

Digital Methods for the Extraction of Phase and Amplitude Information From a Modulated Signal¹

R. S. Lawrence, J. L. Jespersen, and R. C. Lamb

(November 3, 1961)

A description is given of three digital methods which have been used to recover amplitude and phase information from a modulated sinusoidal signal sampled at equal intervals of not more than one-sixth of a period. The first method, the "zero-crossing" method, is economical of computer time and, for modulation which is not too deep and does not contain frequencies near the carrier frequency, accurately recovers the phase and amplitude modulation. The second method, the "filter" method, is more laborious but it gives better accuracy and will operate with deeper and more rapid modulation. The third method, a statistical approach, will work with severely overmodulated signals, but it yields only a statistical summary of the modulation. The methods were designed specifically for analysis of radio-star scintillation records but they may be applied to many other modulated signals.

1. Introduction

Three digital methods are described which have been used to recover amplitude and phase information from a modulated sinusoidal signal sampled at equal intervals of not more than one-sixth of a period. Although these methods were designed specifically for the purpose of analyzing the records described in the preceding paper, "Amplitude and angular scintillations of the radio source Cygnus-A observed at Boulder, Colorado", they may be applied to many other modulated signals. In general, the desired modulation is imposed on a sinusoidal carrier of constant frequency and amplitude. The sample interval and the carrier frequency must be chosen carefully with regard to the expected range of modulating frequencies, in order to avoid biasing the results either by omission of or folding at frequencies of interest.

All three of the methods described herein were designed for use with the IBM-650 digital computer and its peripheral equipment which included an automatic graph plotter. The input data were digital samples of the modulated carrier, sampled to an accuracy of 1½ percent of full scale (six bits) and punched on paper tape. The first method of analysis, the "zero-crossing method", uses the properties of the signal only at the times when it passes through its average value. Phase modulation is derived from variations in the intervals between these times; amplitude modulation is derived from the slope of the signal at these times. The second method, the "filter method", passes the signal through appropriate digital filters which recover the phase and amplitude modulation. The last method is a statistical one, used to analyze records which, because of overmodulation in amplitude and/or phase, are not amenable to treatment by either the

zero-crossing or filter methods. This statistical method cannot give amplitude and phase as a function of time, but produces results which describe some statistical properties of these quantities.

2. Zero-Crossing Method

In this procedure the computer must deal with the modulated carrier at the times when it crosses its average value. The first step is to replace the original samples by deviations from their average value. Since drifts in the digital recording system might cause the carrier to undergo slow changes in average (d-c) level, changes unrelated to the modulation, it is usually better to use a moving average several periods in length than to use an average over the entire record. If a moving average is used, the process of replacing the original samples by deviations from the average is, in fact, a digital high-pass filter which removes the d-c level and its slow variations, but does not appreciably affect the carrier or its modulation. For the radio-star observations, a moving average 30 samples in length (approximately 5 periods) proved satisfactory.

After the filtering operation, the computer can find the points of interest by searching for adjacent samples with opposite signs. Each such "crossing pair" of samples yields, by linear interpolation, a good estimate of the exact time of zero crossing. In order to extract phase information, first differences are taken of the list of times of zero crossings. Each of these intervals is then replaced by its deviation from the mean interval. The cumulative sums of these deviations provide a list of numbers called the phase table. The phase table is simply a tabulation of instantaneous phase variations at half-period intervals.

From the crossing pairs the computer derives another set of numbers called the amplitude table. This tabulation contains values proportional to the amplitude of the modulating signal. It is the sum

¹ Contribution from Central Radio Propagation Laboratory, National Bureau of Standards, Boulder, Colo.

of the absolute values of the two samples involved in each crossing pair, and measures the instantaneous amplitude at half-period intervals. It will be demonstrated later that this simple method, which might be expected to be inaccurate, is, in fact, an effective means for extracting the envelope of the sinusoidal trace. Figure 1 gives examples of actual amplitude and phase plots, made by an automatic plotter from the amplitude and phase tables. Each plot represents 600 points (10 min) of data and has been expanded to occupy the full scale. The full-scale range of each phase plot is shown in electrical degrees. The reliability and sensitivity, together with a brief description of the drawbacks, of the zero-crossing method will be discussed in the following sections.

3. Filter Method

Amplitude and phase tables may also be obtained by passing the data through a system of digital filters. This analysis procedure is based upon a method suggested by N. R. Goodman [1960], of Space Technology Laboratories.

Consider a modulated sine wave which is sampled once every t_0 seconds to produce a time sequence of n data points, which we designate x_i . We are interested in the amplitude (i.e., envelope), and phase of this record. A possible representation of the amplitude $A(t)$ and phase $\alpha(t)$ is

$$A(t) \exp [i\alpha(t)] = X(t) + iY(t)$$

where $X(t)$ and $Y(t)$ are the real and imaginary parts of the complex representation of the power and phase. We may let the x_i represent a set of sampled points from the $X(t)$ process. If we can somehow construct from the x_i a set of y_i which will be a good

representation of the $Y(t)$ process, then we may compute $A(t)$ and $\alpha(t)$ from the relations

$$A(t) = [X^2(t) + Y^2(t)]^{1/2}$$

$$\alpha(t) = \sin^{-1} [Y(t)/A(t)] = \cos^{-1} [X(t)/A(t)].$$

One possible representation of the y_i in terms of the x_i is

$$y_i = \sum_{s=-\infty}^{s=\infty} a_s x_{i+s} \quad (1)$$

where the a_s have suitable values. There are several ways in which one might evaluate the a_s . Goodman shows that for the present purpose the a_s should be chosen in such a way that, over the frequency range of interest, the spectral density matrix of the resulting (x_i, y_i) process is a good approximation to the spectral density matrix of the actual $[X(t), Y(t)]$ process. Following this line of thought, a set of a_s were chosen to satisfy that criterion. For the case where the sampling rate was about 6 samples per period, the following a_s proved satisfactory.

- $a_0 = 0.16$
- $a_1 = .15$
- $a_2 = -.13$
- $a_3 = -.20$
- $a_4 = -.06$
- $a_5 = .03$
- $a_6 = .00$
- $a_7 = -.02$
- $a_8 = .03$
- $a_9 = .05$
- $a_{10} = .02$
- $a_{11} = -.01$
- $a_{12} = -.03$
- $a_s = 0, s > 12$

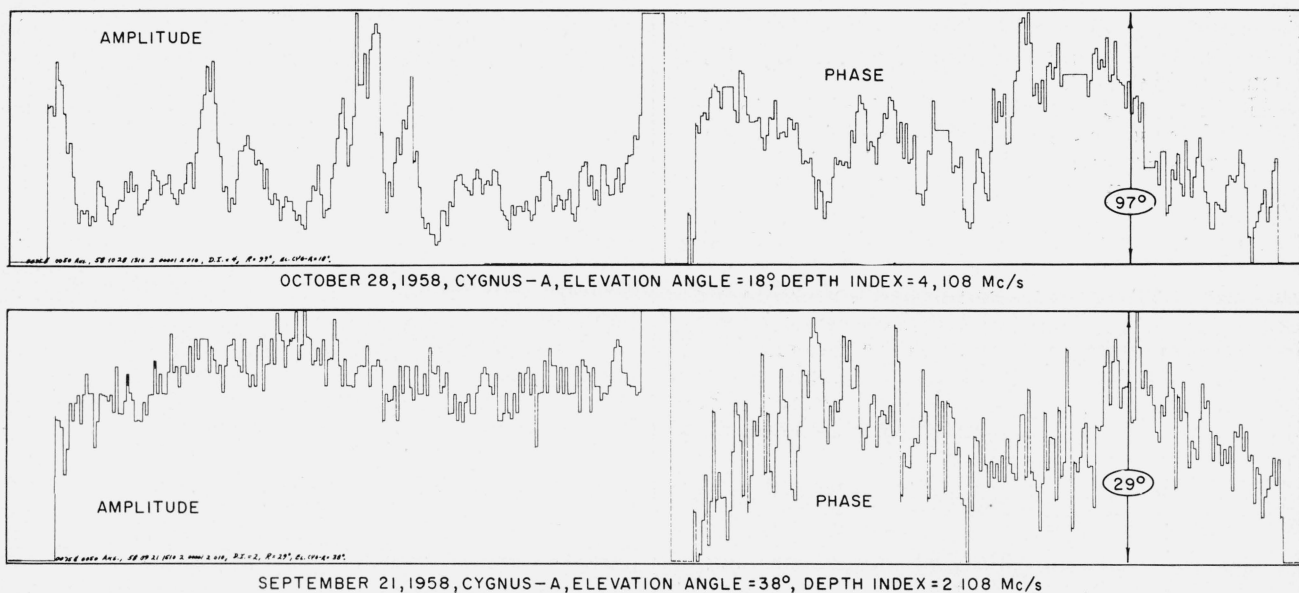


FIGURE 1. Examples of amplitude and phase plots obtained from the computer.

Each record is 600 samples (10 min) in length. The bottom edge of the amplitude plot represents zero signal, and the plot has been normalized to occupy the full scale. The scale of the phase plot, α , given on the figure, is in electrical degrees.

These a_s were then used in conjunction with (1) to evaluate the y_i .

Since the y_i are a moving linear combination of the x_i , they are in fact the result of passing the x_i through a linear digital filter. This filter (y filter) has a narrow pass band, approximately unit gain, and $\pi/2$ phase shift in its pass band. The phase shift, uniform over the pass band, is accomplished by making the filter anti-symmetric; i.e., $a_s = -a_{-s}$, and $a_0 = 0$. Noise outside the pass band of the filter may be removed from the x_i by means of a similar filter (x filter) with zero phase shift. This will be a symmetric filter, $a_s = a_{-s}$.

The amplitude and phase tables produced by either the "zero-crossing" or "filter" methods may be used for such subsequent purposes as spectral analysis and cross correlation between amplitude and phase.

In order to test the ability of the "zero-crossing" and "filter" methods to reproduce accurately the amplitude and phase fluctuations inherent in the data, an artificial record with known amplitude and phase variations was analyzed using each method. The known amplitude and phase variations are shown by the heavy curves in figures 2 and 3. The light curves labeled "two-filter method" show the results of using both the x and y filters. To assess the possibility of noise contamination, the same artificial record was analyzed using only the y filter. The result of this analysis is shown in the figure labeled "one-filter method."

Although both methods reproduce the known amplitude and phase variations quite well, the one-filter method includes a high-frequency component which is not evident in the two-filter method, indicating the existence of noise outside the pass band of the filter.

The heavy curves (input) are tilted slightly with respect to the light curves (output). This is because the input data is passed through a high-pass filter which removes trends.

The figures labeled "zero-crossing method" show the corresponding results produced by the zero-crossing method when it is applied to the artificial data.

The principal difference between the filter method and the zero-crossing method is that the former utilizes consecutive data points whereas the latter uses only pairs of consecutive points which have opposite signs (crossing pairs). For this reason the filter method has a twofold advantage: (1) mod 2π ambiguities are less likely to occur between consecutive points than between crossing pairs, and (2) higher frequency components are preserved by the filter method since its uses data points which are spaced more closely in time. The two-filter method involves about four times as much computing time as does the zero-crossing method, but the fidelity of the more lengthy method is better.

4. Statistical Method

Both of the foregoing methods have been applied to artificial data containing various amounts of amplitude and phase modulation. Although the

filter method is more capable of producing accurate amplitude and phase plots than is the zero-crossing method, both methods fail for some records. These highly modulated records contain significant phase and amplitude modulation with frequencies comparable to the carrier frequency.

In order to analyze such records, a statistical approach was taken. This procedure does not give detailed amplitude and phase plots but instead gives two quantities, P_n and σ_d , related respectively to the magnitudes of the amplitude and phase fluctuations.

P_n is evaluated by first finding the points on the actual modulated sine wave, $P(t)$, which lie above the peaks of a sine wave having the same average amplitude. The average value of these points, divided by the average amplitude, is defined as P_n . The sampled points which enter into the evaluation of P_n all result from amplitude modulation, since phase modulation can never increase the instantaneous amplitude of the record above the peaks of the equivalent sine wave. Thus, P_n is a statistical measure of the amplitude modulation on the record and is unaffected by the phase modulation. The standard deviation, σ_d , of the intervals between zero crossings, is a measure of the phase modulation alone since amplitude modulation cannot affect the time of zero crossings.

It is desirable to relate the quantities P_n and σ_d to quantities which are more commonly associated with amplitude and phase effects. For example, the quantities

$$\left(\frac{\Delta A}{A}\right)^2 = \left(\frac{A - \bar{A}}{\bar{A}}\right)^2$$

and σ_α were of interest in the preceding paper. Here A is the amplitude of the signal, \bar{A} the mean amplitude, and σ_α is the standard deviation of the phase fluctuations, α .

One possible way to relate P_n and σ_d to $(\Delta A/A)^2$ and σ_α respectively, is to compute $(\Delta A/A)^2$ and σ_α from the variations in amplitude and phase for a particular record as given by the "zero-crossing" or "filter" method, and then to find P_n and σ_d for this same record. If enough records are treated in this manner, it is possible to obtain regression curves of $(\Delta A/A)^2$ versus P_n and of σ_α versus σ_d . This procedure not only gives the desired relations but yields confidence limits for P_n and σ_d as measures of $(\Delta A/A)^2$ and σ_α respectively. However, it cannot be applied to the highly modulated records for which the statistical quantities P_n and σ_d were originally introduced. A better method is to compute P_n and σ_d from artificial records which have a predetermined value of $(\Delta A/A)^2$ and σ_α , for the particular example under consideration.

The artificial records were constructed in the following way. A pair of random normal deviates (m_i , n_i) are used to compute the quantity

$$P_i = (a + m_i)^2 + n_i^2$$

where a is a constant. P_i may be thought of as the

square of the envelope of a signal at time t_i which consists of the sum of a sine wave with amplitude a and noise with mean square amplitude of unity. The P_i are then passed through a low-pass filter which auto-correlates them over a length of 30 points. The artificial data points, d_i , are then computed from the following expression:

$$d_i = P_i \sin(\omega_0 t_i + \alpha_i),$$

where α_i , the phase fluctuation, is normally distributed with mean zero, standard deviation σ_α and auto-correlation length of 30; and ω_0 is some constant angular frequency, approximately $\pi/3$ per sample interval. The d_i are then normalized so that the artificial records will have the same range in value of the d_i for any a .

The probability distribution function of the $(P_i)^{1/2}$ constructed in this manner has been derived by S. O.

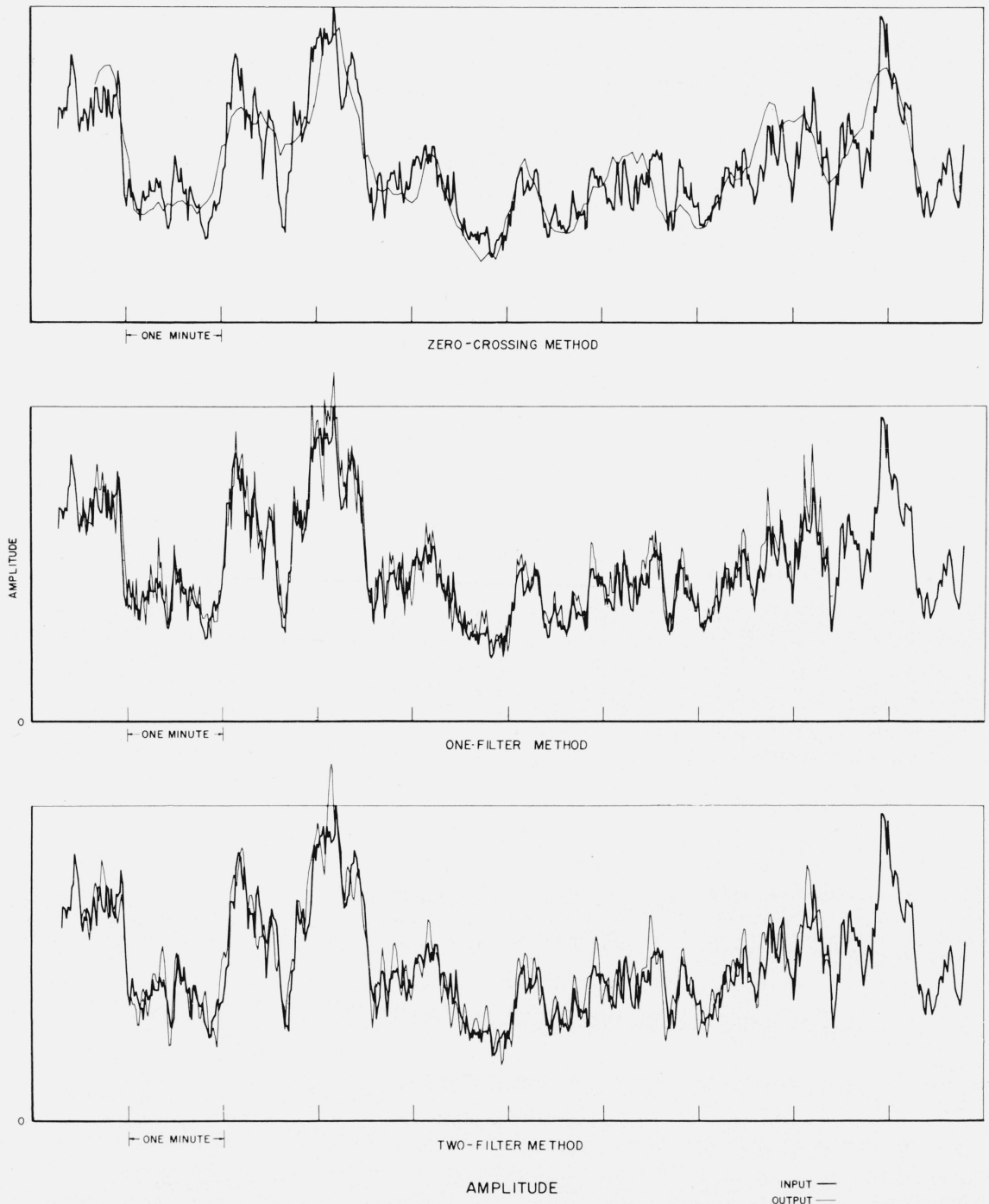


FIGURE 2. An illustration, using artificial data, of the effectiveness of the zero-crossing and filter methods in reproducing amplitude fluctuations inherent in the original record.

Rice [1944 and 1945]. (The reason for constructing the artificial data in this particular way was, of course, dictated by knowledge of the sort of signal to be analyzed. Signals with different modulation characteristics would require different techniques for producing the artificial data.)

The method involving artificial data has the advantage that it cannot be biased by mod 2π ambiguities since σ_α is a predetermined quantity. Thus the artificial records may be used to extend empirically the relation between σ_d and σ_α to regions where σ_α is much greater than 1 radian. P_n , measured on

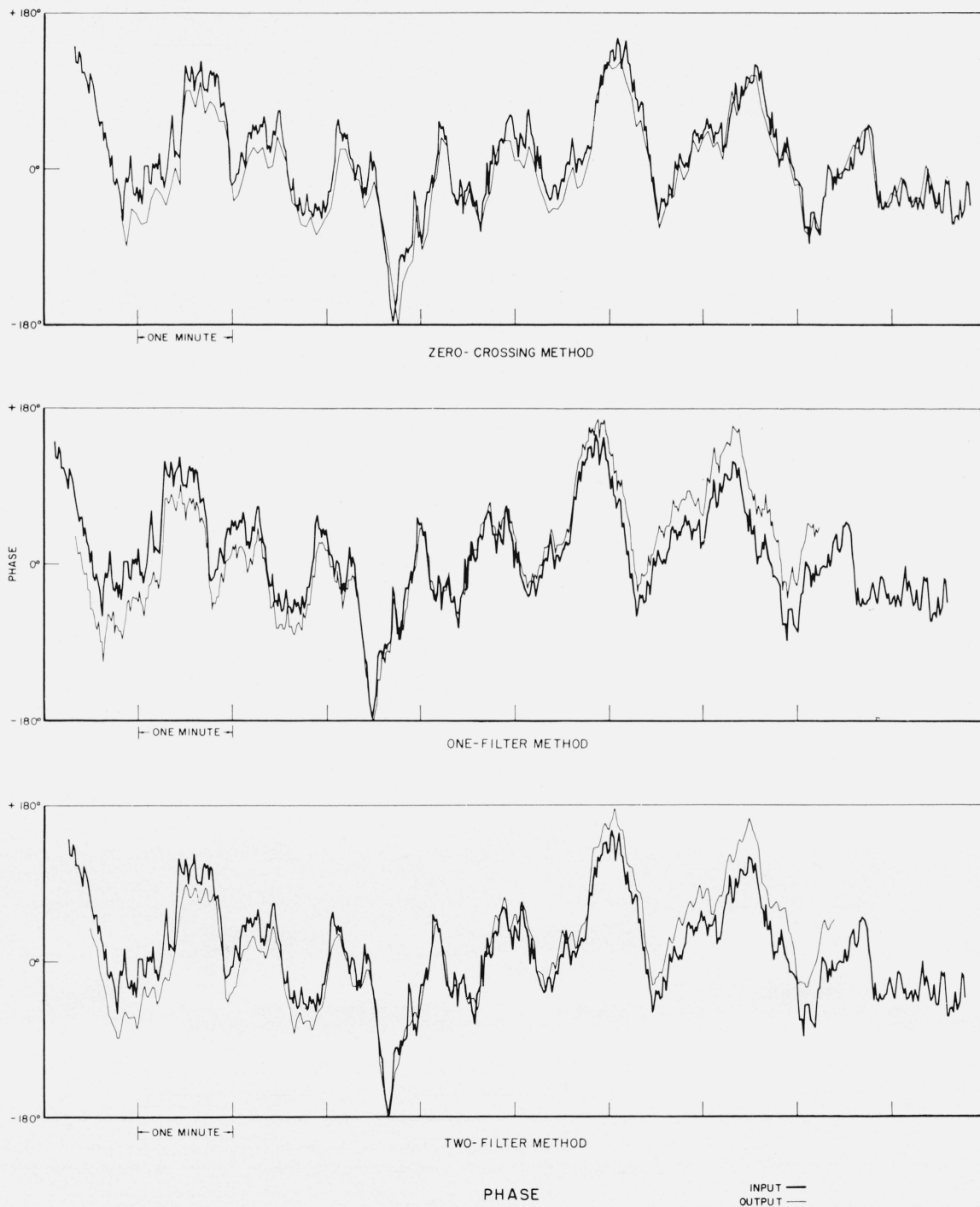


FIGURE 3. An illustration, using artificial data, of the effectiveness of the zero-crossing and filter methods in reproducing phase fluctuations inherent in the original record.

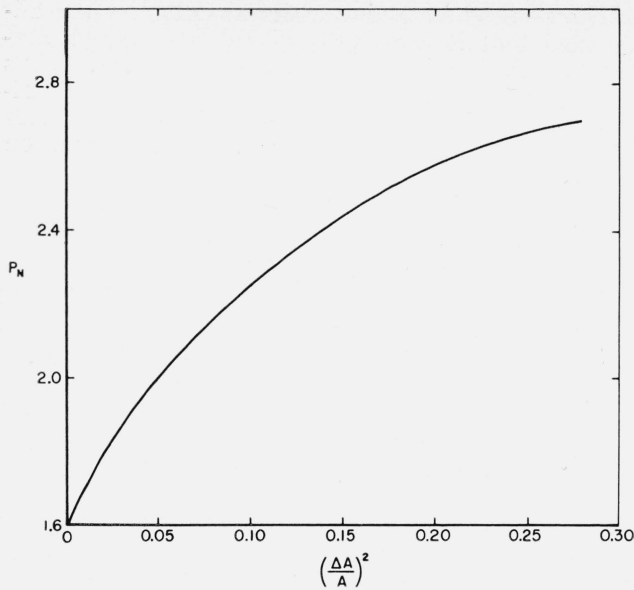


FIGURE 4. The relationship, derived from artificial data, between the statistical quantity P_n and the actual value of $(\Delta A/A)^2$.

the artificial records, can be related to $(\Delta A/A)^2$ through the parameter α which is a known function of $(\Delta A/A)^2$ [Lawrence and Jespersen, 1959].

For values of σ_α as great as 3 radians, the linear relationship

$$\sigma_d = 0.55 \sigma_\alpha$$

appears to hold. Figure 4 shows the observed relationship between $(\Delta A/A)^2$ and P_n .

The development of these digital methods was sponsored by the Ballistic Missile Division of the United States Air Force, under D. O. AF 04(647)-134.

5. References

- Goodman, N. R., *Advances in astronomical sciences*, vol. 5, pt. V (Plenum Press, 1960).
 Lawrence, R. S., and J. L. Jespersen, *A preliminary analysis of amplitude scintillations of radio stars observed at Boulder, Colorado*, Technical Note No. 20, NBS Boulder Labs. (1959).
 Rice, S. O., *Mathematical analysis of random noise*, Bell System Tech. J. **23** & **24**, 282-332, and 46-156 (1944 and 1945).

(Paper 65D4-138)