

OPTICAL IMPLEMENTATION OF SYSTOLIC ARRAY PROCESSING

H. J. Caulfield^{*}, W. T. Rhodes⁺, M. J. Foster[†], and Sam Horvitz⁻

* Innovative Optics, Inc., P.O. Box 1275, Concord, MA. 01742

+ Georgia Institute of Technology, School of Electrical Engineering,
Atlanta, GA. 30332

† Carnegie-Mellon University, Computer Science Department,
Pittsburgh, PA. 15213

- Naval Underwater Systems Center, PA4, New London, CT. 06320

SUMMARY

We show how certain algorithms for matrix-vector multiplication can be implemented using acousto-optic cells for multiplication and input data transfer and using CCD detector arrays for accumulation and output of the results. No 2-D matrix mask is required; matrix changes are implemented electronically. A system for multiplying a 50-component nonnegative-real vector by a 50 x 50 nonnegative-real matrix is described. Modifications for bipolar-real and complex-valued processing are possible, as are extensions to matrix-matrix multiplication and multiplication of a vector by multiple matrices.

INTRODUCTION

During the past several years, Kung and Leiserson at Carnegie-Mellon University (refs. 1,2) have developed a new type of computational architecture which they call "systolic array processing". Although there are numerous architectures for systolic array processing, a general feature is a flow of data through similar or identical arithmetic or logic units where fixed operations, such as multiples and adds, are performed. The data tend to flow in a pulsating manner, hence the name "systolic". Systolic array processors appear to offer certain design and speed advantages for VLSI implementation over previous calculational algorithms for such operations as matrix-vector multiplication, matrix-matrix multiplication, pattern recognition in context, and digital filtering. This paper grew out of our desire to explore the possibility of improving systolic array processors by using optical input and output. We will concentrate on describing the particular case of matrix-vector multiplication, but note that many other operations can be performed in an analogous manner.

SYSTOLIC MULTIPLICATION OF A VECTOR BY A MATRIX

The problem we address is that of evaluating a vector \vec{y} given by

$$\vec{y} = \underline{A}\vec{x} \quad (1)$$

where \underline{A} is an n by n matrix, and \vec{x} and \vec{y} are n -component vectors. We assume that \underline{A} has bandwidth w , i.e., all of its non-zero entries are clustered in a band of width

w around the major diagonal. Such matrices arise frequently in the solution of boundary value problems for ordinary differential equations. A systolic array that solves this problem is introduced by Kung and Leiserson (ref. 1,2) and will be reviewed briefly here.

A systolic array for multiplying a matrix of bandwidth w by a vector of arbitrary length has inner-product cells. The array for bandwidth 4 is shown in figure 1. Each of the four heavy boxes represents an inner-product cell, capable of updating the vector component y_i according to the replacement

$$y_i \leftarrow y_i + a_{ij}x_j \quad (2)$$

The cells act together at discrete time intervals, or beats, with half of the cells active on each beat. The elements of the matrix \underline{A} are input from the right, and the vector \vec{x} is input from the top. Zeroes are input from the bottom, and accumulate terms of the vector \vec{y} as they move upward.

Figure 2 traces the action of the array for several beats, or pulsations, showing the terms of \underline{A} and \vec{x} and the partial terms of \vec{y} that are in each cell on each pulsation. Thus on pulsation 1, $y_1 = 0$ is entered. In pulsation 2, x_1 is entered. In pulsation 3, y_1 becomes $a_{11}x_1$. In pulsation 4, y_1 becomes $a_{11}x_1 + a_{12}x_2$. In pulsation 5, y_1 exits. Every other pulse another y_j exits and on that same pulse another y_k is inserted (at an initial value of zero).

OPTICAL SYSTOLIC ARRAY PROCESSING

Key features of the systolic array approach to matrix-vector multiplication are (1) a regular, directed flow of data streams, (2) multiplication, and (3) addition or accumulation. These features are also characteristic of many optical signal processing systems, and it should come as no great surprise that optical implementations of systolic architectures are possible. Since both bulk and surface acoustic waves are routinely used in optical signal processing to produce a moving stream of data and for multiplication of data, it seems natural to use these components for optical systolic array processing.

We choose as our example the simple matrix-vector multiplication

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (3)$$

assuming initially that all quantities in this equation are real and nonnegative. The basic concept is illustrated with the help of figure 3. The system shown consists of an acoustooptic modulator illuminated by the collimated light from three LEDs, a Schlieren imaging system, and three detectors connected to a CCD analog shift register. At the moment illustrated in the figure, modulating signals proportional to x_1 and x_2 have been input to the acousto-optic modulator driver,

producing short grating segments in the acousto-optic cell. As the x_1 grating segment passes in front of LED number 2 (the situation shown in the figure), that LED is pulsed in proportion to matrix coefficient a_{11} . The transmitted light, proportional in intensity to $a_{11}x_1$, is imaged onto CCD detector 2, which sends a proportional charge to an associated "bin" in the shift register.

The x_1 and x_2 grating segments now travel so as to be in front of LED's 1 and 3, respectively. At the same time, the accumulated CCD charge from detector 2 is shifted one bin, in the direction indicated by the arrow labeled "output" in the figure. LED's 1 and 3 are now pulsed in proportion to a_{21} and a_{12} , respectively. Since these LED's illuminate detectors 3 and 1 via grating segments x_1 and x_2 , charge is generated by these detectors in proportion to $a_{21}x_1$ and $a_{12}x_2$, respectively, and accumulated in the corresponding shift register bins.

In the next increment of the system, charges are again shifted, with accumulated charge in proportion to $a_{11}x_1 + a_{12}x_2$, or y_1 being output. The charge packet now associated with detector 2 (already proportional to $a_{21}x_1$) is augmented by a final strobe of LED 2 by an amount proportional to $a_{22}x_2$. A final two shifts of the CCD charge packets bring charge proportional to $a_{21}x_1 + a_{22}x_2$, or y_2 , to the output, and the operation is complete.

The system illustrated is easily expanded to accommodate matrix-vector operations of higher dimensionality. If \vec{y} and \vec{x} are N -component vectors and \underline{A} an $N \times N$ matrix, the maximum number of LED's required is $2N-1$ (the number of diagonals of the matrix), and the number can be smaller if \underline{A} has a smaller bandwidth.

Numerous variations of the system of figure 3 are possible. Figure 4, for example, shows the LED's replaced by a single light source and an array of modulators. The CCD shift register has been replaced by stationary detectors and integrators combined with a second acousto-optic cell, which serves to deflect light to the correct detector/integrator. The acousto-optic deflector approach to sorting output data may facilitate greater system dynamic range than is achievable with CCD detector arrays.

BIPOLAR AND COMPLEX-VALUED COMPUTATIONS

It was assumed in the preceding section that all elements of the matrix and input vectors were nonnegative-real. In practice, most matrix-vector multiplication operations of importance involve bipolar-real or complex-valued vectors and matrices, and some means must be employed for handling them. If the elements are real valued, but not necessarily nonnegative, a two-component decomposition scheme described in ref. 3 can be employed. For complex-valued processing, several schemes have been described (ref. 4). One of these involves a three-component decomposition of complex numbers according to ref. 5,

$$z = z_0 + z_1 \exp [i2\pi/3] + z_2 \exp [i4\pi/3] \quad , \quad (4)$$

where z_0 , z_1 and z_2 are nonnegative-real. Another involves biased real and imaginary components (ref. 6). All such methods lead to some additional processor complexity and to a reduction in the size of the vectors and matrices that can be accommodated.

OPERATING PARAMETERS OF A TYPICAL SYSTEM

Matrix size limitations are imposed by the acousto-optic modulator. Consider a system using for input a bulk acousto-optic cell with a 100 MHz bandwidth and a 10 μ sec time window. We estimate that such a cell should accommodate 100 LED/lenslet combinations operating side by side, allowing multiplication of a 50-component nonnegative-real vector by a 50 x 50 nonnegative-real matrix. Achievable dynamic range depends on CCD detector dynamic range and on the correction of LED and acousto-optic modulator nonlinearities; it is too speculative to suggest numbers at this time. Operating speed is determined by the amount of time it takes to shift the components of \vec{x} through the acousto-optic cell, plus setup and final readout time. For the 10 μ sec window cell under consideration, it takes 5 μ sec to get the x_1 grating segment to the middle of the acousto-optic cell, at which time the first LED pulse occurs. The last LED pulse occurs 10 μ sec later, when x_{50} finally passes the midpoint of the cell. Following that pulse, an additional 50 μ sec are required to read y_{50} out of the shift register. The time required for the 50 x 50 matrix-vector multiplication is thus 10 μ sec processing time and 10 μ sec latency, for a total of 20 μ sec. During the processing interval, a total of 2500 multiplications are performed, at a rate of 2.5×10^8 multiplications per second. With suitable encoding of the data (refs. 3,4), this corresponds to a processing rate of 6.25×10^7 bipolar-real multiplications per second or 2.78×10^7 complex multiplications per second.

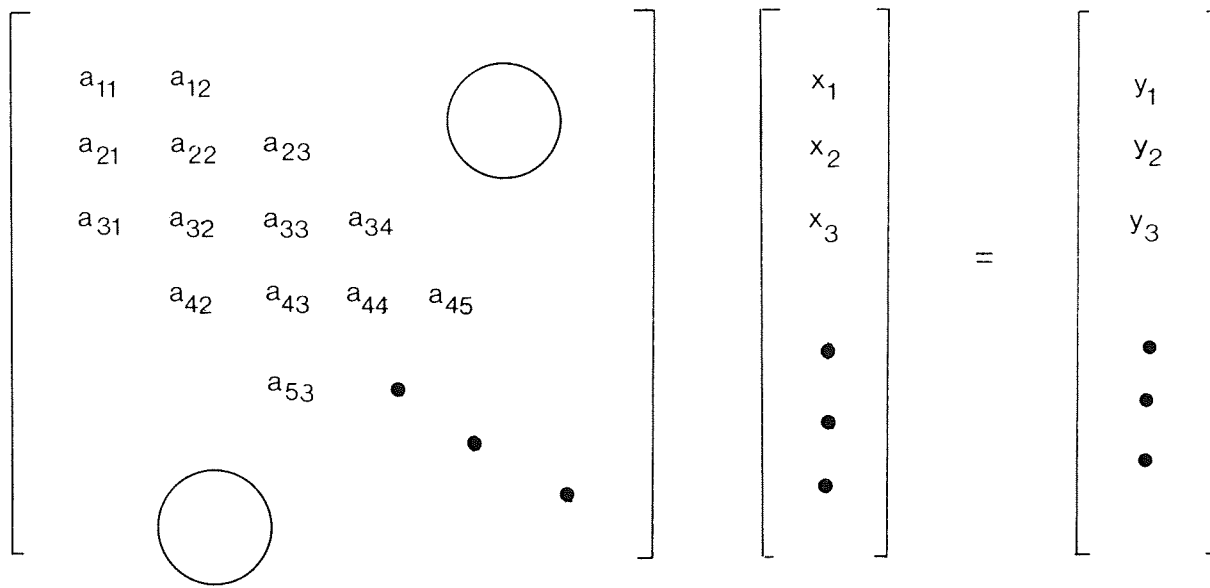
VARIATIONS

The system described does not exploit the two-dimensionality of the optical system. More than one matrix can multiply the same input vector at the same time if the single linear LED/lenslet and detector arrays are replaced with a collection of linear arrays, one above the other. Shear wave acousto-optic modulators, with nearly square window formats, can accommodate perhaps 20 such linear arrays, allowing 20 separate matrices to multiply the same input vector at the same time.

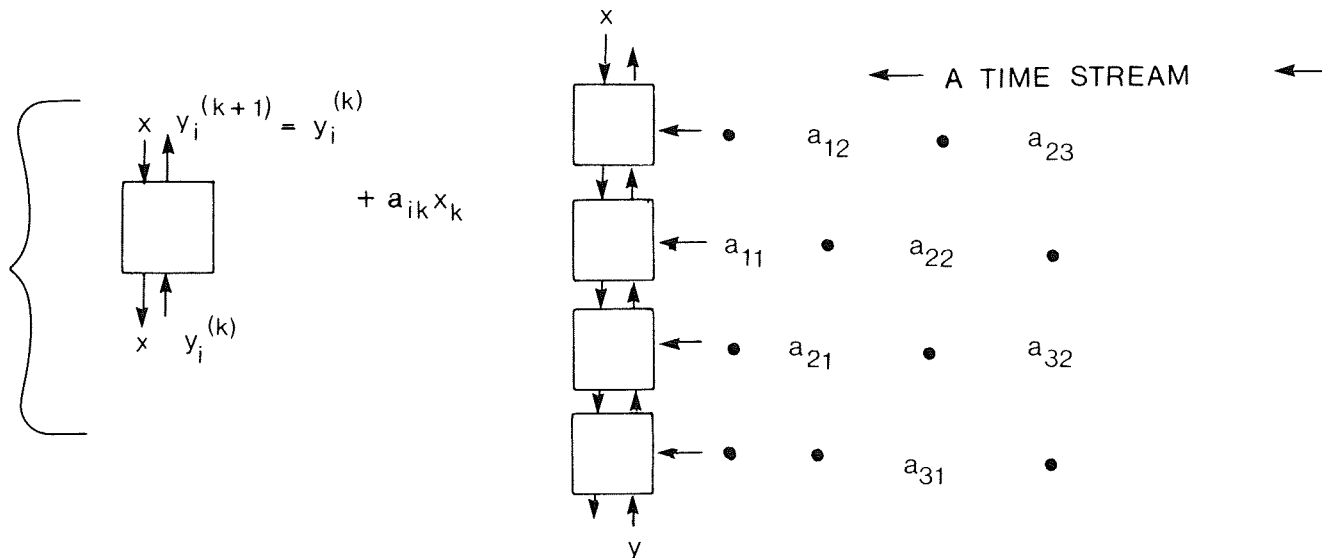
Matrix-matrix multiplication can be performed with related systems using multiple acousto-optic cells, or, alternatively, single cells with multiple driver/transducers. Figure 5 shows one possible arrangement for multiplication of two 2 x 2 nonnegative-real matrices. In general, for such a scheme, multiplication of two N x N matrices requires two multi-transducer acousto-optic modulators with 2N-1 transducers each. Alternatively, one such multi-transducer cell could be used, illuminated by a 2-D array of N^2-2 LED's.

REFERENCES

1. H. T. Kung and C. E. Leiserson, "Systolic Array Apparatuses For Matrix Computations", U.S. Patent Application, Filed December 11, 1978.
2. H. T. Kung and C. E. Leiserson, "Systolic Arrays for VLSI", in Introduction To VLSI, by C. A. Mead and L. A. Conway (Addison-Wesley, Reading, Mass., 1980).
3. H. J. Caulfield, D. Dvornik, J. W. Goodman and W. T. Rhodes, "Eigenvector Determination By Noncoherent Optical Methods", Appl. Opt. 20, (1981) 2263-2265.
4. A. R. Dias, "Incoherent Optical Matrix-Vector Multiplication For High-Speed Data Processing", Ph.D Dissertation, Stanford University, 1980, (University Microfilm No. 8024641).
5. J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully Parallel, High-Speed Incoherent Optical Method For Performing Discrete Fourier Transforms", Opt. Lett. 2, (1978) 1-3.
6. J. W. Goodman, A. R. Dias, L. M. Woody, and J. Erickson, "Some New Methods For Processing Electronic Image Data Using Incoherent Light", In Optica Hoy y Manana (Proceedings Of ICO-11 Conference, Madrid, Spain, 1978), J. Bescos, A. Hidalgo, L. Plaza, J. Santamaria, eds., pp. 139-145.



(a)



(b)

(c)

Figure 1.- Systolic multiplication of vector \vec{x} by banded matrix \underline{A} . Traditional representation of this operation is shown in (a). Basic cell for this operation is shown in (b). Flow of \vec{x} , \vec{y} and \underline{A} data is shown in (c).

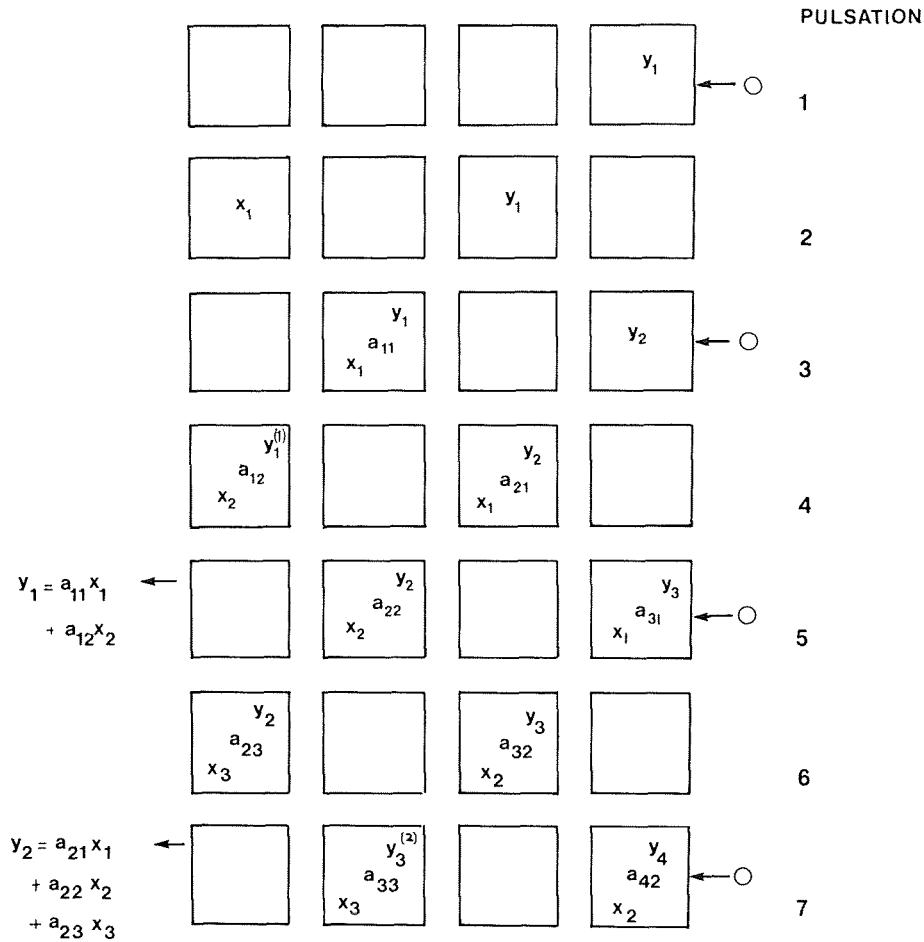


Figure 2.- The first seven pulsations of the processor of figure 1(c) are shown here and described in the text.

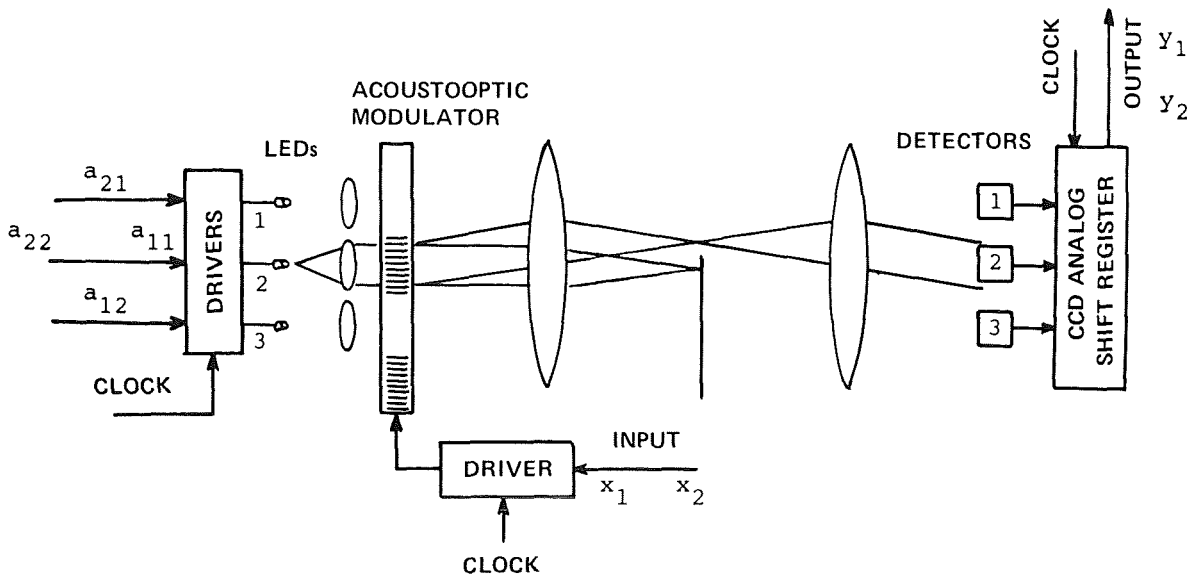


Figure 3.- Optical implementation of systolic array processor of figure 1(c).

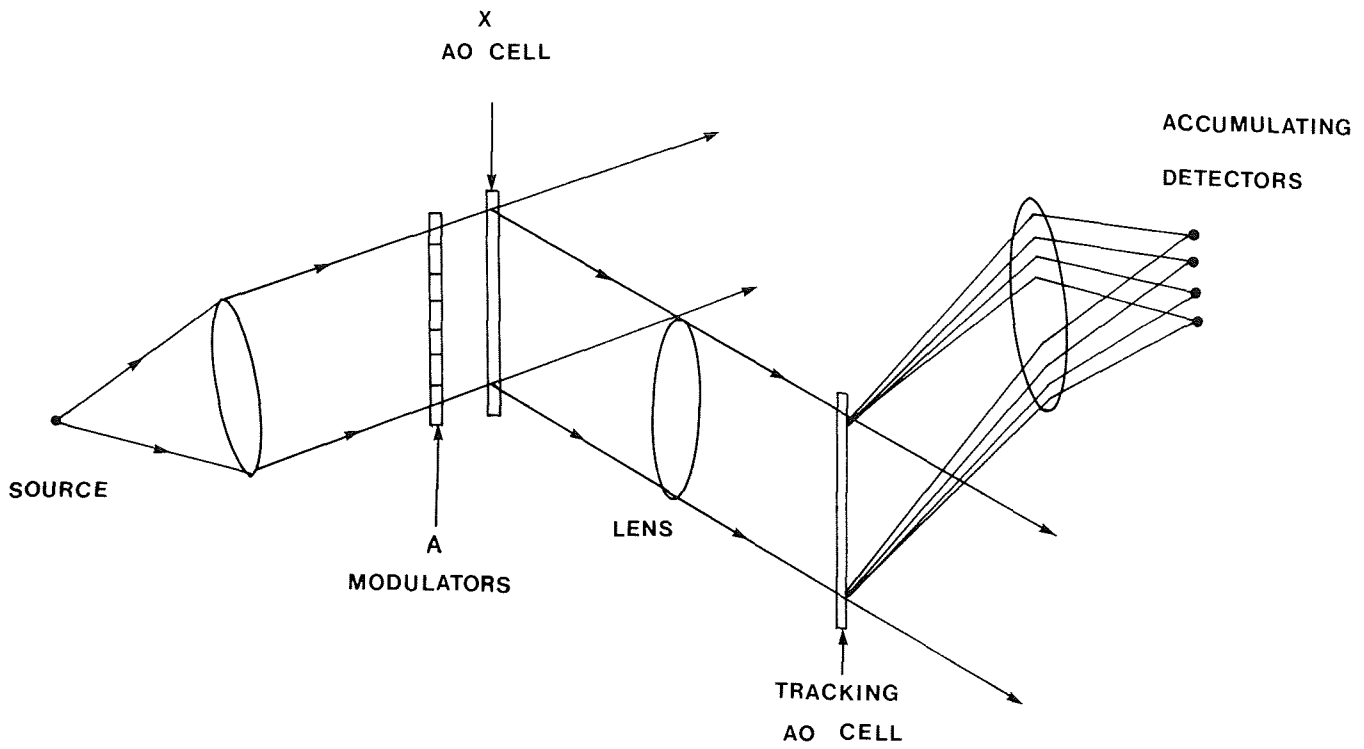


Figure 4.- An alternative optical implementation of the processor of figure 1(c).

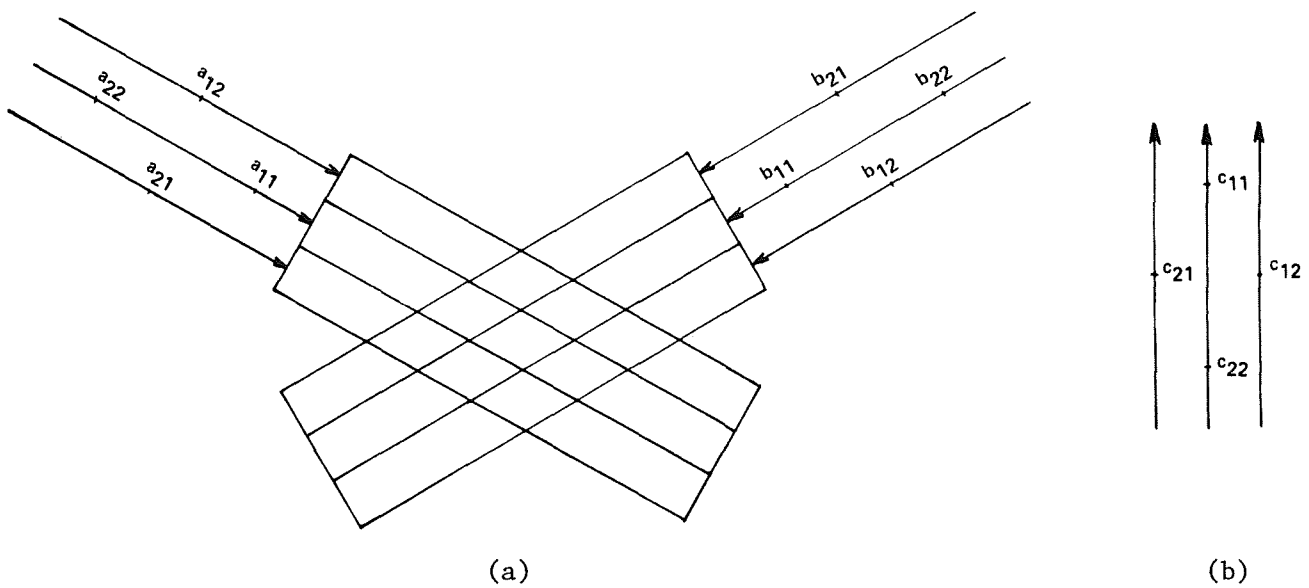


Figure 5.- Use of crossed acousto-optic cells to produce $\underline{A} \times \underline{B} = \underline{C}$. Input information flow is shown in (a), and calculated \underline{C} values are produced as indicated in (b).