Matrix Multiplication by Optical Methods

R. A. Heinz, J. O. Artman, and S. H. Lee

The utilization of coherent optical correlation techniques in the performance of matrix multiplication by optical analog methods has been investigated mathematically. Although the basic concepts have been known for some time, we have not been able to find explicit analyses in the existing literature. Since many correlations other than those corresponding to the desired matrix multiplication terms exist, methods of isolating the desired from the undesired terms are presented. Various spatial configurations for both the input and output arrays are discussed. For the special cases of 2 × 2 matrices the analyses are presented in greater detail. Using simple circ function distributions, the effects of finite-sized array elements and detector apertures are investigated.

I. Introduction

The applicability of optical systems to analog computation involving two independent variables is becoming widely recognized.[1] Suppose we wish to evaluate

$$\iint_{-\infty}^{\infty} f_2(x,y) f_1^*(x,y)dxdy.$$

We can consider this as a special case of the *cross correlation* g of f_1 and f_2:

$$g(x_0,y_0) = \iint_{-\infty}^{\infty} f_2(x,y)f_1^*(x - x_0,y - y_0)dxdy.$$

The two-dimensional fourier transform of $g,\mathfrak{F}[g]$ then is:

$$\mathfrak{F}\{g(x_0,y_0)\} = F_2(\xi_0,\eta_0)F_1^*(\xi_0,\eta_0),$$

where

$$\mathfrak{F}\{f_2(x,y)\} = F_2(\xi,\eta),$$

and

$$\mathfrak{F}\{f_1^*(-x,-y)\} = F_1^*(\xi,\eta).$$

Now

$$\mathfrak{F}[\mathfrak{F}\{g(x_0,y_0)\}] \equiv g(-x_0,-y_0) = \mathfrak{F}\{F_2(\xi_0,\eta_0)F_1^*(\xi_0,\eta_0)\}.$$

Our desired result then is

$$g(0,0) = \mathfrak{F}\{F_2(\xi_0,\eta_0)F_1^*(\xi_0,\eta_0)\}_{x_0=0,\ y_0=0}.$$

An optical analog system which can be used to obtain the above results is shown in Fig. 1. Coherent plane parallel light at the wavelength λ is incident normally upon the plane B, which contains a plate whose amplitude transmission is characterized by f_2. Another plate whose transmittance is characterized by F_1^* is placed in the plane A. f_2 is transformed by lens L_1 to F_2 at the plane A. The amplitude of the resultant transmitted intelligence leaving A is $F_2F_1^*$. This is transformed by lens L_2 to plane C as the cross correlation $g(-x_0, -y_0)$. The desired result, the dot-product, corresponds to the magnitude of the dot at $x_0 = 0$, $y_0 = 0$. The optical system will introduce some additional multiplicative factors. These will be discussed later; they do not affect the basic conditions.

Now we can imagine these functions $f_1^*(x,y)$, $f_2(x,y)$ to correspond to particular two-dimensional distributions—i.e., matrices (much as we can write a matrix upon this two-dimensional piece of paper). Implicit in this development is the conclusion that we can multiply an $l \times m$ matrix by an $m \times n$ matrix optically; i.e., we can form $c = b \times a$ with elements given by

$$c_{ln} = \sum_m b_{lm}a_{mn}.$$

We will now consider this matrix multiplication problem explicitly.

II. Mathematical Preliminaries

We write the two-dimensional fourier transform interrelations between $f(x,y)$ and $F(\xi,\eta)$ as:

$$f(x,y) = \iint_{-\infty}^{\infty} F(\xi,\eta) \exp[i2\pi(\xi x + \eta y)]d\xi d\eta,$$

$$F(\xi,\eta) = \iint_{-\infty}^{\infty} f(x,y) \exp[-i2\pi(\xi x + \eta y)]dxdy.$$

The authors are with the Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

Received 19 January 1970.

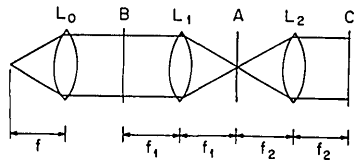

Fig. 1. Coherent optical detection system. The mathematical derivations in the text for simplicity assume all the focal lengths to be the same: $f = f_1 = f_2$.

Where convenient, we will indicate the fourier transform interrelation by a two-headed arrow ($\leftrightarrow$) as follows: $f(x,y)\leftrightarrow F(\xi,n)$. We will assume the matrices under consideration to be composed of arrays of light dots of the form $a\delta(x-x_0)\delta(y-y_0)$. Several fourier transform relations are important here. These are:

(1) Shift Formula: If $f(x)\leftrightarrow F(\xi)$, then

$$f(x - x_0) \leftrightarrow F(\xi)\exp(-i2\pi x_0\xi); \tag{1}$$

(2) Scaling Formula: If $f(x,y)\leftrightarrow F(\xi,n)$, then

$$f(ax,by) \leftrightarrow (1/|ab|)F[(\xi/a),(\eta/b)]; \tag{2}$$

(3) δ-Function Formula: $\delta(x-x_0)\delta(y-y_0)\leftrightarrow \exp[-i\,2\pi(\xi x_0+\eta y_0)]$; (3)

(4) Lens Formula: If $f(x,y)\leftrightarrow F(\xi,\eta)$, then the action of a lens transforming between the front and back focal planes at wavelength λ is given by

$$f(x_1,y_1) \leftrightarrow AF(x_2/\lambda f, y_2/\lambda f) \text{ with } A = (1/\lambda f)\, e^{-i\alpha}, \tag{4}$$

x_2,y_2 are the coordinates on the back focal plane. With no loss of generality we will set α equal to zero.

We will illustrate the method first by examining the multiplication of two 2×2 matrices.

III. Synthesis of the Fourier Transform of the *a* Matrix

Consider the δ-function array of rank two shown in Fig. 2. Note that the rows and columns have been transposed. The reason for this will be transparent later. The role of the additional reference dot r also will become clear later; for simplicity we assume this reference point to be the origin of the coordinate system. We write the amplitude transmission function U_1 of this array as:

$$\begin{aligned} U_1 = {} & a_{11}\delta(x + x_1 + \Delta x)\delta(y - y_1) + a_{21}\delta(x - x_1 + \Delta x)\delta(y - y_1) \\ & + a_{12}\delta(x + x_1 + \Delta x)\delta(y + y_1) \\ & + a_{22}\delta(x - x_1 + \Delta x)\delta(y + y_1) + r\delta(x)\delta(y). \end{aligned}$$

The lens transformation $V_1(x_2,y_2)$ of U_1 at the back focal plane is:

$$\begin{aligned} V_1(x_2,y_2) = {} & (a_{11}/\lambda f)\exp\{-i(2\pi/\lambda f)[-x_2(x_1 + \Delta x) + y_2y_1]\} \\ & + (a_{21}/\lambda f)\exp\{-i(2\pi/\lambda f)[x_2(x_1 - \Delta x) + y_2y_1]\} \\ & + (a_{12}/\lambda f)\exp\{-i(2\pi/\lambda f)[-x_2(x_1 + \Delta x) - y_2y_1]\} \\ & + (a_{22}/\lambda f)\exp\{-i(2\pi/\lambda f)[x_2(x_1 - \Delta x) - y_2y_1]\} + r/\lambda f. \end{aligned}$$

By setting $\beta = (2\pi/\lambda f)x_2$, $\gamma = (2\pi/\lambda f)y_2$, and

$$\begin{aligned} \phi_{11} &= -\beta(x_1 + \Delta x) + \gamma y_1, \\ \phi_{21} &= \beta(x_1 - \Delta x) + \gamma y_1, \\ \phi_{12} &= -\beta(x_1 + \Delta x) - \gamma y_1, \\ \phi_{22} &= \beta(x_1 - \Delta x) - \gamma y_1, \end{aligned} \tag{5}$$

we obtain:

$$V_1 = \frac{a_{11}}{\lambda f}\exp(-i\phi_{11}) + \frac{a_{21}}{\lambda f}\exp(-i\phi_{21}) + \frac{a_{12}}{\lambda f}\exp(-i\phi_{12}) + \frac{a_{22}}{\lambda f}\exp(-i\phi_{22}) + \frac{r}{\lambda f}.$$

We assume that the amplitude transmittance t of a photosensitive medium exposed to V_1 is $t = V_1V_1^*$. We find

$$\begin{aligned} t = {} & \frac{ra_{11}}{(\lambda f)^2}\exp(-i\phi_{11}) + \frac{ra_{21}}{(\lambda f)^2}\exp(-i\phi_{21}) + \frac{ra_{12}}{(\lambda f)^2}\exp(-i\phi_{12}) \\ & + \frac{ra_{22}}{(\lambda f)^2}\exp(-i\phi_{22}) + \frac{ra_{11}^*}{(\lambda f)^2}\exp(i\phi_{11}) + \frac{ra_{21}^*}{(\lambda f)^2}\exp(i\phi_{21}) \\ & + \frac{ra_{12}^*}{(\lambda f)^2}\exp(i\phi_{12}) + \frac{ra_{22}^*}{(\lambda f)^2}\exp(i\phi_{22}) + \frac{r^2}{(\lambda f)^2}, \end{aligned}$$

+ 16 other terms.

The role of the reference r is now clear, it provides terms in t linear in a_{ij} and a_{ij}^*.

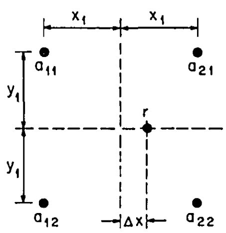

Fig. 2. Optical configuration of a matrix.

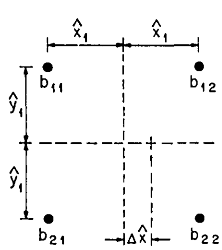

Fig. 3. Optical configuration of b matrix.

IV. Synthesis of the b Matrix and Subsequent $a \times b$ Matrix Multiplication

We construct the b array of rank two as shown in Fig. 3. The coordinate definitions are similar to those used for the a matrix except that now there is no reference at the origin. This b matrix is written in conventional form, however; it is *not* transposed. We have

$$U_2 = b_{11}\delta(x + \hat{x}_1 + \Delta\hat{x})\delta(y - \hat{y}_1) + b_{12}\delta(x - \hat{x}_1 + \Delta\hat{x})\delta(y - \hat{y}_1) + b_{21}\delta(x + \hat{x}_1 + \Delta\hat{x}) \times \delta(y + \hat{y}_1) + b_{22}\delta(x - \hat{x}_1 + \Delta\hat{x})\delta(y + \hat{y}_1).$$

If this array is placed in the plane B of Fig. 1 it is transformed by the lens L_1 to V_2 on the plane A.

$$V_2 = \frac{b_{11}}{\lambda f}\exp(-i\hat{\phi}_{11}) + \frac{b_{12}}{\lambda f}\exp(-i\hat{\phi}_{12}) + \frac{b_{21}}{\lambda f}\exp(-i\hat{\phi}_{21}) + \frac{b_{22}}{\lambda f}\exp(-\hat{\phi}_{22}),$$

where we define

$$\begin{aligned}\hat{\phi}_{11} &= -\beta(\hat{x}_1 + \Delta\hat{x}) + \gamma\hat{y}_1,\\ \hat{\phi}_{12} &= \beta(\hat{x}_1 - \Delta\hat{x}) + \gamma\hat{y}_1,\\ \hat{\phi}_{21} &= -\beta(\hat{x}_1 + \Delta\hat{x}) - \gamma\hat{y}_1,\\ \hat{\phi}_{22} &= \beta(\hat{x}_1 - \Delta\hat{x}) - \gamma\hat{y}_1.\end{aligned} \tag{6}$$

These equations are superficially the same as the previous Eqs. (5), except for the effects of the a transposition. If the plate whose transmittance is t is also in place in plane A of Fig. 1, we obtain the net amplitude transmittance

$$\begin{aligned}V_3 = tV_2 = \; & [rb_{11}a_{11}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{11} - \phi_{11})]\\ &+ [rb_{11}a_{12}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{11} - \phi_{12})]\\ &+ [rb_{11}a_{21}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{11} - \phi_{21})]\\ &+ [rb_{11}a_{22}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{11} - \phi_{22})]\\ &+ [rb_{12}a_{11}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{12} - \phi_{11})]\\ &+ [rb_{12}a_{21}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{12} - \phi_{21})]\\ &+ [rb_{12}a_{12}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{12} - \phi_{12})]\\ &+ [rb_{12}a_{22}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{12} - \phi_{22})]\\ &+ [rb_{21}a_{11}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{21} - \phi_{11})]\\ &+ [rb_{21}a_{21}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{21} - \phi_{21})]\\ &+ [rb_{21}a_{12}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{21} - \phi_{12})]\\ &+ [rb_{21}a_{22}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{21} - \phi_{22})]\\ &+ [rb_{22}a_{11}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{22} - \phi_{11})]\\ &+ [rb_{22}a_{21}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{22} - \phi_{21})]\\ &+ [rb_{22}a_{12}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{22} - \phi_{12})]\\ &+ [rb_{22}a_{22}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{22} - \phi_{22})]\end{aligned}$$

$+$ 84 other terms.

We fourier transform V_3 with lens L_2 to plane C Of the 100 components, the ones of interest to us correspond to the cross correlation of $r \times a$ and b. One such term is the first

$$\begin{aligned}&[rb_{11}a_{11}^*/(\lambda f)^3]\exp[-i(\hat{\phi}_{11} - \phi_{11})]\\ &\quad = A\exp[-i(-\beta\hat{x}_1 - \beta\Delta\hat{x} + \beta x_1 + \beta\Delta x + \gamma\hat{y}_1 - \gamma y_1)]\\ &\quad = A\exp[-i\beta(-\hat{x}_1 + x_1 - \Delta\hat{x} + \Delta x)]\exp[-i\gamma(\hat{y}_1 - y_1)],\end{aligned}$$

where we have set $A = rb_{11}a_{11}^*/(\lambda f)^3$. From Eqs. (1)–(4), this transforms to

$$\begin{aligned}&= A\lambda f\delta(-x_3 + \hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x)\delta(-y_3 - \hat{y}_1 + y_1)\\ &= [rb_{11}a_{11}^*/(\lambda f)^2]\delta(-x_3 + \hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x)\delta(-y_3 - \hat{y}_1 + y_1).\end{aligned}$$

The magnitude of this term is proportional to $b_{11}\,a_{11}$. This component is located on the C plane at $(\hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x, -\hat{y}_1 + y_1)$. We tabulate below the relative magnitudes and locations of this and other terms of interest:

c Component	Relative Magnitude	Location
c_{11}	$b_{11}a_{11}$	$(\hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x, -\hat{y}_1 + y_1)$
c_{12}	$b_{11}a_{12}$	$(\hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x, -\hat{y}_1 - y_1)$
c_{11}	$b_{12}a_{21}$	$(-\hat{x}_1 + x_1 + \Delta\hat{x} - \Delta x, -\hat{y}_1 + y_1)$
c_{12}	$b_{12}a_{22}$	$(-\hat{x}_1 + x_1 + \Delta\hat{x} - \Delta x, -\hat{y}_1 - y_1)$
c_{21}	$b_{21}a_{11}$	$(\hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x, \hat{y}_1 + y_1)$
c_{22}	$b_{21}a_{12}$	$(\hat{x}_1 - x_1 + \Delta\hat{x} - \Delta x, \hat{y}_1 - y_1)$
c_{21}	$b_{22}a_{21}$	$(-\hat{x}_1 + x_1 + \Delta\hat{x} - \Delta x, \hat{y}_1 + y_1)$
c_{22}	$b_{22}a_{22}$	$(-\hat{x}_1 + x_1 + \Delta\hat{x} - \Delta x, \hat{y}_1 - y_1)$

We have assumed here that the a's are real and have dropped the asterisk; this matter will be discussed later.

Now if $\hat{x}_1 = x_1$ and $\Delta x = \Delta\hat{x}$, all these eight components lie on the y axis. Furthermore, if $\hat{y}_1 \neq y_1$ they will be grouped into four pairs:

$$\begin{aligned}c_{11} &= b_{11}a_{11} + b_{12}a_{21} && (0, -\hat{y}_1 + y_1)\\ c_{12} &= b_{11}a_{12} + b_{12}a_{22} && (0, -\hat{y}_1 - y_1)\\ c_{21} &= b_{21}a_{11} + b_{22}a_{21} && (0, \hat{y}_1 + y_1)\\ c_{22} &= b_{21}a_{12} + b_{22}a_{22} && (0, \hat{y}_1 - y_1).\end{aligned}$$

Clearly the transposition of the rows and columns of the a matrix was necessary in order to get the proper combinations entering in the c matrix components. The c matrix components appear on a straight line instead of on a square array; this may be esthetically disturbing but can easily be changed. We must of course make sure that the other ninety-two correlation dots do not interfere with these. Since all the important dots lie on the y axis, the simplest way to do this is to make sure that the remaining dots do not. This will be discussed below for the general matrix multiplication case. Figure 4 indicates the positions of the final correlation dots for $\hat{y}_1 = 4$ and $y_1 = 2$.

V. Multiplication of an $l \times m$ Matrix by an $m \times n$ Matrix

The previous analysis for 2×2 matrices is readily generalized; we wish to form $c = b \times a$, i.e.,

$$c_{ln} = \sum_m b_{lm} a_{mn}.$$

The individual l, m, and n values can be either even or odd. The optical apparatus will produce a large number of correlations corresponding to the fourier transforms of the various elements of $A \times B \times A^*$ ($A \leftrightarrow a + r$; $B \leftrightarrow b$; also $R \leftrightarrow r$). The desired c matrix elements will correspond to certain elements of the subset $R \times B \times A^*$. The R contribution has zero phase. The A^* and B contributions to a given element of c always come from the same *columns* of a and b (we recollect that a is constructed in its *transpose* form). Since the phase of A^* is reversed from that of A, the x component of the phase corresponding to the desired correlations is proportional to $0 - x_1 \mp \Delta x + \hat{x}_1 \pm \Delta\hat{x}$. This is identically zero if $x_1 = \hat{x}_1$ and $\Delta x = \Delta\hat{x}$. The y component of the phase for each of the m contributions to a particular c_{ln} clearly is the same. If we choose $\hat{y}_1$ and y_1 such that $(\hat{y}_1/y_1)^{\pm 1}$ is not an integer, then the y phase components of the various c_{ln} are all different.

We must now insure that undesired $A \times B \times A^*$ type components are shifted off the y axis. The phase of an arbitrary element in the A (or B) array is proportional to $kx_1 + p\Delta x$, where k and p are integers. The phase of an arbitrary element in the $A \times B \times A^*$ product matrix is proportional to $sx_1 + q\Delta x$, where s and q are integers. The allowable p values are 0 and $\pm$ 1. We have the following correspondences:

$$R \rightarrow p = 0,$$

$$\left.\begin{array}{l} A(\text{except } R) \\ B \end{array}\right\} \rightarrow p = \pm 1.$$

(Only one value of p is allowed, depending on whether the origin is to the right or left of the matrix center. The origin is coincident with the optical axis of the system.) Consider the elements in the $A \times B \times A^*$ product matrix for which both (or neither) A and A^* contain the element R. Then $q = \pm 1$. Now consider those elements associated with A containing the element R; for these $q = 0$. Finally, the elements associated with R in A^* correspond to $q = \pm 2$. In summary, we have

$$\begin{array}{ll} q = 0, 1, 2 & \text{(reference to left of matrix center)}, \\ q = 0, -1, -2 & \text{(reference to right of matrix center)}. \end{array} \tag{7}$$

For these elements to be off-axis in the final focal plane, the x component of the phase in $A \times B \times A^*$ must be nonvanishing. Therefore, $sx_1 + q\Delta x \neq 0$. Setting $\Delta x = \epsilon x$, we have $(s + q\epsilon)x_1 \neq 0$ or

$$s \neq -q\epsilon. \tag{8}$$

Clearly, $0 \leq \epsilon \leq 1$; also $|q| \leq 2$ from Eq. (7). Hence the inequality of Eq. (8) is always satisfied for $|s| > 2$. To impose additional restrictions on ϵ we now consider $s = 0, \pm 1$, and ± 2 separately.

If $s = 0$, then $q\epsilon \neq 0$: $q \neq 0 \rightarrow \epsilon \neq 0$, but otherwise arbitrary; $q = 0 \rightarrow$ is excluded since these are the *desired* correlation terms. If $s = \pm 1$, then $q\epsilon \neq \mp 1$: $q = 0 \rightarrow \epsilon$ arbitrary; $q = \pm 1 \rightarrow \epsilon \neq 1$; $q = \pm 2 \rightarrow \epsilon \neq \frac{1}{2}$. If $s = \pm 2$, then $q\epsilon \neq \mp 2$: $q = 0 \rightarrow \epsilon$ arbitrary; $q = \pm 1 \rightarrow \epsilon$ arbitrary; $q = \pm 2 \rightarrow \epsilon \neq 1$.

Therefore, the permissible values of ϵ are: $0 < \epsilon < 1$, $\epsilon \neq \frac{1}{2}$. Recapitulating, matrix multiplication in the general sense is possible for the following parameter values[2]:

$$\begin{array}{ll} x_1 = \hat{x}_1 & (y_1/\hat{y}_1)^{\pm 1} \neq \text{integer}, \\ \Delta x = \Delta\hat{x} & 0 < \epsilon < 1,\ \epsilon \neq \frac{1}{2}. \end{array} \tag{9}$$

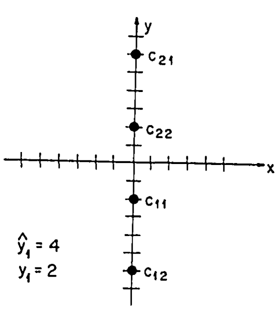

Fig. 4. Positions of the elements of the product matrix $c = b \times a$.

VI. Interpretation of the Output Array

In conventional mathematics, the operational distinction between the two products $b \times a$ and $a \times b$ is obvious. This distinction is made more subtly when performing the analogous optical operations. In the previous examples we formed $c = b \times a$ in the output plane when we transposed the a matrix; to obtain instead the output $c' = a \times b$, we need merely transpose b instead of a. The results for 2×2 matrices are shown in Fig. 5.

We could envisage using our optical analog computer in situations where one matrix was fixed and the other represented sequentially various outputs from another optical system. In such a situation it would be con-

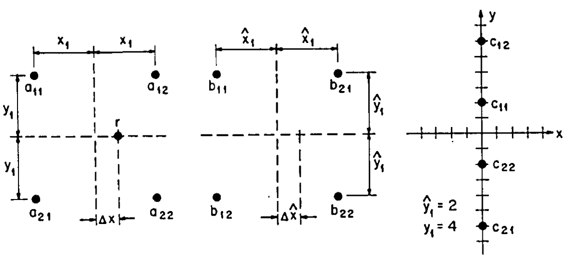

Fig. 5. Optical configurations for obtaining the product matrix $c' = a \times b$.

venient to insert the fixed matrix as a and the others in the position of b. It probably would be simpler to transpose a when making the fourier transform plate and leave b alone; we would then form $c = b \times a$. If we desired $c' = a \times b$, and still did not wish to transpose b, we could transpose a as before and, by choosing the x and y coordinates of the matrix elements appropriately, would be able to form $a \times b$; however, the output matrix dots will lie in a *row* rather than a column. This derivation will not be given here.

The elements of the product matrix were arranged in a column when the input matrices were entered in conventional form. By choosing y_1 and $\hat{y}_1$ appropriately we can separate the column into groups such that each group corresponds either to a row or column of the output matrix. We replace y_1 and $\hat{y}_1$ by y and y_T, where y_T is the y component for the transposed matrix. To make each subgroup of the output represent a row, we choose $y > y_T$ in the 2×2 matrix multiplication case; to group the output by columns, we set $y_T > y$. This is illustrated in Fig. 6. The derivations can be extended to larger matrices.

VII. Arrangement of the Output Array in Conventional Matrix Form; Operations on Vectors

Clearly, because of the nature of optical transformations, the matrices cannot all appear in conventional form. Their dispositions are a matter of taste. If we wish the output c matrix to look like a conventional matrix, we need merely display the a matrix as a row and reflect the b matrix through the x axis ($\hat{y}_1 \leftrightarrow -\hat{y}_1$). We then have

$$\phi_{11} = \beta(-x_1 - \Delta x) \qquad \phi_{12} = \beta(-x_1)$$
$$\phi_{21} = \beta(x_1 - \Delta x) \qquad \phi_{22} = \beta(x_1),$$

and

$$\hat{\phi}_{11} = \beta(-\hat{x}_1 - \Delta\hat{x}) + \gamma(-\hat{y}_1) \qquad \hat{\phi}_{12} = \beta(\hat{x}_1 - \Delta\hat{x}) + \gamma(-\hat{y}_1)$$
$$\hat{\phi}_{21} = \beta(-\hat{x}_1 - \Delta\hat{x}) + \gamma(\hat{y}_1) \qquad \hat{\phi}_{22} = \beta(\hat{x}_1 - \Delta\hat{x}) + \gamma(\hat{y}_1).$$

Proceeding as in Secs. III and IV, we once more obtain the desired c matrix (see the layouts of Fig. 7) now positioned as follows:

$$c_{11} \text{ at } (0,\hat{y}_1) \qquad c_{12} \text{ at } (\Delta\hat{x},\hat{y}_1)$$
$$c_{21} \text{ at } (0,-\hat{y}_1) \qquad c_{22} \text{ at } (\Delta\hat{x}, -\hat{y}_1).$$

The layout of the c elements now agrees with conventional mathematical practice. We have not investigated the locations of the undesired spots for this output configuration.

If the product $a \times b$ rather than $b \times a$ is desired, we need only rearrange the a and b matrices appropriately. The resulting $a \times b$ matrix will then appear in conventional mathematical form; the configurations are shown in Fig. 8. Note that in both of the above cases the a matrix has been transposed.

We did not explicitly discuss the matrix product

$$c = b \times a \qquad (c_{ln} = \sum_m b_{lm}a_{mn}),$$

where the dimensions l, m, and n are all different. There is no need for this since, as in conventional mathematics, we merely supplement our rectangular matrices with additional rows and columns (all bearing 0 elements) until all matrices are square, with dimension given by the maximum of l, m, and n. However, because of their special interest, we will discuss operations involving vectors ($l = 1$ and/or $n=1$).

We will specialize from the 2×2 case discussed earlier in Secs. III and IV. If the lower row of elements of the a matrix of Fig. 2 is discarded, a now represents a column vector in two-dimensional space. (We recollect that the a matrix was prepared optically in the *transpose* form.) Proceeding as before with a_{12}, $a_{22} \rightarrow 0$, $a_{11} \rightarrow a_1$, $a_{21} \rightarrow a_2$, we have

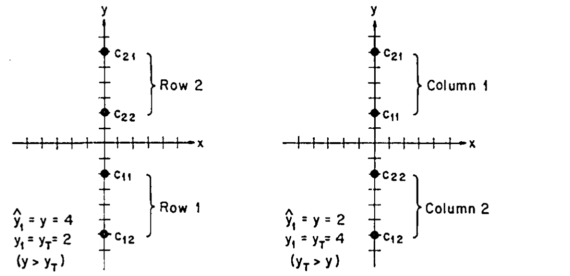

Fig. 6. Separation of the rows and columns of the product matrix.

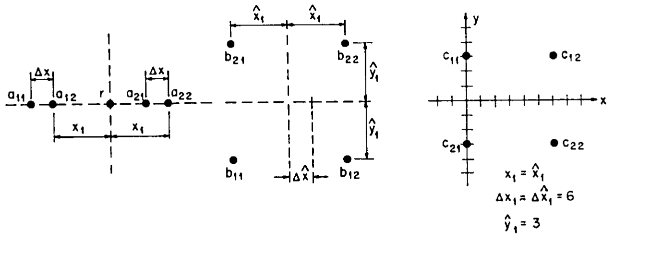

Fig. 7. Optical configurations which yield the product matrix $c = b \times a$ in conventional mathematical form.

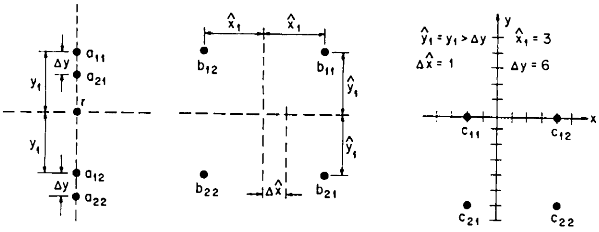

Fig. 8. Optical configurations which yield the product matrix $c' = a \times b$ in conventional mathematical form.

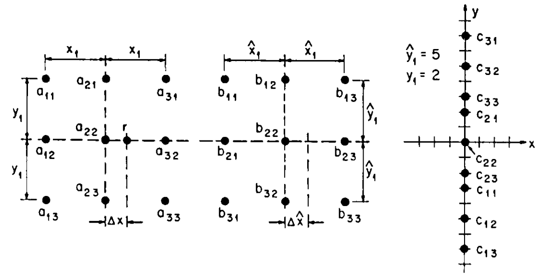

Fig. 10. Optical configurations for multiplication of 3×3 matrices.

$$c_1 = b_{11}a_1 + b_{12}a_2 \qquad (0, -\hat{y}_1 + y_1)$$
$$c_2 = b_{21}a_1 + b_{22}a_2 \qquad (0, \hat{y}_1 + y_1).$$

The output, an inverted column vector, is displayed in Fig. 9. The column vector can be made to appear in standard mathematical form by merely reflecting the b matrix through the x axis. As expected, a matrix operating on a column vector yields a column vector; to get an output columnar vector optically we had to prepare the input column vector optically in the transpose (i.e., row) form. To form the vector dot product

$$c = \sum_m b_m a_m$$

in our two-dimensional vector space, we now set b_{21}, $b_{22} = 0$ and $b_1 = b_{11}$, $b_2 = b_{12}$, thus forming the row vector b. Multiplying by the column vector a previously prepared, we obtain

$$c = b_1a_1 + b_2a_2 \qquad (0, -\hat{y}_1 + y_1).$$

VIII. Multiplication of Two 3 × 3 Matrices

An analysis similar to that of Secs. III and IV was carried out for the product of two 3×3 matrices. The form of the input matrices and the resultant product matrix are presented in Fig. 10. In this case we transposed a and thus have evaluated $c = b \times a$. Since in this example we chose $\hat{y}_1 > 2y_1$ (i.e., $y > 2y_T$), each subgroup of three elements in the output plane represents a row of the product matrix. As in the 2×2 case, we can manipulate the input array so as to obtain an output matrix in conventional mathematical form.

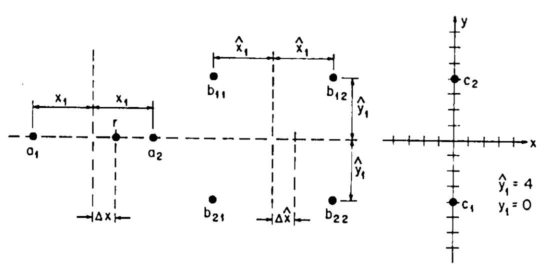

Fig. 9. Optical configurations for operation of a matrix upon a vector.

IX. Output Considerations, Effect of Finite Aperture Dimensions

In the discussions of the preceding sections, the matrices were composed of δ-function arrays. Although this simplified the mathematical analyses considerably, it is a poor representation of actual apertures. A better approximation to an aperture would be the circ function:

$$\mathrm{circ}(r) = \begin{cases} b & 0 \le r \le a \\ 0 & r > a \end{cases} \qquad r = (x^2 + y^2)^{1/2}$$

$$\mathfrak{F}\{\mathrm{circ}(r)\} = \frac{aJ_1(2\pi a\rho)}{\rho},$$

where $\rho^2 = \zeta^2 + \eta^2$. The mathematics is now somewhat more complex; however, we need merely note that the results in the output c plane will now represent correlations or convolutions of several circ functions.

Initially we consider the case where the reference r is a δ-function and the other elements in a and b are circ functions. Since terms in the product matrix correspond to certain elements of $\mathfrak{F}(R \times B \times A^*)$, we will require the convolution of a δ function with the cross correlation of two circ functions. Since both the circ and δ functions are real and centrosymmetric, the terms convolution and correlation are equivalent here. We will further assume that all circ functions have the same radius, a.

The correlation, $C_1(s)$, of two circ functions is merely the overlap area of two circles[3] (see Fig. 11). We find:

$$C_1(s) = \begin{cases} 2abc[a\cos^{-1}(s/2a) - \frac{1}{2}s(1 - s^2/4a^2)^{\frac{1}{2}}] & 0 \le s \le 2a \\ 0 & s > 2a, \end{cases}$$

where a is the radius of the circ functions, b and c are the amplitudes of the circ functions, and s is the distance between circ centers. To obtain the final output $C_2(s)$, we now correlate $C_1(s)$ with the δ function representing r:

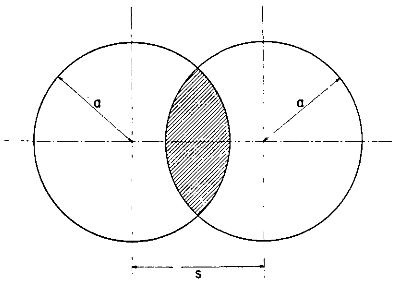

Fig. 11. Geometrical construction equivalent to the cross correlation of two circ functions.

$$C_2(s) = \int_0^\infty p\delta(s_1 - s)C_1(s_1)ds_1 = pC_1(s),$$

where p is the reference amplitude and s the distance between the centroids of $\delta(s_1)$ and $C_1(s_1)$; or simply,

$$C_2(s) = \begin{cases} 2abcp[a \cos^{-1}(s/2a) - \frac{1}{2}s(1 - s^2/4a^2)^{\frac{1}{2}}] & 0 \leq s \leq 2a \\ 0 & s > 2a. \end{cases}$$

Since optical detectors respond linearly to intensity, a plot of $[C_2(s)]^2$ normalized to unity at the origin is shown as the dashed curve in Fig. 12.

The more general case where the reference is also a circ function follows simply. Instead of $C_2(s)$, we now require

$$C_3(s) = \int_0^\infty p \operatorname{circ}(s_1 - s)C_1(s_1)ds_1.$$

$C_3(s)$ was evaluated on a computer. The solid curve in Fig. 12 is a plot of $[C_3(s)]^2$. We normalize to unity at the origin. The triple circ correlation, as expected, does not drop off as rapidly with s as the double circ correlation.

Since optical detectors are finite in area, their response would correspond to some sort of integral over these intensity distributions. The dashed curve in Fig. 13 represents the average intensity seen by a detector (with circ function response) upon which the double circ correlation output is centered. As long as the detector radius d is less than or comparable to the circ function radius a, the average intensity drops off almost linearly with d. When $d > 2a$, the average intensity $\propto d^{-2}$. A similar plot for the triple circ correlation case is shown as the solid curve in Fig. 13. The total detector response, TDR, to the illumination was also computed:

$$TDR = \int_0^d I(s)2\pi s ds.$$

The double circ correlation case is shown as the dashed curve in Fig. 14. For $d \geq 2a$, the TDR has reached its maximum value. The corresponding plot for the triple circ case is shown as the solid curve in Fig. 14. As expected the TDR for the triple circ case rises slower than the analogous curve for the double circ case and assumes its maximum value for $d \geq 3a$.

X. Remarks on the Symmetry and Phase of Individual Matrix Elements

In Sec. IV and thereafter we made the assumption that $a^* = a$. Thus, although our desired output actually evaluates $b \times a^*$, we do compute here $c = b \times a$. However, an array may have nonzero phases assigned to the various members. This difficulty can be disposed of by simply starting with a^* rather than a when synthesizing the a plate.

Alternatively we can use the convolution, h, of f_1 and f_2:

$$h(x_0,y_0) = \iint_{-\infty}^{\infty} f_2(x,y)f_1(x_0 - x, y_0 - y)dxdy.$$

$$\mathfrak{F}\{h(x_0,y_0)\} = F_2(\xi_0,\eta_0)F_1(\xi_0,\eta_0),$$

$$h(-x_0,-y_0) = \mathfrak{F}\{F_2(\xi_0,\eta_0)F_1(\xi_0,\eta_0)\},$$

$$h(0,0) = \mathfrak{F}\{F_2(\xi_0,\eta_0)F_1(\xi_0,\eta_0)\}_{x_0=0,y_0=0}.$$

Now if $f_1(-x,-y) = f_1(x,y)$, then

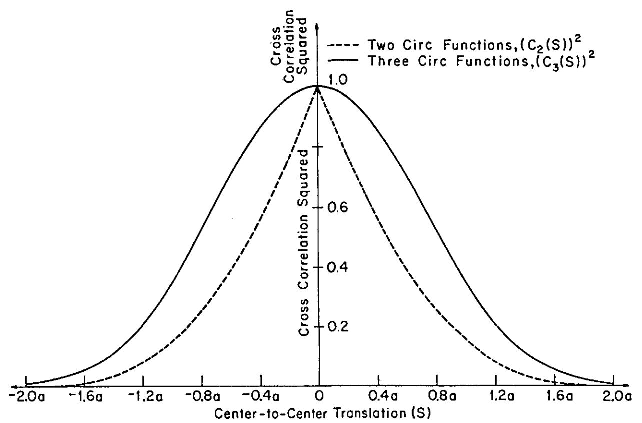

Fig. 12. Intensity distributions corresponding to the double and triple circ cross correlations.

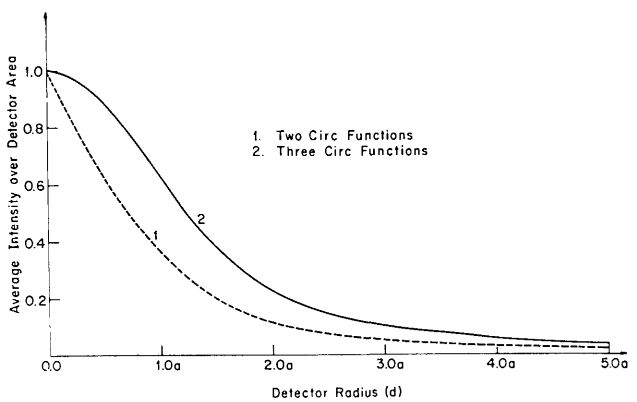

Fig. 13. Average intensity of the cross correlation outputs as a function of detector radius.

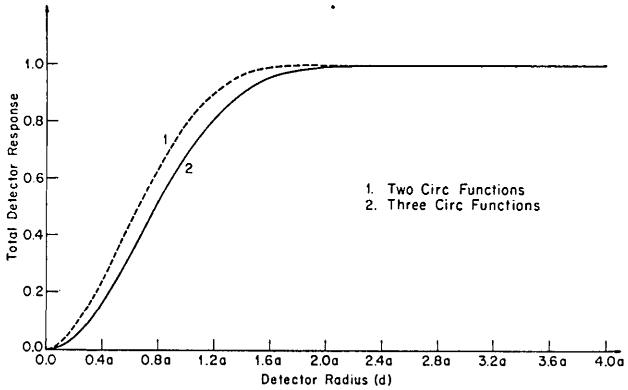

Fig. 14. Total detector response as a function of detector radius.

$$h(0,0) = \iint_{-\infty}^{\infty} f_2(x,y)f_1(x,y)dxdy,$$

which is the desired result. Except in trivial cases, the matrices a and b are not even $[a(-x,-y) \neq a(x,y)]$. Fortunately, the desired matrix product corresponds to a *subset* of the $b \times a$ convolution, and the mathematical restrictions are less severe. It is sufficient here to have the matrix dots be symmetrical about their centroids; this is satisfied by δ-function distribution dots, circ-function distribution dots, and gaussian dots. The desired convolutions do not appear in the same positions on the C plane as the cross correlations. The further analysis required to elucidate the convolution dot positions will not be given here.

One important practical problem arises when negative elements appear in the matrices. The requisite 180° phase shift can be obtained by inserting a binary phase plate adjacent to the matrix plate in plane B (Fig. 1). By selectively etching a flat quartz plate, a phase plate can be produced which will impart the desired 180° phase shift to the appropriate matrix elements. Such a plate has been described by Burckhardt.[4]

References

1. The following texts serve as general references for this material: J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, 1968); A. Papoulis, *Systems and Transforms with Applications in Optics* (McGraw-Hill, New York, 1968); A. Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, New York, 1962). Matrix multiplication is mentioned briefly on pp. 97 and 98 of the article by L. J. Cutrona, in *Optical and Electrooptical Information Processing*, J. T. Tippett *et al.*, Eds. (Mass. Inst. of Technol. Press, Cambridge, 1965).
2. This condition on y_1 and $\hat{y}_1$ is the most general and may be overrestrictive in any particular example. For any case where the order of the matrices is not infinity much less severe restrictions are imposed. For example, for 2×2 and 3×3 matrices we require only $y_1 \neq \hat{y}_1$.
3. Analyses equivalent to some of this material can be found, for example, on pp. 116–120 of Goodman (Ref. 1) and on pp. 32–35 of Papoulis (*Systems and Transforms with Applications in Optics*).
4. C. B. Burckhardt, J. Opt. Soc. Amer. **59**, 1544A (1969).

1971 CORNELL ELECTRICAL ENGINEERING CONFERENCE. HIGH FREQUENCY GENERATION AND AMPLIFICATION—DEVICES AND APPLICATIONS 17–19 AUGUST 1971

The Third Biennial Cornell Electrical Engineering Conference, scheduled for 17–19 August 1971, will present the topic, High Frequency Generation and Amplification—Devices and Applications. Since the original Cornell conference on this subject in 1967, there has been an extensive growth of the field and, in particular, **new areas of application for solid state and quantum electronic devices are now being exploited.** Of special interest is the use of these new devices in the public sector, for example, as applied to marine and vehicular navigational equipment including radar and transponders, the monitoring of pollutants, bioelectronics, and medical treatment, and diagnostics. As in previous conferences of this series, contributed and invited papers will summarize the state of the art through seminar, tutorial, and research presentations, and the Proceedings will be published. The technical program will include general papers on applications, as well as material on microwave semiconductor generator and amplifier devices, and submillimeter and infrared quantum electronic components. Avalanche IMPATT and TRAPATT diodes, gallium arsenide transit time and bulk devices, such as Gunn and LSA diodes, operated as oscillators and amplifiers, and microwave transistors, will be included. In addition to diode and circuit research, the development of applications in microwave systems will also be presented. Means of achieving high power and efficiency, electronic frequency tuning and phase shifting, low noise, frequency and phase locking, wide dynamic range in amplification and other performance properties will be covered. **Quantum electronic topics will include new excitation mechanisms—optical, electrical, chemical—as well as parametric processes in bulk material, together with measurement applications in the millimeter and submillimeter regions.** Both individual and family accommodations will be available in dormitories and area motels. Sightseeing and family entertainment programs will be arranged. The program has as its cosponsor the Office of Naval Research. The Institute of Electrical and Electronic Engineers will be a cooperating society through its individual Groups in Circuit Theory, Electron Devices, and Microwave Theory and Techniques. The Chairman of the Conference is Herbert J. Carlin, Director of the Cornell School of Electrical Engineering. Requests for information and submitted technical paper abstracts should be directed to Lester Eastman, Program Chairman, Cornell School of Electrical Engineering, Phillips Hall, Ithaca, New York 14850.