coating resistance deteriorates appreciably the noise performance of beam-type tubes.

The value of S increases more steeply than is expected from (3) for a beam current above 300 μA [see Fig. 3(a)]. This increase in S may be ascribed to the noise increase at frequencies close to the plasma frequency at the potential minimum as has been theoretically predicted.[10],[11] The computed plasma frequency at the potential minimum for the present gun is 3.7 GHz for $I_0 = 500$ μA.

ACKNOWLEDGMENT

The authors wish to thank Prof. S. Okamura, University of Tokyo, for valuable discussions, C. Shibata and T. Akioka, Japan Radio Company, for preparing the tube, J. M. Hammer, RCA Laboratories, for his advice in improving the tube, and Y. Mizukaga and Y. Sawayama, Tokyo Shibaura Electric Company, who supplied the electron guns.

K. NINOMIYA
T. OKOSHI
Dept. of Elec. Engrg.
University of Tokyo
Tokyo, Japan

REFERENCES

[1] T. Okoshi, "A new method of measuring noise parameters S and Π of an electron beam," *IEEE Trans. Electron Devices (Correspondence)*, vol. ED-11, pp. 37–38, January 1964.
[2] H. A. Haus and F. N. H. Robinson, "The minimum noise figure of microwave beam amplifiers," *Proc. IRE*, vol. 43, pp. 981–991, August 1955.
[3] S. Saito, "New method of measuring the noise parameters of an electron beam," *IRE Trans. Electron Devices*, vol. ED-5, pp. 264–275, October 1958.
[4] S. Saito and Y. Fujii, "Some results from the measurements of the noise parameters in electron beam," *Proc. IRE*, vol. 50, pp. 1706–1707, July 1962.
[5] J. M. Hammer, "Power spectra measurements on ultra-low-noise beams," *J. Appl. Phys.*, vol. 35, pp. 1147–1153, April 1964.
[6] A. Zacharias and L. D. Smullin, "Noise reduction in electron beams," *IRE Trans. Electron Devices*, vol. ED-7, pp. 172–173, July 1960.
[7] W. Kuypers and M. T. Vlaadingerbroek, "Measurement of electron-beam noise," *Philips Research Repts.*, vol. 20, pp. 349–356, 1965.
[8] A. E. Siegman, "Analysis of multivelocity electron beams by the density-function method," *J. Appl. Phys.*, vol. 28, pp. 1132–1138, October 1957.
[9] S. Bloom and B. Vural, "Analysis of the noise spectra of a drifting multi-velocity beam," *Proc. 4th Internat'l Congress on Microwave Tubes* (Scheveningen), pp. 475–478, 1962.
[10] J. R. Whinnery, "High-frequency effects of the potential minimum on noise," *IRE Trans. Electron Devices*, vol. ED-7, pp. 218–230, October 1960.
[11] A. E. Siegman and S. Bloom, "An equivalent circuit for microwave noise at the potential minimum," *IRE Trans. Electron Devices*, vol. ED-4, pp. 295–299, October 1957.

Discrete Fourier Transforms When the Number of Data Samples Is Prime

***Abstract*—The discrete Fourier transform of a sequence of N points, where N is a prime number, is shown to be essentially a circular correlation. This can be recognized by rearranging the members of the sequence and the transform according to a rule involving a primitive root of N. This observation permits the discrete Fourier transform to be computed by means of a fast Fourier transform algorithm, with the associated increase in speed, even though N is prime.**

To compute the discrete Fourier transform of a long sequence of data samples, one of the fast Fourier transform (FFT) algorithms should be used if possible.[1],[2] The limitation common to all of these algorithms is that the number of data samples, N, must be highly composite. In this letter we show how FFT techniques can be applied to the computation of a discrete Fourier transform when N is prime.

I. NOTATION AND DEFINITIONS

We basically restrict our attention to the discrete Fourier transform. Let the data to be transformed be a sequence of N numbers $\{a_i\}$, $i=0, 1, \cdots, N-1$, where the braces indicate the entire sequence, while the symbol a_i represents only the ith member of the sequence. The discrete Fourier transform (DFT) is the sequence $\{A_k\}$, $k=0, 1, \cdots, N-1$, whose members are given by

$$A_k = \sum_{i=0}^{N-1} a_i \exp(-j(2\pi/N)ik). \tag{1}$$

The values of A_k are samples of the z-transform of the finite length sequence $\{a_i\}$ at N points equally spaced on the unit circle.

We need a notation for an integer modulo N. For brevity we will indicate this by superfluous parentheses:

$$((X)) = X \text{ modulo } N. \tag{2}$$

If N is prime there is some number g, not necessarily unique, such that there is a one-to-one mapping of the integers $i=1, \cdots, N-1$ to the integers $j=1, 2, \cdots, N-1$, given by

$$j = ((g^i)). \tag{3}$$

For example, let $N=7$ and $g=3$. The table below gives the mapping of i onto j.

i	1	2	3	4	5	6
j	3	2	6	4	5	1

In number theory, g is called a primitive root of N. A table of primitive roots of primes less than 10^4 is given by Abramowitz and Stegun.[3]

II. COMPUTATION OF $\{A_k\}$ WHEN N IS PRIME

In (1) we have an expression for A_k for all k. The expression for A_0 is particularly simple,

$$A_0 = \sum_{i=0}^{N-1} a_i, \tag{4}$$

and is to be computed directly. For the other A_k we observe that a_0 is not to be multiplied, and we choose to add it last into the summation. We are left with the sequence $\{A_k - a_0\}$, $k=1, 2, \cdots, N-1$ to compute, given by

$$A_k - a_0 = \sum_{i=1}^{N-1} a_i \exp\left(-j\frac{2\pi}{N}ik\right). \tag{5}$$

We permute the terms in the summation, and change the order of the equations via transformations

$$\begin{aligned} i &\to ((g^i)) \\ k &\to ((g^k)) \end{aligned} \tag{6}$$

and, noting that $((g^{N-1})) = ((g^0))$, we can see that

$$(A_{((g^k))} - a_0) = \sum_{i=0}^{N-1} a_{((g^i))} \exp\left(-j\frac{2\pi}{N}g^{(i+k)}\right). \tag{7}$$

We are now able to recognize that the sequence $\{A_{((g^k))} - a_0\}$ is the circular correlation of the sequence $\{a_{((g^i))}\}$ and the sequence $\{\exp(-j(2\pi/N)g^i)\}$. But Stockham has shown how circular (or ordinary) correlation functions can be computed with a greatly reduced number of operations by making use of FFT algorithms.[4] There are two ways we can proceed. Since N is prime, $N-1$ must be composite. Suppose that it is highly composite. Then the $N-1$ point circular correlation (7) may be recognized as the inverse DFT of the product of the DFT of $\{a_{((g^{-i}))}\}$ and the DFT of $\{\exp(-j(2\pi/N)g^i)\}$.[1] All the DFT operations called for are performed by an FFT algorithm.

$$\{A_{((g^k))} - a_0\} = \text{DFT}^{-1}\left\{(\text{DFT}\{a_{((g^{-i}))}\})\left(\text{DFT}\left\{\exp\left(-j\frac{2\pi}{N}g^i\right)\right\}\right)\right\}. \tag{8}$$

[1] It may be shown that this DFT has magnitude $\sqrt{N}$ at all frequencies but the zeroth, where it is -1.

Manuscript received March 4, 1968.

Proceeding in this first way we will be successful only if $N-1$ is highly composite. If $N-1$ is only modestly composite, as with $N=563$, the savings of the FFT algorithm will be overcome by the fact that more than one DFT must be computed. However, there is another way we can proceed which is not subject to these limitations. The second method is based on the observation that a circular correlation or convolution where the number of points is not highly composite can be computed as a part of a circular convolution with a larger number of points. Letting N' be any highly composite integer greater than $2N-4$, we create an N' point sequence $\{b_i\}$ by inserting $(N'-N+1)$ zeros between the zeroth and first points of $\{a_{((g^{-i}))}\}$ and we create a second N' point sequence $\{c_i\}$ by periodically repeating the $N-1$ point sequence $\{\exp(-j(2\pi/N)g^i)\}$ until N' points are present. Then the inverse DFT of the product of the DFTs of $\{b_i\}$ and $\{c_i\}$ contains $\{A_{((g^k))}-a_0\}$ as a subsequence—the first $N-1$ points. Since N' can be chosen to be highly composite, even a power of two, an FFT algorithm can be used to compute the DFTs.

Using either technique, about one-third of the computation can be saved if the transform of $\{\exp(-j(2\pi/N)g^i)\}$ is precomputed. One method requires a computation proportional to $(N-1)$ times the sum of the factors of $(N-1)$ whereas the second method requires a computation proportional to $N'\log N'$. Furthermore, the summation called for in (4) and the addition of a_0 to each other A_k can each be performed with negligible additional computation by operating on intermediate quantities available when the correlation is done by FFT techniques.

III. Conclusions

While the restriction that N be a highly composite number for FFT techniques to be useful has not proved severe, it is interesting to know that it can be removed. On the other hand, the recognition that a DFT can be expressed as a convolution may be useful in itself, as this implies that a single network with fixed parameters can compute all the points of a DFT. It is expected that such diverse applications as radar beam forming and modem design may profitably use this result.

Charles M. Rader
M.I.T. Lincoln Lab.[2]
Lexington, Mass. 02173

References

[1] J. W. Cooley and J. W. Tukey, "An algorithm for the machine computation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, April 1965.
[2] "What is the fast Fourier transform?" *Proc. IEEE*, vol. 55, pp. 1664–1674, October 1967.
[3] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965, p. 827 and pp. 864–869.
[4] T. G. Stockham, "High speed convolution and correlation," *1966 Spring Joint Computer Conf., AFIPS Proc.*, vol. 28. Washington, D.C.: Spartan, 1966, pp. 229–233.

[2] Operated with support from the U. S. Air Force.

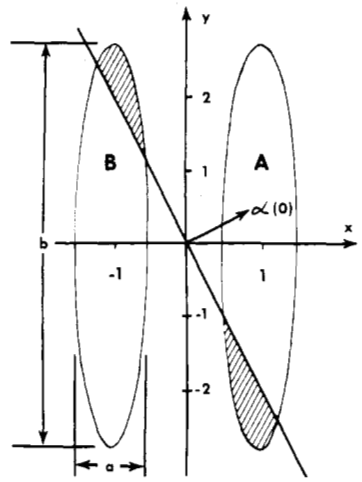

Fig. 1. Two-class ellipsoidal pattern distributions. In this example $a=1.0$ and $b=5.6$. The initial separating hyperplane, which is perpendicular to the weight vector $\alpha(0)$, identifies *incorrectly* one-tenth of the patterns (shaded region), thus $F=90.0$ percent.

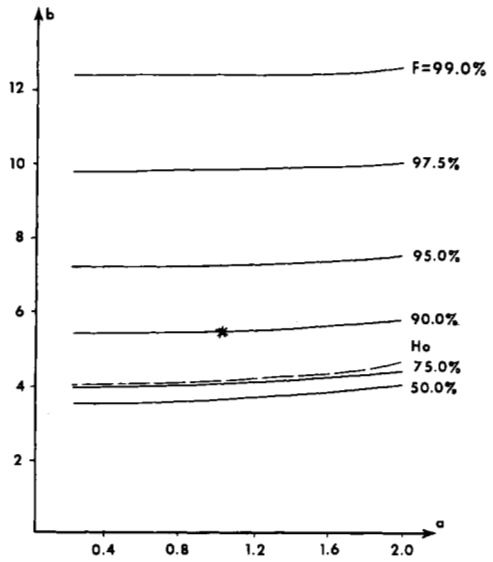

Fig. 2. Critical values of the parameters for two learning algorithms. The region below the dotted line indicates successful convergence of the Ho-Agrawala algorithm. The solid lines mark the upper limits of correct convergence, for a given initial hyperplane, of the Nagy-Shelton algorithm. In the example of Fig. 1, shown with an asterisk, the former would fail, while the latter would suceed provided the initial error rate is lower than 10.0 percent.

On a Theoretical Pattern Recognition Model of Ho and Agrawala

***Abstract*—Two versions of an unsupervised learning algorithm for pattern recognition are compared by means of numerical calculations based on two-dimensional ellipsoidal pattern distributions.**

In a recent letter, Ho and Agrawala[1] describe a theoretical model intended to explain some previously published experimental results in character recognition.[2] They call attention to the introduction of a simplifying assumption, expected to have little effect on the behavior of the recognition algorithm, to render the analysis tractable. Inspired by their observations we have calculated the performance of the algorithm, with and without this modification, for a specific family of distributions also suggested by Ho and Agrawala.

Manuscript received February 19, 1968.

[1] Y. C. Ho and A. K. Agrawala, "On the self-learning scheme of Nagy and Shelton," *Proc. IEEE (Letters)*, vol. 55, pp. 1764–1765, October 1967.
[2] G. Nagy and G. L. Shelton, Jr., "Self-corrective character recognition system," *IEEE Trans. Information Theory*, vol. IT-12, pp. 215–222, April 1966.

Using their notation, the two versions of the unsupervised learning algorithm are

$$\alpha(k+1) = XX^T\alpha(k) \tag{1}$$

$$\alpha(k+1) = X\,\mathrm{sgn}\,(X^T\alpha(k)) \tag{2}$$

where the $\alpha(k)$ are the successive approximations to the weight vector characterizing the hyperplane separating the two classes, and the columns of the X matrix are the pattern vectors to be classified.

Ho and Agrawala show that the procedure described by (1) always converges to the eigenvector associated with the largest eigenvalue of the sample covariance matrix XX^T. It will be seen, however, that the asymptotic behavior of (2) depends strikingly on the initial weight vector $\alpha(0)$.

The family of distributions considered consists of patterns uniformly distributed on two ellipses symmetrically located about the origin, as shown in Fig. 1. Ho and Agrawala's procedure (1) converges to the correct hyperplane (the y-axis) whenever the y component of the variance of the overall distribution is inferior to the x component. This imposes constraints on the relation between the major axis b and the minor axis a of the ellipses, as shown by the dotted line in Fig. 2.