THE APPLICATION OF CORRELATION FUNCTIONS
IN THE DETECTION OF SMALL SIGNALS IN NOISE

Y. W. LEE

T. P. CHEATHAM, JR.

J. B. WIESNER

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Research Laboratory of Electronics

Technical Report No. 141   October 13, 1949

THE APPLICATION OF CORRELATION FUNCTIONS
IN THE DETECTION OF SMALL SIGNALS IN NOISE*

Y. W. Lee

T. P. Cheatham, Jr.

J. B. Wiesner

ABSTRACT

The first part of this paper presents the general concept of the corre-
lation functions. A mathematical reinterpretation of the correlation func-
tion is given in terms of the sum of the products obtained by representing
the correlation function as a discrete time series. An electronic correlator
capable of performing the necessary mathematical operations to obtain corre-
lation functions is described and a description of the application of the
correlator to the detection of small signals is presented. The advantage
of cross-correlation in the detection of a signal of known frequency is
demonstrated.

*This report was presented at the I.R.E. National Convention in New York, March 7-10, 1949.

THE APPLICATION OF CORRELATION FUNCTIONS
IN THE DETECTION OF SMALL SIGNALS IN NOISE

I. Some General Properties of Correlation Functions

Mathematically, the autocorrelation function of a time function $f_1(t)$ is defined as:

$$\phi_{11}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f_1(t)f_1(t + \tau)dt \quad . \tag{1}$$

In an analogous fashion, the cross-correlation functions between two time functions $f_1(t)$ and $f_2(t)$ are defined as

$$\phi_{12}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f_1(t)f_2(t + \tau)dt \tag{2}$$

and

$$\phi_{21}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f_2(t)f_1(t + \tau)dt \quad . \tag{3}$$

Quite simply, correlation functions are a measure of the mean relationship existing between the product of all pairs of points of the time series involved, separated in time by a delay τ. The autocorrelation function and similarly the cross-correlation function are equally well-defined for both periodic and random time functions.

It can be shown that the autocorrelation function $\phi_{11}(\tau)$ is a maximum at $\tau = 0$, being equal to the mean of the square of the time function. In the absence of a periodic component, $\phi_{11}(\tau)$ approaches the square of the mean of the time function as τ increases. These general properties allow us to sketch a rough picture of the bounds or envelope of autocorrelation functions for random time series.

However, if a periodic component such as a sine wave were present in $f_1(t)$, the autocorrelation function would appear roughly as it does in Fig. 2, since component parts of correlation functions add linearly, and the correlation function of a sinusoid is a cosine function of the same frequency. If the random portion of $f_1(t)$ is normally distributed noise, its mean value is zero and therefore its correlation function approaches zero for large values of τ. The correlation for the sinusoid, however, is again periodic and of constant amplitude. It is pointed out, therefore, that theoretically an infinite signal-to-noise ratio can be obtained in the detection of a periodic signal (no matter how small) in noise through the property of autocorrelation alone. One should remember, however, that "theoretically"
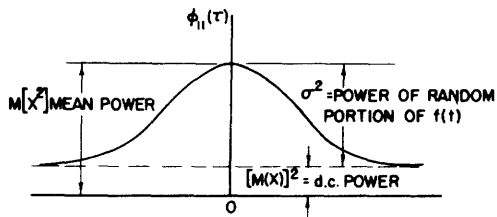
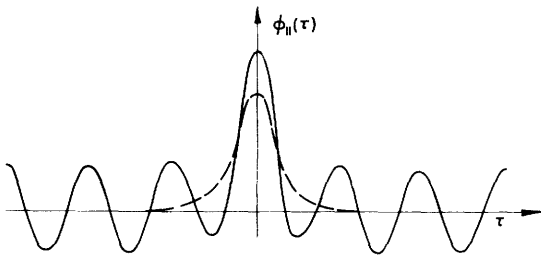Fig. 1 Sketch of a possible autocorrelation function.

Fig. 2 Autocorrelation of $f_1(t)$ = sinusoid + normally distributed noise.

requires an infinite time. It is the purpose of this paper to show what can be accomplished in a practical measurement in a finite time, and also what advantages can be derived from the use of cross-correlation.

The unique Fourier transform relationship between the autocorrelation function $\phi_{11}(\tau)$ and the power density spectrum $\Phi_{11}(\omega)$ of a time series given by Eqs. (4) and (5)

$$\Phi_{11}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{11}(\tau)\cos\omega\tau \; d\tau \quad (4)$$

$$\phi_{11}(\tau) = \int_{-\infty}^{\infty} \Phi_{11}(\omega)\cos\omega\tau \; d\omega \quad (5)$$

makes it possible to consider correlation in the detection of a sinusoid in noise as "filtering in the time domain".

II. Electronic Method for Computing Correlation Functions*

In order to introduce the problems of measuring correlation functions, Fig. 3 shows a portion of a random time function $f(t)$. According to the
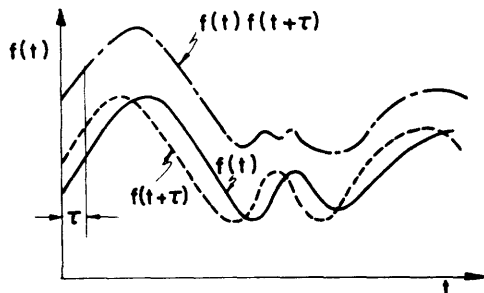

Fig. 3 Graphical method of obtaining an auto-correlation function.

mathematical definition of $\phi_{11}(\tau)$, given in Eq. (1), a graphical method for its computation would be to delay $f(t)$ by an interval τ, giving $f(t + \tau)$, to multiply these two functions continuously, integrate, and average. This interpretation electronically required a means for delaying and storing the entire time function and in addition an a priori knowledge of the bandwidth requirements of the time function — a characteristic which, in general, we are trying to measure. A simpler process is suggested through a re-interpretation of the mathematical expression of the correlation function, where emphasis is placed on the fact that it represents the mean relationship of the product of all pairs of points of the time series separated by

---

*A detailed report on the electronic correlator is given by T. P. Cheatham, Jr. in Technical Report No. 122, Research Laboratory of Electronics, M.I.T. (to be published).

an interval τ. That is, an arbitrary or random selection of a large number of pairs of points separated in time τ_1 seconds, multiplied together, summed, and their mean taken, determines one point on the correlation curve. The important point is that the correlation function can be represented by a summation and averaging of a large number of discrete multiplications, where sampled values of $f(t)$, rather than the entire time function, are delayed and stored. This means that the mathematical definition of Eq. (2) becomes approximated by:

$$\phi_{11}(\tau_k) = \frac{1}{N+1} \sum_0^N a_n b_n \Big|_{\tau_k} \quad . \tag{6}$$

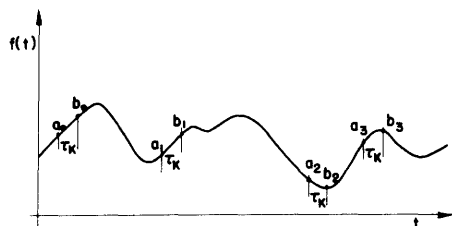The graphical process of Eq. (6) is represented in Fig. 4 by the arbitrary sets of points on the graph.

Fig. 4 Portion of a random time function, f(t), showing graphical representation of Eq. (6).

This simple procedure made it possible to design an electronic correlator in terms of pulse sampling techniques where the delay τ is placed on the sampling voltage. The general technique involved is outlined in the waveforms of Fig. 5.
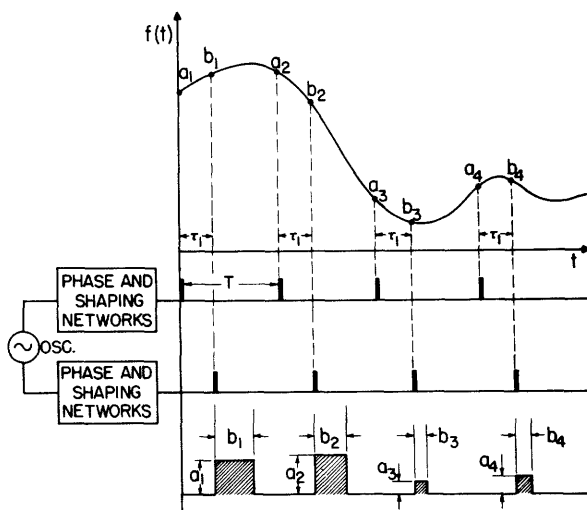

Fig. 5 Characteristic waveforms of electronic correlator.

The first waveform shows a section of a random time function $f(t)$. A sine-wave master oscillator is used to derive the timing and sampling pulses

of the next two lines. These sampling pulses are used to measure the amplitude of f(t) at the a and b points as indicated. These amplitudes are stored over a sampling period T and multiplied in time-amplitude co-ordinates by converting the b amplitudes into linearly equivalent time-width modulated pulses, which are then used to gate the stored amplitudes of the a points. The output of the gating circuit is the series of pulses of varying amplitude and width shown in Fig. 5. The area of each of these pulses is proportional to the product of a pair of points a_n, b_n separated in time τ_1 seconds. The integral or summation of these pulses over a period, say β, which approximates the statistics of the random time function, then gives a measure of $\phi_{11}(\tau_1)$. The value of the integral is recorded, τ shifted, and the process repeated.

III. The Application of Correlation Functions in the Detection of a Sinusoidal Signal in Noise

The practical and quantitative properties of the electronic correlator in the detection of a sinusoidal signal in noise in a finite length of time is presented here. In order to arrive at a clear picture of the basic notions involved, consider first the general case of two time functions $f_1(t)$ and $f_2(t)$, both consisting of a signal plus random noise as indicated by Eq. (7).

$$f_1(t) = S_1(t) + N_1(t)$$

$$f_2(t) = S_2(t) + N_2(t) \quad . \tag{7}$$

By substituting Eq. (7) into Eq. (2), the cross-correlation of these two time functions will be seen to be made up of four component cross-correlation terms:

$$\phi_{12}(\tau) = \phi_{S_1S_2}(\tau) + \phi_{S_1N_2}(\tau) + \phi_{N_1S_2}(\tau) + \phi_{N_1N_2}(\tau) \quad . \tag{8}$$

If the two time functions $f_1(t)$ and $f_2(t)$ are independent and incoherent then each of the four component terms of Eq. (8) will be identically zero as the period of correlation approaches infinity. For a finite time of correlation, however, the standard deviation of the points of the correlation curve will approach zero as a function of the number of discrete independent samples taken by the correlator.

As a first specific case, consider the properties of autocorrelation in the light of a practical problem of detecting a periodic signal in random noise where the only a priori knowledge of the signal is, perhaps, its probable location within a certain broad bandwidth. In terms of the previous general case, $f_2(t)$ is identically $f_1(t)$; both channels of the

correlator operate in parallel from the same input. Here again the correlation function will be made up of four terms:

$$\phi_{11}(\tau) = \phi_{SS}(\tau) + \phi_{SN}(\tau) + \phi_{NS}(\tau) + \phi_{NN}(\tau) \quad . \qquad (9)$$

The first term is the autocorrelation of the signal component — it will be nonzero — the next two terms are the cross-correlations of the signal and noise components. These, theoretically, as T approaches infinity, will be identically zero. The fourth term is the autocorrelation of the noise component. Theoretically, this term will be zero for large τ and can be made zero for all τ by compensation. The first term is a measure of our desired signal, and the last three are essentially noise terms whose dispersion approaches zero as a function of the number of samples computed by the correlator. It is important to note that there are at least three effective noise terms and actually a fourth if one includes the inherent amount of dispersion in the measurement of the signal term in a finite length of time.

It is convenient to evaluate the use of autocorrelation in the detection of a signal in noise in a finite length of time in terms of the rms signal-to-noise ratio improvement at the output of the correlator. This improvement can be determined as a function of the rms signal-to-noise ratio at the input and the length of correlation time.

It should be noted that the cross-correlation terms of Eq. (9), as well as the autocorrelation term for the noise component (for large τ), are computed from the average product of two independent variables. Since the dispersion of the product of two independent variables is equal to the product of the individual dispersions (providing first moments are zero, as in our case), we are able to evaluate the dispersion of each component term of Eq. (9) and hence the aggregate rms noise term at the output of the correlator.

A. At Input to the Correlator

If the signal is $E \cos(\omega t + \phi)$, and the noise is normally distributed with a standard deviation σ, then the input rms signal-to-noise ratio is given by:

$$\left. \frac{\text{Signal}}{\text{Noise}} \right|_{\text{rms}} = \frac{E}{\sqrt{2}\,\sigma} \quad . \qquad (10)$$

B. At Output of the Correlator

The measure of the desired signal is the autocorrelation signal component term of Eq. (9). Analytically, it is determined by substitution in Eq. (1).

$$\phi_{SS}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [E \cos\omega t] \, [E \cos\omega(t + \tau)] dt = \frac{E^2}{2} \cos\omega\tau \ . \quad (11)$$

The above expression for $\phi_{SS}(\tau)$ is its theoretical value. The electronic correlator practically measures a value which approaches close to

$$\phi_{SS}(\tau) = \frac{NE^2}{2K} \cos\omega\tau \quad (12)$$

where

$$N = \text{number of samples computed}$$
$$K = \text{a variable normalizing factor.}$$

The desired output rms signal component as measured by the correlator is therefore

$$\frac{NE^2}{2\sqrt{2} \, K} \quad .$$

The rms contributions of the effective noise component terms at the output of the correlator are evaluated by assuming that minimum τ is set suffi- ciently large so that the correlation of the noise at the discrete sampling points is effectively zero. It has already been pointed out that this is an unnecessary restriction since the correlation of the noise, once measured and known a priori, can be compensated.

Since the four effective noise terms are independent, their aggregate contribution can be evaluated as the square root of the sum of the squares. Therefore, for autocorrelation:

$$\text{output} \left. \frac{\text{Signal}}{\text{Noise}} \right|_{\text{rms}} = \frac{\dfrac{NE^2}{2\sqrt{2} \, K}}{\sqrt{\dfrac{N\sigma^4}{K^2} + \dfrac{2NE^2\sigma^2}{2K^2} + \dfrac{NE^4}{4K^2}}} \quad . \quad (13)$$

If we define:

$$Z = \text{input} \left. \frac{\text{Signal}}{\text{Noise}} \right|_{\text{rms}} = \frac{E}{\sqrt{2} \, \sigma}$$

$$W = \text{output} \left. \frac{\text{Signal}}{\text{Noise}} \right|_{\text{rms}} \quad ,$$

Equation (13) can be simplified symbolically to the following relation:

$$W = \sqrt{\frac{N}{2}} \, \frac{Z^2}{Z^2 + 1} \quad . \quad (14)$$

The derivation of Eq. (14) serves to illustrate the fundamental notions

and experimental techniques involved in the use of autocorrelation functions for the detection of small signals in noise. While the relation of Eq. (14) is explicit and useful, its greatest importance lies in its role as an introduction to and part of a more general concept concerning the evaluation and use of a priori knowledge in the detection of a desired signal.

A sinusoidal signal $E \cos(\omega t + \phi)$ is completely described by a knowledge of the three constants: amplitude, E; angular frequency, ω; and phase angle, ϕ. One knows intuitively from the notions of information theory that if we have an a priori knowledge of any of these three constants of a sinusoidal signal, the knowledge should be of some advantage in the detection of the sinusoid.

A practical demonstration and evaluation of this notion is that with an a priori knowledge of the frequency of the signal we wish to detect we can locally generate as an input to the second channel of the correlator a sinusoid having the same frequency, thus measuring cross-correlation. In this case, we shall find that the cross-correlation $\phi_{12}(\tau)$ is made up of the desired signal correlation component and only one noise term. We have

$$f_1(t) = S_1(t) + N_1(t)$$

$$f_2(t) = S_2(t)$$

$$\phi_{12}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left[S_1(t) + N_1(t) \right] S_2(t + \tau) dt$$

$$\phi_{12}(\tau) = \phi_{S_1 S_2}(\tau) + \phi_{N_1 S_2}(\tau) \quad . \tag{15}$$

We see, therefore, that cross-correlation has reduced the effective number of noise terms from three to one or, if we include the effect of a certain dispersion in the measurement of the signal component in a finite time interval, from four to two. The quantitative advantage of this over auto-correlation is derived below in a manner similar to Eq. (14), but with a parameter $\underline{a}$ introduced to show the transition from auto- to cross-correlation, thus allowing for a corresponding possible physical method of utilizing any a priori knowledge of the probable signal location in frequency.

The generalized problem of the detection of a signal in noise through auto- and cross-correlation as outlined above can be presented in the following manner. Let

$$f_1(t) = S_1(t) + N_1(t)$$

$$f_2(t) = S_2(t) + N_2(t)$$

where $N_1(t)$ is the noise from a bandwidth B, and $N_2(t)$ is the noise from a reduced bandwidth aB, where a < 1. If the rms value of $N_1 = \sigma$, then the rms value of $N_2 = \sqrt{a}\,\sigma$. Also, if $S_1(t) = E\cos\omega t$, let $S_2(t) = mE\cos(\omega t + \phi)$, where $m = |S_2|/|S_1|$, and ϕ = phase angle between $S_1(t)$ and $S_2(t)$.

In a fashion analogous in its reasoning to the specific case of auto-correlation already considered, the effective signal and noise component terms can be evaluated giving the following output signal-to-noise ratio.

$$\text{output } \left.\frac{\text{Signal}}{\text{Noise}}\right|_{\text{rms}} = \frac{\dfrac{mNE^2}{2\sqrt{2}\,K}}{\sqrt{\dfrac{m^2NE^4}{4K^2} + \dfrac{aN}{2}\dfrac{E^2\sigma^2}{K^2} + m^2\dfrac{N}{2}\dfrac{E^2\sigma^2}{K^2} + \dfrac{Na\sigma^4}{K^2}}} \quad . \quad (16)$$

If we define as before

$$Z = \text{input } \left.\frac{\text{Signal}}{\text{Noise}}\right|_{\text{rms}} = \frac{E}{\sqrt{2}\,\sigma}$$

and

$$W = \text{output } \left.\frac{\text{Signal}}{\text{Noise}}\right|_{\text{rms}}$$

we may write Eq.(16) as:

$$W = \sqrt{\frac{N}{2}}\,\frac{Z^2}{\sqrt{Z^4 + Z^2(1 + am^2) + am^2}} \quad . \quad (17)$$

Equation (17) has several interesting interpretations. If we let a = m = 1, we see that it reduces to Eq. (14) for the case of autocorrelation. For the case of cross-correlation, a is equal to zero, and it is interesting to observe that W is then independent of m. That is, the advantage of knowing, a priori, the frequency of our desired signal is exact and complete and independent of the amplitude parameter of the cross-correlated signal. Our advantage is solely determined by the extent of our a priori knowledge. A physical interpretation can be given for values of a other than 0 (cross-correlation) and 1 (autocorrelation). For intermediate values of a, the process is essentially autocorrelation, but a strong a priori probability of the signal's location in a restricted bandwidth makes it possible to shape the bandwidth characteristic of one channel of the correlator to fit this a priori knowledge and hence take advantage of it by reducing a, which increases the input rms signal-to-noise ratio. It should be noted that physical reasoning restricts m to be equal to unity except when a is equal to zero (cross-correlation).

In the case of cross-correlation, it should be noted that since no

knowledge has been assumed of the phase angle constant ϕ, of the desired signal, the maximum of the signal cross-correlation component must be determined. This will require a measurement of the correlation curve at a minimum of two different values of τ.

However, if it is assumed that we know both the frequency and phase angle of the desired signal, there will be no ambiguity as to the position of the maximum since, as already pointed out in section I, it must occur periodically at $\tau = 0$, T, 2T, etc. The advantage of an a priori knowledge of ϕ is then seen to give a constant 3-db improvement in output signal-to-noise ratio.

Equation 17 expressed in db has been plotted in Fig. 6 with $am^2 = 1$ (autocorrelation) and $am^2 = 0$ (cross-correlation) as a function of the input signal-to-noise ratio in db for N (the number of independent samples) equal to 60,000. From Fig. 6, we see that if the input signal-to-noise ratio is —15 db, then the output signal-to-noise ratio is +14.5 db. If we know the frequency of the signal to be detected, we can make the output signal-to-noise ratio +30 db. A knowledge of the desired signal's phase angle will add only 3 db.

The difference between the auto- and cross-correlation curves of Fig. 6 is the advantage to be gained as a result of an a priori knowledge of frequency. In Fig. 7 this is shown to be an approximately logarithmic function of the input signal-to-noise ratio.
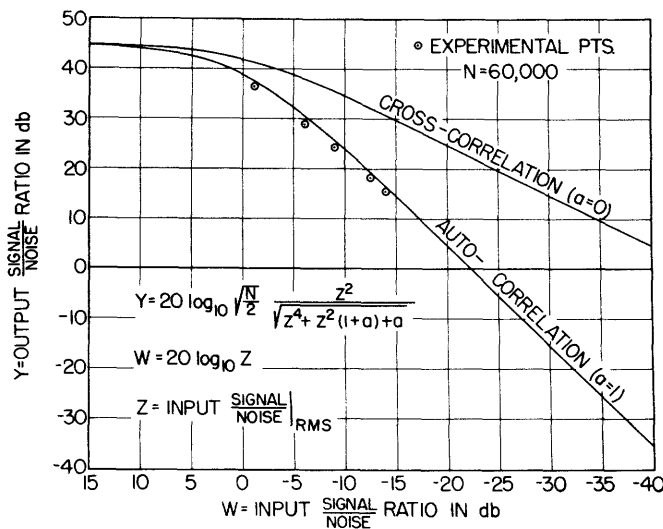

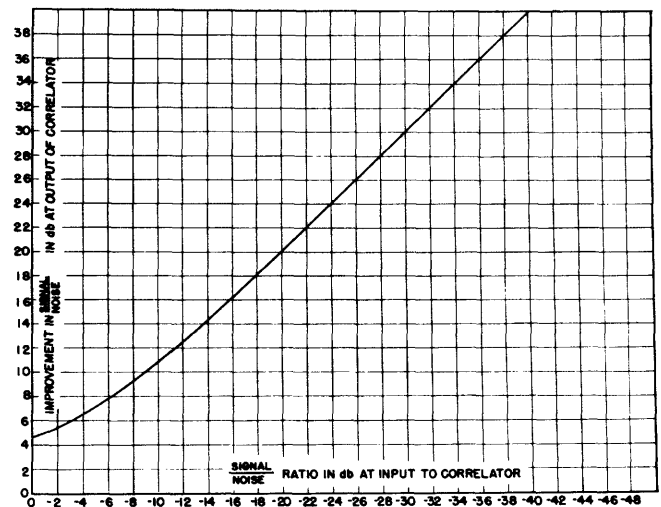Fig. 6 Improvement in signal-to-noise ratio through autocorrelation and cross-correlation.

Fig. 7 Improvement in signal-to-noise ratio through cross-correlation for a pure tone located in white noise.

It is significant to note that this curve is independent of N, the number of samples, and hence the averaging time. For example, if a certain output signal-to-noise ratio in db could have been obtained through
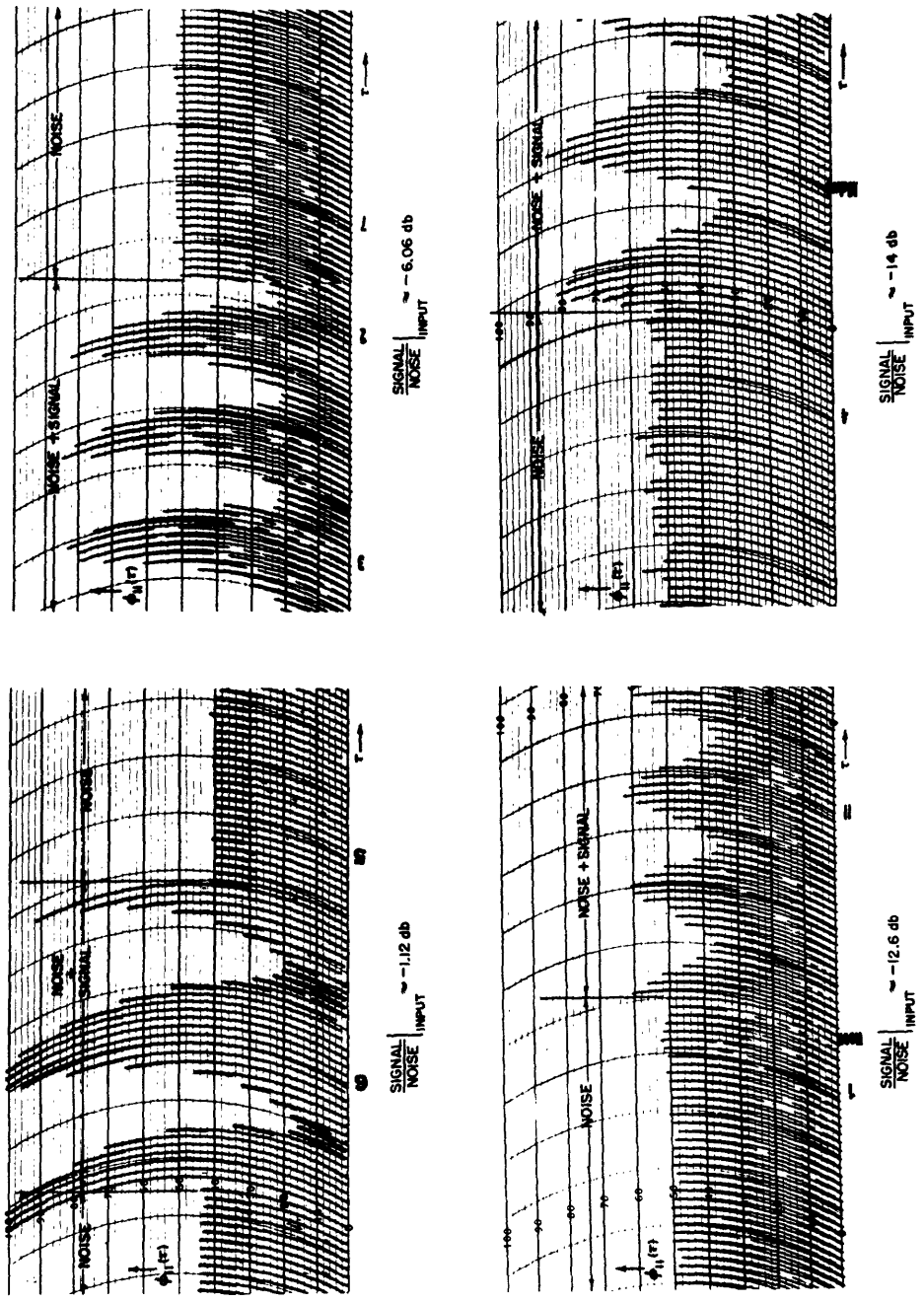
Fig. 8 Experimental results showing use of autocorrelation in the detection of a pure tone (8-kc sine wave) in random noise.

autocorrelation for an input signal-to- noise ratio of, say, —40 db, then
a possibility of having switched to cross-correlation through a knowledge
of the frequency constant will give a +40-db improvement.

As a point of some importance, a few quick calculations will show that
an attempt to use cross-correlation in the detection of a signal of "unknown"
frequency, and hence with a flat a priori probability of location within a
broad bandwidth, by "hunting" will give no advantage over autocorrelation.

Fig. 8 shows some examples of experimentally measured correlation curves
that check fairly well the analytical results above for autocorrelation.

It is well to emphasize that theoretically no claim can be made that
correlation or "filtering in the time domain" is superior to filtering in
the frequency domain in the sense of accomplishing in the time domain what
is impossible in the frequency domain. Such a claim would be false, since
it has already been shown in section I that correlations and power-density
spectrums are uniquely related by the Fourier transform. However, it is
pointed out that many equivalent operations may be more practicable and
feasible from an engineering standpoint in the "time domain". A zero band-
width in the frequency domain, for instance, corresponds in the time domain
to a very stable oscillator. At the present time it is much easier to build
a stable oscillator than to build a network having effectively "zero" width.
Also, the flexibility of control in the time domain of the effective filter
width appears to be a considerable advantage.

Acknowledgment

* * *