Heriot-Watt University
Edinburgh

Implementation of an Optoelectronic Neural Network

Abstract

This proposal examines implementation methodologies for an optoelectronic neural network comparing it to current digital, analogue and hybrid systems. After careful examination of available components, a conclusion is made on the desired characteristics of any demonstrator as well as the final project goal.

Name:	Keith J. Symington
Registration number:	9711710098481
Company/Address:	Physics Dept., Heriot-Watt University, Edinburgh, EH14 4AS
☎:	+44 (131) 451 3040
Fax.:	+44 (131) 451 3136
e-mail:	kjsymington@iee.org
Supervisor:	Dr. J. F. Snowdon
Date:	9th December 1998

1   Contents

2 Proposal Outline

This proposal examines an optical neural network for switching; specifically the hardware used to simulate neurons and associated price/performance issues. It presumes familiarity with the subject area as described in [1].

The proposed implementation (figure 1) would fit a processor between detector and VCSEL arrays to perform the work normally carried out by another type of hardware system. Various alternative systems have been considered and thoughts on these systems are outlined in chapter 6.
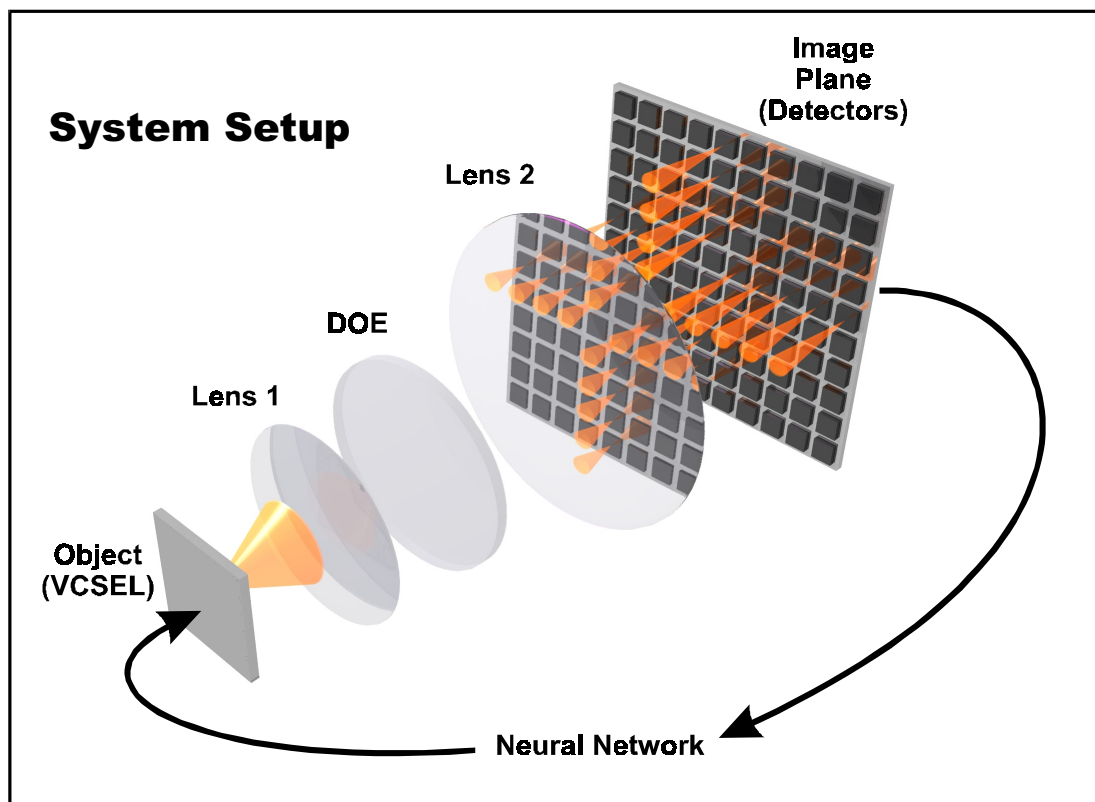

Figure 1

This diagram illustrates how a diffractive optic element (DOE) separates light from a single VCSEL and its consequent imaging onto a detector array.

The complexity in this problem lies in an exponential increase in the number of additions required to perform interconnection as the network size, mxn, grows. The very architecture of this system tackles the problem by executing summation 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 modified by an activation function.

The reason that a microprocessor solution was considered at all is because it offers a flexibility not found in other hardware systems, which will be examined later. Scalability is also not a serious problem since the number of neural activation functions to be calculated scales proportionately with number of neurons in the network, mxn.

---

3   Neural System Performance

There are many ways of classifying a neural network: architecture, network type, number of external inputs and outputs, number of neurons etc. To evaluate system performance, this proposal will use the Connections-Per-Second (CPS) rating as defined in 1991 by M. Holler [2]. This rating is really only applicable in discrete systems.

3.1   Connections-Per-Second

Neural networks consist of a set of interconnected neural processing elements (neurons) which work by calculating the sum of a set of inputs x_i multiplied by a set of stored weights w_i (see figure 2). This sum is then modified by an activation function $f(x)$ to give output y. The inputs to a neuron are known as artificial synapses and calculation of the product of one input x_i and its synaptic weight w_i is referred to as a

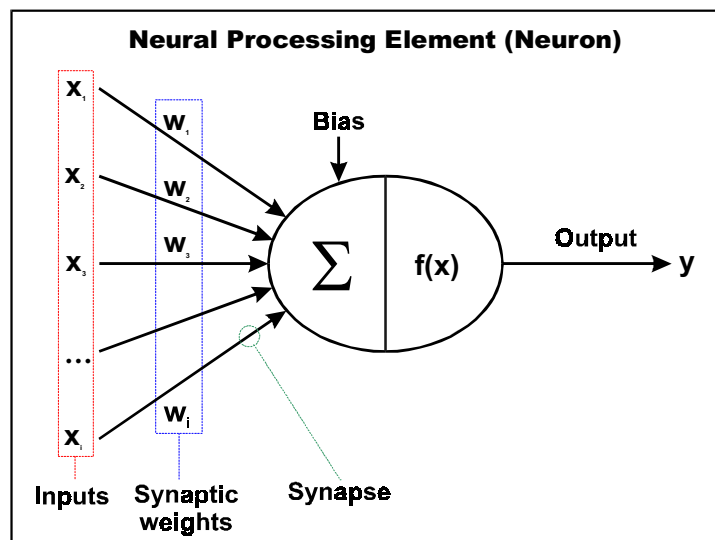**Neural Processing Element (Neuron)**

Figure 2

Diagram of an artificial neuron.

connection. The connection is a basic unit of computation in a neural network and the number of connections per second (CPS) that it can perform a measurement of performance. The CPS is directly related to how fast a network can perform mappings from input to output.

3.2   Neural Network Hardware Review

This section examines existing neural network hardware and reviews their implementations and performance. Such information is useful in that it sets a performance target for any system constructed. For further information on these systems please see [3] or [15].

3.2.1   Analogue Neural Networks

The analogue neural network exploits physical properties to perform operations and thus obtain high speeds and densities. A common output line could sum the currents from several synapses to the neuron inputs. The

major problem though with analogue systems is component tolerances: it can be very difficult to compensate for variations in manufacturing.

Analogue Network	Architecture	Neurons	Synapses	Refs.	CPS
Intel ETANN	FF, ML	64	10280	[4]	2×10^9

Table 1

3.2.2   Digital Neural Networks

A digital network is complete digitalisation of a neural network: the weights are stored digitally and all calculations are made digitally. Although digital summation can be slow, especially with regard to synapses, it is an extremely flexible and a comparatively mature technology.

Digital Network	Architecture	Neurons	Synapses	Refs.	CPS
NeuraLogix NLX-420	FF, ML	16	Off chip	[5]	300
HNC 100-NAP	SIMD, FP	100 PE	512K off chip	-	250×10^6
Hitachi WSI	SIMD, Hopfield	576	32K	[6]	138×10^6
Inova N64000	SIMD, Int.	64 PE	128K	[7], [8]	870×10^6
MCE MT19003	FF, ML	8	Off chip	[9]	32×10^6
Micro Devices MD	FF, ML	1 PE	8	[10]	8.9×10^6
Philips Lneuro-1	FF, ML	16 PE	64	[11]	26×10^6
Siemens MA-16	Matrix ops.	16 PE	256	[12], [13], [14]	400×10^6

Table 2

3.2.3   Hybrid Neural Networks

A hybrid implementation supposedly combines the best of both digital and analogue techniques: analogue summation and digital noise resistance.

Hybrid Network	Architecture	Neurons	Synapses	Refs.	CPS
AT&T ANNA	FF, ML	16-256	4096	[16]	2.1×10^9
Bellcore CLNN-32	Boltzmann	32	992	[17]	100×10^6
Mesa Research Neuroclassifier	FF, ML	6	426	[18]	21×10^9
Ricoh RN-200	FF, ML	16	256	[19]	3×10^9

Table 3

For this reason, hybrid systems reach the highest performance levels of all types of hardware implementations: 21×10^9 CPS having been demonstrated.

3.2.4   Alternative Implementations

It is also possible to implement neural networks using other methods. This usually involves using a dedicated generic processor (e.g. Transputer, Intel i860 or DSP) to simulate the network and its interconnects purely in software. No examples are included here because this proposal is examining a hardware implementation using optical interconnects rather than a software one.

4   Proposed System

This section compares the performance of our proposed system against existing neural networks by calculating its CPS rating in relation to iteration frequency and network size.

4.1   Network Connectivity

Presuming there are enough VCSELs and detectors available for each neuron, where both arrays are of size m inputs by n outputs, we have a connection density c_d as shown in equation 1:

$$c_d = (m + n - 2) \times m \times n$$

Equation 1

The resulting number of connections are graphed in figure 3, showing that an increase in network size can increase the number of interconnections drastically.
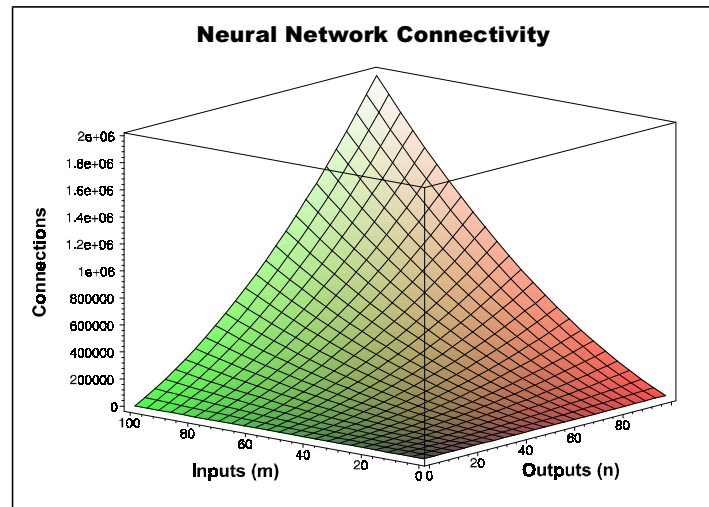

Figure 3

This graph shows how the number of connections grows as the number of inputs and outputs are altered.

4.2   Network Performance

Once we know the number of connections there are we can work out the system's CPS rating by examining its iteration speed. Note that this is not directly related to solutions per second as network convergence requires multiple iterations. This number of iterations is believed to be around 50 per solution. If we
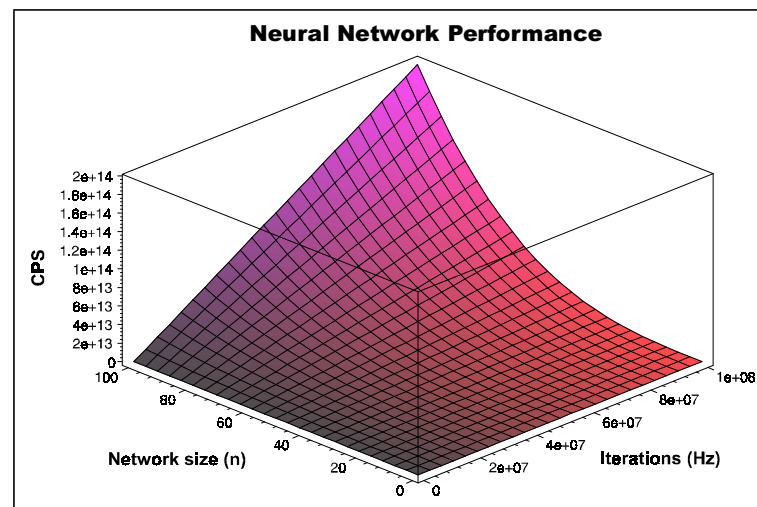

Figure 4

It can be seen that the CPS rating increases rapidly with network size n (n inputs and n outputs giving nxn neurons). This increase is unfortunately only linear when related to iterations per second.

measure iterations f_i in Hz and keep the network (mxn) square (nxn) we can derive the relationship in equation 2 and figure 4:

$$CPS = (2n - 2) \times n^2 \times f_i$$

Equation 2

Keeping the network size square allows all inputs to be connected to all outputs simultaneously. If there were more inputs than outputs then a situation would eventually arise where certain inputs cannot be connected to any output whatsoever as all existing outputs are busy. Thus a square crossbar switch is considered to be of optimal design.

4.3   Proposed System Performance

We can therefore determine the performance of our demonstrator using equation 2. At the present time our demonstrator will be limited to n=8, however the iteration frequency f_i has not been determined. It is hoped that values of 100kHz up to perhaps 1.2MHz would be possible giving CPS ratings of 89.6MCPS to 1.08GCPS respectively. Although not as fast as some neural systems in section 3, it does compare favourably.

What we need to remember is that our neural system design is only partially interconnected. If we were to replace the DOE with one which connects every neuron's output with the input of every other then the network's CPS rating would then be determined by equation 3:

$$CPS = n^4 \times f_i$$

Equation 3

This would give a performance of 410MCPS at 100kHz and 4.9GPS at 1.2MHz: pretty high considering there are only 64 neurons. Finding an application for such a network on the other hand could be awkward.

5 Available Components

At the current stage of development, there are only two components whose characteristics are already known: the detector and VCSEL arrays. Therefore, any optical system must be designed around these components.

5.1 VCSEL Array

The VCSEL array supplied is shown in detail in figure 5 and its characteristics in figures 6, 7 and 8. It can be seen that the mean threshold current is 2.57 ±0.05mA, mean threshold voltage 1.93 ±0.01V and mean optical output power (at 8 mA) 1.25 ±0.02mW. The power conversion efficiency at 8 mA is 6.3 ±0.1%.

The emission wavelength for this array is ~956 nm with a maximum variation across the array of $\Delta\lambda_{max}$=0.25nm. This value has been determined by individual operation of each VCSEL at 8mA. The emission wavelength
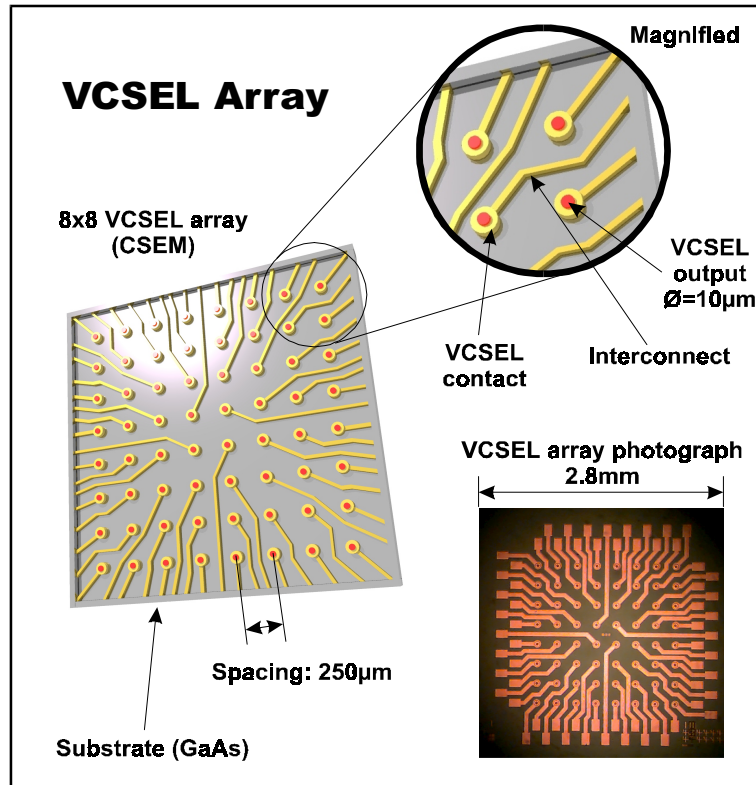
VCSEL Array

8x8 VCSEL array (CSEM)

Magnified

VCSEL output Ø=10µm

VCSEL contact

Interconnect

Spacing: 250µm

Substrate (GaAs)

VCSEL array photograph 2.8mm

Figure 5

VCSEL array was originally fabricated by CSEM as a SPOEC demonstrator.

VCSEL Threshold Current   **VCSEL Threshold Voltage**   **Optical Output Power**

$I_{th(mean)}$=2.57 ±0.05mA $V_{th(mean)}$=1.93 ±0.01V mean $P_{out(mean)}$=1.31 ±0.01mW at I=8mA
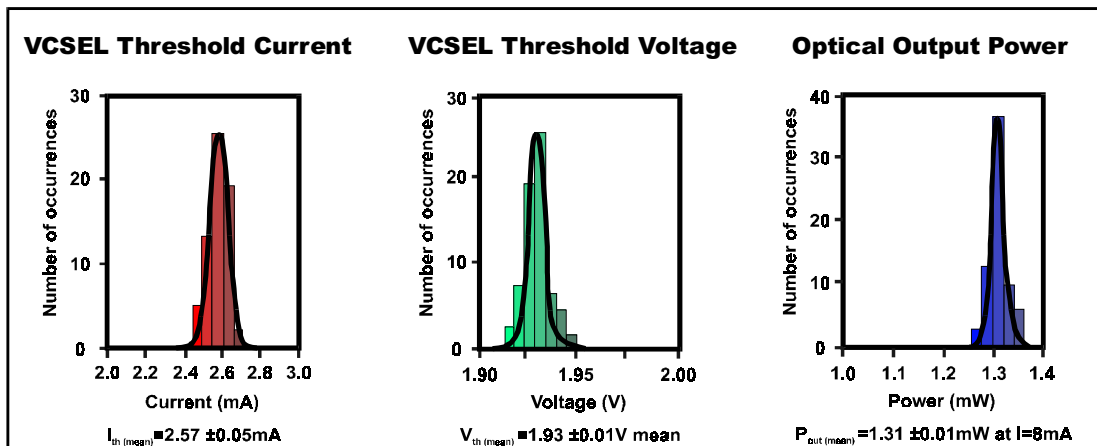
Figure 6

The above values demonstrate how similar all VCSELs in the array are. Such an array reduces the amount of calibration which needs to be done and thus system complexity.
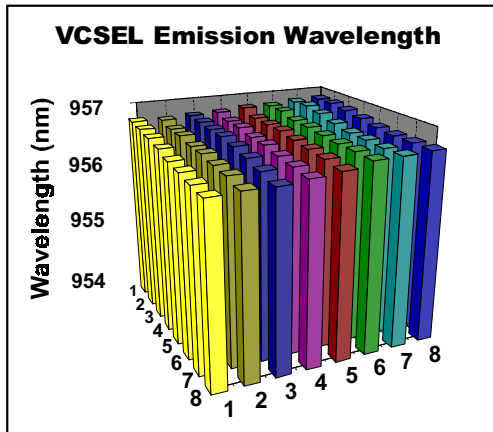
Figure 7

Emission wavelength distribution at a constant drive current I=8mA. Maximum wavelength deviation $\Delta\lambda_{max}$=0.25nm.
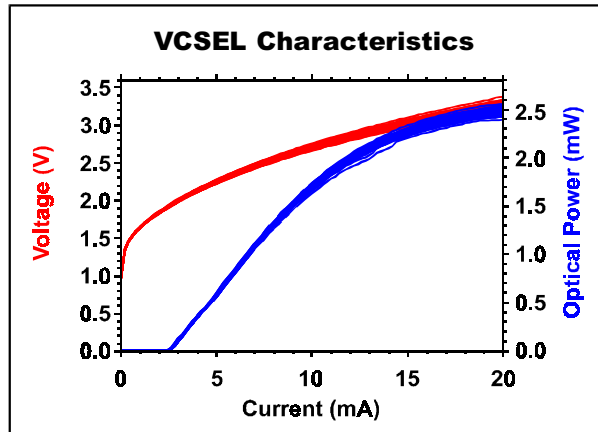

Figure 8

VCSEL optical output power and voltage drop in relation to current.

variation when VCSELs are operated simultaneously can only be determined later, however preliminary experiments with a comparable array ($\Delta\lambda_{max}$=0.8nm for individual VCSEL operation) has shown that this additional wavelength variation is relatively small: $\Delta\lambda_{max}$=1.1nm.

The bias free modulation response of individual array VCSELs can reach data rates of 250MBit/s NRZ with a 1.6ns turn-on delay. Adding a bias of 1.9V reduces the turn-on delay to 0.9ns thus reaching data rates of 500MBit/s with ease.

For further information on the VCSEL array please see [20].

5.2   Detector Array

Two types of detector array were considered for this project: a CCD array and a Photodiode detector array.

Using a CCD array would have given a much higher resolution than a photodiode array perhaps allowing intelligent alignment using signal processing. However, the disadvantages of CCDs in this implementation proved to be far too restrictive: a shutter is required and serial data output from the chip gave frame rates

Figure 9

The above photodiode detector array is to be used in the current project.

of between 10-15Hz (using an inexpensive array). Pixel bining would have improved the frame rate but it is inflexible and enabled chips are expensive (~£15,000).
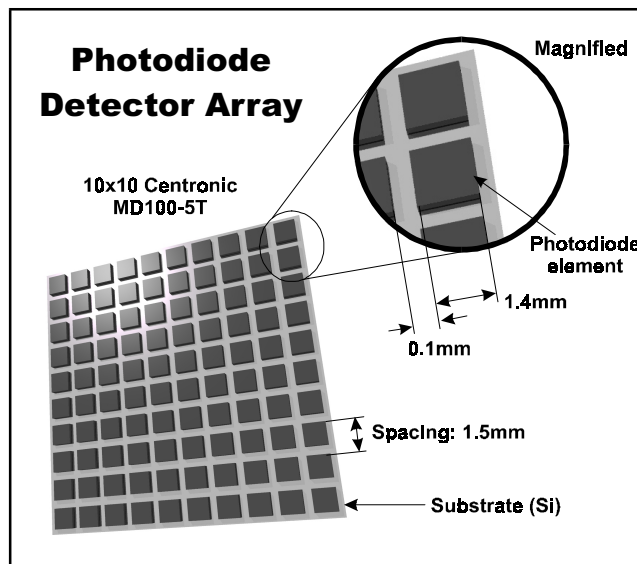
The photodiode detector array shown in figure 9 (Centronic 10x10 element 5T photodiode array) was chosen because it could not only support high data rates (~26ns response) but all elements could be read out simultaneously.
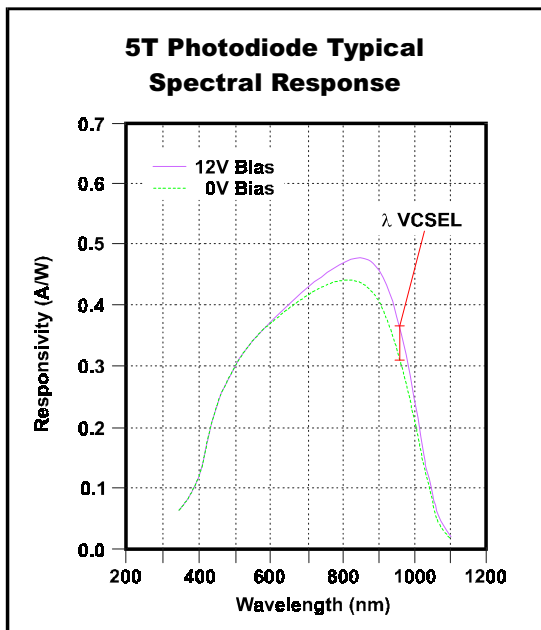

Figure 10

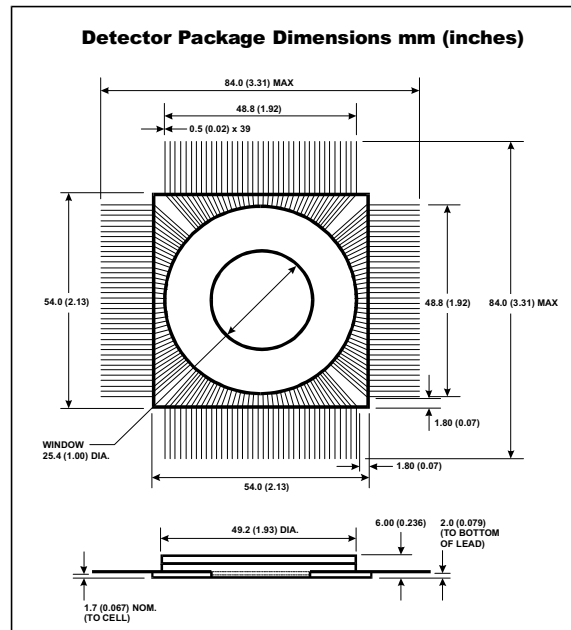The wavelength at which the system operates is clearly marked in red.

Figure 11

The photodiode detector array is placed in the central window. Admittedly, the array is fairly large in comparison to the VCSEL array.

Figures 10 and 11 provide further information on frequency response and packaging of the array respectively. Table 4 examines the detector array in more detail.

Elements	Area mm²	Width mm	Length mm	Separation mm	Responsivity A/W @ ~956nm Minimum	Responsivity A/W @ ~956nm Typical	Dark Current nA Min.	Dark Current nA Max.	Capacitance pF Max. @ 0V Bais	Capacitance pF Max. @ 12V Bais	Shunt Resistance Min. Megaohms	Shunt Resistance Typ. Megaohms
10x10	1.96	1.4	1.4	0.1	0.31	0.36	200	1	55	12	1	400

Table 4

5.3 Optics

Although this proposal does not consider the optical system in detail, the following two points must be kept in mind

- Due to VCSEL choice, the optical system must be designed for 956nm with AR coatings at this wavelength

- Size is a very important consideration: the smaller the better. System size is directly related to DOE working distance and thus it must converge as acutely as possible.

6   Neural Network Implementation

Since we already have a predefined optical solution (figure 1), the only aspect which remains unspecified is the neural network. This section compares and contrasts analogue and digital solutions.

6.1   Analogue Neural Networks

Implementing an analogue solution would mean that the entire system would be classified as an analogue one. The simplest (and cheapest) solution would be to use an operational amplifier which would act as the neurons as described in [1]. Flexibility could also be added by using components such as an EPAC (Electrically Programmable Analogue Circuit) from ITC or FIPSOC (Field Programmable System On Chip).

6.1.1   Advantages

- The component densities of most analogue systems are higher than that of digital.

- Analogue systems can be very fast.

- Cost is very low.

- Easily integrated into any smart pixel implementation.

6.1.2   Disadvantages

- Component tolerances become critical and indeed have already been a problem in a previous system [1].

- Inflexible.   Once designed and built it becomes hard to alter any parameters. This problem could perhaps be circumnavigated by using programmable analogue circuitry.

- Network convergence is highly dependent on components used.

- Tricky to design correctly.

- Signal-to-noise ratio can be low.

6.2   Digital Neural Network

Since there is already an analogue component in the system (weight summation), adding any digital hardware would change the system's classification to hybrid. Although integrating digital components may sound out of place, I believe that the benefits it will bring far outweigh the drawbacks. Unfortunately, the problem with any digital neural implementation is conversion from analogue to digital at the input and digital to analogue at the output. This can not only be slow (1.2MHz ADC, 100kHz DAC) but very

costly. After careful consideration, I believe that some sort of microprocessor solution is very promising.

I do not believe that an FPGA or a digital building block solution is feasible nor sensible since such a system would be as complicated as any analogue implementation if not more so without custom chip design. Any advantages which may have been gained from analogue to digital conversion will therefore be negated and it would be more pertinent and cost efficient to have designed in analogue from the start.

The points made below in both sub-sections 6.2.1 and 6.2.2 thus argue for and against a microprocessor system.

6.2.1   Microprocessor Advantages

- Simplicity. Microprocessor systems are (usually) a plug-in solution so electronic design will be kept to a minimum.

- Flexibility. The neuron activation function f(x) can be reconfigured to anything that can be calculated on a processor.

- Lookup tables can be used for system calibration. A CPU can adjust VCSEL output until it reaches predefined levels on the photodiodes. This can also prevent saturation of the photodiodes; a sort of active calibration.

- Alignment can be made active as a microprocessor can examine light intensities falling on individual detectors to ensure light is only reaching the correct ones. The system is no longer 'dumb'.

- Can rapidly simulate failed neurons and the effects they would have on the system.

- Workload can be divided across multiple processors allowing easy scaling: fortunately the number of calculations per iteration is directly proportional to the number of neurons thus workload does not increase exponentially.

- A microprocessor can judge when the network has converged so any result can be output when the system is finished and not after a predefined time period.

- Proof of principle. A microprocessor could be given measured characteristics from another proposed implementation and replicate them for evaluation purposes: e.g. a proposed analogue implementation.

6.2.2   Microprocessor Disadvantages

- Expensive. Cost is, however, highly dependent on whether it is a turnkey system (£4,000+) or a system built around embedded microcontrollers (from £15+ each).

- DAC can be slow without expensive hardware.

- Overall system could be quite large.

- Any proof of principle is only as good as the model used presuming digital is not going to be the final implementation.

- Just a hardware implementation of the simulation (is that really a disadvantage?).

- Multiple processors required to prevent bottlenecking.

6.3   Component Interfacing

There are three interfacing problems which need to be carefully considered regardless of system design.

6.3.1   Detector to Neural Network

If an analogue neural network is implemented then this is not an issue. However, if a it is digital than D/A conversion over 64 channels will be required. This could also be performed using multiplexing of components but results in serial processing of parallel data.

6.3.2   PC to Neural Network

Presuming that the system to be built will support more than active/inactive inputs, an analogue neural network requires digital to analogue conversion from PC (or any other computer) to network (perhaps vice versa for return - system dependent). A digital network would minimise this problem with some microprocessors capable of direct connection to a PC com port (no extra hardware necessary).

6.3.3   Neural Network to VCSEL

64 VCSELs need to be driven by the neural network. An analogue network would not prove difficult to interface but, again, a digital network would. Digital to analogue conversion would clearly be required which is not only slow and expensive but could result in the multiplexing of several channels.

7   Conclusion

The ultimate goal of this project is to create a flip chip bonded solution which contains VCSEL array, neuron electronics and detector array in a folded system as shown in figure 12.
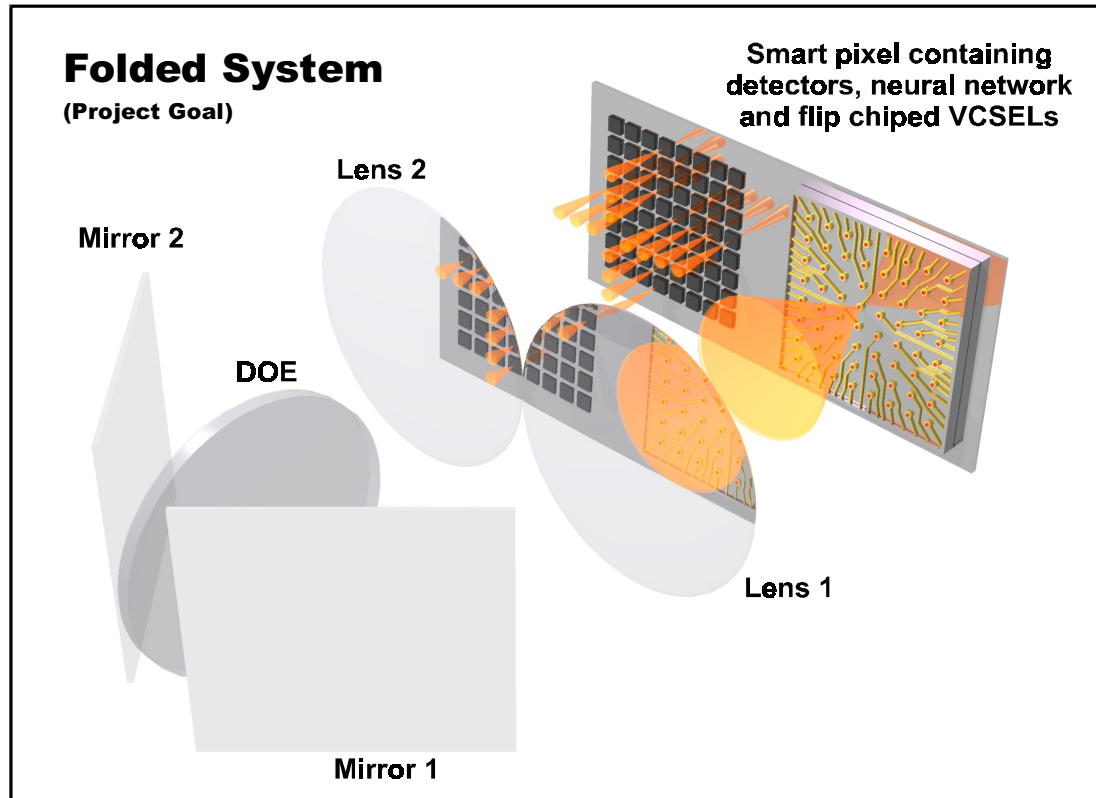

Folded System
(Project Goal)

Smart pixel containing detectors, neural network and flip chiped VCSELs

Lens 2

Mirror 2

DOE

Lens 1

Mirror 1

Figure 12

This folded system has detectors, neurons and VCSELs all fabricated on one component. The lens system still remains the crucial size limiting factor: just how far it can be miniaturised still remains unclear, but it appears to be highly dependent on DOE working distance.

The real deciding factor here is cost. A microprocessor solution is undoubtedly far quicker to implement, a lot more flexible and perfectly placed to examine system nuances, however cost weighs heavily against it. Fortunately, it also has the advantage of reusability: any system which can sample 64 analogue inputs and output 64 analogue voltage levels after signal processing is universally useful, especially with the attached signal processing capabilities.

I believe that implementing a simple analogue solution at this stage would be taking a leap towards a project goal without necessarily understanding it completely. Since we are designing a demonstrator, flexibility is of ultimate importance: the flexibility to alter network configuration, activation functions and compensate for component tolerances. During writing, an analogue demonstrator was completed at BT Labs so I can only conclude that we either examine a discrete solution using microprocessor hardware or tweak the original (using programmable analogue chips rather than op-amps) to additionally support analogue input values.

8   Glossary

ADC	Analogue to Digital Conversion
AR	Anti Reflective
BP	Back Propagation
CCD	Charge Coupled Device
CPS	Connections-Per-Second
CSEM	Centre Suisse d'Electronique et de Microtechnique
DAC	Digital to Analogue Conversion
DOE	Diffractive Optic Element
DSP	Digital Signal Processor
EPAC	Electrically Programmable Analogue Circuit
FIPSOC	Field Programmable System On Chip
FF	Feed Forward
FP	Floating Point
FPGA	Fully Programmable Gate Array
ML	Maximum Likelihood
NRZ	Non-Return to Zero
PE	Processing Elements
SIMD	Single Instruction, Multiple Data
SPOEC	Smart Pixel Optoelectronic Connections
VCSEL	Vertical Cavity Surface Emitting Laser

9   Bibliography

[1]   K. J. Symington, *"Optoelectronic Neural Networks for Switching"*, MSc project dissertation, St. Andrews and Heriot-Watt University, 1998.

[2]   M. A. Holler, *"VLSI Implementations of Learning and Memory Systems: A Review"*, Advances in Neural Information Processing Systems 3, Morgan Kaufmann, San Mateo, Ca. 1991.

[3]   C. S. Lindsey and T. Lindblad, *"Review of Hardware Neural Networks: A User's Perspective"*, Physics Department - Frescati, Royal Institute of Technology Frescativägen 24, 104 05 Stockholm, Sweden, lindsey@particle.kth.se, lindblad@vana.physto.se.
http://msia02.msi.se/~lindsey/elba2html/elba2html.html

[4]   80170NX ETANN, Data Sheet, Intel Corp., Santa Clara, CA. 1991.

[5]   NLX420 Data Sheet, June 1992, Neurologix, Inc., 800 Charcot Av., Suite 112, San Jose, Ca.

[6]   M. Glesner and W. Pöchmüller, *"Neurocomputers: An Overview of Neural Networks in VLSI"*, Chapman & Hall, London, 1994.

[7]   D. Hammerstrom, *"A VLSI Architecture for High-Performance, Low-Cost, On-chip Learning"*, Proc. Int. Joint Conf. on Neural Networks, II, 537-544, San Diego, Ca., June 1990.

[8]   Adaptive Solutions, Inc. 1400 N.W. Compton Dr., Suite 340, Beaverton, Or. 97006.

[9]   MT19003 Data Sheet, MCE, Alexander Way, Tewkesbury, Gloucestershire GL20 GTB, 1994.

[10]   MD1220 Data Sheet, Micro Devices, 30 Skyline Dr., Lake Mary, Fl. 32746, March 1990.

[11]   N. Mauduit, M. Duranton, J. Gobert, *"Lneuro 1.0: A Piece of Hardware LEGO for Building Neural Network Systems"*, IEEE Trans. on Neural Networks, 3, 414-422, May 1992.

[12]   J. Beichter, N. Bruels, E. Meister, U. Ramacher and H. Klar, *"Design of General-purpose Neural Signal Processor"*, 2nd Int. Conf. on Microelectronics for Neural Networks, 311-315, München, Germany, 1991.

[13]   U. Ramacher, *"SYNAPSE - A Neurocomputer that Synthesises Neural Algorithms on a Parallel Systolic Engine"*, Journal of Parallel and Distributed Computing,14, 306-318, 1992.

[14]   Siemens Nixdorf, Informationssysteme AG, D55 Sci. Computing, D-81730 München, Germany.

[15]   S. Haykin, *"Neural Networks: A Comprehensive Foundation"*, Macmillan College Publishing Inc., 1994.

[16] B. E. Boser, E. Sackinger, J. Bromley, Y. LeCun and L.D. Jackel, *"Hardware Requirements for Neural Network Pattern Classifiers"*, IEEE Micro, pp. 32-40, Feb. 1992.

[17] J. Alspector, T. Jayakumar and S. Luna, *"Experimental Evaluation of Learning in a Neural Microsystem"*, Advances in Neural Information Processing Systems - 4, ed. J. Moody et al., 871-878, Morgan-Kaufmann Publishers, San Mateo, Ca. 1992.

[18] P. Masa, K. Hoen, H. Wallinga, *"A 20 Input, 20 Nanosecond Pattern Classifier"*, Proc. Int. Joint Conference on Neural Networks IJCNN '91, Orlando, Fla., USA, July 1994.

[19] H. Eguci, T. Furata, H. Horiguchi, S. Oteki and T. Kitaguchi, *"Neural Network LSI Chip with On-chip Learning"*, Proc. Int. Joint Conf. on Neural Networks IJCNN '90, San Diego, Ca., June 1990.

[20] Heriot-Watt University, Supelec, University of Glasgow, Trinity College and CSEM, *"Supply of 8x8 VCSEL array as input device for demonstrator"*, Deliverable 4, Project No. 22668 SPOEC, April 1998.