An Optoelectronic Implementation of the Adaptive Resonance Neural Network

Donald C. Wunsch II, *Member, IEEE,* Thomas P. Caudell, *Member, IEEE,* C. David Capps, Robert J. Marks II, *Senior Member, IEEE,* and R. Aaron Falk

Abstract— Implementation of the adaptive resonance theory (ART) of neural networks has been a thorny problem for several years. This work presents a novel solution to the problem by using an optical correlator, allowing the large body of correlator research to be leveraged in the implementation of ART. The implementation takes advantage of the fact that one ART-based architecture, known as ART1, can be broken into several parts, some of which are better to implement in parallel. The control structure of ART, often regarded as its most complex part, is actually not very time consuming and can be done in electronics. The bottom-up and top-down gated pathways, however, are very time consuming to simulate and are difficult to implement directly in electronics due to the high number of interconnections. Two facts simplify this. The first is that the pathways are computing a set of inner products. These inner products represent as least 80% of the computation time of the ART1 implementation. The second insight, our contribution, is that implementing the inner products optically, and the rest of the network in electronics, is a very effective marriage of the two technologies to realize the ART1 network. In addition to the design, we present experiments with a laboratory prototype to illustrate its feasibility and to discuss implementation details that arise in practice. This device potentially can significantly outperform alternative implementations of ART1 by as much as two to three orders of magnitude in problems requiring especially large input fields. It should be noted that all of these results apply to just one of the various ART architectures, known as ART1, but that other ART networks and other neural nets in general also use inner products and could benefit from this work as well.

I. INTRODUCTION

ADAPTIVE resonance theory has been published extensively, but because an understanding of it is necessary for the present work, a brief overview is provided in Section II. The particular hardware challenges of ART1 are discussed in Section III to motivate the use of optics instead of electronics. The bulk of our contributions follow these two introductory sections. Section IV shows our design. In this section we address the practical issues that arise with optical filter design for this application. This sets the stage for discussion of

theoretical and experimental results in Section V. Our experimental results are qualitatively similar to Carpenter and Grossberg's original ART1 results, but exhibit differences due to the optical filters we used in the laboratory. The fact that different results occur for various filter choices is another advantage of correlator-based implementations of ART, since the different filter designs have different strengths and weaknesses regarding sensitivity to noise, rotation, or scaling. The effects of the choice of filters in correlators is beyond the scope of this work, but we do define and discuss the filters we use in Sections IV and V. (To explore the implications of choosing other filters, see the references cited in Sections IV and V.) In Section VI we point out a unique advantage of the correlator approach: if one intends to build a hierarchy of ART modules, our implementation has the ability to accomplish this in a single correlator, rather than needing to build a hierarchy in the hardware, too. In this section we reference a few of the many papers that have used ART hierarchies to motivate the observation. Section VII mentions other published ART hardware work and estimates performance if the device were built with state-of-the-art components.

II. MOTIVATION FOR ADAPTIVE RESONANCE THEORY

Adaptive resonance theory (ART) has been of interest to eminent optical computing researchers almost since its inception [1], [2]. However, it has never before, to our knowledge, been fully implemented in optics. This section begins by reviewing ART and identifying the operations that can be performed more effectively in optics. All discussion in this paper refers to the binary−input binary−output version [3], [4] of ART, called ART1, although much of this research is applicable to other ART architectures. ART has been steadily gaining attention in the neural network community for it provides many of the quintessential advantages that the technology, at its best, is expected to offer, while suffering few of the disadvantages or limitations that alternative neural network theories do. This can be seen by considering several areas, such as flexibility of configuring the network into an overall system, stability, required accuracy of computational units (the neurons), locality of computations, speed of learning, and scaling properties.

The first of these attractive properties has been exploited by many workers, notably Healy [5], Caudell *et al.,* [6], and Waxman *et al.* [7]. Basically, because of the network dynamics
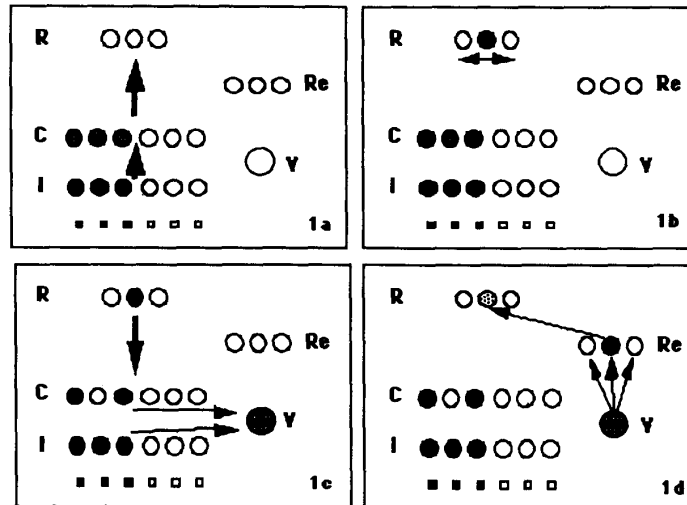
Fig. 1.   The networks dynamics of ART1. Bottom-up pattern matching (1a, b) is balanced by top-down feedback expectancy (1c, d).

constraints in the design of ART, it does not have to be considered as a black box with rigid input/output requirements. Instead, one can tweak the formalization of ART to customize it to a given application. These design constraints are also the reason for its stability, the second property mentioned above. The design of ART began many years before its first formal introduction, with dynamical system analysis of various networks that were to become components of ART. Once understood, these components could be given parameters that made them work well together. Some of the early relevant stability theorems are given in Grossberg [8] and Cohen and Grossberg [9]. Stability of neural networks can affect accuracy requirements, as is well-known with some popular neural models [10]. ART is very competitive with other neural models in this regard, which is one reason why we consider implementing it optically. Another issue affecting processor accuracy requirements, speed, and on-node memory is locality of computations. ART uses only information that is locally available at each node, as opposed to popular neural models with a great deal of nonlocal information transfer [3].

The various works with ART, including but not limited to those we reference here, allow its generalization capability to be compared head to head with supervised neural models. For example, on a representative set of benchmark data, new supervised learning architectures based on ART learned faster than backpropogation [5], [11]. Perhaps the most compelling advantage of ART, however, is its scaling properties. Recent successful applications of ART [12] used input fields in excess of 10^7 nodes. It appears that ART's scaling properties are limited only by hardware and software implementations.

The motivation for ART extends beyond such academic interests, however. Practical applications of the technology include intelligent design retrieval [6], target tracking [7], and automated target recognition [7], laser radar processing [13], and pattern recognition of occluded objects [13]. As hardware becomes increasingly available, the promising proliferation of applications should continue and accelerate. Thus we have significant motivation for the present work.

How does an adaptive resonance unit do all this? The key is that the pattern classification takes place in a feedback loop and that learning does not set in until resonance occurs. If resonance does not occur, a reset mechanism allows a search for a better pattern match, removes all previously considered classifications, and suspends learning until the best answer is found. This is briefly outlined in Fig. 1, where we see the ART unit displayed in several separate layers: R, the recognition layer; C, the comparison layer; I, the input layer; V, the vigilance layer; and Re, the reset layer. This grouping of layers is taken from Ryan et al. [14], and while it does not follow Carpenter and Grossberg's description exactly, it is functionally equivalent. Going left to right we see the ART unit in action. First, the input is registered at the comparison layer and fed up to the recognition layer Fig. (1(a)). In the second frame, the recognition layer's winner-take-all property finds the node corresponding to the initial best guess Fig. (1(b)). This guess is tested by playing back the winning node's previously learned template onto the comparison layer. This is compared with the pattern still on the input layer by competing signals sent to the vigilance node Fig. (1(c)). The final frame shows an example of what happens when the match is not good enough. The vigilance node is now able to activate the reset layer. The reset layer only suppresses nodes at the output that have been recently active and has no effect on the rest Fig. (1(d)). In this example, only the prior winner has been affected. Now with that node removed, the network will reclassify the pattern and continue to do so until it has found a good match.

The network described above can also be represented in algorithmic form [15]. Consider a new n-element binary input vector to be called I. We wish to assign the vector to a category that will have a template associated with it. The unit will classify the input by comparing it to these templates, which
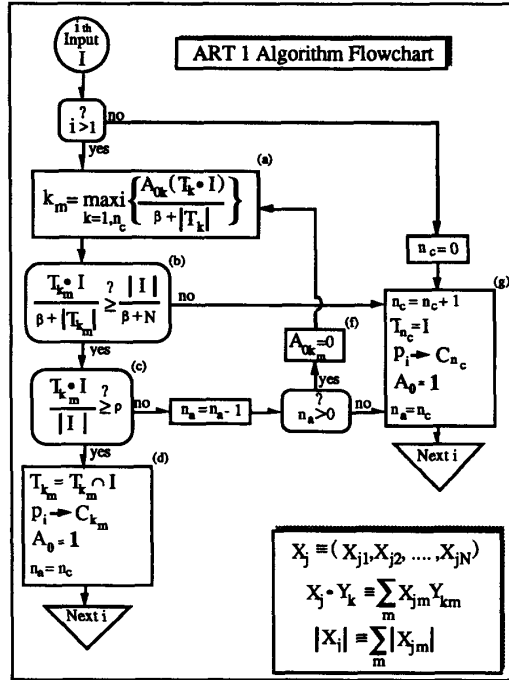
Fig. 2. ART1 as an algorithm. The input pattern is I, the output is the single signal C_{km}, and learning is the side effect $T_{km} = T_{km} \cap I$.

we index and refer to as the T_k, where k is the index number. A total of n_c templates exist from prior learning, and of these, n_a are active at a particular time, as indicated by the active node vector A_O. The final winning output pattern is generated by the signal C_k associated with the kth template. With these definitions in mind, the full ART1 algorithm is given in the flowchart [16] of Fig. 2. The reset test

$$\frac{T_i \cdot I}{I \cdot I} < \rho \tag{1}$$

relies on the parameter ρ, the vigilance threshold, which determines the level of discrimination—coarse of fine—used in grouping. The key point of (1) is that several inner product and norm calculations are required. (See boxes (a) and (c) within Fig. 2. Box (a) is the bottom-up pattern matching, and box (c) is the top-down feedback expectancy.) These inner products are equivalent to correlations that can be performed optically.

III. MOTIVATION FOR OPTICS

Optical computing has the potential to profoundly impact the hardware implementation of neural networks. This is because the implementation of a neural network is beset by a number of challenges, of which the most formidable is the connectivity desired. Furthermore, these are weighted interconnects and, for adaptive networks, the weights are changing by some algorithm that the implementer must control. These constraints, coupled with a desire for a large number of processing elements, can pull the implementer in opposite

directions. The challenge of adaptively weighting the interconnects also places difficult demands because it requires active elements where passive ones would otherwise suffice. The ease with which weights can be calculated electronically carries a large price in terms of speed so that optics can dominate in this area. The importance of adaptive and nonadaptive interconnects is clarified by examining Fig. 3. This shows a simple three-input two-output ART1 unit, yet even at this low level of complexity, the connectivity challenge is daunting. This is a compelling argument for the importance of optical implementations.

The advantage of optics lies in the fact that implementing interconnects electronically is difficult because of electromagnetic interference and the necessity for wires to carry the signals. A basic assumption of the artificial neural systems paradigm is that the importance of interconnects is dominant for a certain class of problems. Psaltis [17] assessed the relationship of computing power to interconnects, independent of the power of the individual processing elements. He suggests that an interconnection-dominated problem " ... has the property that local decisions cannot be made until essential information has been communicated from basically the entire input data. Thus useful computation can progress only when all the input information has been considered by the individual elements. For a parallel processor, this implies that all partial results need to be globally communicated. It is this notion that forms the basis for our conviction that communication capability becomes the dominant factor ... ". The following correlator is a suitable response to these realities in that it marries the strengths of electronics (ease of adaptability) to those of optics (high interconnectivity).

IV. IMPLEMENTATION WITH AN OPTICAL CORRELATOR USING BINARY PHASE-ONLY FILTERS

The previous section set forth reasons for implementing ART1 in optics. In this section, we begin by clarifying what optical correlators have to offer to an ART1 implementation. This is best seen by examining Table I. In this table, we have shown the computational demands (in simulation) of the various parts of the ART1 algorithm as given in Fig. 2. Recall that when the appropriate assumptions are made regarding learning rates, as Carpenter and Grossberg do, the gated interconnections reduce to inner products. As seen in the table inner product computations greatly dominate the processing required to implement ART1. The checkmarks in the column show which computations can be taken over by optics. Over 85% of the processing can be done optically, while the remaining computations that are less amenable to optical processing are handled electronically. This is an effective marriage of the two technologies. Because the optics offers the potential for greater speed than electronics, it is attractive to implement the demanding operations this way while retaining electronics for its flexibility where needed.

The optoelectronic ART1 unit is a novel application of an old device, the 4-f or Vander Lugt [18] correlator, which historically has been used as a fast pattern classifier. Usually the correlation operation is employed as a matched filter so
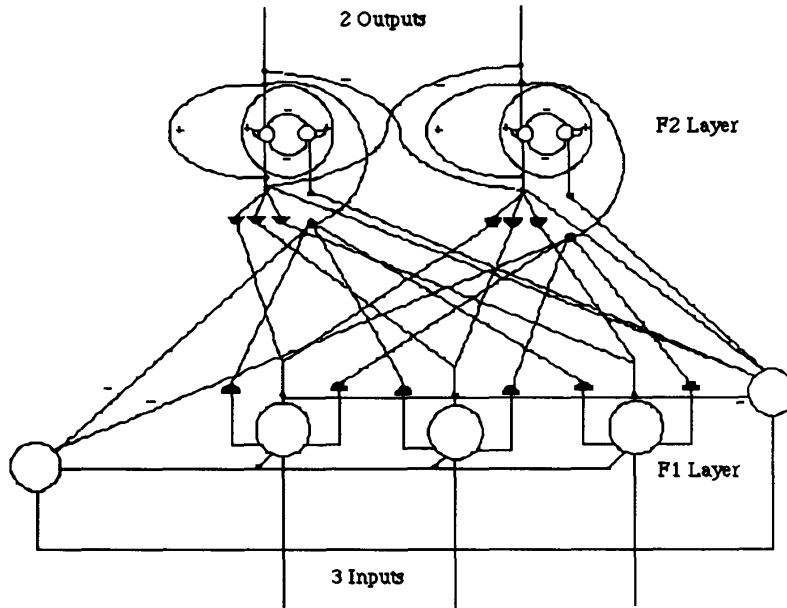
Fig. 3.   Even simple ART units have complicated connectivity patterns. The filled-in semicircles indicate adaptive connections, which are even more challenging to implement.

that a maximum output peak corresponds to a well-matched pattern. A simple thresholding algorithm can be used to detect peaks. The device described here also uses the large peaks but takes specific advantage of the fact that a zero-shift correlation is equivalent to a two-dimensional inner product, which is an essential computation for an ART unit. This device measures relative values of strong and weak correlation peaks and interprets these as inner products. Therefore, for this application of the 4-f correlator, greater measurement accuracy over the correlator's full dynamic range is required.

The device (shown in Fig. 4) consists of two lenses, two magneto-optical SLM's, two polarizers, a laser, a charge-coupled-device camera, and a computer. (The SLM's are 48×48 Semetex Sightmod SMD481, with a maximum frame rate of 30 frames/s. Thus their maximum space-bandwidth product is approximately 7×10^4 b/s.) This experimental setup has demonstrated several kinds of optical computing operations, [18]–[20]. All electronic calculations are done by a DEC MicroVAX, which processes the CCD camera output via a frame grabber and controls the SLM's. The lenses both have focal length f, and all components are placed at a distance f from their neighbors. Reading the figure left to right, notice that the laser light (which is collimated), passes through SLM1, where it is encoded with the input pattern i (referred to as I in the last section), and the templates t_i (referred to as T_i in the last section). (The new notation here is necessary to conform to the convention in optics that the Fourier transform is represented by a capital letter.) The encoded patterns are Fourier transformed by Lens 1 and are multiplied in the Fourier plane by the complex conjugate of the Fourier transform of i (I^*). Reviewed in the next paragraph, the resulting output patterns will be the correlation of the input

pattern with itself and with each template. This is shown only for the case of a single pattern—the result for multiple patterns follows from the Fourier transform's shift-invariant property.

If we define $i(x, y)$ and $t(x, y)$ as spatial patterns, $I(u, v)$ and $T(u, v)$ as their Fourier transforms, * as the complex conjugation operator, and $\mathcal{F}[\cdot]$ as the 2-D Fourier transform operator, then the output $\mathcal{O}$ of the correlator can be expressed in terms of the correlation operator ★ as follows:

$$
\mathcal{O} = \mathcal{F}\left[A_0 I^*(u, v)\, T(u, v)\right] = A_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P^*(u, v)
$$
$$
\cdot \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y)\, e^{-j\frac{2\pi}{\lambda f}(xu + yv)}\, dx\; dy\right]
$$
$$
\cdot e^{-j\frac{2\pi}{\lambda f}(ux_2 + vy_2)}\, du\, dv \tag{2}
$$

$$
= A_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y)\mathcal{F}\left[\mathcal{F}\left[i(x + x_2, y + y_2)\right]\right] dx\, dy \tag{3}
$$

$$
= A_2 i(x, y) \star t(x, y) \tag{4}
$$

which is the desired correlation, except for a coordinate reversal, and ignoring the effects of the constants A_0, A_1, and A_2. (The output coordinate reversal is well known in optics and is handled without difficulty since the value of the correlation peak, centered at $0, 0$, is the only number desired. This is complicated slightly when multiple patterns are processed simultaneously, but it is still a simple matter to keep track of the shifted $0, 0$ locations for each template. To see clearly how the coordinate reversal is handled, examine the output plane in Fig. 4.) The center point of the correlation

TABLE I
LIST OF THE PRINCIPAL COMPUTATIONAL OPERATIONS
REQUIRED TO PERFORM THE ART-1 ALGORITHM

#	Operation	Electronics	Optics	%
1†	I		✓	-
2†	T_k		✓	-
3	$T_k \bullet I$		✓	80
4	$\lvert I \rvert$		✓	5
5	$\lvert T_{k_m} \rvert$	✓		5
6	$T_{k_m} = T_{k_m} \cap I$	✓		5
7	$T_{n_c} - I$	✓		1
8††	$\mathrm{maxi}\{\ \}$	✓		3
9	$\{A_{0i}{=}1;\ i{=}1,\ n_t\}$	✓		<1
10	$P_i \rightarrow C_{n_c}$	✓		<1
11	$\geq$	✓		<1
12	$*$	✓		<1
13	$/$	✓		<1
14	$+$	✓		<1

†: Representation of information as light intensity or transmissivity.
††: This operation can be integrated on the CCD camera for further performance enhancement.
These operations can be partitioned naturally between electronic and optical implementations, as indicated by the check marks. The last column gives an approximate percentage of execution time in the ART-1 algorithm. Note that we implement over 85% of the algorithm, from an execution time point of view, in optics. This is why the device represents an effective marriage of the strengths of optics and electronics.

peak is easily shown to be proportional to the inner product of i and t, like those given in the algorithm of Fig. 2.

Although (2)–(4) are essential to understand the idea behind this device, and the motivation for using it, modern filtering theory allows several alternatives to the filter $I^*(u, v)$. Equation (4) is derived using the conventional matched filter, which is a shorthand way of saying the complex conjugate of the Fourier transform $I^*(u, v)$ of the pattern $i(x, y)$. An active research community has been exploring an alternative filter [21] since the mid-1980's. The filter, known as the binary phase-only filter (BPOF), has been of interest because it achieves an improved peak-to-sidelobe ratio in the correlation output plane (at the cost of a lower signal-to-noise ratio), has lower memory requirements, and can be used with bipolar SLM's such as the magneto-optical SLM. The trade-offs involved in choosing a filter are beyond the scope of this paper. Here we merely introduce the BPOF, because we have used it for our experiments, and point out that different results would be obtained with the conventional matched filter, or other options.

As mentioned before, it is possible to compute the correlation, and therefore the inner products, of multiple patterns simultaneously using correlators because the Fourier transform is a shift-invariant operation, so all the patterns' Fourier transforms will be approximately centered on SLM2. The ability to take inner products of multiple templates with a single input simultaneously is a reason for using a correlator; otherwise one would simply use two cascaded SLM's and a convergent lens. To process all the templates at once, we place them on the spatial plane and compute a filter for the Fourier plane based on the input, as shown in Fig. 4. This results in multiple correlation peaks on the output plane positioned in correspondence with the center location of the input patterns.

This arrangement is especially convenient for an implementation of ART1. The patterns chosen for SLM1 are all the templates known at present, and a copy of the input pattern. SLM2 contains $I_{B\phi}$, the BPOF of the input pattern, defined shortly. The inner product of the input with itself and with all templates is thus calculated in a massively parallel fashion. The choice of the formulation of the BPOF is an area of active research [22]–[24]. One possible definition, which we use, is

$$I_{B\phi}(u, v) = 1 \qquad \text{if } \mathrm{Re}[I(u, v)] \geq 0$$
$$-1 \qquad \text{if } \mathrm{Re}[I(u, v)] < 0. \qquad (5)$$

(Alternative formulations exist.) The filter that is chosen, be it a BPOF, a conventional matched filter, a phase-only filter, or a filter designed for insensitivity to scale and rotation, will affect the behavior of the ART1 model, which can be used to advantage. We have shown [22] that the BPOF as described gives the best possible signal-to-noise ratio, given that a filter that can take on only values of $(+1, -1)$.

V. THEORETICAL AND EXPERIMENTAL RESULTS

The ART1 algorithm of Fig. 2 was mapped onto the hardware and software of the 4-f correlator for testing. The control code for the ART1 implementation experiments begins by defining the data, parameters, and calibration pattern(s) for the system. It then calls up the precomputed BPOF of the input image from a file. (In a practical implementation, the fast Fourier transform and resulting BPOF would have to be calculated on the fly. The implications of this are discussed in the performance analysis later in Section VII.) The user is prompted for the value of ρ, and the calibration routine begins. The SLM's are reinitialized and the first input pattern is presented. Subsequent input patterns are also written to SLM1, in a portion of the SLM reserved for the current input pattern. Then the computer loads the BPOF of the new input pattern onto SLM2. The correlation values are read off the camera and normalized according to the calibrations. The various test values for the ART1 algorithm are then computed. The algorithm causes a reset or records a new or updated template on SLM1 as appropriate. As the process continues, the number of recorded templates increases. The capability to perform reset and to learn updated templates was demonstrated by a number of learning examples with varying values of ρ. An example experiment for $\rho = 0.5$ is shown in Fig. 5. This is a photo of the entire lab setup. The monitor on the left shows
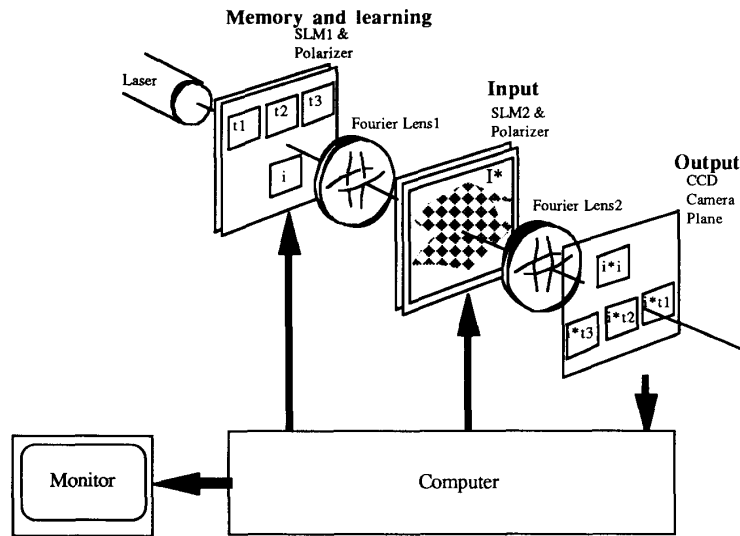
Fig. 4.   The optoelectronic ART1 implementation. Input is the pattern i, transformed (I^*) and loaded into SLM2. Output is chosen by considering the peak values on the camera plane. The computer controls the SLMs and reads the camera plane.
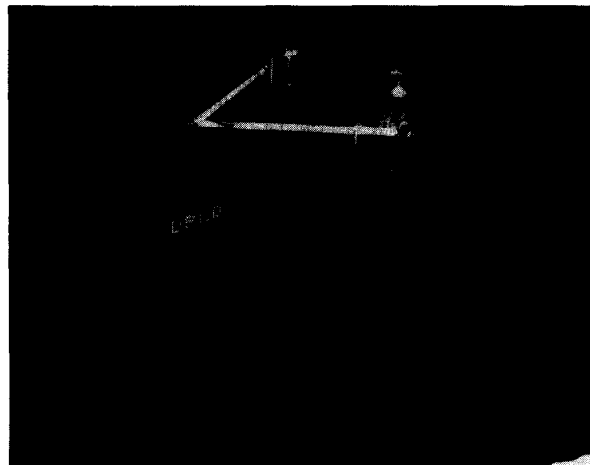

Fig. 5.   An experiment for $\rho = 0.5$.

SLM1 imaged through the correlator. (SLM2 has all pixels turned off.) The pattern in the leftmost corner of the monitor is the current input pattern. All other patterns are the stored templates that have been learned up to the current time. Each time a template is updated by learning, the new template is displayed on the monitor. These experiments verify the unit's ability to learn new templates and to perform the reset function.

In addition to the opportunity to see lab setup, this photo is instructive for several reasons. First, one can clearly see the learning and reset properties discussed in the previous paragraph. Second, one can also see the kind of learning performed by the unit. The second pattern from the left is a template that has captured two inputs, patterns A and B. The unit learns what is common to both patterns—the conjunctive generalization of the input patterns. Third, it clarifies the

operation of the unit. The leftmost pattern will produce the $i \cdot i$ term of the output, and the others will produce the $i \cdot t_i$ terms. New templates can be learned until the system reaches its capacity by filling the screen with templates. Fourth, it is possible to see potential sources of system errors by viewing the photo. For example, the faint repeated images of the patterns are caused by diffraction. SLM2 acts like a grating, whether it is turned on or off, causing the multiple images. (See Figs. 4 and 5.) If a repeated image falls on a measured pixel location, it can potentially introduce enough energy to cause an error. The geometry of the device can be arranged to minimize this problem but probably not enough to eliminate it entirely. Also, the photo shows the discrete pixelation of the device. The dark regions between pixels are also nonideal parts of the system that can cause some error. These errors do

not dominate the system's performance; in fact, the undesired images are very faint. They are common in correlators and are mentioned primarily for completeness.

Some examples of the clusterings performed by the device are given in Table II. The first two examples are for 10×10 alphabet patterns, and the rest are for 5×5 patterns. The clusterings, as expected, are different from those that would be achieved using the conventional matched filter because of our choice to use the BPOF. However, they do show qualitative agreement with the behavior expected of an ART network, especially in the increasing number of clusters generated with increasing values of ρ. Also, some templates formed are later abandoned (indicated by a blank spot in the table) as other templates are formed and capture inputs. This happens only when multiple passes are made through the same data. The 10×10 data shows clusters after just one pass through the system, and the 5×5 data was generated by making four passes in alphabetical order.

To accomplish these experiments, the system had to be carefully calibrated to use peak heights as inner products, which is where the significant differences in the use of the optical correlator becomes apparent. When a Vander Lugt correlator is used simply as a matched filter bank, calibration is usually not done because the signal-to-noise ratio is high enough that the brightest peak will likely be the best correlation even without calibration. In contrast, for an ART1 implementation, the device is used to measure all the inner products given in Fig. 2, large and small, so greater accuracy, and thus calibration, is a matter of critical importance. This required experiments and simulations assessing the accuracy of the BPOF-based correlator as an inner product processor. We have reported theory and experiments relating to this issue elsewhere [22], but it is worth showing some experimental results here to give a flavor of accuracy issues in applying correlators to neural networks. Examples of these are shown in Fig. 6. Fig. 6(a) shows experiments with the correlator emulating ART1 without normalization. Fig. 6(b) shows the same thing using a normalization scheme based on the known value of $I \cdot I$, giving fewer saturated output values, especially for the smaller numbers. In both of these plots, the x axis gives the theoretical inner product values for each computation, and the y axis gives the optically computed values. These plots group all the measurements from an entire data set into one figure. Recall, though, that in a given measurement, the correlator need only find the inner product of a single input pattern with a set of templates.

The normalization scheme used for Fig. 6(b) can best be understood by considering (1) and Fig. 5. In (1), we see that the system needs to compute $I \cdot I$ and $I \cdot T_i$. The former term is cheap to compute, requiring only N adds instead of N multiplies, where N is the number of bits in the pattern. This is therefore calculated and compared to the optically measured value of $I \cdot I$ (recall that I is the upper left template in Fig. 5) to arrive at a normalization factor. This factor is then used on all template locations. The data indicates that it is better than no normalization, giving fewer saturated output values, especially for the smaller numbers. This data was extracted from ART experiments such as those used to generate Table II. This

TABLE II
CLUSTERINGS PERFORMED BY THE OPTOELECTRONIC ART UNIT

Node	$\rho = 0.8$	$\rho = 0.5$
1	A,B	A,B,C
2	C,L,R	D,E,G,L
3	D,G	F,R,S
4	E,F	H,K
5	H,K	I,J,T
6	I,J,T	M,N,Q,P
7	M,N,W	Q,U,V
8	Q,P	W
9	Q,S,U	X,Y,Z
10	V	
11	X,Z	
12	Y	

(a) 10×10 Patterns

Node	$\rho = 0.8$	$\rho = 0.6$	$\rho = 0.4$
1			
2			
3	L		
4	C, E		
5			
6	K	M, R	R
7	T	H	
8		E, F	I, J, T
9	N	N	F, L
10	H	L	P
11	G, Q		
12			
13	M		C, D, E, G, Q
14			H
15	I, J		M, S
16	Q		A, B, K
17			N, Q
18		Q	
19	A, F		
20	R	C	
21	B	D, G	
22	P	A, B, I	
23	Q, S	J, T	
24		K	
25		P	
26		Q, S	

(b) 5×5 Patterns

means that various data points with identical theoretical inner product value correspond to very different pattern correlations. For example, the value of $E \cdot D$ is the same as the value of $P \cdot S$(12), and $G \cdot E$ is the same value as $B \cdot B$ (16).

Therefore, in interpreting these results, it is important to understand that radically different patterns can have the same theoretical inner product. If the device measures the inner product of two similar sets of patterns, it will get two similar results. If the sets of patterns have considerable differences, however, the results can be quite different even if the theoretical inner products are the same. This does not occur with the conventional matched filter—it is only a property of the
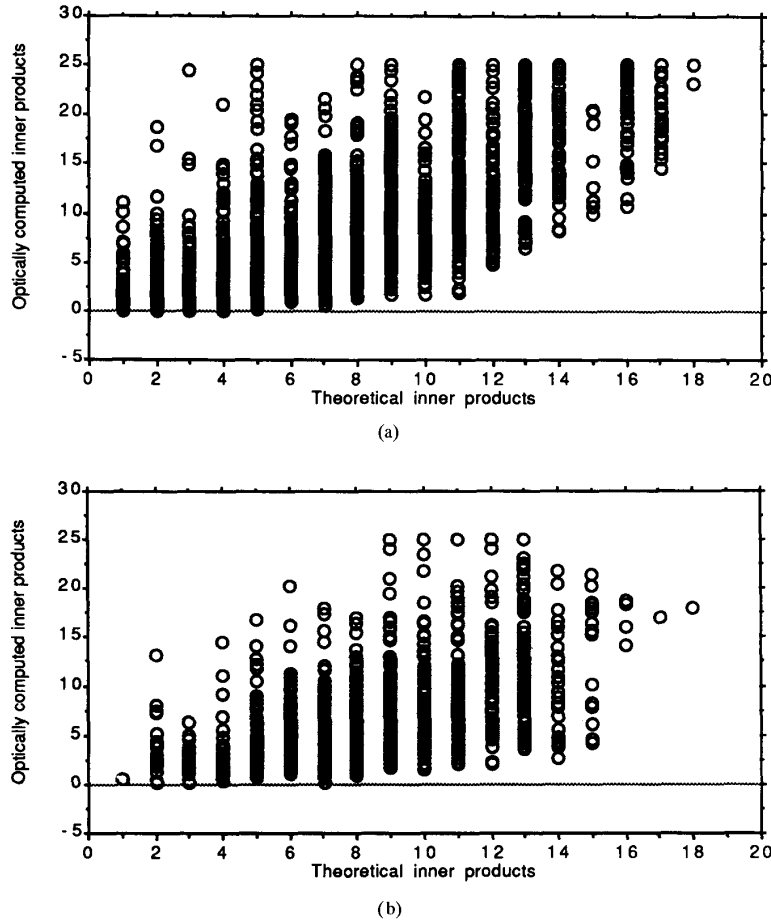
(a)

(b)

Fig. 6. (a) Performance of the correlator as an inner product processor without normalization. Notice the saturation effects that occur even for small theoretical inner product values. Also notice the large spread of optically computed values for the same theoretical inner product value. (b) Normalized correlator performance as an inner product processor. Notice the improvements over the problems mentioned in the previous figure, especially with regard to saturation effects.

BPOF. These plots show that while the BPOF is usable, it is of low accuracy for the types of computations required here. This is not surprising. In other work [22] we have shown that some kinds of BPOF's are of limited utility as an inner product processor. Accuracy with other filters would be higher, although determining precisely how much would require further experimentation.

In addition to the preceding, resolution and scalability must be addressed when using the correlators for calculating inner products optically. Larger patterns actually cause worse resolution problems than smaller ones. This can be seen by considering the simple one-dimensional case of a square pulse. A square pulse of width α has the Fourier transform [19]

$$\frac{\mathrm{Sin}(\alpha\pi u)}{\alpha\pi u} \tag{6}$$

which has its first zero at the point

$$\frac{1}{\alpha} = \frac{x_3}{\lambda f}, \tag{7}$$

where x_3 is the variable in the second SLM plane and λ and f are wavelength and focal length, respectively. This implies that

$$x_3 = \frac{\lambda f}{\alpha}. \tag{8}$$

Now if we take the pixel width Δ, and solve for x_3 in terms of the number of pixels q required in the second SLM plane, we get

$$q\Delta = \frac{\lambda f}{n\Delta} \tag{9}$$

where n is the number of pixels in the first SLM plane.

With $f = 38$ cm, $\lambda = 632.8$ nm, and $\Delta = 0.13$ mm., this means that

$$q = \frac{\lambda f}{n\Delta^2} = \frac{14.2}{n}. \tag{10}$$

Therefore, patterns that are approximately 15 pixels wide or larger in any one direction will cause resolution problems on the Fourier plane. The experiments in this article and others

[20] show better results for 5×5 patterns than for 10×10 patterns for this reason. Notice that as Δ improves, so does the maximum pattern size. A Δ of 0.05 mm allows patterns of up to 96×96 pixels to be processed. Bringing Δ down to 0.01 mm would allow patterns of 2404×2404 pixels to be processed. Further improvements, albeit less dramatic, can be made by increasing f.

VI. OPTICAL IMPLEMENTAION OF ART1 HIERARCHIES

ART hierarchies have been discussed in many publications [3], [6], [7], [12]. One useful type of ART1 hierarchy, called the ART tree [12], can be naturally implemented with the optical correlator. The ART tree is a hierarchy in which the same input pattern is sent to every level. The ART units in a given level that get to look at the input are determined by the winning F2 nodes at a lower level. Thus all F2 nodes in the entire hierarchy see the same input pattern, or nothing at all. This allows ART to perform hierarchical varying-k-means clustering.

Consider the hierarchical array of ART1 units shown in Fig. 7. There are a total of 63 F2 nodes, and thus templates, needed for the entire hierarchy, as can be seen by counting the outputs of each ART unit in the figure. That is, there are twelve third-layer ART units with four F2 nodes each (48 F2 nodes), three second layer units with four F2 nodes each (12 F2 nodes), and one first layer unit with three F2 nodes for a total of 63 nodes. For a 5×5 input pattern, the templates for this entire ART hierarchy can be stored on a 48×48 spatial light modulator, with margins to spare. Since the templates are all correlated with the same input pattern in this type of hierarchy, the multi-ART network maps exactly to the correlator with no hardware changes. Correlator-based implementations of ART are attractive for this property that allows them to implement hierarchies so efficiently.

VII. COMPARATIVE PERFORMANCE ANALYSIS

Other researchers have also proposed interesting ART-related devices. In 1989, Caulfield [2] proposed a design for a high-capacity resonant neural system based on optical hardware that was designed to exhibit an ART-like search property. The device, however, used a fixed hologram and therefore was not capable of learning. Electronic implementations of ART have also been designed. Although these devices would probably be successful if implemented, they each have strengths and weaknesses. A VLSI emulator (reported by Rao et al. [25]) of a 9-b ART1 processor, although capable of performing over 10^7 comparisons per second, requires time multiplexing of the necessary inner product calculations, leaving open the possibility of achieving a higher performance optically. It also allows scaling up to networks with larger input fields than 9 b, with a processing time penalty invoked by doing so. This penalty would not be serious limitation until significantly larger input fields are desired. Input field size is probably the key parameter for determining when it is desirable to switch from an electronic to an optoelectronic implementation. Another electronic device (reported by Tsay and Newcomb [26]) directly implements the long-term and

short-term memory, that is, the bottom-up and top-down interconnection structure and weighted synapses of ART1. This causes severe scaling constraints that exist because direct implementation of interconnections is costly in electronics, using up the majority of real estate on a chip. However, it is attractive if one wishes to tweak the formulation of ART1.

Optimal performance of the correlator-based ART implementation requires the best components. As we discussed, a magneto-optical spatial light modulator (SLM) typically constrains us to use the BPOF. A more flexible choice would be to use another kind of SLM with greater dynamic range. It is possible to do this with ferroelectric liquid crystal light valve SLM's, or with deformable mirror devices, for example. SLM's with a higher number of pixels also will improve processing throughput. Deformable mirror devices and especially ferroelectric liquid crystal SLM's offer a high number of pixels and a high frame rate and, as such, are promising avenues for further refinements to the device. An especially useful critical comparison of various SLM's currently available is given by Johnson and Moddel [27]. Ferroelectric liquid crystal SLM's also offer the potential to integrate photodetectors onto the same device, allowing it to be optically addressed. Another worthwhile improvement is to add some custom-processing electronics. For example, one operation frequently performed is the computation of the maximum element in the output plane, or in a subset of the output plane. Nabet and Pinter [28] have developed an on-center-off-surround neural network, implemented in gallium arsenide, capable of performing this and other useful operations. Such a device could be integrated on the same plane with the photodetector for maximum efficiency. This would radically increase the efficiency of the device by allowing only the important numbers to be processed electronically, rather than the entire output array. This computation is the $\max\{\}$ operation, shown in Table I to occupy 3% of the computational load of an ART1 unit. Three percent may not seem like a radical improvement, but remember that this is one of the major electronic operations left after the optics has taken over 85% of the computational load.

The approximate performance of this correlator-based implementation may be calculated. From Fig. 2, the dominant number of operations is generated by the parallel dot products and the template norms. Assuming that one 8-b MULTIPLY operation is equivalent to approximately 8 ADD operations, an order-of-magnitude estimate of the number of operations necessary to process one input pattern is

$$N_{\text{ops}} \approx 8 \, M_{\text{templates}} N_{\text{bits}'} \qquad (11)$$

where $M_{\text{templates}}$ is the maximum number of possible templates and N_{bits} is the length of the input vector (or the number of input field pixels for two-dimensional patterns). For example, if $M_{\text{templates}} = 961$ and $N_{\text{bits}} = 1024$, then $N_{\text{ops}} \approx 7.9 \times 10^6$. This would require a 1024×1024 SLM, which is certainly achievable. The densest SLM commercially available at this writing has 400×400 pixels, allowing for an $M_{\text{templates}} = 144$ and an $N_{\text{ops}} \approx 1.2 \times 10^6$. Much denser SLM's are expected soon, however, so the 1024×1024 SLM is a conservative projection. From N_{ops}, the approximate cycle

time for the processor can also be calculated. Upon analysis of the steps in Fig. 2, we see that the cycle time is dominated by the slowest of three events: the integration time of the CCD camera, the switching time of the SLM's or the computation of the BPOF's of the new input patterns. The cycle time can be estimated by the following simple formula:

$$T_{\text{cycle}} \approx \text{Max}\{T_{\text{integration}}, T_{\text{switch}}, T_{\text{BPOF}}\}. \quad (12)$$

The integration time $T_{\text{integration}}$ strongly depends on the specific properties of the CCD, the time constants of the analog circuits, the minimum signal-to-noise ratio required by the algorithm, and the specific application in which the network will be used. Since the total signal level is controlled by the intensity of the illumination source, a great deal of latitude exists in estimating the minimum allowable integration times; a value in the range of 10^{-4} to 10^{-3} s is not unreasonable for currently available detectors and light sources. This is also a reasonable value of T_{switch} for the fastest currently available SLM's. The bottleneck in the system would likely be the processing time for the BPOF's (T_{BPOF}) if these are computed electronically. A reasonable value for T_{BPOF} using existing technology would be 3–6 s.

Again, using the example sizes given of $M_{\text{templates}} = 961$ and $N_{\text{bits}} = 1024$, the approximate number of operations per second is found to be

$$R_{\text{op/s}} \approx 1.5 \times 10^6 \text{ to } 3.0 \times 10^6 \text{ operations per second.} \quad (13)$$

Optical computation of the filter (not necessarily a BPOF) on SLM2 is also a possibility. This could potentially be done quickly enough to remove the bottleneck entirely. For example, the filter could be computed in a second optical processor and loaded in parallel into an optically addressed SLM. In this case, the improved performance figure would be

$$R_{\text{op/s}} \approx 3 \times 10^{10} \text{ to } 3 \times 10^{11} \text{ operations per second.} \quad (14)$$

This estimate implies that the processor could potentially operate on over 10^5 input images per second of size 32×32 binary pixels, beating pure electronic implementations by three orders of magnitude.

The Vander Lugt correlator implementation of ART is also attractive from the size and power requirements perspective. Lindberg and Gregory [29] have developed such a correlator that is only a few inches long, has low power requirements, and is rugged. In their report, they mention several target recognition and tracking applications of the device. Horner *et al.* have proposed an optical design that allows the correlator to have a shortened length of $f/5$ instead of $4f$. Other promising miniaturization work is reviewed in Gregory *et al.* [30].

Other ways of implementing ART are to use other correlators, such as the joint transform correlator [31], [32], or the acousto-optical correlator [33], [34] (Molley and Kast have demonstrated processing of up to 1000 templates per second using the latter). These offer most of the advantages pointed out for the Vander Lugt correlator, such as speed and capability of performing hierarchical clustering. ART can also be implemented holographically [37].
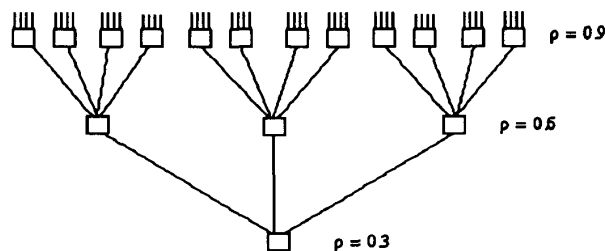

Fig. 7. A hierarchy of ART1 units. The input pattern is fed in at the bottom, and the winning output is read out at the top. This allows much more complex processing, including hierarchical varying k-means clustering.

After the submission of this paper, we learned of work by Kane and Paquin, subsequent to our own, implementing ART on a joint transform correlator [35], [36]. We are pleased to acknowledge this important contribution to the field.

VIII. CONCLUSION

The ART1 neural network performs operations (the $I \cdot T_i$ terms of (1), which require N multiplications apiece) that can be performed more effectively in optics. This observation motivated a discussion of the implementation of ART using a Vander Lugt correlator. This work allows the large body of correlator research to be leveraged in the implementation of ART, by recognizing that the steady-state solutions of the ART equations result in a simple algorithm. This hardware was tested with a categorization problem, allowing assessment of the alternatives for configuring the device, and of other implementations of ART. This is important because of the observation that BPOF's have limited effectiveness for ART implementation, as shown briefly in Fig. 7. The paper concluded with a performance analysis, which indicated that the device shows promise of significantly outperforming electronics, potentially computing 10^{10} to 10^{11} operations, as opposed to 10^7 operations in the fastest electronic alternative.

REFERENCES

[1] D. Z. Anderson and M. C. Erie, "Resonator memories and optical novelty filters," *Opt. Eng.*, vol. 26, no. 5, pp. 434–444, May 1987.
[2] H. J. Caulfield and D. Armitage, "Adaptive resonance theory of optical pattern recognition," *Appl. Opt.*, vol. 28, no. 19, pp. 4060–4061, 1 Oct. 1989.
[3] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," *Cognitive Sci.*, no. 11, pp. 23–63, 1987.
[4] G. A. Carpenter and S. Grossberg, "A massive parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, no. 37, pp. 54–115, 1987.
[5] M. J. Healy, "A logical architecture for supervised learning," in *Adaptive Neural Systems: The 1990 IR&D Technical Report*, Chapter 3, pp. 25–36, Boeing Computer Services Technical Report; The abstract is also published in *Proc. IJCNN-91-Seattle*, vol. 2, p. A-968.
[6] T. P. Caudell, S. Smith, C. Johnson, D. C. Wunsch II, and R. Escobedo, "An industrial application of neural networks to reusable design," in *Adaptive Neural Systems: The 1990 IR&D Technical Report*, Chapter 6, pp. 185–190; The abstract is also published in *Proc. IJCNN-91-Seattle*, vol. 2, p. A-919.
[7] A. Waxman, M. M. Seibert, R. Cunningham, and J. Wu, "Neural analog diffusion-enhancement layer and spatio-temporal grouping in early vision," in *Advances in Neural Information Processing Systems I*, D. Touretzky, Ed. New York: Morgan Kaufmann, 1989, pp. 289–296.
[8] S. Grossberg, "Competition, decision, and consensus," in *Studies of Mind and Brain*, S. Grossberg, Ed. Dordrecht, The Netherlands: D. Reidel, 1982, Chapter 10, pp. 399–424.

[9] M. Cohen, and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Sys., Man, Cybern.* vol. SMC-13, no. 5, pp. 815–826, Sept. 1983.

[10] D. E. Rumelhart, G. Hinton, and R. J. Williams, *Parallel Distributed Processing* vol. 1. Cambridge, MA: MIT Press, 1986, pp. 318–364.

[11] G. Carpenter, S. Grossberg, and J. Marshall, "ARTMAP: A self-organizing neural network architecture for fast supervised learning and pattern recognition," in *Proc. IJCNN-91-Seattle,* vol. 1, pp. 863–868.

[12] T. P. Caudell, S. D. G. Smith, G. C. Johnson, and D. C. Wunsch II, "An application of neural networks to group technology," Invited Paper, *Proc. SPIE,* vol. 1469, *Applications of Neural Networks II,* pp. 612–621.

[13] P. Kolodzy, "Multidimensional machine vision using neural networks," in *Proc. First Int. Conf. Neural Networks,* vol. II, pp. II-747–758, 1987.

[14] T. W. Ryan, and C. L. Winter, "Variations on adaptive resonance," in *Proc. First Int. Conf. Neural Networks,* pp. 767–776.

[15] B. Moore, "ART1 and pattern clustering," in *Proc. 1988 Connectionist Summer School at Carnegie-Melon University,* D. Touretzky and Hinton, Eds. New York: Morgan Kaufmann, 1989.

[16] D. S. Newman and T. P. Caudell, "Characterization of some learning properties of ART machines," *Neural Networks for Automatic Target Recognition,* poster paper, Wang Institute of Boston University, May 1990.

[17] D. Psaltis, and Y. S. Abu-Mostafa, "Theoretical investigation of parallelism and connectivity in optical computers," *Optical Processing: Interim Report,* University of Dayton Research Institute, Report No. UDR-TR-85-148, Dec. 31, 1985, pp. 9.1–9.13.

[18] A. B. Vander Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inform. Theory,* vol. IT-10, no. 2, p. 139, 1964.

[19] J. W. Goodman, *Introduction to Fourier Optics.* San Francisco: McGraw-Hill, 1968, pp. 141–184.

[20] C. D. Capps and S. G. Ferrier, "Toward practical real-time optical correlators," presented at the Int. Topical Meeting on Opt. Computing, paper 10F3, Kobe, Japan, Apr. 8-12, 1990.

[21] J. L. Horner and P. D. Gianio, "Phase-only matched filtering," *Appl. Opt.,* vol. 23, 1984, p. 812.

[22] D. C. Wunsch II, R. J. Marks II, C. D. Capps, and T. P. Caudell, "Limitations of a class of binary phase-only filters," *Appl. Opt.,* vol. 31, no. 26, pp. 5681–5687, Sept. 10, 1992.

[23] M. W. Farn, and J. Goodman, "Optimal binary phase-only matched filters," *Appl. Opt.* vol. 27, no. 21, pp. 4431–4437, Nov. 1988.

[24] R. D. Juday, B. V. Vijaya Kumar, and P. Karivaratha Rajan, "Optimal real correlation filters, *Appl. Opt.,* vol. 30, no. 5, pp. 520–522, Feb. 1991.

[25] A. Rao, "VLSI implementation of neural classifiers," *Neural Computation,* no. 2, 1989, pp. 35–43.

[26] S. W. Tsay and R. W. Newcomb, "VSLI implementation of ART1 memories," *IEEE Trans. Neural Networks,* vol. 2, no. 2, pp. 214–221, Mar. 1991.

[27] K. Johnson and G. Moddel, "Motivations for using ferroelectric liquid crystal spatial light modulators in neurocomputing," *Appl. Opt.,* vol. 28, no. 22, pp. 4888–4899, Nov. 1988.

[28] B. Nabet and R. B. Pinter, *Sensory Neural Networks: Lateral Inhibition.* Boca Raton, FL: CRC Press, 1991.

[29] P. C. Lindberg and D. A. Gregory, "Building an optical pattern recognizer," *Optical Systems Department Report #1470-26,* Teledyne Brown Engineering.

[30] D. A. Gregory, J. C. Kirsch, and J. A. Loudin, "Optical correlators: Optical computing that really works," *SPIE* vol. 1296, *Advances in Optical Information Processing IV,* 1990, pp. 2–19.

[31] F. T. S. Yu, S. Jutamulia, T. W. Lin, and D. A. Gregory, "Adaptive real-time pattern recognition using a liquid crystal TV based joint transform correlator," *Appl. Opt.,* vol. 26, no. 8, pp. 1370–1372, Apr. 1987.

[32] T. Iwaki and Y. Mitsuoka, "Optical pattern recognition of letters by a joint-transform correlator using a ferroelectric liquid-crystal spatial light modulator," *Opt. Lett.* vol. 15, no. 21, pp. 1218–1220, Nov. 1990.

[33] J. Lee and A. Vander Lugt, "Acoustooptic signal processing and computing," *Proc. IEEE,* vol. 77, no. 10, pp. 1528–1557, Oct. 1989.

[34] P. Molley and B. A. Kast, "Automatic target recognition using acoustooptic image correlator," in *Proc. SPIE,* vol. 1471, *Int. Symp. Aerospace Sensing,* Orlando, FL, Apr. 1991, pp. 224–232.

[35] J. S. Kane and M. J. Paquin, "POPART: Partial optical implementation of adaptive resonance theory 2," *IEEE Trans. Neural Networks,* this issue.

[36] M. Paquin and J. Kane, "An optoelectronic neural network utilizing a joint transform correlator," in *Proc. SPIE Optical Information Processing Systems and Architecture,* 1992.

[37] D. C. Wunsch II, D. J. Morris, R. L. McGann, and T. P. Caudell, "Photorefractive adaptive resonance neural network," *Appl. Opt.,* vol. 32, no. 8, pp. 1399–1407, Mar. 10, 1993.
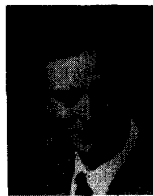
Donald C. Wunsch II (S'86–M'90–S'90–M'91) received the B.S. degree from the University of New Mexico, the M.S. degree from the University of Washington, both in applied mathematics, and the Ph.D. degree from the University of Washington in electrical engineering in 1984, 1987, and 1991, respectively.

He was with International Laser Systems and consulted for Rockwell International before joining the Boeing Company, Seattle, WA, in 1984. He has taught neural networks and linear systems analysis for Boeing and the University of Washington. He was recently a Senior Principal Scientist at Boeing Computer Services, where he was hardware task leader for the adaptive neural systems research and development project. Since January 1993 he has been Assistant Professor of Electrical Engineering at Texas Tech University. His research interests include theory, applications, and hardware implementations of neural networks; decision support systems; and speech processing.

Thomas P. Caudell (M'91) received the B.S. degrees from California State University, Pomona, in physics and mathematics, and the M.S. and Ph.D. degrees in physics from the University of Arizona, Tucson, 1973, 1978, and 1980, respectively.

From 1981 to 1984 he was a Research Assistant Professor in the Department of Physics and Arizona Research Laboratories at the University of Arizona, working in the field of solar seismology. From 1984 to 1989 he was a Senior Staff Physicist at the Hughes Artificial Intelligence Center, Hughes Research Laboratories, Malibu, CA, where he worked in the fields of artificial intelligence, neural networks, optical computing and virtual reality. Currently he is a Senior Principle Scientist in the Research and Technology division of Boeing Computer Services (BCS), Seattle, WA, and an Affiliate Associate Professor in the Department of Electrical Engineering at the University of Washington. Since 1989 he has led the adaptive neural systems research and development project at BCS. This project is developing neural theory, simulations, applications, and optoelectronic hardware for industrial and commercial applications. In August 1993 he is to become a Research Associate Professor in the Department of Electrical and Computer Engineering, the University of New Mexico. His active research interests are in the areas of neural networks, fuzzy logic, genetic algorithms, and virtual reality, including network architectures, hardware and software development, optical computing, and industrial applications of hybrid systems. He is currently writing a book on neural network applications for the IEEE Press, to be published in 1994.

C. David Capps received the B.S. degree from the University of Arkansas, and the M.S. and Ph.D. degrees from Florida State University, all in physics.

He was a guest scientist at the Max Planck Institute for Physics where his research was in experimental particle physics at CERN. In 1977 he joined Hughes Aircraft Company where he worked on the propagation of electromagnetic radiation through the atmosphere and modeling performance of imaging infrared systems. He joined Boeing in 1981. Since then his research areas have included aerosol scattering, optical processing and computing, and electro-optical means of generating and detecting millimeter wave radiation.
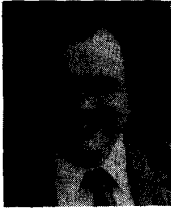
Robert J. Marks II (S'71–M'72–S'76–M'77–SM'83) received the B.S.E.E. and M.S.E.E. degrees from the Rose–Hulman Institute of Technology in 1972 and 1973, respectively, and the Ph.D. degree in electrical engineering from Texas Tech in 1977.

He is a Professor in the Department of Electrical Engineering at the University of Washington, Seattle. He was Chair of the IEEE Neural Networks Committee and was cofounder and first Chair of the IEEE Circuits and Systems Society Technical Committee on Neural Systems and Applications. He was the first President of the IEEE Council on Neural Networks (1990–1991). He was cofounder and first President of the Puget Sound Section of the Optical Society of America and was elected that organization's first Honorary Member. He is cofounder and current President of Multidimensional Systems Corporation in Lynnwood, WA. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS from 1992 to the present. He was topical editor for Optical Signal Processing and Image Science for the *Journal of the Optical Society of America—A* (1989–1991) and a member of the Editorial Board of *The International Journal of Neurocomputing* (1989–1992). He served as North American Liaison for the 1991 Singapore International Joint Conference on Neural Networks (IJCNN) and was International Chair of the 1992 RNNS/IEEE Symposium on Neuroinformatics and Neurocomputing, Rostov-on-Don, Russia, and Organizational Chair for both the 1993 IEEE Virtual Reality Annual International Symposium (VRAIS) in Seattle, WA, and the Organizational Chair of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Victoria, BC, Canada, 1992. He also served as Program and Tutorials Chair for the First International Forum and Applications of Neural Networks to Power Systems, Seattle, 1995. He is cofounder of the Christian Faculty Fellowship at the University of Washington and is faculty advisor to the University of Washington's chapter of Campus Crusade for Christ. He holds two U.S. patents in the field of artificial neural networks.

Dr. Marks was awarded the Outstanding Branch Councilor Award in 1982 by the IEEE and, in 1984, was presented with an IEEE Centennial Medal. He was named a Distinguished Alumnus of the Rose-Hulman Institute of Technology in 1992.

R. Aaron Falk received the B.S., M.S., and Ph.D., degrees from the University of Washington, in 1973, 1974, and 1979, respectively, all in physics.

From 1979 to 1982 he was a postdoctoral research associate at the Joint Institute for Laboratory Astrophysics where he measured fusion-related electron-ion collision cross sections in conjunction with the Oak Ridge National Laboratory, TN. In 1982 he joined the staff at Bell Aerospace where he initiated a laser radar effort based on coherent detection of the return from a chirped laser diode transmitter. In 1984 he joined the technical staff at Boeing Company. His research in the area of optical physics have included integrated and fiber optics, optical computing, and optical generation of microwaves. This work has led to the issuance of over 24 patents. He currently leads a group involved in developing a high-power high-speed photoconductor for use in generation of microwaves.

Dr. Falk is a member of the American Physical Society, the Optical Society of America, and the Puget Sound Chapter of the Optical Society of America and is a past president of the latter.