# Optoelectronic neural-network scheduler for packet switches

Roderick P. Webb, Andrew J. Waddie, Keith J. Symington, Mohammed R. Taghizadeh, and John F. Snowdon

> A novel, to our knowledge, type of packet scheduler that could significantly outperform current state-ofthe-art schedulers is presented. The operation and the design of such a scheduler are discussed, and a fully operational experimental implementation is described. The scheduler uses a neural network in a winner-take-all strategy to optimize decisions on the throughput of both a crossbar and a banyan switching fabric. The problems of high interconnection density are solved by use of a free-space optical interconnect that exploits diffractive optical techniques to generate the required interconnection patterns and weights. © 2000 Optical Society of America

OCIS codes: 200.4700, 200.4650, 050.1970.

#### 1. Introduction

Demands on high-performance networks are increasing rapidly with no slow down yet forseeable. Internet traffic alone has been estimated to grow at more than 30% per month. To operate these networks at their raw-bandwidth capability requires a new generation of switches-routers that are capable of scheduling the traffic while introducing a minimum of latency.

At the core of some of the latest generation of internet routers is a hardware switch that is used to interconnect the line cards. A central scheduler is required for selecting a set of packets from the input queues that can be routed simultaneously to the correct outputs without blocking. The larger the set chosen, the greater the throughput, but the decision must be made within the cycle time of the switch. This assignment of outputs to inputs that is subject to constraints imposed by the switching fabric is an example of a resource-allocation problem and can be solved by a simple neural network. We implemented such a network as a parallel optical system that incorporates a diffractive optical element (DOE) and measured its performance as a scheduler for both crossbar and self-routing switching fabrics.

This paper first discusses the general aspects of the use of optoelectronic neural networks to perform the above kind of optimization problem and then focuses on particular types of network. The actual experimental system is then described in detail, and the results presented. The paper concludes with a section on performance comparisons and considers the future possibilities of such a scheduler.

# 2. Resource Allocation and Optoelectronic Neural Nets

The preferred approach to network-communication problems at present is to use arbitrary-topology cellbased networks such as ATM. These networks have many advantages over bus- or ring-based systems but require high-performance switches to be effective.<sup>1</sup> Optimizing the throughput for a switch is a computationally nontrivial problem for which many alternative strategies exist. In general, it is necessary to provide a fast switching fabric with queuing available at either the inputs or the outputs or both, along with a control structure for matching the inputs to the respective outputs. There are evidently complex trade-offs in terms of performance and hardware costs for all options, particularly in view of the datadependent nature of the problem.

In the implementation described here, we consider both a crossbar and a multistage self-routing switching fabric with random-access input queuing. The novelty in our approach is the use of an optoelectronic neural network<sup>2,3</sup> to perform the input–output matching. The use of neural-network hardware can

R. P. Webb is with British Telecom Laboratories, Martlesham Heath, Ipswich IP5 3RE, UK. A. J. Waddie, K. J. Symington, M. R. Taghizadeh, and J. F. Snowdon (j.f.snowdon@hw.ac.uk) are with Heriot-Watt University, Edinburgh EH14 4AS, Scotland.

Received 14 May 1999; revised manuscript received 24 September 1999.

 $<sup>0003\</sup>text{-}6935/00/050788\text{-}08\$15.00/0$ 

<sup>© 2000</sup> Optical Society of America



Fig. 1. Schematic of the experimental neural-network crossbarswitch controller. On the basis of the connections requested by the incoming packets the neural network chooses an optimal solution, sets the appropriate crossbar switches (cross points), and then selects the packets chosen to pass through.

yield excellent performance on resource-allocation and optimization problems<sup>4–9</sup> at low cost, as it exploits analog circuit capabilities and, more importantly, creates a naturally highly parallel approach to the problem. Such a neural network is, however, intractable to build to any scalable extent in silicon because of the high degree of connectivity required.

Our optical scheme *enables* the deployment of neural-network technology. Because the goal of this research is to supply high connectivity, we use a freespace optical system<sup>3</sup> in which a set of emitters is connected through a diffractive fan-out element to a set of detectors such that each detector integrates over the weighted outputs of several emitters. The system thus exploits several of the degrees of freedom made available to computation by free-space optics: the raw bandwidth of a massively parallel interconnection, the nonlocality that can be achieved in such an interconnection, and the capacity for large fan-out and fan-in. The system we describe can be reconfigured to solve other nonpolynomial problems such as route relocation and path determination.<sup>4,5,10</sup>

# 3. Winner-Take-All Approach

In our crossbar switch the neurons are arranged in a two-dimensional array that represents all possible input-to-output connections such that each neuron corresponds directly to a cross point on the switch (Fig. 1). The neuron outputs can vary continuously between the OFF and the ON levels. To choose a set of connections requires that the neurons representing all the requested connections be enabled simultaneously and set to the same intermediate level. Each neuron has a bias input that tends to increase its output but also receives inhibitory input from



Fig. 2. Schematic of a neural-network controller for a self-routing multistage banyan switch. Here the neural network selects an optimal packet solution and notifies the input buffers: It does not directly control the switching network.

those neurons that represent blocking connections. Crossbar switches can be blocked at their inputs and outputs only, so the neurons are arranged to be inhibited by others in the same row or column. All other possible connections are set to zero.

The dynamics of the network resolve the conflicts between all the mutually excluded neuron pairs, leaving a valid set of neurons in the ON state and the remainder in the OFF state. The network thus behaves as a winner-take-all (WTA) system with a particularly simple interconnect pattern—each neuron sees only its row and column neighbors, each of which is connected to it by a fixed, inhibitory weight. It is apparent that such a pattern is space invariant and therefore highly suitable for implementation in a diffractive optical system. Before we consider our experimental setup in more detail it is worth considering how other types of switch might be controlled in this way (particularly with regard to exploiting shift-invariant interconnection patterns).

Consider the scheduler shown in Fig. 2. In this case the incoming packets have a header address that determines their paths through the banyan network. This feature has the advantage that the scheduler does not need to set the switches explicitly but merely to select packets to be launched into the fabric. The penalty is, of course, that the banyan-type switch shown here is internally blocking, and a more complex task must be performed by the scheduler. The neural-network scheduler copes well with this added complexity in that the correct functionality can be attained by the mere provision of additional inhibitory paths that provide contention between requests for these blocking configurations. How this inhibition can be achieved can be seen by the consideration of the grid of neurons with the left-hand vertical edge corresponding to the input port of a packet and the



Fig. 3. (a) Electronic neuron block diagram: Each neuron can be broken down into a series of electronic components. This diagram indicates how a neuron is represented in a modular form. (b) Neural sigmoid response: The slope of a neuron's activation function is highly dependent on  $\beta$ . This diagram shows how the slope changes by use of three sample values.

lower horizontal edge corresponding to an output port. Because the destination and the input positions are known for any packet, the internal blocking connections are known and can be programmed explicitly in advance.<sup>11</sup> The resultant pattern will perform a WTA optimization and allow the switch to operate.

A novel feature of this study is making these resultant patterns amenable to optics. What is most important from the perspective of a diffractive optical implementation is that the final pattern of optics that is generated be space invariant. The interpretation of one of the edges of the neuron grid in bit-reversed addressing order makes the inhibitory interconnection pattern once more space invariant and suitable. These patterns are described in the context of the optical system layout described in Section 4 and in Ref. 12.

#### 4. Experimental Implementation

The goal of our research was to implement a hardware neural network that was capable of optimally selecting which data packets should be scheduled to pass through a switching fabric. The headers (specifying the desired output port) of queued data packets were examined and used to enable a neuron that corresponds to this connection. The enabled neurons evolve according to a WTA strategy that results in an allowed pattern of neurons being switched on and the remainder OFF. This pattern is then used to select a set of winning data packets and to allow them to pass through the switch fabric. The pattern is also used to close the appropriate switch points in the switching fabric to select the correct path. The communication pathway is illustrated by the long medium-broad arrows shown in Figs. 1 and 2. The 64 circles in the array each correspond to a neuron in the neural network.

The structure and the response of each of the neurons is shown in Fig. 3. In this implementation each of the 48 neurons has an input detector that is followed by a capacitor-coupled inverting amplifier chain and a low-pass filter; the output drives a vertical-cavity surface-emitting laser (VCSEL). (Our VCSEL array was  $8 \times 8$ , but only 48 neurons could be used. Thus our switch control was a  $6 \times 8$  system.) Figure 3(a) shows the modular electronic layout used to generate the required neuron characteristics. Essentially, this arrangement generates the dynamics and the outputs described by<sup>5</sup>

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = I_i \left( -\lambda_i x_i - \sum_{j=0}^{N-1} w_{ij} y_j + t_i \right), \tag{1}$$

$$y(x_i) = o_{\min} + \frac{o_{\max} - o_{\min}}{1 + \exp(\beta x_i)},$$
(2)

respectively, where x is the state of a particular neuron in time, y is its output, and t is the threshold or the bias for the neuron. Equations (1) and (2) evolve such that x, the state of the neuron, follows a negative exponential with the time constant  $1/\lambda$  to reach a steady state with a well-determined value of y. The value of  $\lambda$  is nominally the same for all neurons. The weights w represent the optical connections through the DOE. These weights are also nominally equal and are determined by the attenuation of the optics. In this case it is desirable to maximize the weights.



Fig. 4. VCSEL's that are used are highly sensitive to temperature variations. The VCSEL responses plotted for both high and low ambient temperatures illustrate just how drastic this variance is.

The time constant and t are set by the resistive and the capacitive elements, respectively, within the electronics. The time constant in this implementation is determined largely by the filter circuits shown in Fig. 3(a). No attempt was made to optimize for speed here, but in general one would wish to minimize this time constant.

The threshold *t* determines where on the response curve [Fig. 3(b)] the neuron evolution starts. This parameter is also nominally the same for all the neurons and is used to set the initial state of the network in such a way as to optimize convergence. Different values of *t* would be selected for different attenuations *w* and different values of  $\beta$ . The variable *I* takes the values 0 or 1 and determines whether a given neuron will be allowed to evolve—this process corresponds to a request for a particular connection.  $\beta$  determines the slope of the activation function [Fig. 3(b)] and, in practice, is essentially determined by the gain of the amplifier in the neuron. The terms  $o_{\min}$  and  $o_{\max}$  merely refer to the minimum and the maximum outputs, respectively, of the neuron, which in this case [Fig. 3(b)] can be seen to be 0 and 1, respectively.

Figure 4 shows the relevant characteristics of the VCSEL's. The VCSEL's have typical response times of 1 to 10 ns and exhibit considerable temperature variation. This variation in temperature, both in absolute terms and especially in distribution across the array, generated a source of noise within the system. The consequences of such noise are discussed in Section 5.

The detector array is a commercial photodiode array operated at peak sensitivity with a typical response time of approximately 30 ns. Figure 5 shows the system schematically [two lenses are needed because the DOE's operate in the Fourier plane (see below)]. The following describes a typical operational cycle of the neural-switch scheduler.

Initially all the lasers are set to a fixed output level that is slightly higher than the OFF level. This level sets a stable total power for the array and effectively biases the neurons toward the ON state. When the network is enabled, the lasers of all the requested neurons are connected to their amplifier outputs, and the others are set to the OFF level. Between the laser and the detector arrays are a pair of lenses and a DOE that divides the light from one neuron's laser and focuses it onto the input detectors of the other neurons in the same row and column (for a crossbar) or other required pattern but not to its own input. Because of the inversion in the amplifier chain, light falling on a detector inhibits the neuron, decreasing its output. The WTA nature of the setup guarantees a convergence to a



Fig. 5. Schematic of the experimental optical system setup for the crossbar-switch controller. This diagram shows how the output from a single VCSEL is diffracted by the DOE and imaged onto the detector array.

Table 1. Nonuniformity and Efficiency of Crossbar and Self-Routing DOE's

	Initial Design		Subsequent Design	
DOE	η	$\Delta R$	η	$\Delta R$
Crossbar Self-routing	60 60	4.7 2.9	50.0 $49.9$	0.81 0.83

feasible solution with those neurons that remain ON determining the allowed routes.

A critical component of this system in terms of functionality is the DOE. The inhibitory interconnections between the neurons were implemented by use of far-field scalar DOE's in conjunction with a Fourier lens system. These phase-only diffractive elements were designed by use of a standard iterative Fourier transform algorithm<sup>13</sup> that was followed by a closed-form iterative technique,<sup>14</sup> which was required to produce the uniformity and the signal-to-noise ratio required of the inhibitory interconnections. The initial optimizations had a goal efficiency of 60% for a binary design, but the theoretical nonuniformity  $\Delta R$  of these designs was significantly higher than the

goal nonuniformity of less than 2% after fabrication. Reducing the goal efficiency from 60% to 50% led to the theoretical nonuniformity of each DOE being substantially reduced, as is shown in Table 1.

The DOE designs are e-beam written on chrome masks that are transferred to fused-silica substrates by use of standard photolithographic techniques. The minimum feature sizes that the in-house facilities at Heriot-Watt University can copy accurately and etch are of the order of  $1-2 \ \mu m$ . The optical system used in the neural-network demonstrator required the DOE's to have a period of 96  $\ \mu m$  with minimum features of  $\sim 3 \ \mu m$ .

Figure 6, which contains images of both the e-beam mask and the etched DOE, shows the rounding of the sharp edges that is to be expected from feature sizes that lie so close to the minimum resolvable feature size. The etching of the DOE onto a fused-silica substrate was performed with  $CHF_3$  as the reactive ion for the etching process. This process produces sharp vertical etches without undercutting and has an accuracy in etch depth, and correspondingly in phase, of 0.5%-1% over a 25-mm substrate.

A binary design was used for each DOE to minimize the unavoidable fabrication errors caused by



Fig. 6. (a) Image of an etched DOE phase profile created from (b) the theoretical phase profile.



Fig. 7. Both the crossbar and the self-routing switches require different neural-network inhibitory patterns: (a) crossbar switch and (b) self-routing switch.



Fig. 8. Actual time response of two neurons: One switches itself OFF, and the other switches itself ON.

etch-depth inaccuracies and mask misalignment. These errors were estimated<sup>15</sup> at approximately 1%-2% nonuniformity per mask level, e.g., for a 16-phase-level design the fabrication errors will result in a nonuniformity of approximately 4% for a similar DOE.

The output from the crossbar DOE is shown in Fig. 7. Small zero orders may occur at the centers of the patterns as a result of the fabrication inaccuracies observed from Fig. 6. These orders are not critical in this type of network.

The scheduler was first tested with the crossbar DOE in position. Six sets of requested neurons that represented different combinations of waiting packets were prepared. For each test, the appropriate neurons were enabled, and Fig. 8 shows typical waveforms that were observed at the outputs. Initially, the outputs were at approximately half the maximum power, so most of the detectors were receiving large amounts of light from conflicting neurons. The outputs began to turn off at a rate governed by the low-pass filters. When their inputs were low enough some neurons were able to turn on again, increasing the inhibition of those in conflict with them. After approximately 200 µs all the neurons were clearly either in the ON state or the OFF state. The validity of the output states was checked, and the number of neurons in the ON state, which represents the number of packets routed through the switch, was counted.

The results for the request sets in which the maximum allowable number of neurons in the ON state is six are shown in the form of a histogram in Fig. 9. The experiment was repeated with the self-routing DOE and the same request sets. In some cases the maximum allowable number of neurons in the ON state was reduced by the extra blocking mechanisms in the switch, but those results for which the maximum was still six are shown in Fig. 10. For constructing the histograms, a particular data set was presented 10,000 times to form a trial, and the number of times a particular number of packets was routed out was obtained. (Initially, trials were performed on noncontentious request sets, and these



Fig. 9. Crossbar-switch network convergence: Random-trial request sequences 1 through 6 were input into the neural network 10,000 times, and the number of neurons in the ON state determined. A result with six neurons in the ON state indicates an optimal solution; fewer neurons indicate that the network has failed to converge optimally.

were found to work perfectly.) The six request sets were then produced (randomly), and each was presented 10,000 times to form the six trials shown in Fig. 9. The scheduler never produced an invalid result. Most times it found an optimal result, except with the request sets of trial 3 and trial 4. With these trials it usually routed one fewer packet. Thus the switch would have near-maximum throughput and never block. The self-routing scheduler achieved a better result on trial 4.

No attempt was made to make this demonstration system run rapidly. The decision time was 6 times the time constant of the low-pass filters  $(33 \ \mu s)$ . Reducing this time constant should make it possible to obtain results in tens of nanaoseconds, while still using only off-the-shelf electronics, and to achieve scheduling decisions at a rate compatible with the latest router requirements.



Fig. 10. Self-routing-switch network convergence: Randomtrial request sequences 2 through 4 (as used in the crossbar switch) were input into the network 10,000 times, and the output was observed. Again, six neurons in the ON state indicate an optimal solution.



Fig. 11. Comparison of the neural-network controller with a state-of-the-art scheduler, ISLIP4. The advantage of the neural-network controller is clearly indicated at high levels of the offered load. The output-queuing curve indicates a theoretical optimum value.

## 5. Simulations and Performance Comparisons

Simulations of a  $16 \times 16$  neural-network scheduler were undertaken to make performance comparisons with other scheduler designs possible. The simulations were performed under uniform traffic conditions, and the mean delay (measured in packet periods) was plotted versus the offered load (the probability of a packet arriving at each input). Figure 11 summarizes the results of this exercise. The uppermost curve shows the situation in which the inputs are simply buffered in a first-in-first-out (FIFO) fashion. FIFO queues suffer from the problem of headof-line blocking in that, if the foremost packet in the queue (the next one to go) is blocked by another request, it also blocks all the packets behind it in the queue even if their destinations are not contentious. As might be expected, a scheduler based on FIFO buffering suffers severe performance degradation under an increasing load. The lowest curve (output queuing) represents the theoretical best that can be achieved. This best result is described as output queuing and is calculated under the assumption of an ideal switching fabric (impossible) in which packets have only to wait for a vacant slot on the output line. The middle solid curve represents an algorithm called ISLIP4 (Ref. 16) that could be implemented in complementary metal-oxide semiconductor electronics for a high-speed switch of this size. The curve with the filled circles shows the neural-network scheduler's performance and its favorable throughput at loads as high as and higher than 70%.

It is important to note that the ISLIP4 algorithm could not be implemented at switch sizes much larger than those considered because of hardware constraints, whereas the neural network is highly scalable because of the simple nature of the hardware. In a practical version of the scheduler, we would not use discrete electronics but would employ a smartpixel array.<sup>17,18</sup> Such an array, correctly deployed, would transfer all the communication complexity to the optical domain.

No matter how large the switch becomes there is no increase in the amount of computation that must be done locally, so convergence time changes essentially only in terms of the propagation delay of the optical signals (which is small) and thus will be similar for any desired size of switch. In practice, there will be physical limits to the scheduler size. Increasing the size of the neural network requires an increase in DOE fan-out: Increasing DOE fan-out decreases the signal incident upon a detector with a consequent decrease in the signal-to-noise ratio. A high fan-out will thus bring the signal-to-noise ratio so low that no signals can be detected. DOE efficiency is therefore a bound on the achievable switch size.

A possible way of coping with high fan-out is to consider using higher-power VCSEL's—heat dissipation in the VCSEL array will then bound the switch size. The system is also considered to be scalable with respect to its high tolerance to noise. Indeed, the system requires noise to operate. In any descent-type approach to an optimization problem (e.g., a gradient descent in energy space) there is always a risk that the system will become trapped in a local minimum.<sup>4</sup> In this system such a local minimum may be thought of as a solution that satisfies the switching constraints (i.e., it is a valid solution) but is not a globally optimal solution.

Simulations indicate that the successful convergence to an optimum may be greatly improved by the addition of noise to the system. The same simulations indicate that the system becomes slow to make decisions if there is less than 10-pW rms noiseequivalent power and will begin to make errors at noise levels higher than 100 nW. In the experimental system there is sufficient noise present to assist convergence to the global minimum, as can be seen from the results shown in Figs. 9 and 10.

# 6. Discussion

The increased interconnection density, bandwidth, nonlocality, and fan-out-fan-in offered by optics over conventional electronic technologies makes it a very attractive medium in which to implement communication harnesses for all types of computing engines. This is especially true for neural networks in which the demand for communication resources is extremely high. In this paper, we have described the successful implementation of a neural network that exploits an optical interconnect to perform a *real* task.

Although in this implementation speed was not a goal, impressive performance in terms of convergence and noise tolerance was observed, implying that scalability is good, so large switch sizes could be scheduled at little speed cost. It is also the case that, in terms of speed, the limit in our current implementation is the high time constant of the low-pass filters mainly derived from reasons of cost. We estimate that the network could make decisions in tens of nanoseconds, while still using discrete components. A large switch operating at these speeds would far outperform conventional digital rivals. In addition, it will be possible to push the speeds up still further by the removal of the communication delays in the system with a smart-pixel implementation of the electronics. The use of this technology allows us to envision a miniaturizable scheduler operating on the principles that have been demonstrated herein.

The elimination of the need for special-purpose high-speed electronic hardware combined with the decrease in cost of the optical components as their communications and computing use increases implies that this system might be a cost-effective solution to the ever increasing demands on the world's communications hardware.

## References

- T. Anderson, S. Owicki, J. Saxe, and C. Thacker, "High speed switch scheduling systems for local area networks," ACM Trans. Comput. Syst. 11, 319–352 (1993).
- R. P. Webb and A. W. O'Neill, "Optoelectronic neural networks," Br. Telecom Technol. J. 10, 144–154 (1992).
- R. P. Webb, "Optoelectronic implementation of neural networks," Int. J. Neural Syst. 4, 435–444 (1993).
- J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," Biolog. Cyber. 52, 141–152 (1985).
- T. X. Brown, "Neural networks for switching," IEEE Commun. Mag. 5(11), 72–81 (1989).
- P. W. Protzel, D. L. Palumbo, and M. K. Arras, "Performance and fault tolerance of neural networks for optimization," IEEE Trans. Neural Net. 4, 600–614 (1993).
- J. Ghosh, A. Hukkoo, and A. Varma, "Neural networks for fast arbitration and switching noise reduction in large crossbars," IEEE Trans. Circuits Syst. 38, 895–904 (1991).
- A. Marrakchi and T. Troudet, "A neural net arbitrator for large crossbar packet switches," IEEE Trans. Circuits Syst. 36, 1039–1041 (1989).
- 9. S. B. Aiyer, M. Niranjan, and F. Fallside, "A theoretical inves-

tigation into the performance of the Hopfield model," IEEE Trans. Neural Net. 1, 204–215 (1990).

- R. D. Brandt, Y. Wang, A. J. Laub, and S. K. Mitra, "Alternative networks for solving the travelling salesman problem," in *Proceedings of the IEEE International Conference on Neural Networks* (Institute of Electrical and Electronics Engineers, New York, 1998), pp. 673–681.
- T. X. Brown and K. H. Liu, "Neural network design of a banyan network controller," IEEE J. Select. Areas Commun. 8, 1428– 1438 (1990).
- R. P. Webb, A. J. Waddie, K. J. Symington, M. R. Taghizadeh, and J. F. Snowdon, "A neural network scheduler for packet switches," in *Digest of the Topical Meeting on Optics in Computing* (Optical Society of America, Washington, D.C., 1999), pp. 422–426.
- F. Wyrowski, "Iterative quantization of digital amplitude holograms," Appl. Opt. 28, 3864–3870 (1989).
- A. Vasara, M. R. Taghizadeh, J. Turunen, J. Westerholm, E. Noponen, H. Ichikawa, J. M. Miller, T. Jaakkola, and S. Kuisma, "Binary surface-relief gratings for array illumination in digital optics," Appl. Opt. **31**, 3320–3336 (1992).
- P. Blair, "Diffractive optical elements: design and fabrication issues," Ph.D. dissertation (Department of Physics, Heriot-Watt University, 1995).
- N. McKeown, M. Izzard, A. Mekkittikul, W. Ellersick, and M. Horowitz, "The tiny tera: a packet switch core," IEEE Micro. 17, 26–33 (1997).
- J. Gourlay, T. Yang, J. A. B. Dines, J. F. Snowdon, and A. C. Walker, "Development of free-space digital optics in computing," Computer **31**, 38-44 (1998).
- M. P. Y. Desmulliez, B. S. Wherrett, A. J. Waddie, J. F. Snowdon, and J. A. B. Dines, "Performance analysis of self-electrooptic effect device-based (SEED-based) smart-pixel arrays used in data sorting," Appl. Opt. 35, 6397-6416 (1996).