

# OPTOELECTRONIC NEURAL NETWORKS

K. J. Symington<sup>1</sup>, J. F. Snowdon<sup>1</sup>, A. J. Waddie<sup>1</sup>, T. Yasue<sup>1</sup> and M. R. Taghizadeh<sup>1</sup>

## Abstract

A general purpose neural network demonstrator is presented along with its application specific predecessor which employs a winner take all strategy to optimise decisions on the throughput of both a crossbar and a banyan packet switching fabric. The problems of high interconnection density in neural networks are solved by using free space optical interconnects which exploit diffractive optical techniques to generate the required interconnection patterns. The design, construction and operation of the general purpose network is discussed along with the fully operational experimental application as a packet switch scheduler which could significantly outperform current state of the art schedulers.

## 1 Introduction

A neural network is intractable to build to any scalable extent in silicon because of the high degree of connectivity required. The object of using optical interconnection is to supply very high connectivity using a free space optical system [1] in which a set of emitters are connected through a diffractive optic fan-out element to a set of detectors. The very architecture of this system also tackles the problem of weight summation by executing it in an analogue manner. Optical interconnects, of an appropriate intensity, converge onto a detector associated with each neuron, the output of which is inherently proportional to the sum of all incident light. Thereafter, all that needs be performed in electronics is calculation of the activation function. Our optical scheme enables the deployment of neural network technology in a scaleable manner.

This paper will discuss the optical neural network demonstrator currently under construction and conclude by giving a specific example of its complete and working predecessor.

## 2 A General Purpose Optoelectronic Neural Network Demonstrator

A neural network consists of a set of neurons, interconnected in an application specific manner, which perform some transfer function on the summation of a set of incoming weights. This demonstrator uses a fixed set of weights defined using a diffractive optic element (DOE) [2] to perform interconnection and summation with a digital signal processor (DSP) calculating the transfer function. The system currently under construction consists of 64 neurons in an 8×8 array. A DSP solution was adopted to provide flexibility.

### 2.1 The Electronic System

The electronic system can be considered to consist of five stages, each performing a specific task (figure 1). At the optical input end there is the detection system which converts a current generated by incident light into a voltage of magnitude specified by the amplification of the transimpedance amplifier. The second stage is an analogue-to-digital converter (ADC) which converts the voltage received from the first stage into digital information (normally 8 bits) and multiplexes 16 analogue channels through two octal ADC chips. The third stage consists of a Texas Instruments DSP which takes the digital information from the second stage and performs a transfer function on it based on previous and requested values. There are four DSPs in this system each handling 16 neurons (or channels) with each DSP under the control of a master DSP. The fourth stage consists of two octal digital to analogue converters (DACs) which are fed the new activation levels from the third stage

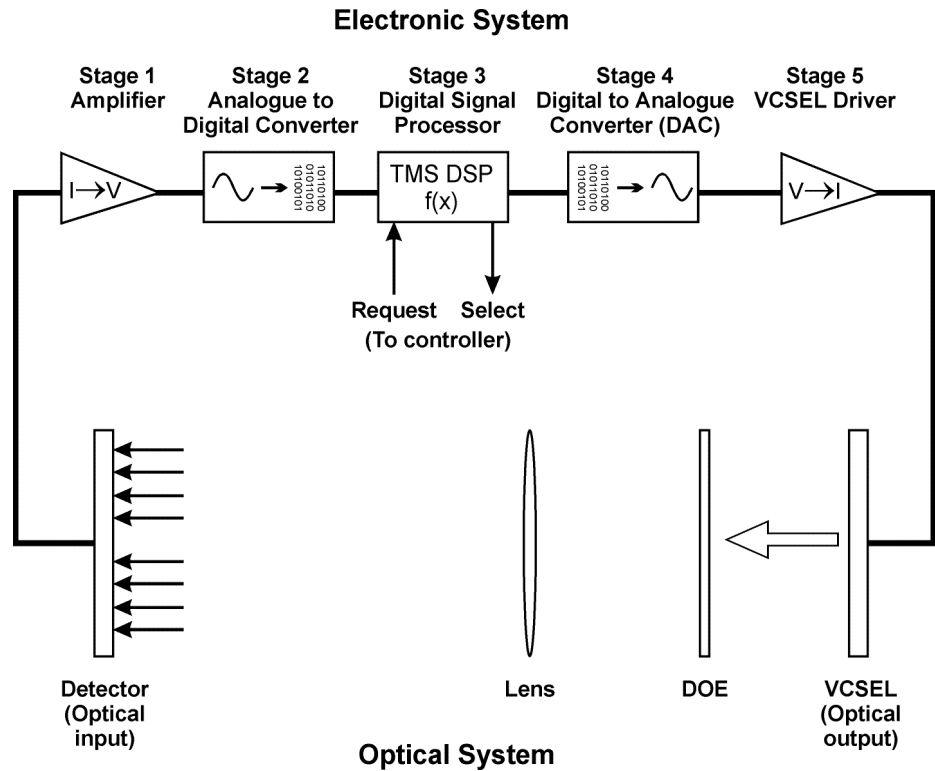
---

<sup>1</sup> Department of Physics, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS.

and convert this information into voltages. The fifth and final stage takes the analogue values from the fourth stage and converts the voltage into an appropriate drive current for the vertical cavity surface emitting lasers (VCSEL) thus returning the signal into the optical domain. A single chip solution for stage five is currently being fabricated.

The use of off-the-shelf digital signal processors to provide the neural thresholding functions allows the functionality of the network to be altered and new applications to be tackled with minimal alteration to the

optoelectronic hardware: essentially any thresholding function that can be written in C can be implemented. The DSPs are operated in a time-multiplexed manner enabling each of the four DSPs to handle 16 input and 16 output channels.



**Figure 1:** Electronic and optical system components in one diagram. There are 64 parallel electronic channels in the electronic system whereas only one optical system.

## 2.2 The Optical System

The optical components have been mounted on an optomechanically designed baseplate. This packaging scheme facilitates focus adjustments and mechanical and thermal stability for the VCSELs whose total power consumption is estimated to be 10W. The entire optomechanical package holds all the optical components within a space of approximately 12cm×15cm×20cm.

The role of the DOE is to provide fixed and evenly weighted interconnection between the neurons in the system. Increasing the size of the neural network requires an increase in DOE fan-out: increasing DOE fan-out decreases the signal incident on a detector with a consequent decrease in the signal to noise ratio. Achievable network size is therefore tightly bound to DOE fan-out.

The pattern of neurons inhibited by a given active neuron is shift invariant. That is, it remains the same relative to the position of the active neuron. An electrical system would require a separate wiring network for each output. The system has a high tolerance with respect to noise - indeed the system requires noise in order to operate.

## 2.3 Extending the System

There are a number methods which can be used to extend system functionality. Firstly, as summation is performed in the optical domain, it is not normally possible to add additional weight to a single input before summation. This problem can be overcome by time-multiplexing all neurons which output a similar weight and multiplying this weight by the appropriate value. Partial summation is performed by the optical system but this also results in an extra addition and multiplication instruction per neuron in electronics for every time slice used. Alternatively, one other solution is to use a weighted DOE pattern, but this presumes that the weightings are already known and will never change.

Secondly, by using either a spatial light modulator (SLM) or computer generated hologram (CGH), both of which can be controlled by a DSP and reconfigured according to the problem, neural interconnection can be altered as desired.

The final enhancement is that of multi-layer networks. This is actually very simple to implement by using pipelining in the DSPs. Thus for a three layer network there would have to be three iterations in the optical system. This method is feasible as long as neural interconnection between all the layers is the same.

The key to utilising the parallelism of a neural network is matching the network topology to the problem as closely as possible. The choice of DOE element and flexibility given by the programmability of the DSPs allows us to achieve this for a number of interesting problems, e.g. Travelling Salesman (and related optimisation), feature extraction and process control [3-5]. We have mapped examples of these problems and validated the mappings using an accurate simulator of the physical network. Experimental results will be used to verify the theoretical predictions made by simulation.

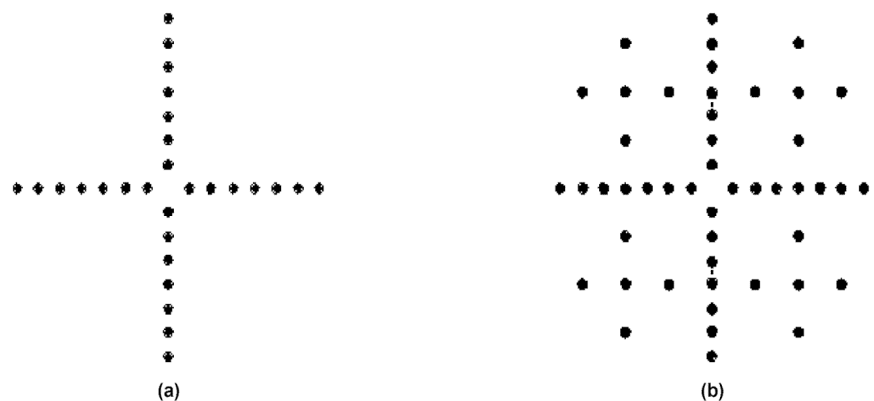
### 3 Neural Network for Packet Switching

This section considers specifically the assignment problem in a crossbar switch for packet routing [6, 7]. These switches are present in many telecommunication systems and computer networks, one good example being ATM (Asynchronous Transfer Mode) networks. The problem of packet routing in crossbar switches is known to be analogous to the travelling salesman problem (TSP) [3, 8-11]. The TSP problem is a renowned NP complete problem which means that although it can be solved by linear programming techniques, it is computationally intensive and its complexity grows exponentially as its order increases. Thus, a simple single processor solution will not provide satisfactory scalability.

One alternative is to apply a neural network to the TSP problem. The advantage of a neural net lies in the speed obtained through its inherent parallel operation, especially when dealing with large problems. Such an implementation will easily outperform any other method at higher orders of network size, providing a very good, but not always optimal, solution. It has been shown that, at lower orders of network size, the average solution is within 3% of optimal. However, as the network size grows this figure improves slowly and begins to approach the optimal solution.

In this implementation we consider both a crossbar and a multistage self-routing switch fabric with random access input queuing. A two-dimensional array of neurons represents all possible input to output connections. In the case of a crossbar, the neurons correspond directly to the crosspoints of the switch. The neuron outputs can vary continuously between the off and on levels. In order to choose a set of connections, the neurons representing all the requested connections are enabled simultaneously and set to the same intermediate level. Each has a bias input that tends to increase its output, but also receives inhibitory inputs from those neurons which represent blocking connections.

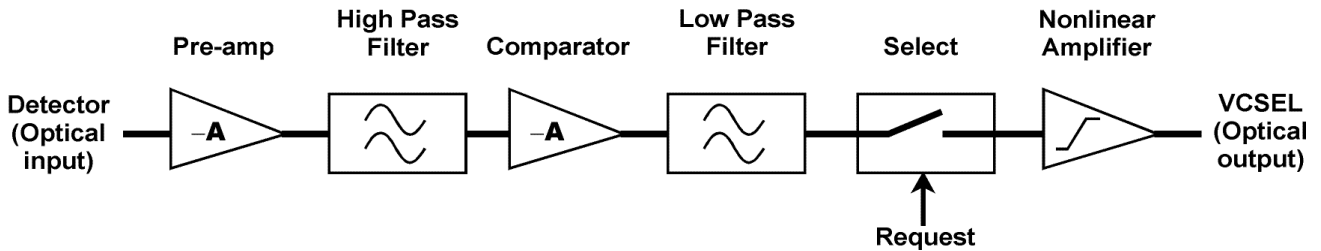
Crossbar switches can be blocked at their inputs and outputs only, so the neurons are arranged to be inhibited by others in the same row or column. Other types of switch can be blocked internally, for these switches additional inhibitory connections are provided between the appropriate pairs of neurons. The dynamics of the network resolve the conflicts between all the mutually exclusive neuron



**Figure 2:** Inhibitory DOE interconnection patterns. (a) illustrates that for a crossbar switch, (b) for a banyan network. All spots indicate an inhibitory interconnect.

pairs, leaving a valid set of neurons in the on state and the remainder off. The network is thus behaving as a winner take all (WTA) system with a particularly simple interconnect pattern – each neuron sees only its row and column neighbours, each of which are connected to it by a fixed inhibitory weight. Sample interconnect patterns are shown in figure 2.

In this implementation, each of the 48 neurons (6×8) has an input detector followed by a capacitor-coupled inverting amplifier chain and a low-pass filter, and the output drives a VCSEL (figure 3). Initially all the lasers are set to a fixed output level, slightly higher than the off level. This sets a stable total power for the array and effectively biases the neurons towards 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 detector arrays are a pair of lenses and a DOE that divide the light from one neuron’s laser and focus it onto the inputs of the other neurons in the same row and column, but not its own input. Due to inversion in the amplifier chain, light falling on a detector inhibits the associated neuron, decreasing its output.



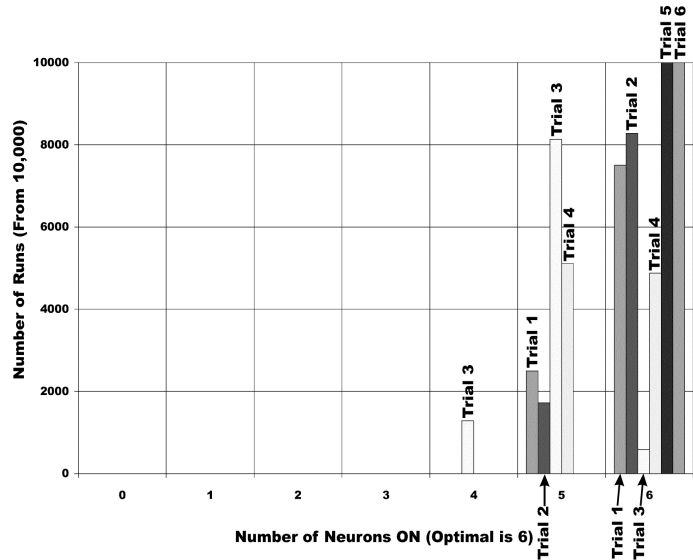
**Figure 3:** The electrical system used in the test system is much simpler than that of the general purpose network.

### 3.1 Packet Switch Results

This system has been implemented and its performance measured as a scheduler for both crossbar (figure 4) and self-routing (figure 5) switch fabrics. The scheduler never produced an invalid result. Most times it found an optimal result except with the requests TRIAL6 and TRIAL8 (crossbar only, figure 4). With these it usually routed one fewer packets. Thus the switch would have near maximum throughput and never block. No attempt has been made to make this demonstration system run fast but nevertheless it provided a decision within 33μs: a rate compatible with the latest router requirements.

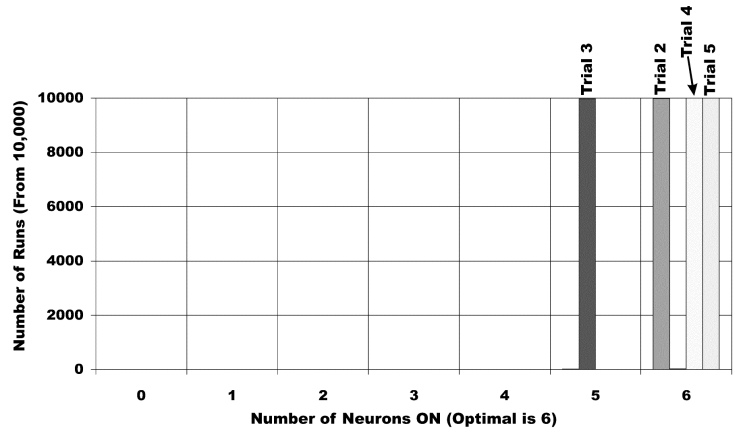
Simulations of a 16x16 neural network

scheduler were undertaken in order to make performance comparisons with other scheduler designs. The simulations were performed under uniform traffic conditions and the mean delay (measured in packet periods) was plotted against the offered load (the probability of a packet arriving at each input). Figure 6 summarises the results of this exercise. The uppermost curve shows the situation when the inputs are simply buffered in a first in first out (FIFO) fashion. FIFO queues suffer from the problem of head of line (HOL) blocking in that if the foremost packet in the queue (the next one to go) is blocked by another request but this also blocks all the packets behind it in the queue even if their destinations are not in contention. As might be expected, a scheduler based on FIFO buffering suffers severe performance degradation under increasing load. The lowest curve represents the

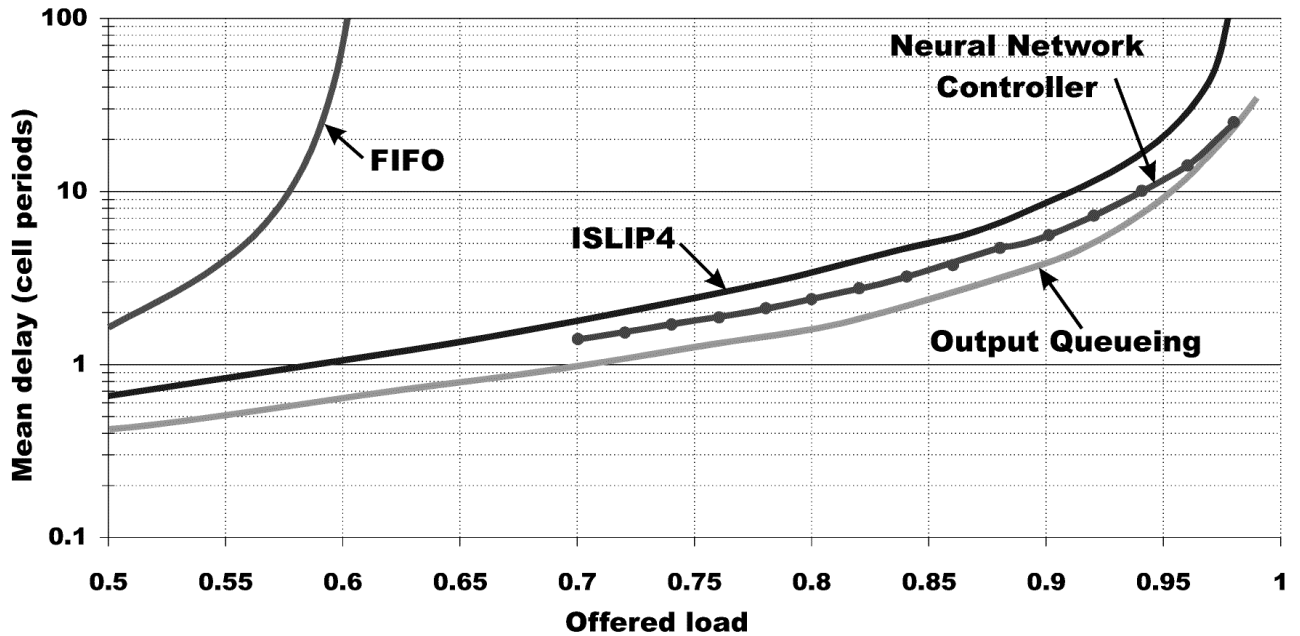


**Figure 4:** Histogram of packets routed in the crossbar switch.

theoretical best that may be achieved. This is described as output queuing and is calculated assuming an ideal switch fabric (impossible) where packets have only to wait for a vacant slot on the output line. The solid line represents an algorithm called ISLIP4 [12] which can be implemented in CMOS electronics for a high speed switch of this size. The dotted line shows the neural network scheduler performance and its favourable throughput at loads from 70% upwards.



**Figure 5:** Histogram of packets routed in the self-routing switch.



**Figure 6:** A comparison of the neural network controller against a state of the art scheduler, ISLIP4, clearly indicates its advantage at high levels of offered load. The output queuing curve indicates a theoretical optimum value.

#### 4 Conclusion

In this paper we have described the successful implementation of a neural network which 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 and therefore very large switch sizes could be scheduled with little cost in speed. In addition to this it will be possible to push the speeds up still further by removing the communication delays in the system with a smart pixel based implementation of the electronics where each detector/VCSEL combination has its own processing element (PE). This work [13] demonstrates the principles of this system using discrete components along with an application where the system excels.

## 5 References

- [1] R. P. Webb, "Optoelectronic Implementation of Neural Networks", *International Journal of Neural Systems*, **4**, pp.435-444, (December 1993).
- [2] P. Blair, "Diffractive Optical Elements, Design and Fabrication Issues", Ph.D. Thesis, Department of Physics, Heriot-Watt University, (1995)
- [3] T. X. Brown, "Neural Networks for Switching", *IEEE Communications Magazine*, **27**, pp. 72-81, (November 1989).
- [4] J. J. Hopfield and D. W. Tank, "'Neural' Computation of Decisions in Optimisation Problems", *Biological Cybernetics*, **52**, pp. 141-152, (1985).
- [5] R. D. Brandt, Y. Wang, A. J. Laub and S. K. Mitra, "Alternative Networks for Solving the Travelling Salesman Problem", *IEEE International Conference on Neural Networks*, San-Diego 24<sup>th</sup>-28<sup>th</sup> February 1998.
- [6] T. X. Brown, K. H. Liu, "Neural Network Design of a Banyan Network Controller", *IEEE Journal on Selected Areas in Communications*, **8**, 8, pp. 1428-1438, (October 1990).
- [7] R. P. Webb, A. J. Waddie, K. J. Symington, M. R. Taghizadeh and J. F. Snowdon, "A Neural Network Scheduler for Packet Switches", *Technical Digest of Optics in Computing, OSA International Conference*, Snowmass, Colorado, pp. 193-195, April 13<sup>th</sup>-16<sup>th</sup> 1999.
- [8] J. J. Hopfield, D. W. Tank, "Neural Computation of Decisions in Optimisation Problems", *Biological Cybernetics*, **52**, pp. 141-152, (1985).
- [9] P. W. Protzel, D. L. Palumbo and M. K. Arras, "Performance and Fault-Tolerance of Neural Networks for Optimisation", *IEEE Transactions on Neural Networks*, **4**, 4, pp. 600-614, (July 1993).
- [10] J. Ghosh, A. Hukkoo and A. Varma, "Neural Networks for Fast Arbitration and Switching Noise Reduction in Large Crossbars", *IEEE Transactions on Circuits and Systems*, **38**, 8, pp. 895-904, (August 1991).
- [11] A. Marrakchi and T. Troudet, "A Neural Net Arbitrator for Large Crossbar Packet Switches", *Circuits and Systems Letters, IEEE Transactions on Circuits and Systems*, **36**, 7, pp. 1039-1041, (July 1989).
- [12] N. McKeown, M. Izzard, A. Mekittikul, W. Ellersick and M. Horowitz, "The Tiny Tera: A Packet Switch Core", *IEEE Micro*, **17**, 1, pp. 26-33, (January/February 1997).
- [13] R. P. Webb, A. J. Waddie, K. J. Symington, M. R. Taghizadeh and J. F. Snowdon, "Optoelectronic Neural-Network Scheduler for Packet Switches", *Applied Optics*, **39**, 5, pp. 788-795, February 2000.