Journey From Optical Neural Networks To Photonic Chips

Neha Soni, Enakshi Khular Sharma, Amita Kapoor

Abstract: In recent years, there has been a rapid expansion in two fields, photonics and artificial neural networks (ANNs). ANNs based on the basic property of a biological neuron, has become the solution for a wide variety of problems in many fields, such as prediction, modeling, control, recognition, etc. and many of them have reached to the hardware implementation phase. Photonics on the other hand, with several advantageous features like inherent parallelism, high speed of information processing (photon), high capacity data storage, etc. has become a natural choice for researchers for the implementation of ANNs. This combination of photonics and ANNs has resulted in novel realizations of various ANN models. In this paper, we attempt to survey the optical realizations of various neural network models made in last the 30 years. We focus on self-organizing neural networks, associative memories, and perceptron neural networks. We also survey the state-of-the-art photonic chips for the realization of ANNs.

Index Terms:Feed-Forward Networks, Hopfield Neural Networks, Neuromorphic Chips, Optical Neural Networks, Self-Learning Algorithms.

1 INTRODUCTION

An optical neural network (ONN) is a physical implementation of an artificial neural network (ANN) with optical components like laser, lens, light modulators, mirrors, liquid crystals etc. In 1987, Mostafa and Psaltis [1] for the very first time focused on the need for optical neural computers. They believed that due to bandwidth limitation, using only electrical means, artificial intelligence tasks like pattern recognition cannot be duplicated. They proposed to arrange optical elements in the way as neurons are arranged in the human brain. Psaltis [1], in 1990 with his team contributed remarkable analogies between neural network models and simple holography, which made it easy to construct large densely connected multilayer network in the laboratory. Since then, optical implementation of neural networks has widened resulting in novel optical realizations of various neural networks. Photonics have proved to be promising candidature for ANN implementation. ANNs, unlike conventional computing systems, are based on the basic property of a biological neuron and perform in two stages: weighted summation of inputs and a nonlinear activation function. An artificial neuron has n inputs $I_1, I_2, I_3, \dots I_n$ each connected with weights $W_1, W_2, W_3, \dots W_n$ respectively and an output y. Artificial neuron can be described mathematically using these equations:

$$h = \sum_{i=1}^{N} (I_i * W_i) \qquad (1)$$

$$y = g(h) \qquad (2)$$

In its nascent years, ANNs have been implemented using optical devices. With the success of ANNs in intelligent tasks like pattern recognition, voice recognition etc. [2], [4], [5], [6], [7], [8], researchers were also motivated to implement ANNs in integrated chips (ICs) commonly called as photonic and neuromorphic chips. Most of the neuromorphic chips as of now are realized using the technologies of fabrication of electronic integrated circuits. But with this approach it is difficult to construct neural networks with large neural node density [1], [9], [10], [11], [12].

- *Neha Soni, Dept. of Electronic Science, University of Delhi, India. Email: soni.neha2191@gmail.com.*
- *Enakshi Khular Sharma, Dept.of Electronic Science, University of Delhi, India.*
- *Amita Kapoor, Dept. of Electronics, SRCASW, University of Delhi, Delhi, India.*

In contrast, photonics has numerous benefits for surpassing the constraints of electronic realizations. It has inherent parallelism and uses photons to transport information instead of electric signals. In optical systems, holograms can be used for storage. Holograms are capable of storing a high amount of information and establishing three dimensional massive interconnections [1], [11], [13]. This is the reason why the researchers from photonics and ANNs are combining the two resulting in ONNs and photonic chips.

2 OUR CONTRIBUTION

In this paper, we make the following contributions: 1. First, we have reported significant methods for optical matrix-matrix multiplication [14], [15], [16]. We choose to explore matrix multiplication optically as it is one of the most important mathematical operations in neural networks. 2. Report the success of ONNs for the realization of associative memories [12], [19], [20], [20], [22] self organizing neural networks [13], [17], [18], [24] and perceptron neural networks [17], [18], [22]. 3. The state-of-the-art photonic chips [25], [26] for the realization of ANNs. 4. Challenges and opportunities in making photonic chips.

3 OPTICAL MATRIX-MATRIX MULTIPLICATION

Matrix multiplication is one of their most significant operations in neural networks. In this paper, we have reported some of the successful real time optical realizations of matrix-matrix multiplication [14], [15], [16]. Consider A and B as two matrices; C as their product. This can be written as: $AB = C$ If A and B are considered as 2x2 matrices, the above equation can be represented in terms of their elements as:

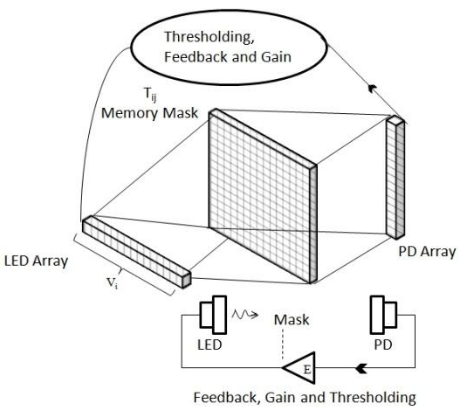
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$ where

$$c_{11} = a_{11}b_{11} + a_{12}b_{21}$$
$$c_{12} = a_{11}b_{12} + a_{12}b_{22}$$
$$c_{21} = a_{21}b_{11} + a_{22}b_{21}$$
$$c_{22} = a_{21}b_{12} + a_{22}b_{22}$$

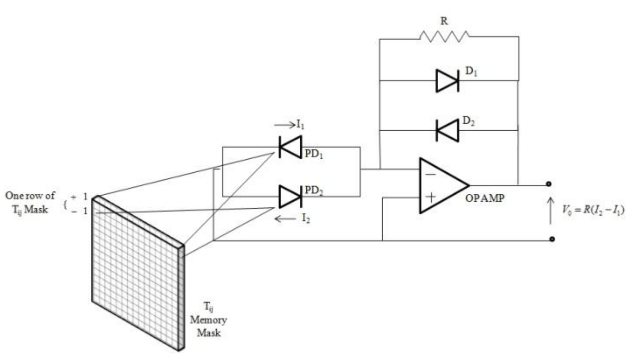
To obtain all four elements of C simultaneously, each value of A and B are needed two times. In real time, matrix-matrix multiplication is difficult than vector matrix multiplication because of the need of reusable modulators, detector arrays and space variant optical element. In 1984, Liang and Liu demonstrated a method of a multi-focus dichromated-gelatin holographic lens as a space-variant optical element. They performed a 3X3 matrix multiplication and compared the theoretical and experimental results. Their results show the maximum percentage error as 0.9%. The method was proved very useful and can be extended to large dimensions [16].

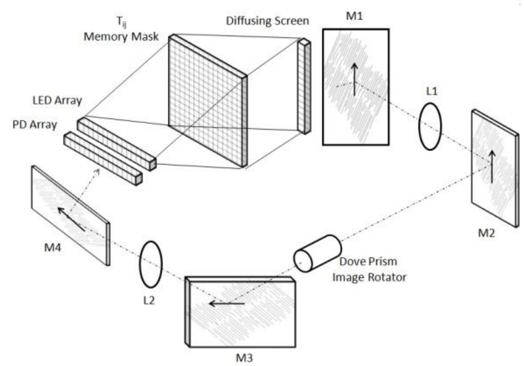
4 ASSOCIATIVE MEMORIES

Neural associative memories (NAM) have its origin from human memory, if we allow our mind to stroll, it moves from one topic to another based on a chain of mental links. Humans, use this associative ability to recollect a past memory. In NAMs, the association can be of two forms: auto-association and hetero-association. In an auto-associative neural network like HNN, a set of patterns (vectors) is stored, the task is to retrieve a particular pattern from stored patterns whose distorted (noisy) version is presented to the network. A hetero-associative neural network like bi-directional associative memory (BAM), pairs a set of arbitrary input patterns with another arbitrary set of output patterns [17], [18]. In this paper, we explore the optical implementation of HNN because of its following properties. One, the network doesn't require synchronism i.e. it works efficiently whether one element of the distorted vector is fed at a time or all of its elements are fed together. However, the later is preferable. Two, the network is fault tolerant i.e. quite insensitive to noise, distortions and even imperfections in the input vector and weight matrix. Fig 1 illustrates the optical implementation of Hopfield neural network.



(a)





(c)

Fig 1 *Optical implementation of Hopfield Neural network (a) Matrix-vector multiplier incorporating non-linear feedback. (b) Realization of binary bipolar memory mask in incoherent light. (c) Optical feedback*

One of the implementations of Hopfield model realized optically [19] by Farhat et al, used the state (on or off) of an array of light emitting diodes (LEDs) to represent the input vector and an array of photodiodes (PDs) to detect the output vector (fig 1(a)). Global interconnection between the elements is realized through an optical vector-matrix multiplier and nonlinear feedback. Horizontal imaging and vertical smearing (using an anamorphic lens system) of the input vector (displayed by the LEDs on the plane of the interconnection mask) perform the multiplication of the input vector by the weight matrix. One more anamorphic lens system is used to collect the light emerging from each row of the interconnection mask on individual photo sites of the PD array. The bipolar interconnection matrix is realized in incoherent light by dividing its each row into two sub rows one for positive and one for negative values (fig 1(b)). The light emerged from each sub row is focused on two adjacent photo sites of the PD array that are electrically connected in opposite direction and the feedback is achieved using electronic wiring. This optical feedback can preferably be replaced by electronic feedback (fig 1(c)). This simple optical system can be used for many applications. Farhat et al simulated a network of 32 neurons with a variation [9] in above optical system. An array of 32 LEDs and two multichannel silicon PD arrays, each consisting of thirty-two elements was constructed for three binary state vectors (1, 2, 3) with mean hamming distance of 16 between them. In the experiment performed, a binary photographic transparency (32 × 64) was generated computationally from the weight matrix (obtained with the technique discussed above) to assign the positive values in any given row of weight matrix to transparent pixels in one sub row of the mask and negative values to transparent pixels in the adjacent sub row. Also, they split the memory mask in two halves to ensure that the image of the input LED array is uniformly smeared over the memory mask and use the resulting sub masks in two identical optical arms. The size of the sub rows of the memory sub masks was made exactly equal to the element size of the PD arrays in the vertical direction which were placed in register against the masks. They used electronic amplification and thresholding box which contain all electronic circuits (amplifiers, thresholding comparators, LED drivers, etc.) in the thirty-two parallel feedback channels to make the system compact. Once an input vector is selected it appears displayed on the composing box and on the input LED box simultaneously. A single switch is then thrown to release the system into operation with the composed vector as the initializing vector. The final state of the system, the output, appears after a few iterations, is displayed on the input LED array and the display box simultaneously. The evaluation of the performance of the system is made by making variation in hamming distance, they concluded that optically implemented Hopfield network is working as accurately as the digital simulations. In 1989, Shariv and Friesem [20] realized HNN optically with only inhibitory connections. They modified the interconnection matrix by discarding (set to zero) all the positive weight matrix elements to obtain a positive unipolar interconnection matrix that can be implemented by optical means only. As per their computer simulation results for 150 neurons and 15 stored states (repeated 50 times for each of the stored states) the performance of the network remains the same with this modification in weight matrix. They build an all optical Hopfield neural network with this modified weight matrix that supports two stable states – the numerals 4 and 6 (4 × 4) which could be retrieved from noisy optical inputs. This is achieved by using read side of liquid crystal light valve (LCLV) to incident 16 (4 × 4 numerals, 4 and 6) equally intense plane waves (1 mm in diameter and 2 mm apart) derived from an argon laser (input for the network). The output is represented by back reflected plane waves from the LCLV and global interconnections are formed using 3-D holographic interconnection method. Light is dispersed in different directions when every output (reflected plane waves) is fed back to illuminate its corresponding sub hologram. The diffracted intensities from all the holograms are incoherently summed at an array of 4 × 4 circles (each 1 mm in diameter) on the write side of the LCLV. These circles are located precisely across from the 16 plane waves incident on the read side and therefore determined the local reflectivity (input light intensity) of the LCLV at each of these points. This all optical implementation demonstrated associativity and robustness. In 1989, Ohta and his team [12] fabricated a device consisting of a light-emitting-diode array, binary-valued interconnection matrix made of Cr/Au, and a photodiode array, integrated into a hybrid-layered structure on a GaAs substrate that simulates a 32-neuron system. As per our knowledge, this optical neurochip for synaptic interconnections is reported for the first time by this group, all other optical neural network systems have been constructed by discrete devices. They took English alphabets A, J and E (number of patterns, m=3; number of neurons, i=32) as input patterns with a mean hamming distance 14. To perform the vector matrix multiplication, they used two optical synaptic chips for positive and negative elements of the weight matrix, whose outputs were further processed by differential amplifiers and comparators. They performed recognition of English alphabets and demonstrated that the recognition rate is a function of hamming distance. This optically synaptic interconnection chip demonstrated an associative memory based on the Hopfield model.

5 PERCEPTRON NEURAL NETWORKS

Perceptron relies on a straightforward structure of weighting and thresholding in single or multiple layers. The single layer perceptron is a simple model capable of exhibiting the features important for neural classification system. Multilayer perceptrons (MLPs) are generalization of single layer perceptron that constitute the input layer, one or more hidden layers and an output layer. MLP algorithm based on error correction learning rule has been applied successfully to solve problems like pattern classification, function approximation etc [17], [18]. In 1992, Saxena and Fiesler [12] described an application independent adaptive multilayer ONN (MONN) based on LCLV as thresholding device. The 3-layer MONN described has 256 input neurons (presented by liquid crystal television (LCTV1) through fanned out argon-ion laser beam (480 nm)), 256 hidden layer neurons and 16 output neurons. The 16×16 input array is replicated 256 times onto transmissive LCTV2. The interconnection weight matrix (IWM) between layers is represented by the transmission values of the pixels of the LCTVs. The vector matrix product of the inputs and weights is presented by the integration of intensity of light transmitted by LCTV2 and micro lens array. LCLV then threshold this product after biasing for the production of activation values for hidden layers to be read by different light source. LCTV3 represents the 256×16 IWM between the hidden and output layer. The final outputs are detected by 4×4 photo detector array. Authors have discussed the sigmoidal thresholding in MONNs and evaluated practical LCLVs for performing thresholding.

6 SELF-ORGANIZING NEURAL NETWORKS

Self-organized maps are motivated by how visual, auditory or other sensory information is handled in separate parts of the cerebral cortex in the human brain. The most popular self-organizing algorithm was given by Kohonen in 1984 [13] with similar inputs mapped to nearby nodes, located in one two-dimensional layer. It finds first the closest matching unit to a training input, and then increases the similarity of that unit and those in the neighboring proximity Then it slowly decreases the neighborhood in size in order to minimize the number of connections between nodes. Lateral excitation and inhibition in the neighborhood proximity is applied using a "Mexican hat"-function. In 1990, Taiwei and Francis [13] demonstrated optical Kohonen self-organizing feature map algorithm capable of organizing a feature map and preserving old memory while learning new knowledge. This adaptive (64, 8×8 interconnected neurons) ONN used 80W xenon arc lamp (an inherent coherent light source), liquid crystal television (LCTV1), diffuser, lenslet array, LCTV2, imaging lens and CCD array. LCTV1 is used for the generation of interconnection weight matrix (IWM), which consists of 8×8 array of sub matrices and each sub matrix has 8×8 elements, which gets displayed on a diffuser placed immediately behind LCTV1. LCTV2 is used as an input device for the generation of the input patterns. The interconnections between the IWM and the input pattern are provided by the lenslet array (8×8 lenses), when each lens of the lenslet array projects each of the IWM sub-matrices onto the input LCTV2 to establish proper interconnections. This represents, CCD array detector picks up the output result from LCTV2. Four 8×8 pixel binary input patterns (a tree, a dog, a house, and an airplane) were sequentially presented at the input LCTV2 for experimental demonstrations. The final result is obtained by a microcomputer. This proposed LCTV optical neural network is capable of performing pattern recognition task.

3 PHOTONIC CHIPS

The above optical implementations of different neural network architectures have become possible with the use of special purpose optical devices. For the real-time ONN implementation, the number of neurons and the complexity of the neural network architecture increases. Therefore, these bulk optical components become a major barrier. Integrated photonic chips seem to be a solution to this problem. Shen et al [25] demonstrated a programmable nano-photonic processor composed of cascaded inferometeres in silicon photonic integrated chip. This has demonstrated its usefulness for vowel recognition. This architecture has high detection rate, sensitivity and speed in comparison to the state-of-the-art electronic computer architectures. These advancements in optical computing prove photonics as the promising candidature for the implementation of deep networks which can replace electronic chips in the future computers. However, photonic chips are complicated to design, more error prone, requires more expertise and only few engineers are trained to design such systems [26]. Moreover, propagation and coupling loss need to continue to improve to allow continued progress towards large scale systems [27]. Researchers are on their progressive path to resolve these issues.

7 CONCLUSION

The most important components of ONNs are interconnection and storage systems (planar and volume holograms, optical disks etc.), thresholding devices (LCLVs, spatial light modulators etc.), and optical processing configurations (inferometer, photorefractive crystal etc.). Numerous possibilities of realizing above components together give rise to various ONN models which have been constructed and reported successfully in last three decades. Researchers are working on the development of neurochips with remarkable features, such as good mechanical stability, compact size, and compatibility with electronic circuits for real time ONNs that can perform complex operations in an easy and attractive way. We can conclude that in last 30 years various interesting and promising ONN systems have been produced and they pave way to building better reliable and more importantly compact photonic chips.

ACKNOWLEDGMENT

One of the authors, Ms. Neha Soni, wants to thank Department of Science & Technology, Ministry of Science & Technology, New Delhi, India for sponsoring this work.

REFERENCES

[1] Y.S. AbuMostafa, and D. Psaltis, "Optical neural computers," Scientific American 256(3), March 1987: pp. 88-95.
[2] D. Psaltis, D. Brady, X. G. Gu, and S. Lin, "Holography in artificial neural networks," Nature 343.6256, January 1990: pp.325-330.
[3] Soni, Neha, et al. "Face recognition using cloud Hopfield neural network." 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2016.
[4] Soni, Neha, et al. "Low-Resolution Image Recognition Using Cloud Hopfield Neural Network." Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore, 2018. 39-46.
[5] Soni, Neha, et al. "Impact of Artificial Intelligence on Businesses: from Research, Innovation, Market Deployment to Future Shifts in Business Models." arXiv preprint arXiv:1905.02092 (2019).
[6] Gulli, Antonio, and Amita Kapoor. (2017) "TensorFlow 1. x Deep Learning Cookbook: Over 90 unique recipes to solve artificial-intelligence driven problems with Python" Packt Publishing Ltd.
[7] Jain, Ankit, Armando Fandango, and Amita Kapoor. (2018) "TensorFlow Machine Learning Projects: Build 13 real-world projects with advanced numerical computations using the Python ecosystem." Packt Publishing Ltd.
[8] Kapoor Amita (2019) "Hands-On Artificial Intelligence for IoT: Expert machine learning and deep learning techniques for developing smarter IoT systems." Packt Publishing Ltd.
[9] Soni, Neha, et al. "Success of Optical based Networks for Deep Learning", 2018, Computing For Sustainable Global Development (INDIACom 2018), IEEE.
[10] Soni, Neha, et al "A survey of the existing optical neural networks", 2017, Recent Developments in Electronics (NCRDE 2017).
[11] C. Denz, Optical neural networks. Springer Science & Business Media, 2013.
[12] J. Ohta, M. Takahashi, Y. Nitta, S. Tai, K. Mitsunaga and K. Kyum, "GaAs/AlGaAs optical synaptic interconnection device for neural networks," Optics letters 14, 16 August 1989,pp.844-846.
[13] T.T. Lu, F. Yu, and D.A. Gregory. "Self-organizing optical neural network for unsupervised learning," Optical Engineering 29.9,1990,pp.1107-1113.
[14] Heinz, R. A., J. O. Artman, and S. H. Lee. "Matrix multiplication by optical methods." Applied optics 9.9 (1970): 2161-2168.
[15] Tamura, Poohsan N., and James C. Wyant. "Matrix multiplication using coherent optical techniques." Optical Information Processing: Real Time Devices & Novel Techniques. Vol. 83. International Society for Optics and Photonics, 1977.
[16] Liang, Yin-Zhong, and Hua-Kuang Liu. "Optical matrix–matrix multiplication method demonstrated by the use of a multifocus hololens." Optics letters 9.8 (1984): 322-324.
[17] S. Haykin, Neural Network A comprehensive foundation,Neural Networks2004.
[18] P.D. Wasserman, Neural computing. Van Nostrand Reinhold, New York, 1989.
[19] N. Farhat, and P. Demetri, "New approach to optical information processing based on the Hopfield model (A)," Journal of the Optical Society of America A 1,1984: 1296.
[20] N. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical implementation of the Hopfield model," Applied Optics 24.10, 1985: pp.1469-1475.
[21] I.Shariv, and A. A. Friesem. "All-optical neural network with inhibitory neurons," Optics letters 14.10,1989:pp. 485-487.
[22] S. Jang, S. W. Jung, S.Y. Lee, and S.Y. Shin, "Optical implementation of the Hopfield model for two-dimensional associative memory," Optics letters 13.3, 1988: pp.248-250.

[23] I.Saxena, and E. Fiesler, "Adaptive multilayer optical neural network with optical thresholding," Optical Engineering 34.8, 1995: pp.2435-2440.
[24] T. Kohonen, Self-organization and associative memory. Vol. 8. Springer Science & Business Media, 2012.
[25] Tait, Alexander N., et al. "Neuromorphic Silicon Photonics." arXiv preprint arXiv:1611.02272 (2016).
[26] Tait, Alexander N., et al. "Broadcast and weight: an integrated network for scalable photonic spike processing." Journal of Lightwave Technology 32.21 (2014): 3427-3439.
[27] Tait, Alexander N., et al. "Microring weight banks." IEEE Journal of Selected Topics in Quantum Electronics 22.6 (2016): 312-325.
[28] Shen, Yichen, et al. "Deep learning with coherent nanophotonic circuits." Nature Photonics (2017).