Adaptive Multilayer Optical Neural Network with Optical Thresholding

I. Saxena and E. Fiesler

IDIAP, Case Postale 592, 1920 Martigny, Switzerland
E-mail (Internet): ISaxena@IDIAP.CH

Abstract

An adaptive multilayer optical neural network with optical thresholding by a liquid crystal light valve (LCLV) and all-optical forward propagation is described. Its modular structure allows the cascading of layers, each with its own light source, and has a large number of modifiable optical interconnections which are implemented by liquid crystal television screens.

Sigmoid fits to response curves of four LCLVs are evaluated and their suitability as optical thresholding functions is examined on the basis of neural network simulations.

1 Introduction

The promise of massive parallelism in optics has led to the investigation of several optical neural network (ONN) implementations. Amongst these, Hopfield neural networks (NNs) have been most frequently implemented optically due to their effective single layeredness. However, they are mainly limited to associative memory applications, whereas multilayer neural networks are capable of addressing a broad class of problems. Neural networks of the latter kind require a differentiable thresholding function, such as the conventional sigmoidal activation function [Rumelhart-86].

Previous multilayer opto-electronic NNs (MONNs) performed thresholding electronically (in software [Jang-93, Yu-90, Kranzdorf-89] or hardware [Ohta-90, Yuk-93]), and it is primarily the vector-matrix multiplication which was done optically. To progress towards all-optical recall and to reduce the overhead involved in conversion of optical signals into electronic, and re-conversion (after electronic thresholding) into optical signals for the next MONN layer, *optical* thresholding is essential. A liquid crystal light valve (LCLV) is a potential device for optical thresholding as its non-linear response is sigmoid-like. To evaluate the usefulness of LCLVs for optical thresholding in MONNs, the closeness of their response curves to sigmoidal functions is studied here.

In the following section, a MONN is described in which optical thresholding is implemented by an LCLV. Next, the properties of an ideal optical thresholding function for incoherent, all-positive, optical systems are examined. In the light of these properties, sigmoid curve-fits of four LCLV response curves are compared with ideal sigmoids, and their thresholding characteristics discussed. Finally, neural network simulations with the sigmoid-fits are analyzed to make a qualitative assessment of the most suitable LCLV for optical thresholding.

2 Optical Multilayer Neural Network

In ONNs where information is coded in light intensity, all variables including the interconnection weight matrix (IWM) elements have to be positive. (In general, IWMs contain both negative and positive weights, which are often referred to as bipolar weights.) An application independent, adaptive MONN architecture without negative weights is described here, in which optical thresholding is implemented by an LCLV. Three possible solutions allow working with unipolar weights in MONNs. In two-layer[1] systems (without hidden layer), the second and final layer can be electronic, and bipolar weights may be separated into two sets of unipolar weights, which, after being photodetected separately, either (i) *spatially* [Kasama-90] or (ii) *temporally* [Yu-90], are subtracted electronically. Bipolar weights may alternatively (iii) be transformed to unipolar weights by adding a bias term and compensating accordingly. In the MONN described here, the third possibility is adopted, and the argument of the activation function is biased at the hidden layer (with the aid of a personal computer), before optical thresholding.

Our three layer adaptive MONN has 256 input neurons, 256 hidden layer neurons, and 16 output neurons. This size enables addressing a wide range of applications. The information flow through the MONN can be described as follows. Figure 1 shows the optical system whereby the inputs are presented by a Liquid Crystal Television

[0] Invited paper for *Optical Engineering*; under review.

[1] In this report, a *layer* of a neural network is defined as a layer of neurons (cf. [Fiesler-94.1]). In the three layer neural network described in the text, the input layer is considered number one, the hidden layer number two, and the output layer as layer number three.

Table 1: Sigmoid parameter definitions

Properties	Standard Sigmoid	Generic Sigmoid
Y-axis intercept	$\frac{1}{2}$	$\alpha + \frac{\gamma}{1+e^{-\delta}}$
Inflection Point (x,y)	$(0, \frac{1}{2})$	$(\frac{\delta}{\beta}, \alpha + \frac{\gamma}{2})$
Slope at inflection point	$\frac{1}{4}$	$\frac{\gamma\beta}{4}$
Steepness	1	$\gamma\beta$
$S_{max}(x)$= asymptote for $x \to +\infty$	1	$\alpha + \gamma$
$S_{min}(x)$= asymptote for x $\to -\infty$	0	α

(LCTV1) [Collings-94] to the fanned-out laser beam, and the 16×16 input array is replicated 256 times onto the transmissive LCTV2, following the fan-out design of [Weible-90, Weible-92]. The IWM between layers is represented by the transmission values of the pixels of the LCTVs. The intensity of light transmitted by LCTV2 is integrated by a micro-lens array to represent the vector-matrix product of inputs and weights[2]. This product, after biasing, is thresholded by the LCLV to produce the activation values of the hidden layer which are read by a different light source. The hidden layer outputs are similarly presented to LCTV3 which represents the 256×16 IWM between the hidden and output layer. Following LCTV3 the micro-lens array output gives the second vector-matrix product. The final outputs are detected by a 4×4 photodetector array. The hidden layer non-linearity, concerning 16×16 neurons, implemented optically by a transmissive LCLV, shall be discussed in section 4.

In the MONN described above, light emitted by an Argon-ion laser (480nm) transmits through LCTV1, the fan-out optics, LCTV2, and the fan-in optics, and then writes to the LCLV. A helium-neon (633nm) light source reads the optically thresholded hidden layer output from the LCLV. Provided that it has sufficient gain, this method can amplify light intensity to compensate for light losses in the first two layers of the neural network. Subsequently, each additional layer can be modularly stacked, each with a new light source, onto the previous module, to build an ONN with multiple layers. A suitable LCLV at the hidden layer, therefore offers a two-fold advantage of gain in light intensity in addition to optical thresholding.

3 Ideal Optical Thresholding

A sigmoid of the form $S(x) = \frac{1}{1+e^{-x}}$, to be referred to as the *standard sigmoid* in this paper, is the most often used thresholding or activation function in multilayer neural network simulations based on the backpropagation learning rule [Rumelhart-86]. The most interesting part of this function is centered about the y-axis for which the argument (of neural inputs) can assume both positive and negative values and can generate only positive outputs, $S(x)$, in a range from 0 to 1. However, in an incoherent optical processor based NN, all variables (inputs and outputs) can only be represented by all-positive light intensities. This constraint requires the interesting part of an optical thresholding function to be in the all-positive quadrant, which can be obtained by translating a standard sigmoid along the x-axis by an amount δ that is represented by $S(x) = \frac{1}{1+e^{-x+\delta}}$. The shift, δ, could be determined by the range of NN inputs for the problem at hand, akin to dynamic or input-dependent thresholding [Jang-88] in all-positive neural networks.

To allow a comparison with what real devices have to offer, a generic sigmoid, translated along the x- and y-axes with respect to the standard sigmoid by amounts δ and α respectively, and where both the range, γ ($= S_{max}(x) - S_{min}(x)$), and slope at the inflection point can have any value, is given as:

$$S(x) = \alpha + \frac{\gamma}{1 + e^{-\beta x + \delta}}. \tag{1}$$

Table 1 provides a comparison of the characteristics of this generic sigmoid with those of the standard sigmoid. Note that the *shift* of a generic sigmoid is defined as the x-coordinate of its inflection point, and depends on the parameters δ and β.

To satisfy the all-positive nature of thresholding in an incoherent optical system, the response of an optical device should approximate a generic sigmoid satisfying the following conditions:
(i) $S_{min}(x)$ should be non-negative, that is, for all $x \geq 0$, the corresponding $S(x) \geq 0$, and, (ii) the magnitude of the difference between the y-intercept and α is small, typically within a few percent of γ.

For sigmoids having unity range γ, that are most commonly used in neural network simulations, the slope at the inflection point differs only by a numerical constant from β, which is usually taken as the *steepness* which plays

[2] When light of unit intensity transmits through a material, a fraction of it is transmitted corresponding to the product of the incident intensity with the transmittance of the material.

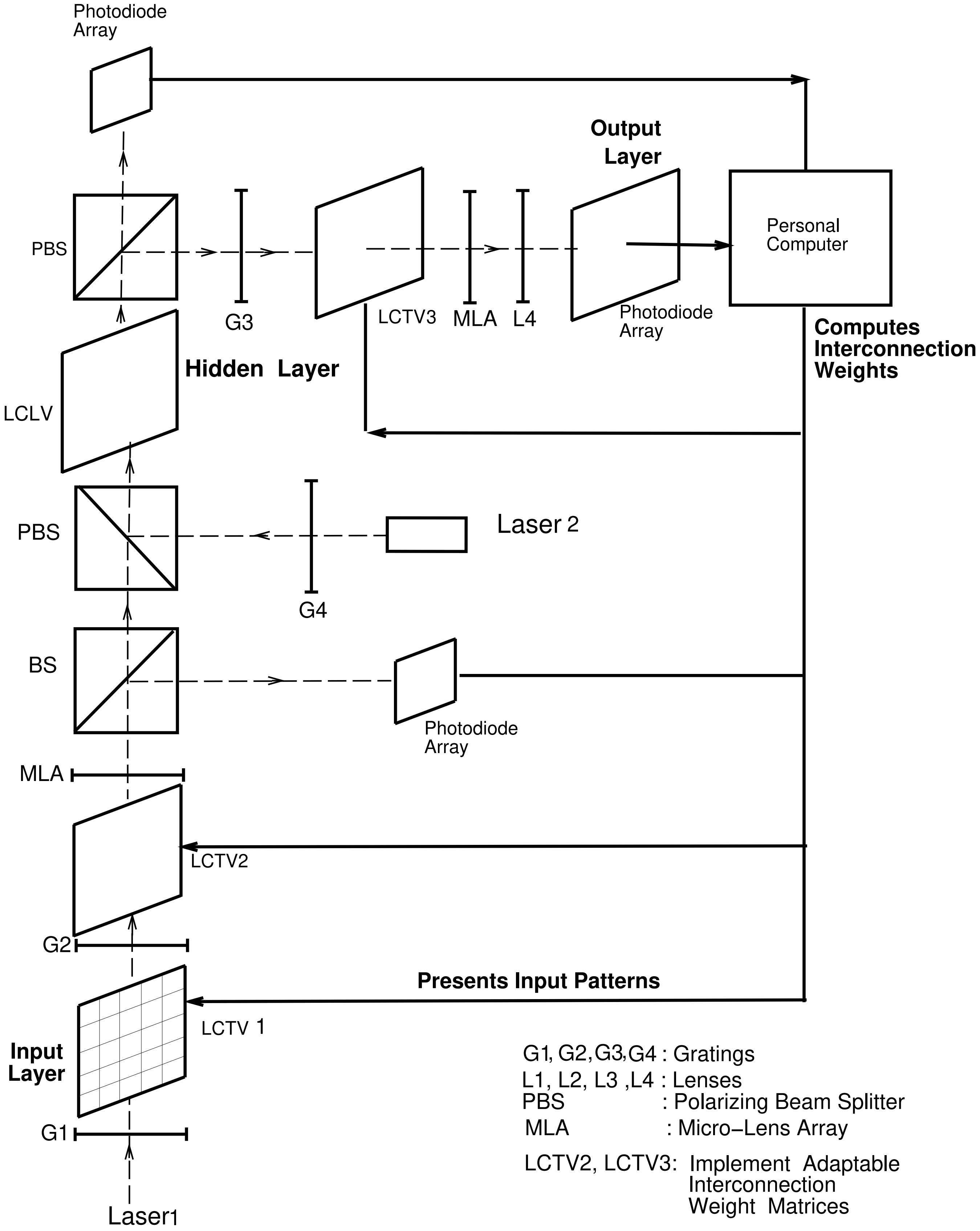

Figure 1: Schematic of the Multilayer Optical Neural Network.

Table 2: Sigmoid curve fit parameters for LCLV transfer curves

LCLV #	α	$\alpha' = \frac{\alpha}{\alpha+\gamma}$	β	γ	$\gamma' = \frac{\gamma}{\alpha+\gamma}$	δ
1	-28.2	-0.41	8.7×10^{-2}	97.7	1.41	0.93
2	-82.7	-2.34	6.2×10^{-3}	118.0	3.34	-0.82
3	0.46	0.015	4.3×10^{-2}	31.0	1.0	3.20
4	-0.43	-0.019	1.40	23.1	1.0	4.57

Table 3: Properties of scaled sigmoid-fits of LCLV response curves.

Properties	LCLV1	LCLV2	LCLV3	LCLV4
Y-axis intercept	-0.01	-0.024	0.05	-0.01
Inflection Point (x,y)	(10.79, 0.3)	(-132.71, -0.67)	(75.26, 0.51)	(3.26, 0.49)
Slope at inflection point	0.03	0.005	0.01	0.36
Steepness	0.123	0.021	0.042	1.4
$S_{max}(x)$	1.0	1.0	1.0	1.0
$S_{min}(x)$	-0.41	-2.34	0.015	-0.019

an important role in the learning rule (see section 5). However, the steepness of a generic sigmoid, (equation 1), should be explicitly defined to encompass the vertical scaling factor γ, thereby rendering it invariant to vertical (and horizontal) scaling, as follows:

$$Sigmoid\ Steepness = \gamma\beta. \tag{2}$$

This definition also holds for the hyperbolic function $tanh(\beta x)$, (which is sometimes used instead of the standard sigmoid as the activation function), whose slope at the origin is β, and whose range (distance between asymptotes) is 2. If its range is multiplied by a factor γ, its slope becomes $\gamma\beta$.

4 LCLV thresholding

In addition to offering benefits of optical over electronic thresholding, LCLVs enable MONNs to be scalable in size, offering very high numbers of effective pixels [Bahadur-92], as compared to devices for electronic thresholding which are hardware limited by the number of pixels, as in smart SLMs (spatial light modulators) or photodetector arrays.

Response curves or transfer characteristics for three LCLVs were obtained by Xue [Xue-94]. In order to find their useful functional approximations for NN implementation, they were fitted to a generic sigmoid (equation 1). The three transfer characteristics and their generic sigmoid fits [Xue-94] are shown in figures 2 to 4, where the y-coordinate represents the actual optical read-out efficiency. In general, these transfer characteristics are truncated and asymmetric in comparison to generic sigmoids. The corresponding curve-fit parameters, both with optical coordinates, and with scaled y-coordinates, are given in table 2. The scaling, obtained by dividing α and γ by $(\alpha+\gamma)$, normalizes $S_{max}(x)$ to unity, and provides a comparison with the standard sigmoid in addition to facilitating their integration into NN simulations. Characteristic properties of the sigmoid fits are given in table 3, together with those of another LCLV, LCLV4 [Micro-Optics]. The response curve of LCLV4 and its sigmoid fit is shown in fig 5. The useful parts of these curve-fits are the truncated parts for which $x \geq 0$ and $y \geq 0$. These truncated parts closely approximate the actual LCLV transfer characteristics, as can be seen in figures 2 to 5. It should be noted, however, that the y-coordinate of the lower asymptote, $S_{min}(x)$, of the LCLV2 curve-fit is decidedly negative due to the asymmetry of its transfer characteristics (see table 3).

The response curve of a given LCLV varies significantly with electrical and optical operating parameters like the frequency and magnitude of applied voltages and the wavelength of the write light [Xue-94, Davis-91]. Desired response curves can theoretically be generated by combining various available response curves [Caulfield-91], at the cost of increased hardware and speed. Alternatively, the operating parameters may be tuned to obtain a more desirable transfer characteristic to optimize system performance. Transfer curves at a write light wavelength of 480nm were obtained for the higher write sensitivity LCLV4 by varying the applied voltage. Hence, it is noted that while the transfer characteristic of LCLV4 at a lower voltage could not be fit properly to a generic sigmoid, a good sigmoid fit could be obtained at a higher voltage of 15 volts.

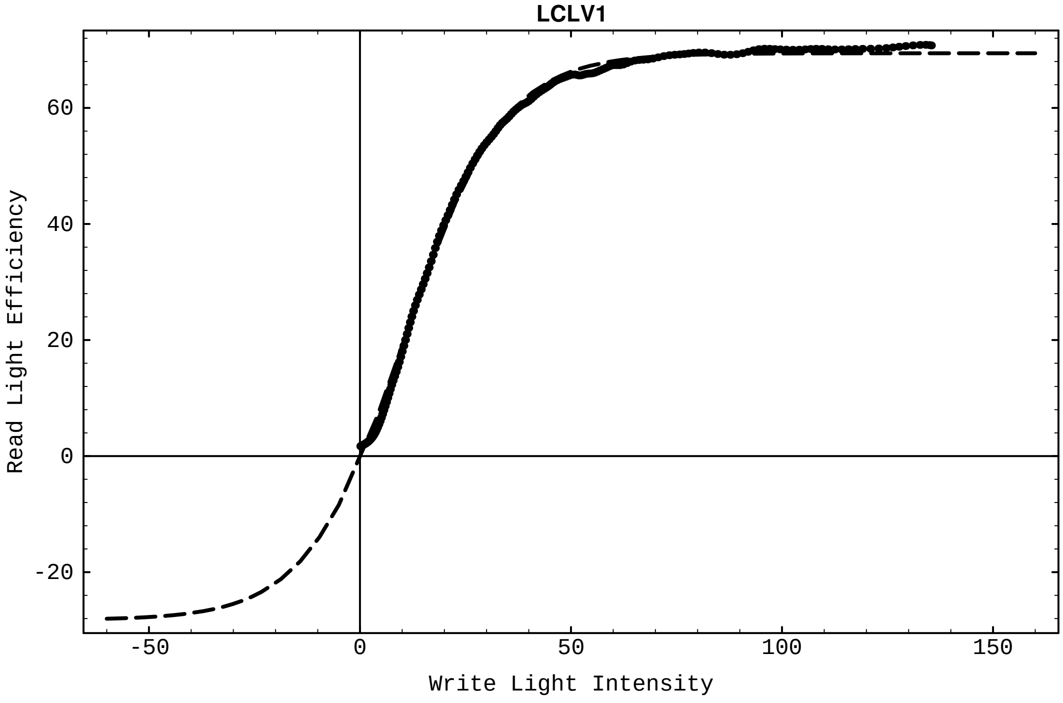

Figure 2: Transfer Curve of LCLV1 and curve fit.

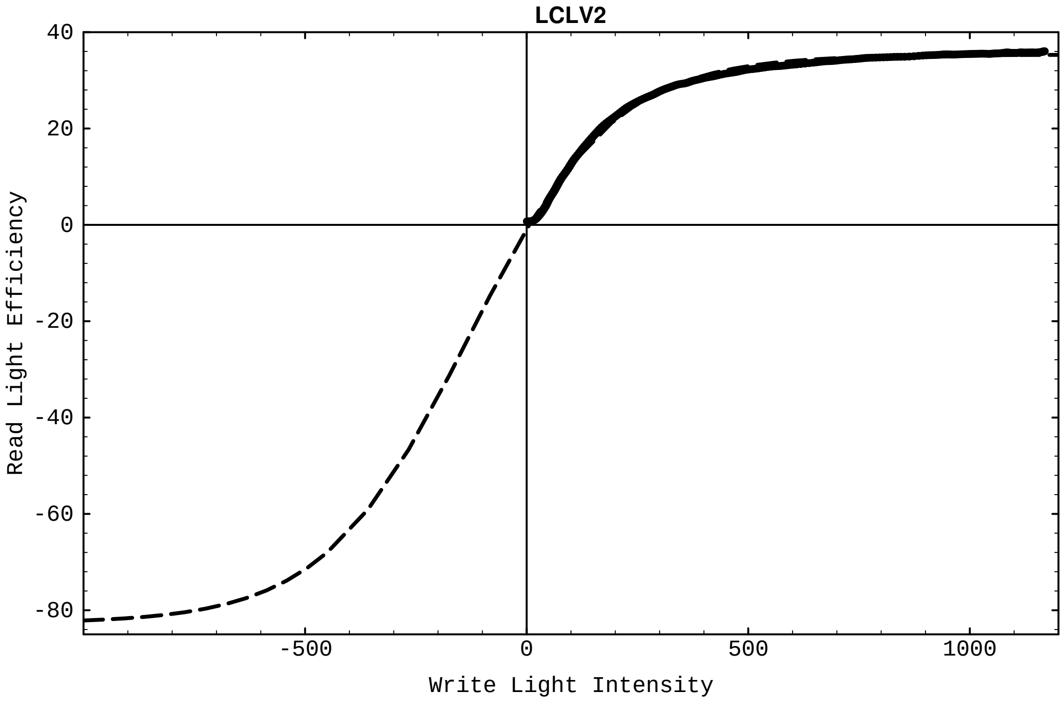

Figure 3: Transfer Curve of LCLV2 and curve fit.

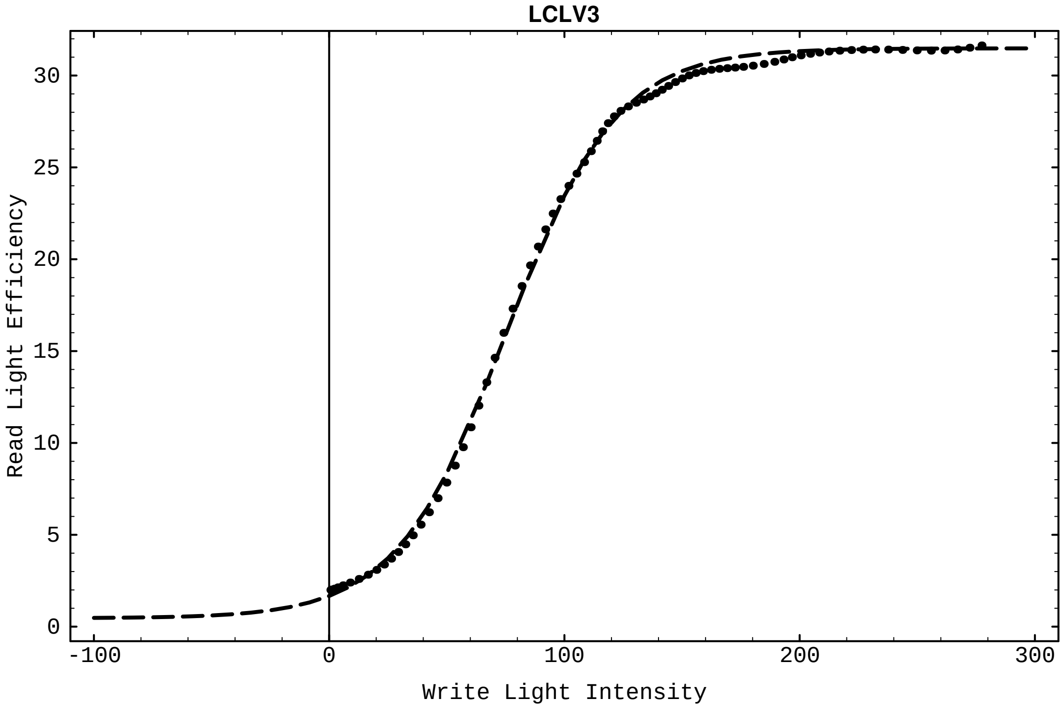

Figure 4: Transfer Curve of LCLV3 with curve fit.

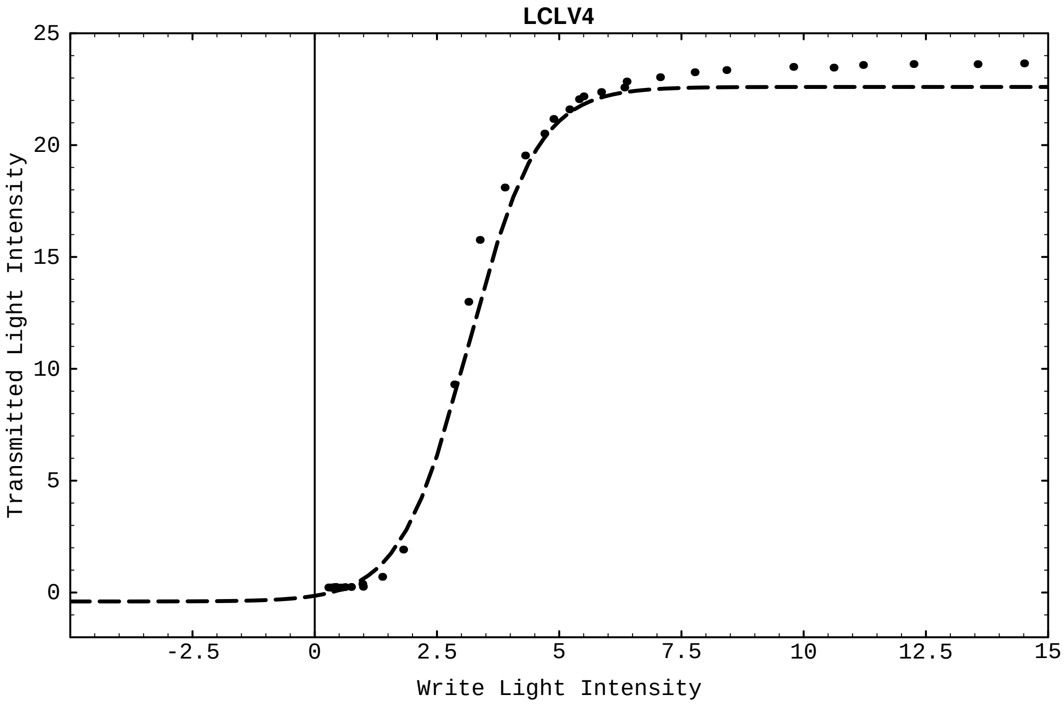

Figure 5: Transfer Curve of LCLV4 with curve fit.

An important aspect concerning the suitability of an LCLV as an optical thresholding device for an MONN is its write sensitivity which is defined as the light intensity required to switch it on. The transmissive LCLV4, operating at a wavelength of 480nm, has the lowest intensity requirement, and hence, the highest write sensitivity of $0.1\mu W/cm^2$. The minimum gain requirement [Psaltis-88] at each neural layer is $\sqrt{N}/2$, for incoherent optical systems, (N being the number of neurons). A layer with 256 neurons would then require a gain of approximately 8, which can easily be achieved with a read-out beam having $0.8\mu W/cm^2$ for LCLV4. With the second highest write sensitivity of LCLV1, of $45\mu W/cm^2$ at 555nm, the read-out light must have $360\mu W/cm^2$ over the entire area of the LCLV. For even less sensitive LCLVs [Xue-94] the read-out beam power requirement increases to more than one mW/cm^2 for LCLV3, and more than $4mW/cm^2$ for LCLV2; both at 515 nm.

5 Simulation Results

In order to assist in the selection of the most appropriate LCLV for optical thresholding, an extensive set of computer simulations was performed in which the LCLV transfer curves are represented by the sigmoid functions obtained from their curve fits.

The training of the optical neural network is performed by the weight discretization and update algorithm described in [Fiesler-94.2]. Besides a functional representation of the transfer curve itself, in this case the LCLV curve fit, the information needed for incorporating it into this training procedure are, its derivative, its steepness, and its shift. The simulated neural networks are fully interlayer connected and initialized with small random weights.

Four well-known benchmark data sets, exclusive OR, parity, addition, and the two output sonar problem, were used for training; see [Fiesler-94.2] for details concerning these benchmarks. The first stage of each simulation consists in pre-training the NN with continuous weights using the backpropagation learning rule. The convergence of this pre-training is a useful criterion for determining whether a given transfer function is suitable, since this is an important indicator for successful training of the network with discrete weights.

The results of a complete simulation were labeled *good* if the network was able to learn all the benchmarks with a maximum error of ten percent, using a maximum of twenty neurons in the hidden layer, and less than ten weight discretization levels. As a point of reference, a set of simulations has been performed using the standard sigmoid under conditions identical to those to be used for LCLV curve-fits, whose results were good for all the four benchmarks [Fiesler-94.2].

The first step towards adapting the training algorithm to enable the use of LCLV threshold curve fits was translating the standard sigmoid four units along the positive x-axis, such that the relevant part of the sigmoid was shifted into the all-positive quadrant. It was subsequently truncated at the y-axis to render it an all-positive sigmoid with a small y-intercept of about 0.018. It was possible to adapt the learning rule in such a way that it compensated for this translated and truncated sigmoid with only a negligible loss of performance.

Next, this adapted learning rule was used to train the neural network using each of the four LCLV sigmoid-fits. For each set of simulations with a given benchmark data set, the same learning rate, weight discretization step size, and initial weight range were used for all four sigmoids.

The results of the pre-training using the sigmoid fits for LCLV1, LCLV2, and LCLV3 did not converge and those of the complete simulation were also not satisfactory. Only the results of the simulations based on LCLV4, whose transfer characteristic is almost symmetric and closely approximated by its curve fit, were good.

It was then noted that the steepness of LCLV4 was closest to the standard value of one (see table 3) and that the steepness could therefore play an important role in the simulations. Analysis of a series of subsequent experiments yielded the following heuristic:

> *Divide the neural network learning rate, the weight discretization step size, and the initial weight range by the parameter β of the activation function.*

Using this heuristic, the simulation results for all LCLVs improved considerably, especially those for LCLV1 and LCLV3. The results based on LCLV1 and LCLV3 are now good for three out of the four benchmarks, failing only on the sonar problem.

6 Summary and Conclusions

A large and application independent adaptive MONN, based on the LCLV as a thresholding device and using available opto-electronic hardware, is presented. Sigmoidal thresholding in MONNs has been discussed and practical LCLVs for performing the thresholding have been evaluated based on sigmoid-fit approximations to their response curves. Translated and truncated sigmoids have been successfully incorporated as activation functions in our weight discretization learning rule for training multilayer neural networks. With this adapted learning rule, in which sigmoid fits have been used as thresholding functions, a large number of simulations have been performed to enable the selection of a useful optical thresholding LCLV. Simulations show that this adapted learning rule performs well with truncated generic sigmoids like the curve fit of LCLV4.

The adapted learning rule has been further extended by a heuristic to improve its capability of handling activation functions with a non-standard steepness, like the LCLV1 and LCLV3 transfer curve fits. Based on these simulations, the LCLVs can be arranged in order of suitability for optical thresholding, as follows. The LCLV4 response at 15 volts is the most desirable because of its similarity to a shifted standard sigmoid, and its high write sensitivity for providing sufficient gain. Good simulation results are obtained using its curve fit for all four benchmark problems. LCLV3 and LCLV1 are both a close second, with good results for three of the four benchmarks, with LCLV1 having an additional asset of its high write sensitivity at 555nm. The highly asymmetric LCLV2 response curve did not yield a desirable sigmoid fit for optical thresholding in NN simulations, which does not perform well on the benchmarks.

7 Acknowledgements

We would like to thank Neil Collings for his contribution towards the optical design, Perry Moerland and Jeroen van Valburg for their neural network simulations, and Wei Xue and Christoph Berger for their assistance concerning the data-acquisition and curve-fitting software. We appreciate the overall support from Daniel Osherson and René Dandliker.

This work is partially supported by the 'Fonds national suisse de la recherche scientifique', project number 21-36497.92.

References

[Rumelhart-86] David E. Rumelhart, James L. McClelland, and the PDP Research Group, 'Parallel Distributed Processing: Explorations in the Microstructure of Cognition', Volume 1: Foundations, The MIT Press, Cambridge, Massachusetts, 1986, ISBN: 0-262-18120-7.

[Jang-93] Jang, Shin, Yuk, Shin and Lee, 'Dynamic Optical Interconnections Using Holographic Lenslet Arrays for Adaptive Neural Networks', *Optical Engineering*, Vol. 32, No. 1, p. 80–87, January 1993.

[Yu-90] F. T. S. Yu, T. Lu, and X. Yang, and D. A. Gregory, 'Optical Neural Network with Pocket-Sized Liquid-Crystal Televisions', *Optics Letters*, Vol. 15, No. 15, pp. 863–865, August 1990.

[Kranzdorf-89] M. Kranzdorf, B. J. Bigner, L. Zhang, and K. M. Johnson, 'Optical Connectionist Machine with Polarization-Based Bipolar Weight Values', *Optical Engineering*, Vol. 28, No. 8, pp. 844–848, August 1989.

[Ohta-90] J. Ohta et al., 'Optical Neurochip Based on a Three-layered Feed-Forward Model', *Optics Letters*, Vol. 15, 1362–1364, 1990.

[Yuk-93] Seong-Won Yuk, Hyuek-Jae Lee, Soo-Young Lee, and Sang-Yung Shin, 'Optical Neural Networks Based on Error Back Propagation Learning for Hetero-Association of Two-Dimensional Patterns', *International Journal of Optical Computing*, Vol. 2, pp. 397–407, 1991.

[Kasama-90] N. Kasama, Y. Hayasaki, T. Yatagai, M. Mori, and S. Ishihara, 'Experimental Demonstration of Optical three Layer Neural Network', *Japanese Journal of Applied Physics*, Vol. 29, No. 8, p. L1565–L1568, August 1990.

[Collings-94] Both LCTV1 and LCTV2, having 440×480 pixels, are from a VPJ-200 Seiko Epson video projector, see: N. Collings, 'Design Considerations for a Useful Two-Layer Neural Network', presented at *Euro-American Workshop on Optical Pattern Recognition* at La Rochelle, France, June 1994.

[Weible-90] K. J. Weible, G. Pedrini, W. Xue, and R. Thalmann, 'Optical Implementation of a Neural Network Associative Memory Using Diffraction Gratings', *Japanese Journal of Applied Physics*, Vol. 29, No. 7, pp. L1301–L1303, July 1990.

[Weible-92] K. J. Weible, N. Collings, and A. Pourzand, 'Initial Results of a Fully Interconnected Neural Network with Modifiable Interconnects', $8^{th.}$ *Workshop on Optics in Computing*, Paris, $8^{th.}$–$9^{th.}$ September 1992, European Optical Society: Topical Meetings Digest series, Vol. 1, No. 2, pp. 157–157, 1992.

[Jang-88] Ju-Seog Jang, Su-Won Jung, Soo-Young Lee, and Sang-Yung Shin. 'Optical Implementation of the Hopfield Model for Two-Dimensional Associative Memory.' *Optics Letters*, Vol. 13, No. 3, pages 248–250, March 1988.

[Bahadur-92] Birendra Bahadur (editor), 'Liquid Crystals Applications and Uses', Vol. 3, World Scientific Publishing Co., 1992, Chapter on 'Applications of Liquid Crystals in Optical Computing' by N. A. Clark and K. M. Johnson.

[Xue-94] Wei Xue, 'Characterization of Liquid Crystal Light Valves for Neural Network Applications', Ph.D. dissertation at the University of Neuchâtel, Switzerland, 1994.
Also in W. Xue, N. Collings, and K. J. Weible, 'The Characteristics of Three Kinds of LCLVs: Measurement and Comparison', IMT Report, January 1992.
And in, N. Collings and W. Xue, 'Liquid-Crystal Light Valves as Thresholding Elements in Neural Networks: Basic Device Requirements', *Applied Optics*, Vol. 33, No. 14, p. 2829–2833, May 1994.

[Micro-Optics] Micro-Optics Technologies, 8608 University Green # 5, Middleton, WI 53562, U.S.A.

[Davis-91] Jeffrey A. Davis, 'Binary Operation of a Liquid Crystal Light Valve', *Applied Optics*, Vol. 30, No. 3, p. 267, January 1991.

[Caulfield-91] H. John Caulfield, Janine Reardon, and Bahram Javidi, 'Simulating Arbitrary Response Curves with Available Response Curves for SLMs', *Proceedings of the SPIE*, Vol. 1562, p. 103–105, 1991.

[Psaltis-88] Demetri Psaltis, David Brady, and Kelvin Wagner, 'Adaptive Optical Networks Using Photorefractive Crystals', *Applied Optics*, Vol. 27, No. 9, May 1988.

[Fiesler-94.1] E. Fiesler, 'Neural Network Classification and Formalization', *Computer Standards & Interfaces*, Vol. 16, No. 3, special issue on Neural Network Standards, John Fulcher (editor), pp. 231–239, North-Holland/Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 1994, ISSN: 0920-5489.

[Fiesler-94.2] E. Fiesler, A. Choudry, and H. J. Caulfield, 'A Universal Weight Discretization Method for Multi-Layer Neural Networks', *IEEE Transactions on Systems, Man, and Cybernetics (IEEE-SMC)*, (Accepted for publication). See also: E. Fiesler, A. Choudry, and H. J. Caulfield,' A Weight Discretization paradigm for Optical Neural Networks', in *Proceedings of the International Congress on Optical Science and Engineering*, volume SPIE 1281, pp. 164–173, SPIE, Bellingham, Washington, 1990, ISBN: 0-8194-0328-8.