

# POPART: Partial Optical Implementation of Adaptive Resonance Theory 2

Jonathan S. Kane and Marc J. Paquin, *Member, IEEE*

**Abstract**— Adaptive resonance architectures are neural nets that are capable of classifying arbitrary input patterns into stable category representations. In this paper, a hybrid optoelectronic implementation utilizing an optical joint transform correlator is proposed and demonstrated. The resultant optoelectronic system is able to reduce the number of calculations compared to a strictly computer-based approach. The result is that, for larger images, the optoelectronic system is faster than the computer-based approach.

## I. INTRODUCTION

ADAPTIVE resonance theory (ART) provides a neural network architecture for self-organizing arbitrary input binary patterns into stable recognition codes [1]–[3]. One recent advance of these general architectures is adaptive resonance theory 2 (ART2), which is capable of categorizing either binary or analog patterns [2]. In this paper we utilize a model of ART2 known as ART2-A (algorithmic-ART) [4] assisted by a joint transform optical correlator (JTC) to produce an optoelectronic neural network. This hybrid system has the advantages of parallelism and translation invariance offered by the optics as well as the ability to process two-dimensional images directly without first partitioning the inputs into a one-dimensional vector space.

There are a few related hardware implementations of ART. Caulfield and Armitage [5] proposed one implementation that is not strictly the ART model as described by Carpenter and Grossberg [1]–[3]. Rather, it is an adaption inspired by ART that utilizes a fixed page hologram. Our work is most closely related to that presented by Wunsch *et al.* [6]. In their work, they implemented ART1 by utilizing a four focal length correlator architecture. This type of architecture necessitated calculating the binarized Fourier transform each time a new input image was presented. Our work differs in that we utilize a JTC that does not require constructing a matched filter for each input. In addition, the model we use is based on ART2, which is capable of processing either binary or analog imagery. In this paper we present a brief overview of our variation of the ART2-A algorithm, which incorporates translation invariance, and discuss the operation of our proposed device.

Manuscript received March 20, 1992; revised July 24, 1992. M. J. Paquin's research was supported by AFOSR summer-hire program and by Parke Mathematical Contract FY9628-90-C0140.

The authors are with the Optical Signal Processing Branch/EROP, Rome Laboratory, Hanscom AFB, Lexington, MA 01731-5000.  
IEEE Log Number 9203739.

## II. OVERVIEW OF ART2-A

Adaptive resonance theory first emerged from a study of feedforward adaptive coding structures by Grossberg [7], [8]. These theories were later refined and presented as a system of differential equations by Carpenter and Grossberg [1]–[3].

ART2-A approximates and simplifies the dynamics of an ART2 system. The simplification arises because in many cases the result can be logically calculated from a steady state solution rather than from a numerically integrated dynamical equation. The payoff is that the ART2-A model is much simpler to implement and can be as much as 150 times faster than the full ART2 system of equations [4].

A block diagram of ART2-A is shown in Fig. 1. Physically it consists of three layers, a "representation" layer, F0, a "matching" layer F1, and a "category" layer, F2. Inputs are received at the representation layer (F0) where they are contrast enhanced and normalized and passed to the matching layer (as the normalized vector,  $I$ ). The matching layer calculates the inner product between the bottom-up input and the top-down stored categorizations. The "best match," or highest inner product is then compared against an external parameter,  $\rho^*$ . If the value of the inner product is greater than or equal to  $\rho^*$ , then only those weights associated with the "best match" stored categorization are modified via the learning laws (discussed shortly). Otherwise, the input vector is placed into a "uncommitted" node, or a node whose associated weights  $Z_{ij}$ , have yet to be modified by the system.

The operation of the algorithm below follows [4] closely.

### A. Input

First, an  $X$ -dimensional input vector,  $I^o$  (we assume  $X = 2$  for image inputs) is presented to the F0 layer where it undergoes normalization and contrast enhancement:

$$I = NF_{\theta}NI^o \quad (1)$$

where  $N$  stands for a normalization operator, or

$$Nx = \frac{x}{\|x\|} \quad (2)$$

and

$$(F_{\theta}x)_{ij} = \begin{cases} x_{ij}; & \text{if } x_{ij} > \theta \\ 0; & \text{otherwise.} \end{cases} \quad (3)$$

The threshold  $\theta$  satisfies the inequalities [3]

$$0 < \theta \leq \frac{1}{\sqrt{M}} \quad (4)$$

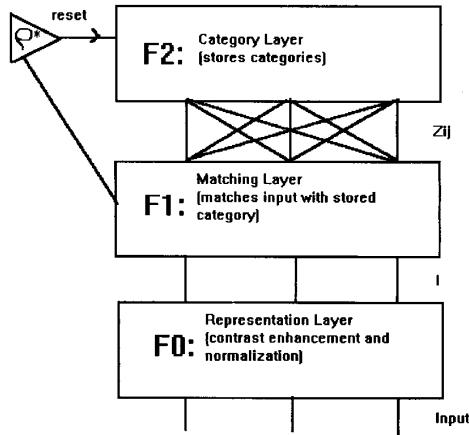


Fig. 1. Block diagram of ART2-A model.

where  $M$  is the number of pixels in the image. The foregoing assumes that the input is nonzero [1]. In our case we set  $\theta$  equal to the upper bound.

### B. Activation

The F1 layer takes the F0 output and gates it to the F2 layer through weighted connections,  $Z_{ij}^n$ . The input to the  $n$ th F2 node is given by

$$T_n = \begin{cases} \alpha \sum_i \sum_j I_{ij}; & \text{if } n \text{ is an uncommitted node} \\ \sum_i \sum_j I_{ij}^* Z_{ij}^n; & \text{if } n \text{ is a committed node} \end{cases} \quad (5)$$

where

$$I_{ij}^* = I_{i+k, j+l} \quad (6)$$

and the indices  $k$  and  $l$  represent any possible shift in position of the input pattern relative to the stored pattern and are chosen such that  $T_n$  is maximum. This incorporates translation invariance into the computer portion of the algorithm. For a given committed node,  $T_n$  is the inner product of the stored pattern in node  $n$  and the input pattern positioned so that this inner product is largest.

The parameter  $\alpha$  is generally chosen to satisfy

$$\alpha \leq \frac{1}{\sqrt{M}} \quad (7)$$

Initially the  $Z_{ij}^n$  weights are set to a uniform value  $\alpha$  [2]. As different patterns are categorized the  $Z_{ij}^n$  weights change according to the learning rules discussed as follows.

### C. Choice Function

For our network, only the weights associated with the best match are changed in learning a new pattern. Thus the winning node  $w$  has the inner product such that (after [4]):

$$T_w = \max_n (T_n). \quad (8)$$

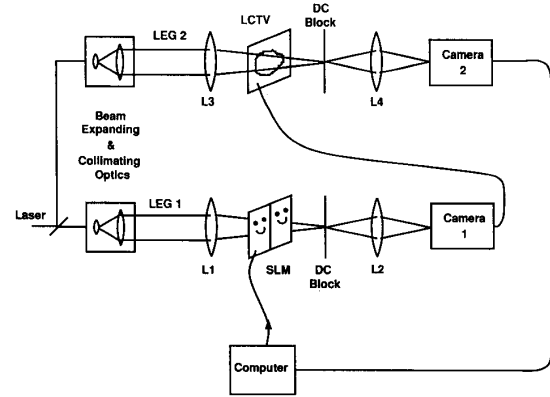


Fig. 2. Sketch of experimental setup.

As mentioned earlier, the subscript  $n$  refers to the  $n$ th "node" or category. If more than one node is maximum, then we choose the lowest number maximum node. Once a node is chosen (after a given input is presented) then the node is said to become "committed."

### D. Resonance or RESET

Each  $T_n$  is the dot product between the input  $I^*$  and each stored vector,  $Z^n$ . When the input precisely correlates with the given categorization, the value of  $T_n$  is 1, since both  $I$  and  $Z^n$  are normalized. A completely orthogonal representation yields a value of 0 for  $T_n$ . The resultant "best match" can be compared to a preset "vigilance" parameter,  $\rho^*$ , chosen such that

$$0 \leq \rho^* \leq 1. \quad (9)$$

If  $T_w \geq \rho^*$  then the match passes the vigilance test and the input is incorporated into the winning node using the learning rule discussed shortly. If the best match fails vigilance, then the input is stored directly into the weights associated with a new node.

### E. Learning

A given pattern is encoded by changing the weights associated with the node of interest. If the input is to be incorporated into an uncommitted node, the input values are simply incorporated directly into the weights  $Z_{ij}^w$ . If the winning node has some previously learned information stored in it, the weights that are stored become an amalgamation of the input and the stored values. Mathematically,

$$Z^{w(\text{new})} = \begin{cases} I; & \text{if } w \text{ is an uncommitted node} \\ N(\beta N \Psi + (1 - \beta) Z^{w(\text{old})}); & \text{if } w \text{ is a committed node} \end{cases} \quad (10)$$

where  $\beta$  is a constant between 0 and 1, which weights new versus old information and  $\Psi$  is defined as

$$\Psi_{ij} = \begin{cases} I_{ij}^*; & \text{if } Z_{ij}^{w(\text{old})} > \theta \\ 0; & \text{otherwise.} \end{cases} \quad (11)$$

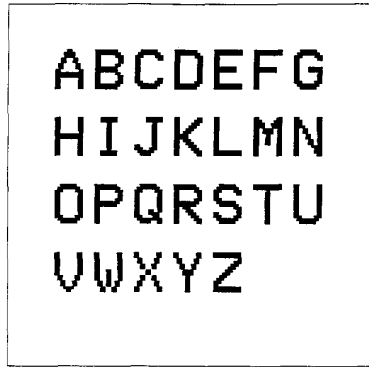


Fig. 3. Binary letter set used in experiment.

The importance of the  $\Psi$  parameter is central to the ART2-A algorithm. Once a given value in a stored pattern decays enough that it falls below the threshold set by  $\theta$ , then it can no longer grow again and will eventually decay to zero. This prevents supersets from recoding subsets [3]. For all simulations  $\beta = 0.3$ .

### III. JOINT TRANSFORM CORRELATOR

The correlation described earlier is performed numerically via the translation-invariant dot product (see Section II-B). Each computation requires at least  $M^2$  multiplications due to the translational invariance of the algorithm. Correlation can also be performed in real-time through the use of an optical JTC. The JTC offers a method of two-dimensional translation-invariant real-time correlation, and can offer a time savings over the electronic method, as the setup time of the correlator is nearly independent of the size of the inputs.

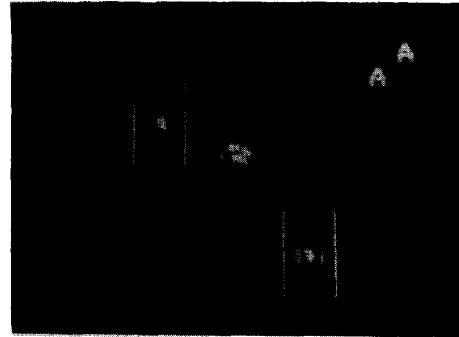
The details of the JTC have been covered extensively elsewhere [9]–[11], however, a brief review follows. A schematic of the JTC can be found in Fig. 2. The coherent beam from the laser source is expanded and collimated to illuminate the spatial light modulator (SLM). Binarized versions of the input  $I$  and a stored pattern  $Z^n$  are encoded side by side onto the SLM as an optical transmission function:

$$t(x_1, y_1) = f(x_1 - a_1, y_1 - b_1) + h(x_1 + a_1, y_1 + b_1) \quad (12)$$

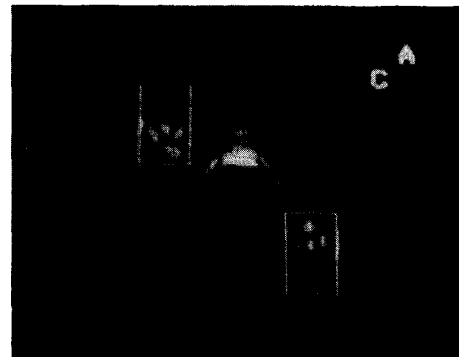
where  $f(x_1, y_1)$  represents the stored pattern and  $h(x_1, y_1)$  represents the input pattern. The Fourier transform (FT) of  $t(x_1, y_1)$  is created by lens L1 and magnified by lens L2, appearing at camera 1. The magnitude of the FT is encoded as the magnitude of light at the Fourier plane, and the FT phase arises from the relative phase differences of coherent light falling on this plane. This FT is given by

$$T(x_2, y_2) = F(x_2, y_2)e^{-j2\pi\beta_1(a_1x_2 + b_1y_2)} + H(x_2, y_2)e^{j2\pi\beta_1(a_1x_2 + b_1y_2)} \quad (13)$$

where  $\beta_1$  is a scaling factor dependent on the wavelength of laser light used and the focal length of the Fourier and magnifying lenses. The symbols  $T(x_2, y_2)$ ,  $F(x_2, y_2)$ , and  $H(x_2, y_2)$



(a)



(b)

Fig. 4. Example of experimental operation of optoelectronic implementation. Letters A and C have been previously learned by the network (see text for details). (a) The correlation plane for the two letter A's (one stored vector and one input). (b) The cross-correlation between the second stored vector (letter 'C') and the input (letter A).

represent the Fourier transform of  $t(x_1, y_1)$ ,  $f(x_1, y_1)$ , and  $h(x_1, y_1)$  respectively. Since the camera is a square-law detector (not sensitive to phase), the magnitude squared of the Fourier plane is recorded:

$$|T(x_2, y_2)|^2 = |F(x_2, y_2)|^2 + |H(x_2, y_2)|^2 + F(x_2, y_2)H^*(x_2, y_2)e^{-j4\pi B_1(a_1x_2 + b_1y_2)} + F^*(x_2, y_2)H(x_2, y_2)e^{j4\pi B_1(a_1x_2 + b_1y_2)}. \quad (14)$$

This image, referred to as the joint power spectrum (JPS), is sent to the liquid crystal television (LCTV) as a transmission function where another FT is calculated optically. The electric field distribution (assuming 1 : 1 scaling) at camera 2 is given by

$$E(x_3, y_3) = f(x_3, y_3) * f^*(x_3, y_3) + h(x_3, y_3) * h^*(x_3, y_3) + f(x_3, y_3) * h^*(-x_3, -y_3) * \delta(x_3 - 2a_1, y_3 - 2b_1) + f^*(-x_3, -y_3) * h(x_3, y_3) * \delta(x_3 + 2a_1, y_3 + 2b_1). \quad (15)$$

TABLE I  
RESULTS WITH  $\rho^* = 0.80$

Node #	Letter Category		
	Trial 1	Trial 2	Trial 3
0	A	A	A
1	B, C, D, E, G, O, Q, S	B, C, D, G, O, Q, S	B, C, D, G, O, Q, S
2	F, P	E, F, P	E, F, P
3	H, M, N	H, M, N	H, M, N
4	I, J, T	I, J, T	I, J, T
5	K	K	K
6	L	L	L
7	R	R	R
8	U, W	U, W	U, W
9	V	V	V
10	X	X	X
11	Y	Y	Y
12	Z	Z	Z

The number of categories have decreased a factor of 2 from the  $\rho^* = 1$  case. Note that the letter E switches categories in the second run, but then stabilizes into Node 2. All other nodes are stable after a single iteration. The confidence ratio for this example was set to 0.80.

The first two terms are autocorrelations that appear in the center of the correlation plane and are not of interest. The third and fourth terms are the cross-correlations and are offset from the center in the correlation plane. In other words, if  $f(x, y)$  correlates with  $h(x, y)$  then there will be two bright spots representing the correlations, one that appears at coordinates  $2a_1, 2b_1$  and the other at  $-2a_1, -2b_1$ . Since both locations contain the same information, either area can be monitored.

There were several factors to consider in setting up the JTC with LCTV's and SLM's. The trade-off is that binary SLM's have superior optical characteristics (such as flatness and contrast ratio) compared to LCTV's. However, the binary SLM's are only capable of displaying binary images. Thus the use of a SLM in the second leg (Fourier plane) would require binarizing the joint power spectrum before displaying it, which would add additional complexity. Our compromise was to encode the input information onto a SEMETEX Sight-Mod  $128 \times 128$  binary spatial light modulator and display the resultant joint power spectrum using a CASIO model TV-430 television set fed with the image signal directly from the camera. Note also that in this experiment we are limiting ourselves to displaying binarized versions of the images. This is a limitation imposed purely by the optics as the neural network is capable of processing either binary or analog imagery.

#### IV. POPART

The operation of the neural network can now be discussed. The experimental setup is shown in Fig. 2. As mentioned earlier, it consists simply of a Unisys 386 personal computer with an imaging board installed and a JTC. A new pattern is first presented to the system, and is preprocessed according to

TABLE II  
RESULTS WITH  $\rho^* = 0.70$  AND CONFIDENCE RATIO SET TO 0.70

Node #	Letter Category	
	Trial 1	Trial 2
0	A	A
1	B, C, D, E, F, G, L, O, P, Q, R, S	B, C, D, E, F, G, L, O, P, Q, R, S
2	H, M, N, U	H, M, N, U
3	I, J, T	I, J, T
4	K	K
5	V, W	V, W
6	X	X
7	Y	Y
8	Z	Z

TABLE III  
RESULTS WITH  $\rho^* = 0$

Node #	Letter Category		
	Trial 1	Trial 2	Trial 3
0	A	A	A
1	B, C, D, E, F, G, H, I, J, L, O, P, Q, R, S, T	B, C, D, E, F, G, I, J, L, O, P, Q, R, S, T	B, C, D, E, F, G, I, J, L, O, P, Q, R, S, T
2	K, X	K, X	K, X
3	M, N, U, V, W	H, M, N, U, V, W	H, M, N, U, V, W
4	Y	Y	Y
5	Z	Z	Z

Note that this represents the coarsest categorization for this network. The letter H switches categories in the second trial, but then stabilizes in node 3. The rest of the network remains stable after a single iteration. The confidence ratio was set to 0.75 in this example.

(1)–(4) electronically. The resultant vector is binarized around the average value and the result is displayed on one side of the JTC. The stored categories are also binarized and displayed on the other side of SLM1 either one at a time or several at once.

The SLM is placed after the 60-cm focal length FT lens (L1) to allow for variable scaling of the FT. A 5-cm focal length relay lens (L2) used in conjunction with this allows magnification of the joint power spectrum for adequate sampling. In addition, the variable scaling allows us to effectively aperture the unwanted replicated Fourier orders caused by the pixelation of the SLM. A DC block is used in order to block the lower Fourier orders that tend to saturate the camera [10]. The joint power spectrum is then displayed on the second leg via the LCTV. Note that the contrast ratio of the LCTV is approximately 30 : 1 and is sufficient to represent the JPS accurately. The resultant correlation plane appears at camera 2 where it is frame-grabbed and stored in the computer. This stored image is scanned in two predefined regions for the maximum intensity value. The maximum value is then used to calculate an optical measure of how well the images correlate.

It should be noted that the resultant intensity values for each cross correlation are extremely sensitive to the input

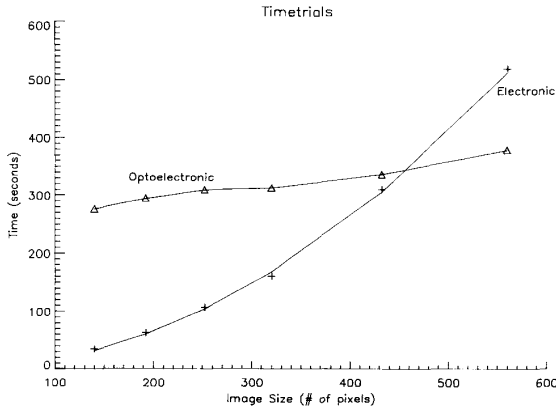


Fig. 5. Results of time trials for different size images. The pluses represent the experimentally measured computer time for completion of two successive presentations of the input. The triangles represent the experimentally measured optoelectronic time. The line following the pluses is a theoretical curve fit to the data demonstrating that the number of calculations increases with the square of the image size (see text). The line connecting the triangles is provided for clarity.

energy, especially given the original binarization of the input and stored patterns for display purposes. Thus in order to calculate the "best match," or largest cross-correlation, the disparate energies between different stored patterns must be considered. In order to achieve this, we used a metric based on the well-known Schwartz inequality [11] or,

$$\left| \int_a^b g_1(t)g_2(t) dt \right|^2 \leq \int_a^b |g_1(t)|^2 dt \int_a^b |g_2(t)|^2 dt \quad (16)$$

,which led us to choose, the metric

$$\text{optical metric} = \sqrt{\frac{P_{\max}}{E_1 \cdot E_2}} \quad (17)$$

where  $P_{\max}$  is defined as the peak intensity value in the autocorrelation plane,  $E_1$  equals the energy of the binarized input pattern and  $E_2$  equals the energy of the binarized stored pattern. In our case the given energy for each pattern is calculated by counting the number of "on" pixels in a given binarized input image.

Once each value has been processed via the metric calculation, the pattern corresponding to the largest resultant value is chosen as the "best match."

At this point the computer value  $T_w$  corresponding to the optical "best match" is calculated electronically and compared to the vigilance parameter (see (8)). Since the node with the "best match" has already been identified, (5)–(10) operate on a single node (or small subset) instead of on the entire set. The optical system, therefore, acts simply to filter out unnecessary calculations that would otherwise have to be calculated through CPU-intensive calculations electronically.

For our experimental demonstration, in order to minimize crosstalk, we decided to present a single stored pattern against a single input on the JTC for each optical correlation test

(as opposed to simultaneous multiple stored patterns against a single input). This was decided on after placing identical patterns in several positions of the SLM (see Fig. 2) and measuring the resultant correlation peaks. Ideally for multiple repetitions without any noise added to the original patterns, the correlation peak values should be identical for all patterns. However, we found that in the best case, when only a single target image and a single unknown image were simultaneously presented, the correlation peaks varied by approximately 8% over 100 trials. For this reason, we chose to serially present the stored images and store the resultant values in an output vector. This also led to the introduction of a "confidence ratio," which indicated our confidence in the optical system. If a given optical metric was within the confidence ratio of the highest optical metric of all the stored patterns, then it was double checked via a computer calculated metric (see (5)). In this way we account for noise intrinsic to our optoelectronic setup.

For all of our tests, a simple binary letter set shown in Fig. 3 was chosen for simplicity. A typical example of operation is shown in Fig. 4. The letters A and C have already been learned by the network. The letter A is input into the system as the unknown. In Fig. 4(a), the input vector (the letter A) is compared against the vector stored in the first node (also the letter A). A picture of the correlation plane output is shown in the figure. The resultant correlation peaks are then framegrabbed and used to calculate the optical metric. A similar operation occurs for the input and the next stored vector (see Fig. 4(b)). In this case the letter C is correlated against the input letter A. When the values from each optical metric are compared, the correlation of the letter A as expected, is larger than the correlation against the letter C. The information that the first node had the best match out of all possible nodes is then used by the rest of the network, which is purely electronic. In this case, the system correctly categorizes the letter A as belonging to the first node.

As discussed in [10], the joint power spectrum consists of fringes with period dependent on image separation. In order for the fringes to be adequately sampled, the spatial frequency of the camera pixels must be at least twice the spatial frequency of the joint power spectrum fringes (the Nyquist criterion). If the separation of the JTC inputs is too large, the camera can not adequately resolve the joint power spectrum fringes. If on the other hand the spacing of the JTC inputs is too small, then there are fewer fringes resulting in a lower correlation peak output. In addition, for small spacing, the lower order noise in the joint power spectrum corrupts the potential correlation peaks. For our system, we discovered that the best performance occurred for our  $20 \times 28$  pixel image when each image was separated by 30 pixels from the centerline of the input SLM.

## V. RESULTS AND ANALYSIS

### A. High-Vigilance Tests

In the first series of tests, we set  $\rho^* = 1$ . The result (not shown) was as expected: 26 separate classifications were formed, each corresponding to a different letter. It should be noted that two separate tests were performed, the "electronic"

TABLE IV  
COMPARISON OF NUMBER OF COMPUTER CORRELATION CALCULATIONS USING A PURELY COMPUTER BASED APPROACH VERSUS AN OPTOELECTRONIC APPROACH

(a) Small Data Set (9 elements)				
$\rho^*$	Number of Computer Correlations		Confidence Ratio	Savings (%)
	Computer	Optoelectronic		
1.00	81	16	0.90	80.25
0.85	54	17	0.85	68.12
0.70	36	13	0.90	63.89

(b) Large Data Set (26 elements)				
$\rho^*$	Number of Computer Correlations		Confidence Ratio	Savings (%)
	Computer	Optoelectronic		
1.00	676	55	0.90	91.86
0.85	520	134	0.80	74.22
0.70	234	124	0.70	47.01

$$\text{Savings} = \frac{\text{computer} - \text{optoelectronic}}{\text{computer correlation calculations}}$$

A sample comparison of the number of computer-calculated metrics, with the computer alone versus with the optoelectronic system. The resultant savings in terms of number of calculations offered by the optics is shown in the column marked "Savings." Clearly, the number of calculations is reduced when the optoelectronic system is used, a) Demonstration for a reduced data set (only 9 inputs) and b) large data set (26 letters). (see text for details).

test with the computer working alone without the optics, and the "optoelectronic test" or with both computer and joint transform correlator working together as a system. The electronic and optoelectronic tests should produce identical results given the same operating parameters (identical  $\alpha, \beta, \rho^*$ ). In fact, we required that the optoelectronic system agree perfectly with the electronic system and that it be self-consistent.

### B. Lower Vigilance Test

A typical lower vigilance test with  $\rho^* = 0.80$  is shown in Table I. The result as expected is that some of the categories have combined into an amalgamation of letters. This trend continues as shown for the case where  $\rho^* = 0.70$  (Table II). The limit occurs when  $\rho^* = 0.0$  where a minimum of six categories are formed (Table III). This behavior is to be expected, since lower vigilance means less discrimination of inputs.

### C. Time Trials

In addition to recording the categorization on each test, the amount of time it took for a given data set of 9 inputs to be processed when the 9 inputs are presented to the system twice, was compared for both the electronic and the optoelectronic cases. The results are shown graphically in Fig. 5. For small images ( $10 \times 14$  pixels), the electronic implementation is faster than the optoelectronic implementation. However as the size of the image increases ( $20 \times 28$  pixels) the optoelectronic implementation becomes faster.

The behavior plotted in Fig. 5 can be understood by considering the number of calculations that each implementation performs. The most time consuming of these calculations is that of the computer correlation between the input and each

stored node. Thus if the optoelectronic system can yield a savings in number of computer calculated correlation metrics, it will fundamentally enhance the operation of the overall device. In the case shown in Fig. 5, the number of operations of the optoelectronic implementation remains approximately constant while the computer implementation increases with the square of the image size. Thus for larger images, the optoelectronic implementation is faster. Of course, the speed ultimately depends on the computer processor and the read-write capability of the optical portion of the system.

The number of computer-calculated correlations the optoelectronic system will execute is highly dependent on several factors, such as the vigilance, the confidence ratio and the data set. The overall number of computer correlation calculations for two different data sets, one with nine elements and the other with 26 elements, is tabulated for several different values of  $\rho^*$  in Table IV. Note that when  $\rho^* = 1$  the optoelectronic system saves as much as 80% in the number of computer correlations for the smaller data set and approximately 90% for the larger data set. In other words, the optical system eliminates unnecessary calculations leaving only those within the "confidence ratio." As the input data sets gets larger, the overall savings will continue to increase.

When the vigilance decreases however, the number of categories also decreases, which diminishes the advantage afforded by the optics. In addition, for low vigilance, each grouping becomes more of an amalgamation of different inputs and less representative of any single input. Therefore, a given input presented a second time to such a system is less likely to match the appropriate stored amalgamation. In many cases this necessitates lowering the confidence ratio in order to increase the error tolerance of the system. This in turn increases the number of overall calculations by requiring more frequent computer correlation doublechecks. As shown in Table IV,

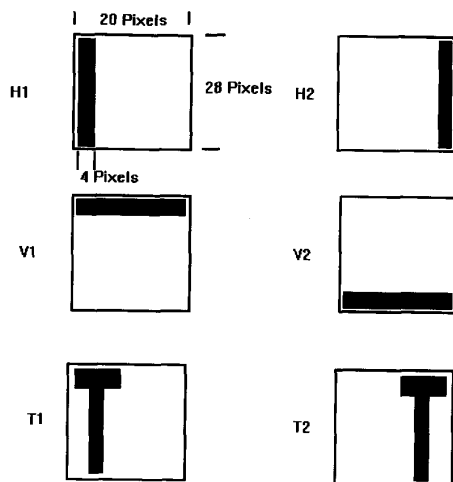


Fig. 6. Sketch of inputs to system for translation invariance tests. H1 and H2, which are four pixel-wide vertical lines, are categorized into the same node regardless of horizontal translation. Tests were also demonstrated for horizontal lines vertically translated (V1, V2) and one horizontal-vertical combination (T1, T2). See Table V for results.

TABLE V  
TRANSLATION INVARIANCE RESULTS

Node Number	Trial 1	Trial 2
0	H1, H2	H1, H2
1	V1, V2	V1, V2
2	T1, T2	T1, T2

Tabulation of results after patterns in Fig. 6 were presented to system (see text for details).

when  $\rho^* = 0.70$ , the performance with the larger data set has fallen by 40% (as compared to the  $\rho^* = 1$  case).

#### D. Translation Invariance

Translation invariance is a desirable characteristic for classifiers, since recognition should occur for a given object regardless of position. The translation invariance of our system is demonstrated as shown in Fig. 6 and Table V, by presenting the same object in different positions and then observing the resultant categorization. Specifically, a 4 pixel wide vertical bar is translated from the left of a  $20 \times 28$  pixelated image to the right side and is consistently categorized into the same node. Similar results are obtained for a horizontal bar and for a combination of horizontal and vertical lines.

It should be noted that although optical correlation is invariant to translation, the actual correlation spot shifts when a translation is introduced, thus the search regions where correlations can be located in the autocorrelation plane are large enough to encompass any potential peaks resulting from a spatial shift in the input.

#### E. Generality of System

As a final example of the generality of the system, several figures not contained in the letter set were constructed and

TABLE VI  
TABULATION OF GENERAL CLASSIFICATION RESULTS

Node Number	Trial 1	Trial 2
0	T1, T2	T1, T2
1	R1	R1
2	H3, H4	H3, H4
3	R2	R2
4	V1, V2	V1, V2
5	H1, H2	H1, H2
6	R3	R3
7	R4	R4

Tabulation of results after presentation with patterns shown in Fig. 7. All nodes were consistently categorized. Note the sensitivity to scale. Patterns H3 and H4 are twice the thickness of H1 and H2 and are grouped together in a separate node. In addition, patterns R1 and R3 are grouped differently although they are identical except for a single  $4 \times 4$  pixelated block.

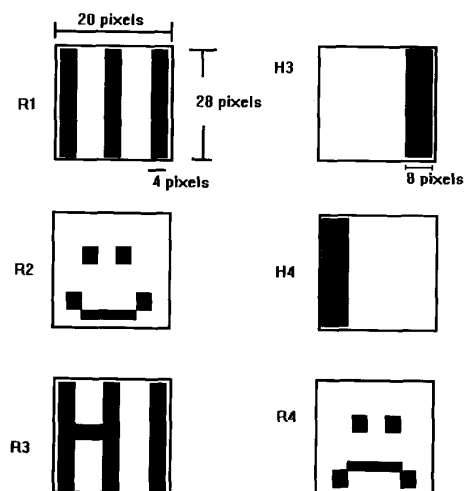


Fig. 7. Demonstration of generality of system. Sketched is several general figures presented to the system (not to scale). See Table VI for results.

presented to the system with  $\rho^* = 1$  (see Fig. 7). As shown in Table VI, the system consistently categorizes all of these patterns. In addition, the high discrimination ability of the system is demonstrated by observing that the patterns stored in nodes 1 and node 6 only vary by a single  $4 \times 4$  pixelated block.

#### VI. CONCLUSIONS

An optoelectronic implementation of ART2-A has been proposed and demonstrated. In order for optoelectronic systems to be useful they must offer the potential for increased performance over a completely electronic approach. The results presented in this paper indicate that for small images, an electronic system would most likely be preferable. However, as the data sets get larger, an optoelectronic implementation would potentially be preferable. The savings arise because the number of operations to calculate the correlation via a JTC is approximately constant, regardless of the size of the

input image, while the number of operations required by a translation-invariant computer-based correlation is proportional to the size of the image (given in pixels) squared.

The advantage of the proposed system is that it combines some of the useful properties of optics and electronics. The parallelism of optics reduces the number of operations and computation time while the computational power of the electronics allows for fault tolerance and system integration.

#### VII. FUTURE WORK

We are continuing to explore the use of ART for categorization. One current research drive is to use a CCD camera to generate the input images instead of computer generated data. Other work includes improving the signal to noise ratio of the JTC, and utilizing purely LCTV's so that analog data can be processed optically in addition to binary data.

#### ACKNOWLEDGMENT

The authors are indebted to the many fruitful discussions with K. Fielding of the Air Force Institute of Technology (AFIT) and V. Ryan of Tufts University concerning the Joint Transform Correlator. They also thank J. Horner for the use of his spatial light modulator. Furthermore, they are indebted to G. Carpenter and T. Kincaid of Boston University, M. Farn of MIT Lincoln Laboratories, D. Wunsch of Boeing Computer Systems, and P. Gianino of Rome Laboratory for their helpful comments and suggestions.

#### REFERENCES

- [1] G. A. Carpenter and S. Grossberg, "ART2: Self-organization of stable category recognition codes for analog input patterns," *Appl. Opt.*, vol. 26, pp. 4919-4930, 1987.
- [2] G. A. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*, pp. 77-88, Mar. 1988.
- [3] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics and Image Proc.*, vol. 37, pp. 54-155, 1987.
- [4] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "ART2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, vol. 4, 1991.

- [5] J. H. Caulfield and D. Armitage, "Adaptive resonance theory of optical pattern recognition," *Appl. Opt.*, vol. 28, no. 19, pp. 4060-4061, 1989.
- [6] D. C. Wunsch II, T. P. Caudell, C. D. Capps, and R. A. Falk, "An optoelectronic adaptive resonance unit," in *Proc. Int. Joint Conf. Neural Networks*, Seattle, WA, July 8-12, 1991, vol. I, pp. 541-549.
- [7] S. Grossberg, "Adaptive pattern classification and universal recoding I: Parallel development and coding of neural feature detectors," *Biol. Cybern.*, vol. 23, pp. 121-134, 1976.
- [8] S. Grossberg, "Adaptive pattern classification and universal recoding II feedback, expectation, olfaction, and illusions," *Biol. Cybern.*, vol. 23, pp. 187-202, 1976.
- [9] B. Javidi and J. Wang, "Binary nonlinear joint transform correlation with median and subset median thresholding," *Appl. Opt.*, vol. 32, pp. 867-980, 1991.
- [10] K. H. Fielding and J. L. Horner, "1-f binary joint transform correlator," *Opt. Eng.*, vol. 29, no. 9, 1990.
- [11] C. S. Weaver and J. W. Goodman, "Technique for optically convolving two functions," *Appl. Opt.*, vol. 5, no. 7, p. 1248, 1966.
- [12] S. Haykin, *Communication Systems*. New York: Wiley, 1978, Appendix 1.



**Jonathan S. Kane** received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1987. In 1989 he received the M.S. degree in electrooptics from the Electro-Optics Technology Center at Tufts University, Medford, MA.

Currently, he is pursuing the Ph.D. degree at Boston University, part-time while working for Rome Laboratory in the optical signal processing branch. His interests include neural networks, analog signal processing, applications of photore-

fractive crystals, and optical correlators.



**Marc J. Paquin** (S'91-M'91) received the B.S. degree in electrical and computer engineering in 1989 and the M.S. degree in electrooptics in 1991, both from Tufts University, Medford, MA.

His research interests include optical signal processing and communications. He is currently with Lancet Technology of Cambridge Massachusetts developing medical and communications software.