is the data, the second from left in each packet is the priority, and the right most two bits are the least significant and most significant address bits, respectively. Pairs of packets which enter the switch simultaneously are; 1 and 4, 2 and 1, 3 and 2, and 4 and 3. This CP interprets addresses "00" and "11" as desiring to go to output port 0 and "01" and "10" to port 1. Each packet requests to route to output port 1 and is routed according to the priority rule discussed in Section II.

At the switch node, data and control are optically demultiplexed by wavelength. The data bits are directed to the inputs of a 2 × 2 $LiNbO_3$ directional coupler switch. A fiber delay of 225 m is used at each input to compensate for the CP latency. The routing bits are demultiplexed using a diffraction grating and focused onto a pin detector array. These signals then address an EPROM whose output generates the switch state, synchronous to a global clock, according to a preprogrammed lookup table. All possible combinations of address and priority are accounted for in the CP. In order to verify correct contention resolution and routing, the data bits are detected by pigtailed InGaAsP receivers.

V. Results

The results are summarized in the logic analyzer output shown in Fig. 6. The detected sequences for each programmed contention state are indicated by arrows at the top. The associated resulting switch states are indicated by the

arrows at the bottom. Traces 00–02 show the control bits detected at switch input 0: [110, 010, 010, and 101]. Traces 05–07 show the control bits for switch input 1: [101, 110, 010, and 010]. Trace 09 shows the global clock signal. Trace 11 shows the switch state control signal with a high indicating a cross state and a low indicating a bar or straight through state. Traces 13 and 14 show outputs 0 and 1 of the photonic switch with the appropriately routed data bits from packets 1–4. Note the allowable timing error window of ± 150 ns, equivalent to 75% of a packet, due to the encoding technique. Contention states 1 and 2 demonstrate routing requests to the same output port with different priorities, while contention states 3 and 4 demonstrate routing requests to the same output port with the same priority.

The effective throughput for transmission of one data bit per packet is $T = 2.5$ Mb/s. Using the current electronics with laser arrays, 64 bit wide words can be transmitted, increasing the throughput to 160 Mb/s. The node latency $D = 400$ ns + (225 m)/(2×10^8 m/s) = 1.525 μs. Optical link latencies, limited by the speed of light, make this node appropriate for internode distances of > 1 km (assuming a 3:1 ratio between link and node latency is acceptable). These performance parameters are scalable using current technology, as discussed in Section VI. Note that the current implementation of one data bit has throughput equal to a time multiplexing system, however, as additional data bits per packet are added, the performance for BPW improves over that of serial multiplexing.

VI. Scaling Issues and Future Work

The primary scaling issues involve an increase in the number of nodes, separation between nodes, and aggregate capacity. The number of bits scale as $O(\log_2 N) + M$ where N is the number of users and M is the number of data bits in a packet. For networks on the order of one thousand nodes and computer words of 64 bits, approximately 85 channels are required (including additional communications overhead). This capacity can be met by either direct intensity modulation of individual lasers or subcarrier multiplexing of significantly fewer lasers [12]. An important limitation to transmission of many simultaneous channels is nonlinear crosstalk in the optical fiber. With 85 lasers spaced > 20 GHz apart, transmission of approximately 1mW per laser over distances greater than 10 km results in interchannel crosstalk of less than 0.5 dB [13]. The limitation on internode distance is a complex function of inter-data-bit dispersion, switching time, duty cycle and other factors.

In future experiments, the control will be encoded at 1300 nm and the data at 1550 nm. This allows the data to be amplified using erbium doped fiber amplifiers [14], while the control is regenerated at each node and reinserted on the network with the data as in Fig. 2. Additionally, the control processor can operate synchronously with a globally distributed clock [15], at the packet race, collecting all requests over the period T_p. This technique reduces sensitivity to thermal fiber variations since the clock travels with the data and control. The maximum phase error between the two inputs due to slowly varying temperature changes is $\pm 0.5\ T_p$

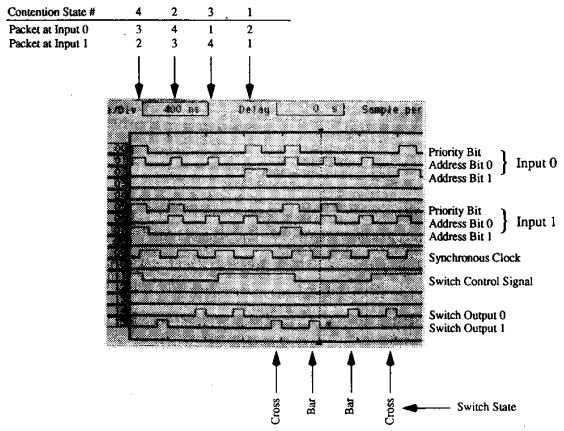

Fig. 6. Demonstration of self-routing with contention resolution.

and can be accommodated for at the CP by deflecting packets. Thermal variations affect processing on a node-to-node scale. The incoming control bits are averaged over the period.

$$T = T_p - 2\tau_s - \tau_{\text{jitt}} - \tau_{\text{synch}} \tag{1}$$

where τ_s is the switch rise or fall time, τ_{jitt} the electronic jitter, and τ_{synch} the synchronization error. The number of data bits will determine the throughput or latency of the switch. The source to destination distance is relevant for interdata-bit dispersion. For example, if we transmit 64 data bits, each 1 ns in duration, and contained within the erbium amplifier bandwidth of 25 nm over 64 km of dispersion shifted fiber (dispersion of 1.5 ps/nm · km [16]), the end bits will separate by approximately 2.25 ns (this is a conservative calculation). Assuming the parameters in (1) are $\tau_s = 1$ ns, $\tau_{\text{jitt}} = 0.5$ ns, and $\tau_{\text{synch}} = 1.5$ ns, the packet rate is $T_p = 6.25$ ns. The effective throughput is $T = (64 \text{ bits})/(6.25 \text{ ns}) \approx 10$ G/sec and the node latency $D \approx 8.25$ ns. This effective data rate can be increased by introducing a negative dispersion element at the switch to realign the data bits, also increasing the switch latency by approximately 1 ns.

The above operating times are reasonable given current technology. Operation of $LiNbO_3$ with 100 MHz circuits in latency sensitive applications has been reported [17] as well as operation of 50 MHz fiber optic delay line memories kept synchronous for periods of days [18]. Static RAM's can be used as a table lookup, with state of the art access times of 2 ns (Gigabit Logic GaAs SRAM) accommodating the above packet rate where SRAM's may be operated in parallel and cascaded to scale to larger networks. Additional logic is needed to perform updating of priority bits and reinsertion to the network, adding several nanoseconds to the latency, but not limiting the throughput of the pipelined CP.

VII. Discussion and Summary

We have demonstrated for the first time, electro-optic deflection routing in a 2×2 photonic switch. The minimal latency, flow-through optical protocol used is well suited for large scale tightly-coupled computer networks which transfer single words. Transmission of control and data in multiple

wavelengths improves tolerance to timing uncertainties by occupying a small portion of the packet period. Control processor throughput is increased due to bit per wavelength (BPW) encoding with parallel optical input and the feedforward, pipelined control processor architecture. The packet period scales with the bit width, not the number of bits per packet, and therefore so does the processing rate. The EPROM is the pipeline stage which limits the packet processing rate. State-of-the-art EPROM technologies currently process lookup times to within a few nanoseconds allowing a significant speedup in throughput and decrease in latency. This architecture can optically route computer wide words with the electronics required to run only at the effective packet rate. Integration of this switch into a fully operational node with three or four self-routing photonic switches is currently underway. Also being investigated are alternative architectures and implementations of the control processor.

References

[1] A. L. Decegama, *The Technology of Parallel Processing: Vol. 1.* Englewood Cliffs, NJ: Prentice Hall, 1989.

[2] N. R. Dono, P. E. Green, Jr., K. Liu, R. Ramaswami, and F. F.-K. Tong, "A wavelength division multiple access network for computer communication," *IEEE J. Select. Area. Commun.*, vol. 8, no. 6, pp. 983-993, 1990.

[3] M.-S. Chen, N. R. Dono, and R. Ramaswami, "A media-access protocol for packet-switched wavelength division multiaccess metropolitan area networks," *IEEE J. Select. Area. Commun.*, vol. 8, no. 6, pp. 1048-1057, 1990.

[4] M. G. Hluchyj and M. J. Karol, "ShuffleNet: An application of generalized perfect shuffles to multihop lightwave networks," *J. Lightwave Technol.*, vol. 9, no. 10, pp. 1386-1397, Oct. 1991.

[5] J. R. Sauer, "A multi Gbit/s optical interconnect," *Proc. SPIE*, Digital Optical Comp. II, Los Angeles, CA, conf. 1215, paper 22, 1990.

[6] P. R. Prucnal, D. J. Blumenthal, and P. A. Perrier, "A self-routing photonic switching demonstration with optical control," *Opt. Eng.*, vol. 26, pp. 473-477, 1987.

[7] A. Budman, E. Eichen, J. Schlafer, R. Olshansky, and F. McAleavey, "Multigigabit optical packet switch for self-routing networks with subcarrier addressing," *OFC '92*, San Jose, CA, Feb. 1992.

[8] P. Baran, "On distributed communications networks," *Trans. Commun. Syst.*, pp. 1-9, Mar. 1964.

[9] B. J. Smith, "A pipelined, shared resource MIMD computer," in *Proc. 1978 Int. Conf. Parallel Processing*, Bellaire, MI, 1978, pp. 6-8.

[10] A. Krishna and B. Hajek, "Performance of shuffle-like switching networks with deflection," *IEEE INFOCOM '90*, pp. 473-480, 1990.

[11] Z. Zhang and A. S. Acampora, "Performance analysis of multihop lightwave networks with hot potato routing and distance-age priorities," in *IEEE INFOCOM '91*, Miami, FL, Apr. 1991.

[12] R. Olshansky, V. A. Lanzisera, and P. M. Hill, "Subcarrier multiplexed lightwave systems for broad-band distribution," *J. Lightwave Technol.*, vol. 7, no. 9, pp. 1329-1342, 1989.

[13] A. R. Chraplyvy, "Limitations on lightwave communications imposed by optical-fiber nonlinearities," *J. Lightwave Technol.*, vol. 8, no. 10, pp. 1548-1557, 1990.

[14] *Fiber Laser Sources and Amplifiers, Proc. SPIE*, M. J. F. Digonnet Ed. Boston, MA, vol. 1171, Sept. 6-8, 1990.

[15] D. H. Hartman, P. J. Delfyett, and S. Z. Ahmed, "Optical clock distribution using a mode-locked semiconductor laser diode system," *OFC '91*, San Diego, CA, paper FC3, OSA, Washington, DC, 1991.

[16] S. E. Miller and I. P. Kaminow, Eds., *Optical Fiber Telecommunications II.* New York: Academic, 1988, p. 35.

[17] R. J. Feuerstein, T. Soukup, and V. P. Heuring, "100-MHz optical counter that uses directional coupler switches," *Opt. Lett.*, vol. 16, no. 20, p. 1599, 1991.

[18] T. Soukup, R. J. Feuerstein, and V. P. Heuring, "Implementation of a fiber-optic delay-line memory," *Appl. Opt.*, 1991.

A Novel Twin-Ridge-Waveguide Optical Amplifier Switch

J. Singh, I. D. Henning, P. S. Mudhar, M. A. Fisher, S. Perrin, D. A. H. Mace, and M. J. Adams

Abstract—**A novel twin ridge-waveguide optical amplifier switch is reported. We have applied the technique of hydrogen passivation of acceptors to decrease the current spreading in the region between the ridges. The incorporation of a passive waveguide below the active waveguide leads to the result that using 1.53 μm TE polarized light, through and cross states, each with only 1 dB insertion loss, fiber-to-fiber, can be selected in a device of 360 μm length by varying the currents to each ridge with a total current of 140 mA. A minimum crosstalk of less than −33 dB was achieved when the cross state was selected. This is the first report of a twin ridge-waveguide amplifier switch with such a low-loss, low-polarization sensitivity, and low crosstalk.**

Manuscript received October 29, 1991; revised December 4, 1991.

J. Singh was with British Telecom Laboratories, Ipswich IP5 7RE, UK. He is now with the Department of Electrical Engineering, University of Maryland, College Park, MD 20740.

I. D. Henning, P. S. Mudhar, M. A. Fisher, S. Perrin, D. A. H. Mace, and M. J. Adams are with British Telecom Laboratories, Martlesham Heath, Ipswich IP5 7RE, UK.

IEEE Log Number 9106093.

Introduction

SWITCHING of an optical signal for applications in optical communications systems can be achieved by the use of active directional couplers with gain (twin ridge-waveguide optical amplifiers) in order to overcome coupling and device losses [1], [2]. Fig. 1 shows a schematic cross-section of a twin ridge-waveguide optical amplifier. Switching in such a device is achieved by varying the currents in the two guides. In a typical twin ridge-waveguide amplifier structure without a passive guide below the active layer [1], [2], the