

Optical Architecture for Finite Impulse Response Neural Networks

Paulo E. X. Silveira and Kelvin H. Wagner

Optoelectronic Computing Systems Center
Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309-0525
Phone: (303) 492-7916. Fax:(303) 492-3674. E-mail: paulo@colorado.edu

Abstract

An optical architecture for multi-layer non-linear finite impulse response neural networks is suggested, complete with forward propagation and on-line learning in the form of back-propagation.

Optical Architecture for Finite Impulse Response Neural Networks

Paulo E. X. Silveira and Kelvin H. Wagner

Optoelectronic Computing Systems Center
Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309-0525
Phone: (303) 492-7916. Fax: (303) 492-3674. E-mail: paulo@colorado.edu

1 Introduction

A finite impulse response neural network (FIRNN) is equivalent to a neural network whose interconnects are composed of finite impulse response (FIR) adaptive filters.¹ Some interesting applications of FIRNNs are multi-dimensional time-series prediction, real-time speech recognition² and system classification, an application which consists of using adaptive FIRs to model different systems and a non-linear combination of the FIRs to classify the modeled systems.

FIRNNs can take advantage of the parallelism and capacity of arbitrary global interconnects of optics. The required time delays can be implemented using a scrolling spatial light modulator (SLM) on the input and a space integrating (SI) architecture, or a scrolling and accumulating detector on the output in a time integrating (TI) architecture. In the SI architecture the spatial-temporal input $\underline{x}$ is represented by a two-dimensional plane, where one of its dimensions represents the spatial dimension and the other contains a sliding window of delayed versions of $\underline{x}(k)$. The input plane is then diffracted by a volume hologram representing the three-dimensional weight matrix $\underline{w}$, producing the linear output $\underline{s}$. A rotating photorefractive crystal (PRC) delay line enabled the implementation of a single-layer time-delay neural network with an associative learning algorithm.³ In the TI approach the instantaneous input $\underline{x}(k)$ is diffracted to a spatio-temporal plane by the weights represented by a volume hologram. The output plane undergoes a shift and sum of its rows in the temporal dimensional, performed by a device such as a time-delay and integrate charge coupled device (TDI-CCD). Such an architecture has been proposed in the implementation of multi-layer non-linear time-delay neural networks, but without the implementation of on-line learning.⁴

2 FIRNN analysis

The complete derivation of the relations for forward propagation and error back-propagation equations for FIRNNs is credited to Wan.¹ Forward propagation is given by

$$x_{j,k}^{l+1} = f \left(\sum_{i=1}^{N^l} \sum_{\tau=0}^{T^l} w_{ji,\tau}^l x_{i,k-\tau}^l \right), \quad (1)$$

where $x_{i,k}^l$ is the input to node i at layer l at the time step k , $f(\cdot)$ is the non-linear discriminant function and $w_{ji,\tau}^l$ is the filter weight connecting node i in layer l to node j in layer $l+1$ after τ delays. Notice that equation 1 consists of an array of two-dimensional dot-products followed by a non-linear function.

Weights are update by a delta rule,

$$w_{ji,\tau}^{L-1-n}(k+1) = w_{ji,\tau}^{L-1-n}(k) - \eta \delta_j^{L-n}(k - \Upsilon^n) x_{i,k-\Upsilon^n-t}^{L-1-n}, \quad (2)$$

which consists of an outer-product between a delayed version of the input $x_{i,k-\Upsilon^n-t}^{L-1-n}$ and the gradient term $\delta_j^{L-n}(k - \Upsilon^n)$, given by $\delta_j^{L-n}(k - \Upsilon^n) =$

$$\begin{cases} -2e_j(k) f'(s_j^L(k)) & n = 0 \\ f'(s_j^{L-n}(k - \Upsilon^n)) \sum_{m=1}^{N^{L-n+1}} \sum_{c=0}^{T^{L-n}} w_{mj,c}^{L-n} & 1 \leq n \leq L-1 \\ \times \delta_m^{L-n+1}(k - \Upsilon^n + c) & \end{cases} \quad (3)$$

where L is the total number of layers and the delay $\Upsilon = \sum_{i=0}^{n-1} T^{L-i}$ represents the sum of all the delays from the present layer to the network output.

Notice that the last term of equation 3 is also a two-dimensional dot-product. $e_j(k) = d_j(k) - x_j^L(k)$ is the instantaneous error of node j at time step k , defined as the difference between the desired signal $d_j(k)$ and the output signal $x_j^L(k)$. $f'(\cdot)$ is the derivative of the discriminant function with respect to its input, and $s_j^L(k) = \sum_i \sum_{\tau} w_{ji,\tau}^L x_{i,k-\tau}^L$.

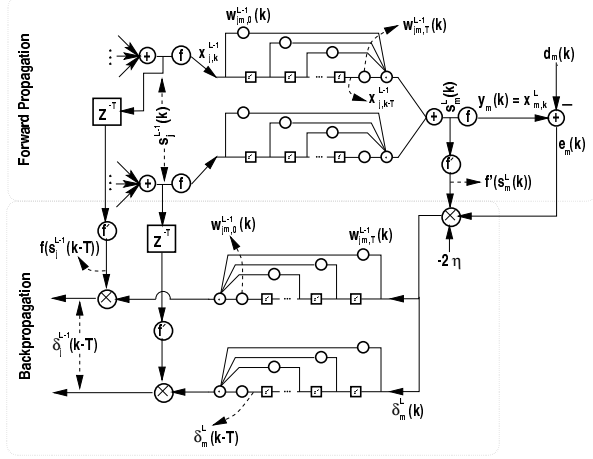


Figure 1: Forward propagation (top) and backward propagation (bottom) in a FIRNN. The squares represent delays and the small circles represent weights. The weights are identical at the top and bottom, but used in reverse order during back-propagation. Notice that signals $s_j^{L-1}(k)$ and $x_{i,k}^{L-1}$ must be stored and delayed, but signal $\delta_j^{L-1}(k - T)$ is already delayed during back-propagation through its respective FIR delay line.

The top half of figure 1 describes forward propagation. The left side of the top portion implements equation 1, and the right side implements the error calculation. Following the signal flow, we see the implementation of equation 3 for $n = 0$, producing the term $-2e_j(k) f'(s_j^L(k))$. Next, back-propagation takes place, when the term $\delta_m^L(k)$ is fed through the FIR interconnects to the previous layer, producing the term $\delta_i^{L-1}(k - T)$ (for simplicity of notation, we are assuming that $T^{L-1} = T$ and, therefore, $\Upsilon^1 = T$).

3 Optical architecture

The complete FIRNN implementation requires the storage of information over Υ^n time steps, the calculation of the non-linear functions $f(\cdot)$ and $f'(\cdot)$, and the evaluation of linear functions (outer-products and dot-products). Combining the input plane and output plane architectures described earlier, it is possible to implement all these operations in one single system, where the input plane is implemented by a rotating volume hologram in the phase conjugate geometry, such as used by Gan,³ but unlike Gan's approach, we use the signal beam perpendicular to

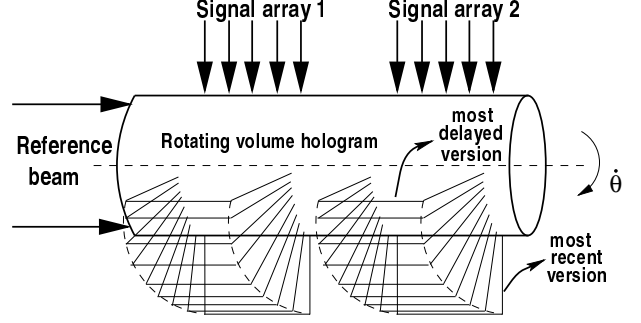


Figure 2: Multiple signal arrays in a perpendicular implementation of a rotating volume hologram. One single device can perform a scrolling SLM for several signal arrays.

the reference beam, as shown in figure 2. The output beam in this case would describe a cylinder with its symmetry center in the rotation axis of the volume hologram and more than one signal beam could be present at a time, where the signals would be placed along the symmetry axis of the volume hologram, as depicted in figure 2. We are currently working on implementing this multiple-scrolling SLM using a photo-polymer cast in this cylindrical shape as the holographic material.

Figure 3 depicts one possible optoelectronic architecture of a two-layer FIRNN capable of performing forward propagation, backward propagation and weight updating. Notice that for the two-layer case we have that $\Upsilon^n = 0$ for the last layer ($n = 0$), and $\Upsilon^n = T$ for the first layer ($n = 1$). This system can be expanded to an arbitrary number of layers.

Starting from the SLM 1, the uni-dimensional input signal $\underline{x}^{L-2}(k)$ is recorded in the rotating hologram 1, producing the two-dimensional array $\underline{\underline{x}}^{L-2}(k)$, which is diffracted by the grating recorded in the volume hologram PRC 1, performing the input plane implementation of forward propagation. The diffracted signal, $\underline{s}^{L-1}(k)$, is detected by the smart-pixel array (SPA) 1, which performs the non-linear functions $f(\cdot)$ and $f'(\cdot)$, producing signals $\underline{x}^{L-1}(k)$ and $f'(\underline{s}^{L-1}(k-T))$, respectively. $\underline{x}^{L-1}(k)$ is recorded in the rotating hologram 2, producing signal $\underline{\underline{x}}^{L-1}(k)$, which is diffracted by the volume hologram PRC 2, producing signal $\underline{s}^L(k)$, which is detected by the SPA 2. This SPA generates the final network output $\underline{x}^L(k)$ and uses the electronic input $\underline{d}(k)$ (the desired signal) to generate the weight update term

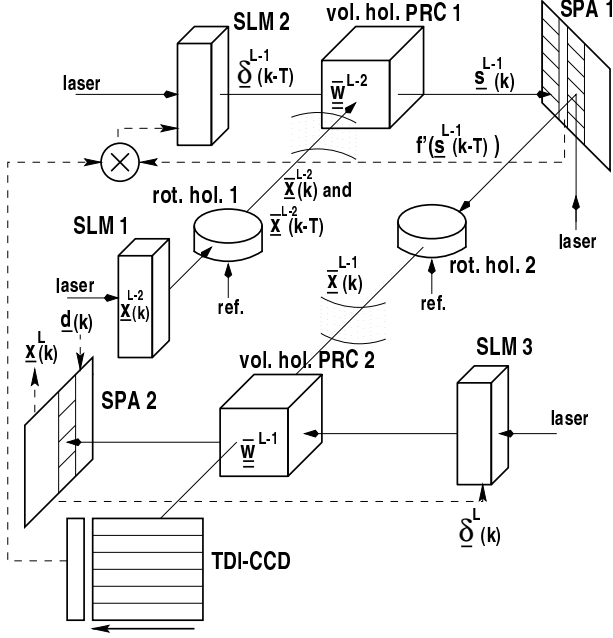


Figure 3: Optical architecture of a complete two-layer FIRNN, including forward propagation, backward error propagation and weight updating. Optic signals are represented by solid lines and electronic signals are represented by dashed lines. The cylindrical surfaces represent the spatio-temporal signals in the SI architecture, where the phase conjugate architecture is assumed.

$\underline{\delta}^L(k)$, fed to SLM 3. Notice that the rotating hologram 2 displays $\underline{\bar{x}}^{L-1}(k)$. The interference of this array with the error gradient signal in SLM 3 produces the outer-product used for updating the weights in the volume hologram PRC 2. Simultaneously, the signal $\underline{\delta}^L(k)$ in the SLM 3 is back-propagated through the volume hologram PRC 2, being diffracted to the TDI-CCD, which time integrates the back-propagated error in a conjugate plane of the rotating hologram. The TDI-CCD outputs are multiplied by the delayed derivative of the discriminant function, $f'(\underline{s}^{L-1}(k-T))$, stored in the SPA 1, producing the error gradient $\underline{\delta}^{L-1}(k-T)$. Setting the angular velocity $\dot{\theta}$ of the rotating hologram 1 to less than or equal to π/T , we have that this scrolling SLM will be displaying simultaneously $\underline{\bar{x}}^{L-2}(k)$ and $\underline{\bar{x}}^{L-2}(k-T)$, depending on which part of the output plane is used. Using signal $\underline{\bar{x}}^{L-2}(k-T)$ to perform an outer-product with signal $\underline{\delta}^{L-1}(k-T)$ from SLM 2, the weights in the volume hologram 1 are updated, finishing the learning cycle.

Notice that the spatio-temporal signal planes

$\underline{\bar{x}}^{L-2}(k)$ and $\underline{\bar{x}}^{L-2}(k-T)$ must be perfectly aligned in the volume hologram PRC 1, since the first reads the grating written by the second in combination with the error gradient $\underline{\delta}^{L-1}(k-T)$. Also, the two rotating volume holograms in figure 3 can be combined in a single rotating hologram using the multi-signal architecture depicted in figure 2. This would guarantee the synchronization of the delayed signals in the input plane between the first and second layers of the network, besides reducing the number of optic components required in the system.

The double-dot products performed by the SI and TI architectures are operations of the order $O(N^3)$, since they consist of the summation over two indices of the product between a $N \times T$ matrix and a $N \times 1$ vector. All other operations required for the implementation of FIRNNs are of a smaller order of complexity. Therefore, the proposed optical architecture assures us that all $O(N^3)$ operations are performed optically and only the less computationally intensive operations are performed electronically in the SPAs. For a sufficiently large array the optical parallel architecture will thus inevitably outperform an all-electronic implementation.

We greatly acknowledge the support by Dr. William Miceli of ONR and OSD through the MURI program grant no N00014-97-1-1006.

References

- [1] E. A. Wan, "Time series prediction by using a connectionist network with internal delay lines," in A. S. Weigend and N. A. Gershenfeld, eds., (*Time Series Prediction: Forecasting the Future and Understanding the Past*). Addison-Wesley, 1993.
- [2] A. H. Waibel, T. Hanazawa, G. E. H. K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 3, pp. 328-339, 1989.
- [3] Z. Gan and D. Z. Anderson, "Acoustic signal recognition with a photorefractive time-delay neural network," *Optics-Lett*, vol. 19, no. 9, pp. 655-657, 1994.
- [4] P. E. X. Silveira and K. H. Wagner, "Time delay neural networks," *Int. Conf. on Optics in Computing*, vol. 34, no. 90, pp. 266-269, 1998.