ESD-TR-67-220

SECOND QUARTERLY TECHNICAL REPORT ON LARGE APERTURE SEISMIC ARRAYS

February 1967

DIRECTORATE OF PLANNING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Sponsored by: Advanced Research Projects Agency, Washington D. C.
ARPA Order Number 800

(Prepared under Contract No. AF19(628)-5981 by General Atronics Corp., 1200 E. Mermaid Lane, Philadelphia, Pennsylvania 19118).

LEGAL NOTICE

When U. S. Government drawings, specifications or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

OTHER NOTICES

Do not return this copy. Retain or destroy.

Second Quarterly Technical Report
Contract AF 19(628)-5981

LARGE APERTURE SEISMIC ARRAYS

Sponsored by:

ADVANCED RESEARCH PROJECTS AGENCY
ARPA Order No. 800

GENERAL ATRONICS CORPORATION
February 1967

SYNOPSIS

This report describes three topics considered during the second quarter of Contract AF19(628)-5981.

Three three topics reported on are:

a) the estimation of the spectra of transient signals with additive noise,

b) the theoretical aspects of clustering seismometers in a large aperture seismic array,

c) the results of tests of a revised automatic pP phase detection technique.

A method has been devised to estimate the energy density spectra of transient events in such a way that the variance of the spectral estimates produced by additive noise is minimized.

The study of clusters shows:

1) greater spacing between sensors would improve the S/N performance of the arrays,

2) a way in which to determine the amount of clustering possible without S/N degradation depending on noise conditions at the receiving site.

Tests with the revised automatic pP detector have shown 100% correct results with surface focus events (6), 83% correct results for earthquakes (12), with the remaining earthquakes producing anomalies.

TABLE OF CONTENTS

	Page
SECTION I - Introduction	1
SECTION II	
A. Introduction	3
B. A Comparison of DS Processing with FS Processing	5
C. Seismic Array as a Diversity System	6
D. A Comparison of DS Processing with WDS Processing	8
E. Array Gain with DS Processing	14
F. Conclusions	26
SECTION III	
A. Introduction	27
B. The Time Window	30
C. Noise-Alone Case	31
D. Signal Plus Noise Case	39
E. Recommended Method	41
F. Improved Method Based on Subtracting an Estimate of the Noise Waveform	44
G. Some Comments on the Papers of McIvor and by Larrowe and Crabtree	50
SECTION IV - Automatic Test for pP in Shallow Earthquakes	54
References	59

APPENDIX A - Details and Derivations for Section III

APPENDIX B - Weighting Factors

SECTION I

Introduction:

This is the Second Quarterly report on Contract AF19(628)-5981. It describes, in part, the studies conducted during the period from 15 June to 15 September.

In addition to the special task mentioned above, three topics were considered in detail. These topics are:

a) The development of a computation technique for the estimation of the spectra of transients corrupted by additive noise which minimizes the variance in the spectral estimate,

b) A theoretical examination of the effects of clustering on the signal-to-noise properties of LASA's for a variety of noise conditions,

c) An extension and expansion of the earlier work on the automatic identification of the pP phase for shallow earthquakes.

In the study of estimation of the spectra of transients imbedded in noise, a "smoothing" technique was developed, somewhat analogous to that of Blackman and Tukey for estimating the spectra of random processes.

The cluster study results indicate that larger clusters than those currently used in Montana should be used in order to increase the signal-to-noise ratio gain

of an array. In addition, a method is given to choose the maximum amount of clustering which can be used without degrading performance. This limit is set by the distribution of local and teleseismic noise at a given receiving site.

The revised pP detection scheme was tested with 12 earthquakes and six surface events. All of the earthquakes studied had depth picks from other sources. None of the surface focus events indicated a pP phase, while ten of the quakes showed depths in good agreement with the previous data. One of the remaining quakes showed a peak in the test statistic for a depth appreciably shallower than the reported depth, while the other showed three peaks in the test statistic, one of which corresponded to the reported depth.

SECTION II

A. Introduction

In the last quarterly report, it was noted that microseismic noise was independent from cluster to cluster in the Montana LASA. Further, it was observed that signal correlation was quite high (mean value of 0.97) across the array. As a result, the signal-to-noise enhancement which should be obtained from phased summation of the clusters of the array is very nearly equal to the number of clusters summed. Actual experience at LASA bears out this estimate, since the on-line gain achieved is approximately 13 dB, and the number of clusters summed is 21. The work reported on in this section represents an extension of this previous effort to include the effects of clustering elements of an array on the signal-to-noise enhancement which can be achieved.

Specifically, three types of signal processing will be discussed in conjunction with clusters for a LASA.

1. Delayed sum (DS) processing, in which the element outputs are added with each output delayed in order to have coherent addition of signals.
2. Weighted delayed sum (WDS) processing, which is the same as DS processing with the addition that each element output is given a weight that is inversely proportional to its noise power.
3. Filter and sum (FS), in which each element is filtered before summing to form a maximum - likelihood estimate of the noise field in order to remove as much noise as possible from the summed output.

Further details of these processing methods are presented elsewhere (4).

Results of experiments with these three processing techniques obtained by UED and Lincoln Labs. (1-3) will be used here to arrive at recommendations for signal processing on a LASA. Also, some theory dealing with DS processing is developed. Furthermore, it is shown that a seismic array is actually a diversity system so that much of the current technical literature dealing with diversity systems may be applied to seismic arrays.

B. A Comparison of DS Processing with FS Processing

Experiments performed by UED on actual events show that for the existing LASA, FS is somewhat better than DS (1). Lincoln Lab's experimental results show that FS for the present LASA is much better than DS (2). This is true even if prefiltering of the signals is used to remove noise that is out of the signal band, even though the FS processing gain is reduced.* However, present computational facilities cannot handle on-line FS processing of LASA.

The poor results of DS processing on the present LASA are to be expected because the small subarrays maintain high noise correlation between the elements, preventing power gains anywhere near the number of elements in the array. Recently, experiments at Lincoln Labs (3) have shown that by increasing the size of a subarray from a 7 km diameter to a 22 km diameter will raise the DS gain to almost the same level (about 3 dB less) as the FS gain, whereas the FS gain is increased only slightly. The conclusion, then, is that FS processing is not worth the additional complexity and time that it requires when it performs only slightly better than DS processing.

---

* The overall gain of FS processing with prefiltering is actually increased over that of FS processing alone, but the amount of gain due to the FS processing is reduced. The rest of the gain is a result of prefiltering.

C. Seismic Array as a Diversity System

In this section it is shown that DS and WDS processing are two types of diversity combining. This is a field that has recently received much attention in the technical literature, particularly with respect to troposcatter communications.

By the term diversity system, Brennan (5) refers to "a system in which there are two or more closely similar copies of some desired signal." With m such copies

$$f_i(t) \qquad i=1,\ldots,m \ ,$$

each copy composed of signal and noise,

$$f_i(t) = s_i(t) + n_i(t)$$

a general linear combination will be considered

$$f(t) = a_1 f_1(t-\tau_1) + \ldots + a_m f_m(t-\tau_m) = \sum_{i=1}^{m} a_i f_i(t-\tau_i).$$

Each $f_i(t)$ is weighted by a combining coefficient a_i and delayed by time τ_i so that when the a_i's and τ_i's are properly chosen we can expect $f(t)$ to be "better" in some sense than any $f_i(t)$.

There are three common linear combination, or diversity combining, techniques. In all three the τ_i's are chosen so that the signals, the s_i's are aligned along the time axis at the point of summation. The first type is equal-gain diversity, in which all the a_i's are equal. This is the same as the DS processing mentioned previously.

Secondly, there is maximal-ratio diversity. In this technique the weighting factors, the a_i's, are chosen to maximize the signal-to-noise ratio of the sum. If P_{si} is the signal power at the i^{th} element, and P_{Ni} the noise power at the i^{th} element, then the optimum weighting factors for noise which is uncorrelated between elements can be shown to be (5)

$$a_i = \frac{P_{si}}{P_{Ni}} .$$

The signal-to-noise ratio of the sum, then, is the sum of the signal-to-noise ratios. This type of diversity becomes weighted delayed sum (WDS) processing for the case where all signals are equal, so that the weighting of each element is inversely proportional to its noise power. The equal-gain, or DS, technique is a special case of this technique when all element signal powers P_{si} are equal and all element noise powers P_{ni} are equal.

The third type of diversity is selector diversity. In this case, the input with the highest signal-to-noise ratio is chosen, and all the others are removed. In other words, if element j has the highest signal-to-noise ratio, the weighting factors are:

$$a_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} .$$

The problem involved with using maximal-ratio diversity or selector diversity in a LASA is that the signal amplitude, as well as the noise power, must be known. The signal amplitude at any element is not constant for different events, nor is it constant over the array for any single event. (experiments conducted at Lincoln Labs (3) have unearthed events in which the signal amplitudes differed as much as 9 to 1 over the LASA.) The difficulty in estimating the short duration signal accurately at each element with a noisy background makes the use of these two diversity techniques too difficult.

Of the two remaining techniques, DS (equal-gain combining) and WDS (the variation on maximal-ratio combining), WDS requires that an estimate of the noise power be made at each element and used there as its weighting factor. This means that WDS processing is more complex than DS processing. Further comparison of the two techniques is presented in Section D.

D. A Comparison of DS Processing with WDS Processing

Because the signal amplitude at each element is not measured in the LASA processing one must use a processing technique that does not require this information. In its absence, it is reasonable to assume that all elements have the same signal amplitude. Under this assumption, and the assumption of uncorrelated noise from element to element, WDS processing is the optimum combining technique. However, to the extent that the equal signal assumption does not hold for an event WDS processing gains begin to fall off.

Experiments conducted at Lincoln Labs (3) have shown that the WDS processing gain in the LASA is significantly better than the DS processing gain when the 7 km. subarrays are used. When the subarray size was expanded to 22 km., the two gains both increased because of the decrease in noise correlation between the elements. The more significant gain increase was in DS processing, so that for the larger subarrays the DS processing gain was, on the average, within 1 dB of the WDS processing gain. With only a 1 dB difference, WDS does not seem to be worth the extra complexity.

For the case of a LASA with large subarrays (large enough so that the noise between elements can be assumed to be decorrelated), we can derive an expression for the loss in gain by using DS processing instead of WDS processing. For the same set of input signals and noises to the two processing schemes, the loss L is simply the ratio of the output signal-to-noise ratio of DS (SNR_{DS}) to the output signal-to-noise ratio of WDS (SNR_{WDS}). Assuming the signal to be perfectly correlated over the array, we have

$$SNR_{DS} = \frac{\left(\sum_{i=1}^{m} \sqrt{P_{si}}\right)^2}{\sum_{i=1}^{m} P_{Ni}}$$

where P_{si} is the signal power at the i^{th} element

P_{Ni} is the noise power at the i^{th} element

m is the number of elements in the array.

Also
$$SNR_{WDS} = \frac{\left(\sum_{i=1}^{m} \frac{\sqrt{P_{si}}}{P_{Ni}}\right)^2}{\sum_{i=1}^{m} \frac{1}{P_{Ni}}} .$$

Then the loss is given by

$$L = \frac{SNR_{DS}}{SNR_{WDS}} = \frac{\left(\sum_{i=1}^{m} \sqrt{P_{si}}\right)^2 \sum_{i=1}^{m} \frac{1}{P_{Ni}}}{\left(\sum_{i=1}^{m} \frac{\sqrt{P_{si}}}{P_{Ni}}\right)^2 \sum_{i=1}^{m} P_{Ni}}$$

We can consider the signal power across the array as a random variable, of which each P_{si} is a sample. The noise power can be regarded in the same manner, with its samples P_{Ni}. With $E\{\ \}$ indicating expectation, we can approximate

$$\sum_{i=1}^{m} \sqrt{P_{si}} \approx mE\left\{\sqrt{P_{si}}\right\}$$

$$\sum_{i=1}^{m} \frac{1}{P_{Ni}} \approx mE\left\{\frac{1}{P_{Ni}}\right\}$$

$$\sum_{i=1}^{m} P_{Ni} \approx mE\left\{P_{Ni}\right\}$$

$$\sum_{i=1}^{m} \frac{\sqrt{P_{si}}}{P_{Ni}} \approx mE\left\{\frac{\sqrt{P_{si}}}{P_{Ni}}\right\}$$

Assuming P_{si} and P_{Ni} to be independent random variables, we can write

$$E\left\{\frac{P_{si}}{P_{Ni}}\right\} = E\left\{P_{si}\right\} E\left\{\frac{1}{P_{Ni}}\right\}.$$

The equation for the loss in gain then becomes

$$L \approx \frac{1}{E\left\{P_{Ni}\right\} E\left\{\frac{1}{P_{Ni}}\right\}} \qquad (1)$$

or

$$L \approx \frac{m^2}{\left(\sum_{i=1}^{m} P_{Ni}\right)\left(\sum_{i=1}^{m} \frac{1}{P_{Ni}}\right)}.$$

Note that the above approximation is exact when the signal power across the array is constant. The accuracy of the approximation begins to deteriorate as the samples of signal power begin to spread. The approximation itself, however, depends on the samples of noise power only.

If we assign a reasonable probability density to the variable P_{Ni}, we can compute the loss L of equation (1). For simplicity we shall use a symmetrical triangular density, shown in Figure 1 . Computation of the loss L gives

$$L \approx \frac{1}{E\left\{\frac{1}{P_{Ni}}\right\} E\left\{P_{Ni}\right\}} = \frac{\frac{w^2}{2\eta}}{(2\eta + w)\ln\left(\frac{\eta + \frac{w}{2}}{\eta}\right) - (2\eta - w)\ln\left(\frac{\eta}{\eta - \frac{w}{2}}\right)}.$$

where η is the mean of the density and w is its width. This loss is plotted in Figure 2 as a function of $\frac{\eta}{w}$, a $\frac{\text{mean}}{\text{spread}}$ parameter. Note that even at $\frac{\eta}{w} = .5$, the lowest reasonable value allowed by the assumed distribution, the loss is only 1.4 dB.

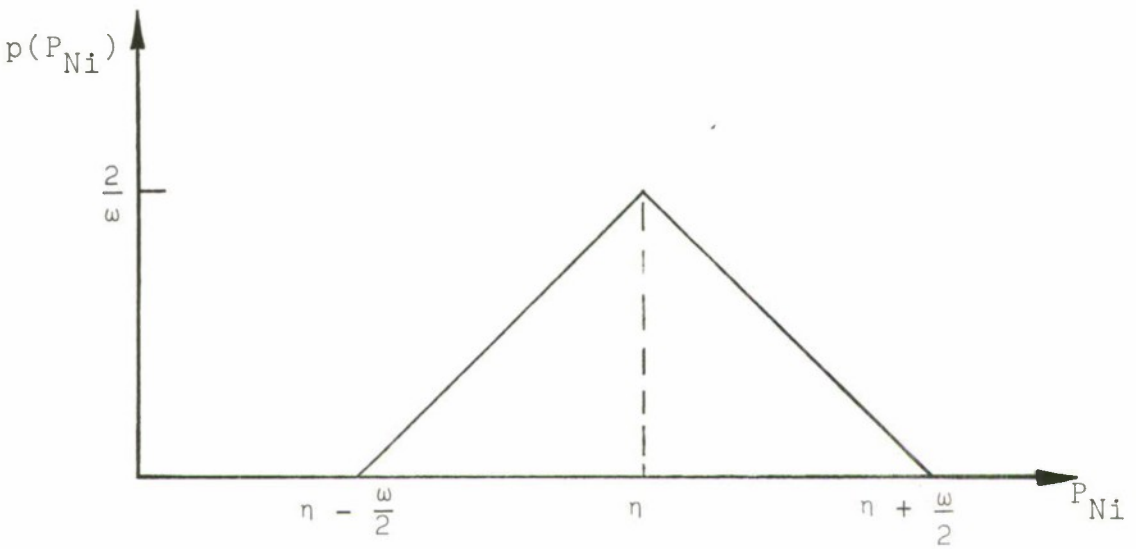


FIGURE 1 - ASSUMED PROBABILITY DENSITY OF NOISE POWER ACROSS THE ARRAY

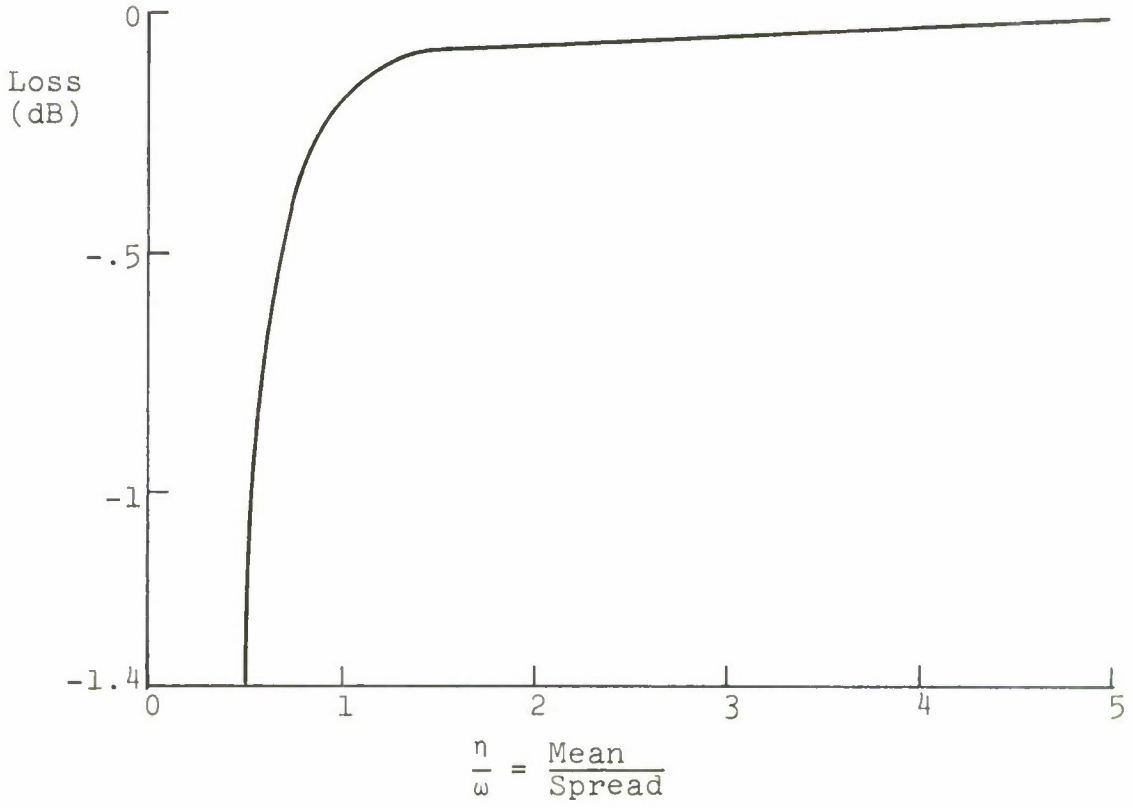


FIGURE 2 - LOSS OF DS PROCESSING BELOW WDS PROCESSING VS. MEAN/SPREAD PARAMETER OF DENSITY IN FIGURE 1

Within the accuracy of the approximation to the loss in gain, then, WDS processing offers negligible improvement over DS processing in exchange for a sizable increase in processing complexity.

E. Array Gain with DS Processing

The primary reason for building an array is to improve the signal-to-noise ratio over that present at a single element. This improvement, or gain, should be written in a functional form to determine which variables affect it. In this section the gain equation is derived based on certain s stated assumptions, the most restrictive being that the signal amplitude is the same at all element outputs and that the noise power is the same at all element outputs. Although these assumptions have frequently been violated in actual events recorded at the LASA, the gain derived in this section may be used as a first approximation to what is expected in a seismic array and as a basis for the design of the array when DS processing is employed.

In an array that uses DS processing (all element outputs appropriately delayed and then summed), the array gain may be easily computed as a function of the number of elements in the array, the average signal cross-correlation between all pairs of elements, and the average noise cross-correlation between all pairs of elements. With m elements, let the output of the i^{th} element be x_i. The array output then is simply the sum of all the element outputs, $\sum_{i=1}^{m} x_i$. The average power in the array output is:

$$P = \overline{\left(\sum_{i=1}^{m} x_i\right)^2}$$

$$P = \sum_{i=1}^{m} \sum_{j=1}^{m} x_i \, x_j$$

$$P = \sum_{i=1}^{m} \sum_{j=1}^{m} \overline{x_i \, x_j}$$

If we assume stationarity of the outputs, at least for the duration of a signal, we can write

$$\overline{x_i\, x_j} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)\, x_j\ (t)\, dt.$$

$$= R_{ij}(o),$$

where $R_{ij}\ (\tau)$ is the cross-correlation function of x_i and x_j. Under the assumption that the output power is the same at each element ($\overline{s_i^2} = \overline{x^2}$), we can write

$$\overline{x_i\, x_j} = R_{ij}(o) = \rho_{ij}\ \overline{x^2}\ ,$$

where ρ_{ij} is the correlation coefficient between the outputs of the i^{th} and j^{th} elements. Thus,

$$P = \sum_{i=1}^{m} \sum_{j=1}^{m} \rho_{ij}\ \overline{x^2} = (m^2 - m)\bar{\rho}\ \overline{x^2} + m\overline{x^2}\ , \quad (2)$$

where $\bar{\rho}$ is the average correlation coefficient of all possible element pairs.

If we now consider the element outputs to be composed of two uncorrelated components, signal and noise, the average power of the array output is the sum of the signal power P_s and noise power P_N, so that applying equation (2) gives

$$P_s = (m^2 - m)\ \bar{\rho}_s\ \overline{s^2} + m\overline{s^2}$$

and $$P_N = (m^2 - m)\ \bar{\rho}_n\ \overline{n^2} + m\ \overline{n^2},$$

where $\overline{s^2}$ is the signal power at each element

$\overline{n^2}$ is the noise power at each element

$\bar{\rho}_s$ is the average signal correlation between all element pairs

$\bar{\rho}_N$ is the average noise correlation between all element pairs

The gain in signal-to-noise ratio, then, is

$$G = \frac{P_{s/P_N}}{\overline{s^2} \,/\, \overline{n^2}} = \frac{(m-1)\,\bar{\rho}_s + 1}{(m-1)\,\bar{\rho}_N + 1} \quad . \qquad (3)$$

The above equation can be applied to any summed array where the signal amplitude at each element is constant, the noise power at each element is constant, and the signal is uncorrelated with noise. In the case of a seismic array like the LASA, however, it is useful to break the noise into its two contributors: local noise and teleseismic noise.

These two components differ, insofar as the array is concerned, by the fact that the wavelengths in the plane of the array of the local noise are much smaller than those of the teleseismic noise. Thus, one would expect the local noise correlation to fall off more quickly with distance than that of the teleseismic noise. A second difference is that, if the noise power is normalized to a signal a given distance away, the normalized local noise power varies more from site to site than does the normalized teleseismic noise power.

Because of these differences between local noise and teleseismic noise, a gain formula will be developed that separates these two noise components. In this derivation several assumptions will be made:

1. Signal amplitude is the same at all seismometer outputs.
2. Local noise power is the same at all seismometer outputs.
3. Teleseismic noise power is the same at all seismometer outputs.

4. Signal, local noise, and teleseismic noise are all uncorrelated.
5. The average cross-correlation for local noise over any cluster is constant for all clusters, but local noise is uncorrelated between clusters.
6. The average cross-correlation for teleseismic noise over any cluster is constant for all clusters.
7. Signal is perfectly correlated between all seismometers (6).

The first five assumptions are essentially the same as were made in deriving formula (3).

Some additional definitions are needed before deriving the gain formula in its two-component noise form. These are:

$\overline{n_L^2}$ = local noise power at each element

$\overline{n_T^2}$ = teleseismic noise power at each element

$R = \overline{n_L^2} / \overline{n_T^2}$

P_L = local noise power at the array output

P_{Tc} = teleseismic noise power at the output of each cluster.

P_T = teleseismic noise power at the output of the array

$\bar{\rho}_L$ = average correlation of local noise across any cluster.

$\bar{\rho}_{T1}$ = average correlation of teleseismic noise across any cluster

$\bar{\rho}_{T2}$ = average correlation of teleseismic noise between clusters.

c = number of elements per cluster

$\frac{m}{c}$ = number of clusters in the array.

If filtering is used following the DS, the linearity of the technique allows us to picture each element output as being prefiltered before summing. The noise powers and correlations considered here are those present after this prefiltering.

The signal power after the phased sum is

$$P_s = m^2 \overline{s^2} ,$$

using the assumption that the signal is perfectly correlated over the array. The local noise power out of the array, based on equation (2) and the assumed zero correlation between clusters, is

$$P_L = c \left[(c-1) \bar{\rho}_L + 1 \right] \overline{n_L^2} \cdot \frac{m}{c} .$$

The teleseismic noise power out of any cluster is

$$P_{Tc} = c \left[(c-1) \bar{\rho}_{T1} + 1 \right] \overline{n_T^2}$$

and the teleseismic noise power out of the entire array is

$$P_T = \left(\frac{m}{c}\right) \left[\left(\frac{m}{c} - 1\right) \bar{\rho}_{T2} + 1 \right] P_{Tc} ,$$

giving

$$P_T = \frac{m}{c} \left[(m-c) \bar{\rho}_{T2} + c \right] \left[(c-1) \bar{\rho}_{T1} + 1 \right] \overline{n_T^2} .$$

the gain in signal-to-noise ratio can then be written as

$$G = \frac{P_S}{P_L + P_T} \quad \frac{\overline{n_T^2} + \overline{n_L^2}}{\overline{s^2}}$$

$$= \frac{(\overline{n_T^2} + \overline{n_L^2})\, m}{\left[(c-1)\bar{\rho}_L + 1\right] \overline{n_L^2} + 1/c \left[(m-c)\, \bar{\rho}_{T2} + c\right] \left[(c-1)\bar{\rho}_{T1} + 1\right] \overline{n_T^2}}$$

Using R, the ratio of local noise to teleseismic noise at the site, we have:

$$G = \frac{(1 + R)\, m}{\left[(c-1)\bar{\rho}_L + 1\right] R + \frac{1}{c} \left[(m-c)\bar{\rho}_{T2} + c\right] \left[(c-1)\bar{\rho}_{T1} + 1\right]} \qquad (4)$$

The importance of this equation can be seen if it is applied to the present LASA, where c = 25 and m = 525. In this case the clusters are small enough to consider the teleseismic noise to be perfectly correlated over a cluster, $\overline{\rho}_{T1} = 1$. We shall also assume that the clusters are separated enough so that teleseismic noise is uncorrelated between clusters, $\overline{\rho}_{T2} = 0$. Thus, for LASA

$$G_{LASA} = \frac{525\ (1 + R)}{(24\overline{\rho}_L + 1)\ R + 25} \quad . \qquad (5)$$

Equation (5) is plotted in Figure 3 for several values of R. The dotted line indicates the gain achievable when the noise between elements is completely uncorrelated. The curves show the degradation in gain as $\overline{\rho}_L$ increases and/or as R decreases. The degradation in gain is primarily due to the correlation of teleseismic noise across a cluster, and it is increased by increasing amounts of local noise correlation across a cluster. The effect of local noise correlation is, of course, diminished by a smaller ratio of local noise to teleseismic noise, but then the effect of teleseismic noise correlation is enhanced.

Clustering of seismometers in the LASA was originally proposed (7) to ease signal transmission problems over the entire 200 km diameter array. The size of a cluster (7km.) was chosen so it would reduce local noise, which has an average wavelength of about 4 km. Teleseismic noise, with an average wavelength of about 16 km. is well correlated over the 7 km. cluster, so that the action of a cluster is indeed to reduce only local noise. Expanding a cluster sufficiently allows it to strongly reduce teleseismic noise as well as local noise. (This is supported by an UED experiment (1) in which the power gain of an array formed by DS processing the outputs of one element from each LASA subarray was found to be the

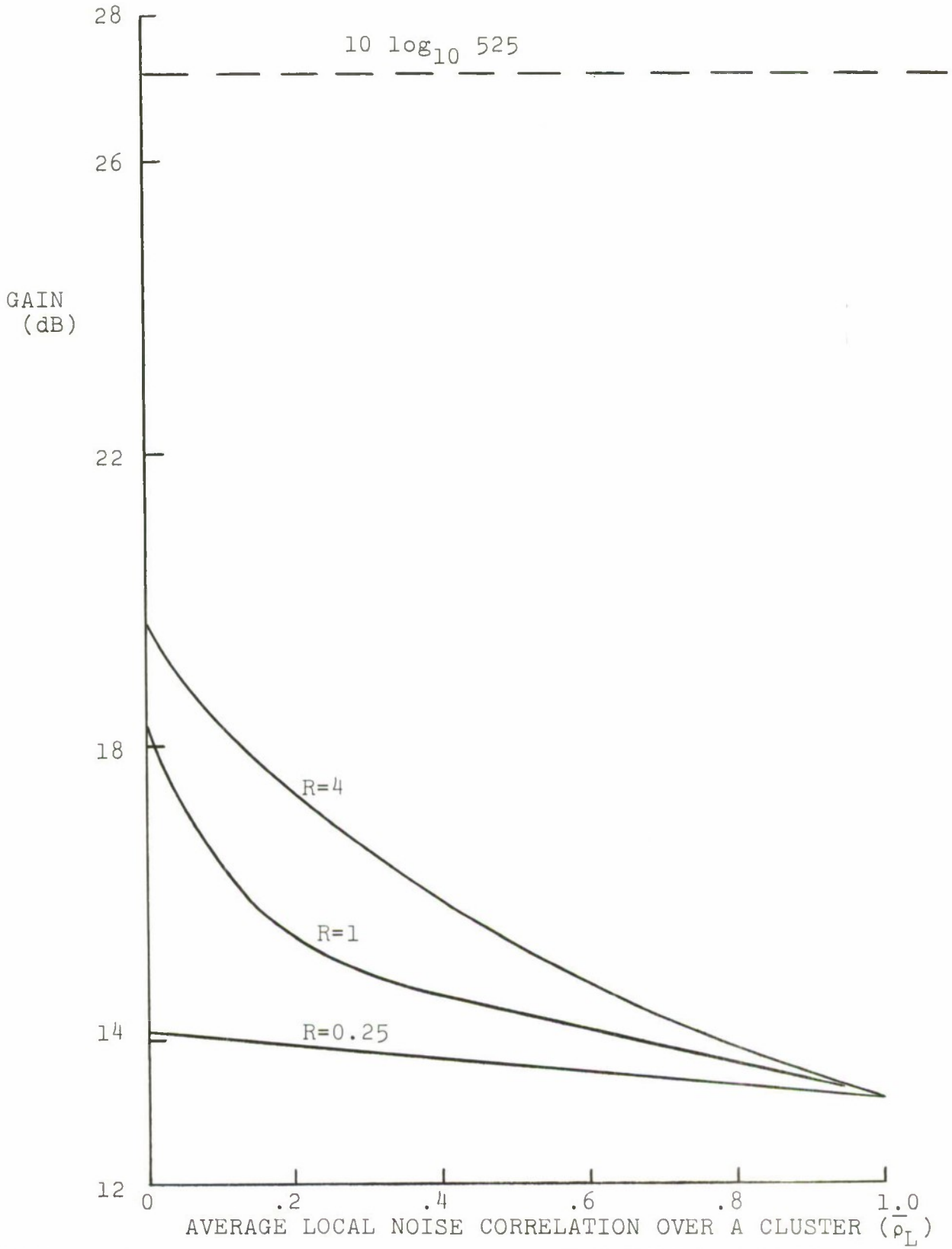


FIGURE 3 - DS GAIN OF LASA VS. AVERAGE LOCAL NOISE CORRELATION FOR SEVERAL RATIOS OF LOCAL NOISE TO TELESEISMIC NOISE

number of elements in that array, indicating no noise correlation between the elements.) The action of an expanded cluster cannot be viewed solely as suppression of local noise. The cluster concept now becomes elements grouped together only for cabling efficiency.

Since an expanded subarray reduces teleseismic noise as well as local noise, at a quiet site such as the Montana LASA where teleseismic noise is dominant, one would expect to see a very large gain improvement in DS processing by using the expanded cluster. A Lincoln Labs experiment (3) shows that this is indeed the case when a subarray diameter of 22 km, instead of 7 km, was used. The ratio of local noise to teleseismic noise at an array site, in fact, has been shown above to be a strong factor in array performance with DS processing and noise correlation between the elements so that it must be considered in array design and performance.

Considering a LASA with expanded clusters, we see that at a quiet site ($R<1$) such as the one in Montana, further decorrelation of local noise by expansion of cluster size does not improve the gain much. However, with a cluster diameter of 22 km. or greater the teleseismic noise correlation across a cluster $\bar{\rho}_{T1}$ will decrease appreciably from 1, providing a large gain improvement over the array with smaller clusters. With the present concentration of clusters near the center of the Montana LASA, however, as the cluster size is expanded the outer elements of each cluster will come close to each other, causing an increase in the teleseismic noise correlation between clusters $\bar{\rho}_{T2}$ and a corresponding decrease in gain. To take full advantage of the additional array gain provided by cluster expansion, then, the dense array center must be thinned by removing several clusters from that region and placing them in the less dense outlying regions of the array. In other words, a more uniform distribution of clusters is desired.

The relatively high density of elements in the center of the array might, in fact, have a more detrimental effect on the array gain, an effect that is not immediately obvious. It is conceivable that thinning the center of the array by removing elements, but not placing them anywhere else in the array, could actually increase the array gain. In other words, including these elements in the array not only wastes seismometers, but also reduces the array gain.

As a simple example of a case in which the addition of an element can actually reduce the array gain, consider an array composed of three equally weighted elements that are separated so that there is no noise correlation between them ($\rho_N = o$). If the signal is perfectly correlated ($\rho_s = 1$), then application of equation (3) gives a gain of G=3. If a fourth element is added with perfect signal correlation near enough to one of the original elements so that its noise correlation with that element is 0.9, but its noise correlation with the remaining two elements is zero (See Figure 4), the average noise correlation becomes $\rho_N = .15$. The gain of equation (3) then becomes

$$G = \frac{4}{3(.15) + 1} = 2.76,$$

less than the gain G=3 with the three element array.

A general condition for the deterioration of the array gain by the addition of an element can easily be derived. Assuming that the signal is perfectly correlated over the array, equation (2) gives the m-element array gain as

$$G_m = \frac{m}{1 + (m-1)\,\bar{\rho}_m}$$

Adding an element gives

$$G_{m+1} = \frac{m+1}{1 + m\,\bar{\rho}_{m+1}}$$

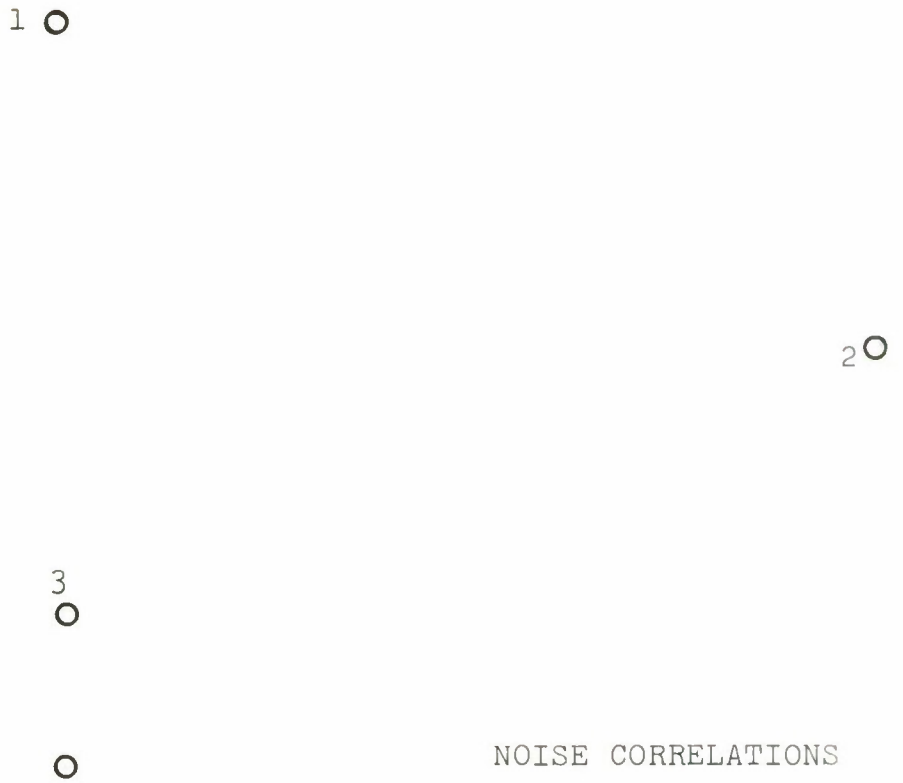


FIGURE 4 - SKETCH OF AN ARRAY IN WHICH AN EXTRA ELEMENT DECREASES THE GAIN

The additional element causes a drop in gain

when $\frac{G_{m+1}}{G_m} < 1,$

which results in

$$\bar{\rho}_{m+1} - \bar{\rho}_m = \Delta\bar{\rho} > \frac{1-\bar{\rho}_m}{m^2} \quad . \qquad (6)$$

Equation (5) is the condition for loss of gain by addition of an element to the array. An equivalent condition may be derived by noting that

$$\bar{\rho}_m = \frac{2}{m(m-1)} \sum_{j=2}^{m} \sum_{i=1}^{j} \rho_{i,j} \quad ,$$

where $\rho_{i,j}$ is the noise correlation between the i^{th} element and the j^{th} element. Also

$$\bar{\rho}_{m+1} = \frac{2}{(m+1)m} \left[\sum_{j=2}^{m} \sum_{i=1}^{j} \rho_{i,j} + \sum_{i=1}^{m} \rho_{i,m+1} \right] .$$

Using equation (5), we have

$$\frac{2}{(m+1)m} \left[\sum_{j=2}^{m} \sum_{i=1}^{j} \rho_{i,j} + \sum_{i=1}^{m} \rho_{i,m+1} \right] > \frac{2}{m(m-1)} \sum_{j=2}^{m} \sum_{i=1}^{j} \rho_{i,j}$$

$$\frac{1 - \frac{2}{m(m-1)} \sum_{j=2}^{m} \sum_{i=1}^{j} \rho_{i,j}}{m^2} ,$$

which results in

$$\sum_{i=1}^{m} \rho_{i,m+1} > \frac{m+1 + (2m+1)(m-1)\bar{\rho}_m}{2m} \qquad (7)$$

Equation (7) is a condition for loss of gain with an additional element that puts a bound on the sum of the noise correlations between the added element and the existing ones in terms of the number of elements and the average noise correlation in the original array. For an array with many elements we can approximate equation (7) by

$$\sum_{i=1}^{m} \rho_{i,m+1} \approx m\,\bar{\rho}_{i,m+1} > \tfrac{1}{2} + m\bar{\rho}_{m} \,. \qquad (7a)$$

If the array already has fairly high noise correlations between its elements (such that $m\bar{\rho}_m >> \frac{1}{2}$)m the condition for a decrease in gain with an additional element is simply that the average noise correlation between the additional element and the other array elements be greater than the average noise correlation across the original array.

As far as the LASA is concerned, this detrimental situation might very conceivably occur if an element were added in the dense center of the array, where its average correlation with the present elements would be relatively high. Conversely, the D^S array gain might very conceivably be increased by removing an element from the array center, and it would be further increased by relocating that element in the sparsely populated outskirts of the array. Applying this reasoning to more than one of the central elements leads to the requirement that the array elements be distributed more evenly.

F. Conclusions

Several conclusions can be drawn from this report. The first is that the present 7km diameter cluster size is too small, and expansion of the cluster size will increase the signal-to-noise ratio gain of the LASA. With expanded clusters DS processing gains are close to those of FS processing and WDS processing without the additional complexity of these two techniques. DS processing with expanded clusters, then, is the type of processing to use.

Along with cluster expansion it is also desirable to spread the clusters far enough apart so that there is very low noise correlation between elements of adjacent clusters, because noise correlation decreases the array gain. It is shown that removal of an element whose noise is relatively highly correlated with the noise at other elements may actually decrease the array gain, so that low noise correlation is indeed desirable. The combination of cluster expansion with cluster separation in effect means a more uniform arrangement of elements than is present in the existing LASA.

A gain formula for DS processing is derived that considers the local noise and the teleseismic noise separately because the two have different spatial variations in their correlations and different power. The formula is given by equation (4).

Finally, it is shown that a seismic array with DS or WDS processing may be viewed as a diversity system with a commonly used diversity combining technique. The benefit of this approach is that it can take advantage of the important developments and theory that are appearing in the technical literature.

SECTION III

A. Introduction

It has been suggested that earthquakes and nuclear blasts result in seismograms that differ significantly in their spectral energy content. If this is the case, the analysis of the spectral distribution of energy might provide useful statistics for discriminating between these two events. This chapter is concerned with the estimation of the energy density spectrum of a transient signal in the case where the desired signal is corrupted by additive, stationary, Gaussian noise. Although this investigation is motivated by the potential seismic applications, it is felt that the results may be of interest for other applications as well.

There is a considerable literature associated with spectral estimation, but the majority of the papers are concerned with the problem of estimating the power density spectrum of a stationary random process on the basis of one sample function observed during an interval of limited duration. Although these papers are relevant, they consider a problem that is fundamentally different from the one of interest here -- that of estimating the energy density spectrum of a transient which is corrupted by additive stationary noise. We are aware of two papers, by McIvor [8] and by Larrowe and Crabtree [9], that have attempted to apply a concept of a time-varying spectrum to transient seismic data; however, we do not feel that these papers have handled the problem satisfactorily. These two papers will be discussed in a later section.

The problem under consideration can be formulated as follows. There is a finite energy "signal", x(t), for which the Fourier Transform,

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt, \qquad (1)$$

is assumed to exist. It is desired to calculate the energy density spectrum (EDS), $|X(f)|^2$ for this signal. The available data consist of a record of $x(t)+n(t)$, where $n(t)$ is stationary Gaussian noise, with autocorrelation function,

$$R_n(\tau) = \overline{n(t)n(t+\tau)} = \int_{-\infty}^{\infty} S_n(f)e^{j2\pi ft}\,df, \qquad (2)$$

where $S_n(f)$ is the power density spectrum (PDS) of the noise. It is assumed that the approximate onset time of $x(t)$ is known, but that its duration is somewhat uncertain. The problem is then to devise a method of estimating $|X(f)|^2$. In doing this a major problem will be the variability in the estimates caused by the stochastic nature of the noise, and, in trying to reduce this variability, the resulting distortion in the "signal" contribution to the estimate will have to be considered.

The "direct" method of estimating spectra, which essentially consists of calculating the magnitude-squared of the Fourier Transform of a finite-time sample of the process, will first be considered. It will be observed that this method has serious statistical deficiencies in that the variance of the noise contribution to the estimate is always greater than the square of the expected value of the noise contribution. Furthermore, this will be the case even if the observed sample is multiplied by a "time window" before transforming. One approach to alleviating this variability problem is the indirect method discussed at length by Blackman and Tukey [10]. This method was proposed by them as the best means of estimating the PDS of a stationary process and it has since been widely applied to that problem. However, as will be discussed below, it does not appear to be a useful method of estimating the EDS of a transient corrupted by additive noise. From these discussions it will be concluded that a direct method, despite

its inherent statistical shortomcings, is the appropriate method for this problem. Suggestions will also be made for improving the statistical quality of the method as further experimental data become available. A potentially valuable method of reducing variability will then be discussed. This scheme is based on subtracting, from the observed $x(t)+n(t)$, an estimate of the noise that is based only on the noise preceding the onset of the signal. The last section consists of a critical discussion of the papers of McIvor [8] and of Larrow and Crabtree [9]. Finally, details of derivations are presented in Appendix A.

There are a few topics that are omitted from this discussion that would have to be considered before implementing these recommendations with actual data. For convenience, continuous representations of the waveforms are employed throughout the report. As calculations will presumably be done on digital computers, it is necessary to consider the problem of discrete representations of these waveforms. This should not be particularly difficult, but there are some issues to be resolved. In applying these calculations there will be several parameters to be chosen and varied experimentally. No detailed recommendations for choosing and varying these parameters are included in this chapter. Finally, there are potential problems of approximation that should be considered. The noise prediction scheme involves operations closely related to "whitening". Some consideration should be given to the sensitivity of the overall results to the approximations and compromises that will of necessity be involved in the computer implementation of the predictor.

B. The Time Window

It is assumed that all calculations are performed on

$$z(t) = a(t)[x(t)+n(t)] \tag{3}$$

where a(t) is called the "time window", and is assumed to have a Fourier Transform

$$A(f) = \int_{-\infty}^{\infty} a(t)\, e^{-j2\pi ft}\, dt \tag{4}$$

Specifically, $|Z(f)|^2$ will be calculated, where

$$Z(f) = \int_{-\infty}^{\infty} a(t)[x(t)+n(t)]e^{-j2\pi ft}\, dt$$

$$= Z_x(f) + Z_n(f) \tag{5}$$

Perhaps the simplest choice of a(t) would be 1 for the time interval that is thought to contain "most of" the signal, and 0 outside. However, there are two reasons for choosing a smoother off-on transition for a(t). In the application under consideration, the "duration" of x(t) is not known. Presumably, a(t) should be "turned off" when x(t) becomes small compared to the noise level. An intuitively appealing way to incorporate the uncertainty about an appropriate turn-off time is to turn off a(t) gradually. In this way, times with good signal-to-noise ratios would be weighted more heavily than those without. The signal contribution to Z(f), $Z_x(f)$,is simply

$$Z_x(f) = A(f) \circledast X(f) \tag{6}$$

where $\circledast$ indicates convolution. To reduce the distortion in going from X(f) to $Z_x(f)$, A(f) should be a narrow pulse at the origin of the frequency domain. If a(t) is a rectangular pulse, A(f) will have relatively large "side lobes". These can be reduced by a more "rounded" choice of a(t).

In the discussion which follows,the "width" of A(f) and the size of its sidelobes will be important parameters. To illustrate the relation between the shapes of A(f) and a(t)

two examples are presented in Fig. 5. The first example involves a rectangular shape for a(t), which leads to relatively large sidelobes in A(f). The second example is the Hanning Window, which has been recommended by Blackman and Tukey as a "lag window". (The lag window will be defined and discussed below.) The Hanning window is more rounded than the rectangular one and, hence, has relatively small sidelobes and a somewhat wider central peak. It will be convenient in much of what follows to assume

$$A(f) = 0 \qquad |f| > 1/T \tag{7}$$

The two examples just considered indicate the degree of this approximation.

C. Noise-Alone Case

1. Direct Method

In order to characterize the variability problems that result from the noise it is easiest to begin by considering the case of noise alone, i.e., the case of x(t) = 0. In Section D these results will be generalized to include the signal contribution and any interaction terms.

$$|Z_n(f)|^2 = \int d\alpha \int d\beta a(\alpha)a(\beta)n(\alpha)n(\beta)e^{-j2\pi f(\alpha-\beta)} \tag{8}$$

Taking expected values, and interchanging the order of integration and expectation,

$$E[|Z_n(f)|^2] = \overline{|Z_n(f)|^2} = \int d\alpha \int d\beta a(\alpha)a(\beta)R_n(\alpha-\beta)e^{-j2\pi f(\alpha-\beta)} \tag{9}$$

Recognizing that the result of the β integration is simply $a(\alpha) \circledast R_n(\alpha)e^{-j2\pi f\alpha}$

$$\overline{|Z_n(f)|^2} = \int d\alpha a(\alpha) \int du\, e^{j2\pi u\alpha} A(u)S_n(u+f) \tag{10}$$

Interchanging the order of integration, and integrating over α,

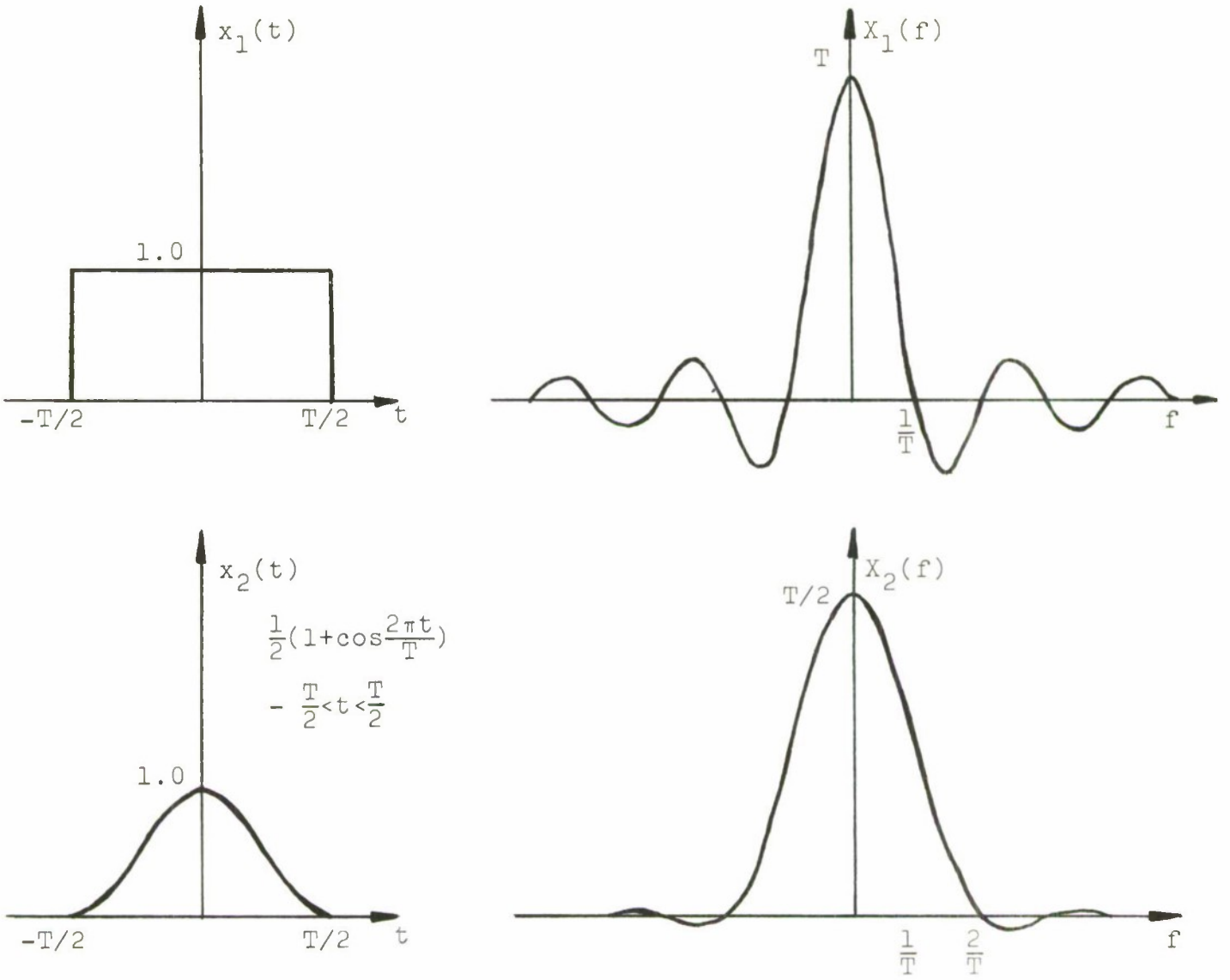


FIGURE 5 - EXAMPLES OF TWO TIME-WINDOWS AND THEIR TRANSFORMS

$$\overline{|Z_n(f)|^2} = \int du\ A^*(u)A(u)S_n(u+f)$$
$$= \int dv\ S_n(v)|A(v-f)|^2$$
$$= S_n(f) \otimes |A(f)|^2 \qquad (11)$$

Thus, the expected value of $|Z_n(f)|^2$ is a "smoothed" version of the true PDS, $S_n(f)$, where the "smoothing" is accomplished by a convolution in the frequency domain.

Similar calculations, which employ the well-known property of zero-mean Gauss variables,

$$\overline{n(t)n(\tau)n(\alpha)n(\beta)} = \overline{n(t)n(\tau)}\ \overline{n(\alpha)n(\beta)} + \overline{n(t)n(\alpha)}\ \overline{n(\tau)n(\beta)} + \overline{n(t)n(\beta)}\ \overline{n(\alpha)n(\tau)} \qquad (12)$$

are presented in Appendix A, and yield

$$\mathrm{cov}[|Z_n(f_1)|^2, |Z_n(f_2)|^2] = |\int du S_n(u)A^*(u-f_1)A(u+f_2)|^2 + |\int du S_n(u)A^*(u-f_1)A(u-f_2)|^2 \qquad (13)$$

which specializes to

$$\mathrm{var}[|Z_n(f)|^2] = |\int du S_n(u)|A(u-f)|^2|^2 + |\int du S_n(u)A(u+f)A^*(u-f)|^2$$
$$= (E[|Z_n(f)|^2])^2 + |\int du S_n(u)A(u+f)A^*(u-f)|^2 \qquad (14)$$

This last result is a generalization of the standard "periodogram" result (see, e.g., Davenport and Root, [11], pp. 107-108), which may be obtained from the above by choosing

$$A(t) = 1/\sqrt{T} \qquad 0 < t < T \qquad (15)$$
$$= 0 \qquad \text{elsewhere}$$

The important observation from Equation (14) is that, no matter what a(t) is chosen

$$\mathrm{var}[|Z_n(f)|^2] \geq (E[|Z_n(f)|^2])^2 \qquad (16)$$

In particular, using longer and longer observation times will

never reduce the variance to less than the square of the mean. Thus, $|Z_n(f)|^2$ cannot provide a "consistent" estimate of $S_n(f)$.

Equation (16) implies that the direct method, without any modifications or further calculations, would not be a satisfactory method for estimating the PDS of a stationary random process. There are, of course, ways to modify the method that would improve its statistical features, and these provide one way of motivating the indirect method. Before considering these modifications, a few properties of Equations (11), (13) and (14) should be noted.

For the present discussion, it is assumed that a(t) is zero outside of an interval of length T. It is further assumed that within this interval a(t) has been chosen in such a way as to minimize the width of A(f). As illustrated by example in the previous section, this could result in an A(f) having a total width of roughly 2/T. Assuming, therefore, that

$$A(f) = 0 \qquad |f| > 1/T \tag{17}$$

Equation (13) implies that $|Z_n(f_1)|^2$ and $|Z_n(f_2)|^2$ are uncorrelated for

$$||f_1|-|f_2|| > 2/T \tag{18}$$

i.e., for frequencies separated by more than 2/T. And, Equation (14) implies that

$$\mathrm{var}[|Z_n(f)|^2] = (E[|Z_n(f)|^2])^2 \qquad |f| > 1/T \tag{19}$$

Considering only the samples $|Z_n(2K/T)|^2$ for $K \geq 0$, these results may be summarized by

$$\overline{|Z_n(\tfrac{2K}{T})|^2} = [S_n(f) \circledast |A(f)|^2]_{f=(2K/T)} \tag{20}$$

$$\mathrm{cov}[|Z_n(\tfrac{2K}{T})|^2, |Z_n(\tfrac{2\ell}{T})|^2] = 0 \quad \text{for } |K| \neq |\ell| \tag{21}$$

$$\text{var}\left[\left|Z_n\left(\frac{2K}{T}\right)\right|^2\right] = \begin{cases} \overline{\left|Z_n(2K/T)\right|^2}^2 & K \neq 0 \\ 2\overline{\left|Z_n(0)\right|^2}^2 & K = 0 \end{cases} \qquad (22)$$

2. Direct Method with Further Smoothing

In considering these equations an important question is how rapidly $S_n(f)$ varies with f. If it varies slowly compared to 2/T, then the convolution of Equation (20) will not distort the power density spectrum, i.e.,

$$S_n(f) \circledast |A(f)|^2 \simeq cS_n(f) \qquad (23)$$

If, furthermore, it varies slowly compared to frequency intervals of several times 2/T, then, the samples $|Z_n(2K/T)|^2$, will have the same expectation for several successive values of K. In this case, a local average of these statistics would be an appropriate estimate of this expectation and, since successive values of $|Z_n(2K/T)|^2$ are uncorrelated, the local average would be better statistically than the individual samples. A convenient measure of the variability of a single statistic is the "coefficient of variation", which is defined as the ratio of the standard deviation of the mean. It is easy to show that adding together M uncorrelated random variables with identical means and variances results in a new random variable with a coefficient of variation that is $1/\sqrt{M}$ times that of the original variable.

To summarize, for K>0, $|Z_n(2K/T)|^2$ has a coefficient of variation of unity. If T is large enough that $S_n(f)$ is essentially constant over frequency regions of length 2M/2T, then the local averages taken over M of the samples $|Z_n(2K/T)|^2$ will yield spectral estimates with coefficients of variation equal to $1/\sqrt{M}$. The calculation just described -- the "direct" method, followed by further "smoothing" in the frequency domain -- could be a perfectly satisfactory way of estimating the PDS of a stationary random process, provided it were possible to

choose T large enough to achieve a low coefficient of variation for the spectral estimates. This method is essentially equivalent to the "indirect" method of Blackman and Tukey, as will be discussed below. A more careful, and thorough discussion may, of course, be found in Blackman and Tukey.

An operation similar to that of adding M adjacent values of $|Z_n(2K/T)|^2$ would be convolving $|Z_n(f)|^2$ with Q(f), where Q(f) is a pulse of width 2M/T located at the origin of the frequency domain. Because this operation does in a sense combine M uncorrelated samples, one would expect it to result in coefficient of variation of roughly $1/\sqrt{M}$. Defining the result of this further smoothing as P(f) yields

$$P(f) = |Z_n(f)|^2 \circledast Q(f) \tag{24}$$

and, therefore,

$$\begin{aligned} \overline{P(f)} &= \overline{|Z_n(f)|^2} \circledast Q(f) \\ &= S_n(f) \circledast |A(f)|^2 \circledast Q(f) \\ &\approx c\, S_n(f) \circledast Q(f) \end{aligned} \tag{25}$$

where the last approximate equality is based on $|A(f)|^2$ being narrow compared to Q(f), which will be the case if Q(f) is sufficiently wide compared to 2/T that a low coefficient of variability results. P(f) is sometimes called a "smoothed periodogram".

3. Indirect Method

To relate this method to the indirect method of Blackman and Tukey, it is easiest to make a few specific assumptions and definitions.

Let

$$a(t) = \begin{cases} 1/\sqrt{T} & 0 < t < T \\ 0 & \text{elsewhere} \end{cases} \tag{26}$$

and

$$\phi_n(\tau) = \frac{1}{T}\int_0^{T-|\tau|} n(t)n(t+|\tau|)\,dt \qquad (27)$$

Then it is easy to show that

$$\Phi_n(f) = \int \phi_n(\tau)\, e^{-j2\pi f\tau}\, d\tau = |Z_n(f)|^2 \qquad (28)$$

Let $D_i(\tau)$ be a symmetric function of τ that is zero for $|\tau| > T_M$, and

$$Q_i(f) = \int D_i(\tau)\, e^{-j2\pi f\tau}\, d\tau \qquad (29)$$

Then

$$P_i(f) = |Z_n(f)|^2 \circledast Q_i(f) = \Phi_n(f) \circledast Q_i(f) = \int d\tau\, e^{-j2\pi f\tau} D_i(\tau)\phi_i(\tau) \qquad (30)$$

This last equation states that the direct method followed by smoothing in the frequency domain is equivalent to calculating an estimate of the noise autocorrelation function, $\phi_n(\tau)$ [Equation 27], multiplying this by a "lag window", $D_i(\tau)$, and then calculating the Fourier Transform of the result. By assumption $D_i(\tau)$ is zero for $|\tau| > T_M$. T_M is called the "maximum lag". If $D_i(\tau)$ is chosen appropriately, $Q_i(f)$ will have a width of approximately $2/T_M$, and the coefficients of variability of the resulting spectral estimates will therefore be approximately $\sqrt{T_M/T}$.

The label "indirect method" is applied to the Blackman-Tukey approach because it begins with the estimation of the autocorrelation function. Actually, Blackman and Tukey recommend

$$C_{oo}(\tau) = \frac{1}{T-|\tau|}\int_0^{T-|\tau|} n(t)n(t+|\tau|)dt = \frac{T}{T-|\tau|}\,\phi_n(\tau) \qquad (31)$$

as the estimate, rather than $\phi_n(\tau)$. $C_{oo}(\tau)$ has some appeal as the "natural" way to use all of the available data to estimate the autocorrelation function, and it is an unbiased estimate in that

$$E[C_{oo}(\tau)] = R_n(\tau) \qquad (32)$$

However, there is some dispute as to which is preferable, and Parzen, for example, recommends the biased estimator $\emptyset_n(\tau)$. According to Parzen [12], $\emptyset_n(\tau)$ has a smaller mean square error than $C_{oo}(\tau)$, and the Fourier Transform of $\emptyset_n(\tau)$ will always be positive [see Equation 28], which is not the case for $C_{oo}(\tau)$.

In practice, the lag window is applied in two steps. The first, and usually the most significant in its effect, is setting the estimated autocorrelation function to zero for all values of the argument larger than the maximum lag. The second is the multiplication of the remaining function by the lag window. In the discrete formulation appropriate for digital computers it is often easier to implement the second step by an appropriate convolution in the frequency domain after transforming the autocorrelation function (which has already been truncated at the maximum lag). The problem of selecting a lag window is discussed at some length, with several illustrations, by Blackman and Tukey. There is no straightforward method of determining a "best" window, but it does not seem to matter much which of several "good" possibilities are used.

In using these calculations to estimate the PDS of a stationary random process, it is first necessary to make some guesses about the rate of change of $S_n(f)$ with frequency in order to select the maximum width of $Q(f)$ that could be tolerated without distortion. It is also necessary to specify the desired coefficient of variation for the estimate; for the sake of illustration assume this is 1/3. Then T must be chosen so that about nine intervals of $2/T$ are included within the width of $Q(f)$. Specifying the width of $Q(f)$ is equivalent to specifying the maximum lag, T_M. To achieve a coefficient of variation of 1/3, T must be at least as large as $9T_M$.

To summarize, the direct method by itself leads to spectral estimates with a coefficient of variation of at least unity and an expected value,

$$\overline{|Z_n(f)|^2} = S_n(f) \circledast |A(f)|^2 \qquad (33)$$

And, this will be the case no matter what time window, a(t), is chosen. To reduce variability it is necessary to perform additional smoothing, after calculating $|Z_n(f)|^2$ (or, equivalently, in the indirect method, by multiplying the autocorrelation function by a lag window). This additional smoothing leads to a spectral estimate P(f), with expected value,

$$P(f) = S_n(f) \circledast |A(f)|^2 \circledast Q(f) \qquad (34)$$

Typically Q(f) is much broader than $|A(f)|^2$, so that

$$|A(f)|^2 \circledast Q(f) = cQ(f) \qquad (35)$$

and hence

$$\overline{P(f)} = cS_n(f) \circledast Q(f) \qquad (36)$$

Furthermore, the coefficient of variability of P(f) will be proportional to the square root of the ratio of the widths of $|A(f)|^2$ and Q(f). This ratio may be expressed as T_M/T, where T_M is the "maximum lag" and T is the total observation interval.

D. Signal Plus Noise Case

1. Direct Method

The observed waveform z(t) is now assumed to include the transient "signal", x(t).

$$z(t) = a(t)[x(t)+n(t)] \qquad (37)$$

As shown in Appendix A, this leads to

$$E[|Z(f)|^2] = |Z_x(f)|^2 + \overline{|Z_n(f)|^2} \qquad (38)$$

$$\text{var}[|Z(f)|^2] = \text{var}[|Z_n(f)|^2] + 2|Z_x(f)|^2\overline{|Z_n(f)|^2} + 2\text{Re}Z_x^2(f)\int du S_n(u)A^*(u+f)A(u-f) \qquad (39)$$

And, as before, for frequencies f_1 and f_2 separated by more than the width of A(f), $|Z(f_1)|^2$ and $|Z(f_2)|^2$ are uncorrelated random variables. (In Equations 38 and 39, $Z_x(f)$ is as defined earlier in Equation 6, $Z_n(f)$ is the quantity discussed in Section C and the mean and variance of $|Z_n(f)|^2$ are given by Equations 11 and 14.) By examining Equations (38) and (39), it may be observed that the expected value of the estimate is just the sum of the contributions of signal and noise, but that the variance includes an interaction term proportional to both $|Z_x(f)|^2$ and $\overline{|Z_n(f)|^2}$ plus another interaction term that will be zero for f larger in magnitude than half the width of A(f). The implications of these results will be discussed after considering the indirect method.

2. Indirect Method

To dismiss the possibility of applying the indirect method, with the customary choice of parameters, to the problem of estimating the EDS of a transient, it is sufficient to consider the effect of the indirect calculation on the signal term alone. In Section C the case of noise alone was considered. The total spectral estimate includes both of these plus an interaction term. The signal contribution, obtained by setting n(t) = 0, is

$$P_x(f) = |Z_x(f)|^2 \otimes Q(f) = |X(f) \otimes A(f)|^2 \otimes Q(f) \qquad (40)$$

where a(t) is a rectangular pulse of duration T, and A(f) therefore has its first zero crossings at $\pm 1/T$, and a width of approximately 2/T. The customary choice of parameters for the indirect method leads to Q(f) having a width of approximately ten times that of A(f). In this application, that would involve smoothing over a frequency interval of 20/T, which could easily be comparable to the frequency interval covered by the entire spectrum of X(f). In order to view the problem in the time domain, it

should be recalled that T is chosen to correspond roughly to the duration of x(t). A pulse of length T has an autocorrelation function extending from -T to +T. Applying the indirect method with $T_M = (1/10)T$ corresponds to truncating this autocorrelation function at 1/10 of its duration. In short, using the indirect method with parameters chosen to constrain the standard deviation of the spectral estimate to be a small fraction of the expected value of the noise contribution to the spectral estimate, results in a potentially severe distortion of the signal's EDS.

E. Recommended Method

The recommended method is essentially the direct one, except that it could be followed by further smoothing of the kind accomplished by the indirect method. However, the particular smoothing that would be appropriate depends on the detailed statistics of this process, and cannot at this time be specified. To understand the considerations that should go into this smoothing, it is useful to review the properties of the direct method.

As before,

$$z(t) = a(t)[x(t)+n(t)] \tag{41}$$

Defining

$$y_K = |Z(2K/T)|^2 \tag{42}$$

the earlier results may be summarized as follows:

$$\overline{y_K} = |Z_x(2K/T)|^2 + \overline{|Z_n(2K/T)|^2} \tag{43}$$

where

$$|Z_x(f)|^2 = |A(f) \circledast X(f)|^2 \tag{44}$$

$$\overline{|Z_n(f)|^2} = S_n(f) \circledast |A(f)|^2 \tag{45}$$

$$\mathrm{cov}(y_K, y_\ell) \simeq 0 \quad \text{for } |K| \neq |\ell| \tag{46}$$

$$\text{var } y_K = [\overline{|Z_n(2K/T)|^2}^2 + 2|Z_x(2K/T)|^2\overline{|Z_n(2K/T)|^2}](1+\delta_{K,0}) \tag{47}$$

The term

$$\delta_{K,0} = \begin{matrix} 1 & K=0 \\ 0 & K\neq 0 \end{matrix} \tag{48}$$

results from the integral $\int du S_n(u)A^*(u+f)A(u-f)$ in Equations (14) and (39). This integral is zero for $|f| > 1/T$ and equals $|Z_n(0)|^2$ for $f=0$.

Defining

$$\eta_K = \overline{|Z_n(2K/T)|^2} \tag{49}$$

$$x_K = \overline{|Z_x(2K/T)|^2} \tag{50}$$

gives

$$\overline{y_K} = x_K + \eta_K \tag{51}$$

$$\text{var } y_K = \eta_K^2 + 2x_K\eta_K](1+\delta_{K,0}) \tag{52}$$

In the application under consideration the η_K may be assumed known. In practice these could be obtained by performing the (indirect) Blackman-Tukey calculation on a long section of noise preceding the signal. In this calculation the lag window would be chosen so that the expected value of the noise contribution to both calculations would be the same. This implies

$$D_i(\tau) = a(\tau) \circledast a(-\tau) \tag{53}$$

or equivalently

$$Q_i(f) = |A(f)|^2 \tag{54}$$

The remaining statistical problem is then, given the random variables y_K, and assuming the η_K are known, how should the x_K be estimated? An interesting aspect of this estimation problem is that not only the mean of y_K, but also its variance, depend on x_K.

If $\eta_K << \chi_K$ little difficulty would be expected in estimating χ_K, but in the case where $\eta_K \approx \chi_K$, the problem becomes difficult. There are two general ways to proceed. The first is to consider each y_K separately and develop some rule for estimating χ_K from it. An obvious rule, but certainly not the only possibility, would be to use the unbiased estimator $y_K - \eta_K$, which, of course, has the same variances as y_K, $(\eta_K^2 + 2\chi_K\eta_K)(1+\delta_{K,0})$. This approach would give good results where the signal-to-noise ratio (perhaps defined as χ_K/η_K) is large, but might not be satisfactory where it is small. The other approach would be to try to improve the statistics having high coefficients of variability by in some way combining adjacent statistics. Since the y_K are uncorrelated, rules for combining that would lead to a reduced coefficient of variability in the new statistic could be specified. Of course, the new statistic would be a measure of the energy spectrum in a broader region of frequencies, and it probably would not be simply the total energy in this region of frequency.

If it is desired to combine adjacent statistics, there are again two alternatives. The first, and simplest, is simply to add them. If they had comparable means and variances, this would result in a new statistic with a smaller coefficient of variation (by a factor of $1/\sqrt{2}$ if pairs are combined). In general, however, it would be better <u>statistically</u> to take a weighted sum, where the weighting would depend on both the mean and variance of each statistic.* The disadvantages of this

* For example if it were desired to choose α and β in such a way as to minimize the coefficient of variation of $z = \alpha x + \beta y$, where x and y are uncorrelated random variables, a suitable choice would be

$$\alpha = (1/\bar{x})[u_y/(u_x+u_y)],$$

and similarly for β, where

$$u_x = \mathrm{var}x/\bar{x}^2$$

This results in $u_z = (u_x u_y)/(u_x+u_y)$

weighting are twofold. To choose the weights it is necessary to know, or make some assumptions about, both x_K and n_K, and the statistics that result from a weighted sum would correspond roughly to some unequal weighting of different frequencies in the EDS.

For the present it seems reasonable to defer development of rules for combining adjacent statistics until some further information concerning the detailed statistics of these variables is obtained. Once this is available, it may turn out that some compromise between simple summation and an optimally weighted sum seems appropriate. It may also be the case that signal-to-noise ratios for some of the y_K are typically good enough that no combining seems necessary, while for others some combining always seems necessary.

F. Improved Method Based on Subtracting an Estimate of the Noise Waveform

Because the "signal" is probably confined to a relatively short time interval (e.g., 2 or 4 seconds) and because the noise appears to be correlated over times comparable to this, it should be possible to reduce the variability of the spectral estimate by subtracting from the observed signal an estimate of the noise that is based on the noise preceding the observation interval. Mathematically, this turns out to be very similar to the standard minimum-mean-square-error filtering problem (see, e.g., Bode and Shannon, [13]), but there are some important differences. Specifically, the optimum predictor for this application will not be time invariant, as it is in the standard smoothing or prediction problem.

For this discussion, it is convenient to assume that the observation interval is (0,T). The problem is to devise an estimate $\hat{n}(t)$, which is based on n(t) for t<0. Before an optimum estimate can be formulated it is necessary to specify

some criteria. One possibility would be to choose n(t) in such a ways as to minimize

$$E[(n(t)-\hat{n}(t))^2]$$

for each t. A more appropriate criterion might be to minimize $\text{var}[|Z(f)|^2]$ where z(t) is now defined by

$$z(t) = [x(t)+n(t)-\hat{n}(t)]\ a(t) \tag{55}$$

Unfortunately this latter criterion is not only difficult to apply analytically, but it cannot be applied without knowledge of $|Z_x(f)|^2$. A related, but simpler criterion would be to minimize $\overline{|Z_{n-\hat{n}}(f)|^2}$, the expected value of the noise contribution to the spectral estimate. In the stationary case (without $\hat{n}(t)$) the square of this term was one contribution to the variance (Equation 14). The same result holds here as well (see Appendix A and details below). This criterion will be discussed after considering the simpler one of minimizing $E[(n(t)-\hat{n}(t))^2]$. It will turn out that both criteria lead to the same estimate.

1. Choosing $\hat{n}(t)$ to Minimize $\overline{(n(t)-\hat{n}(t))^2}$

In addition to assuming that n(t) is a zero-mean, stationary Gaussian process, it is now assumed that $S_n(f)$ is such that a realizable whitener, with a realizable inverse, exists. [For a discussion of the restrictions on $S_n(f)$ implied by this assumption, see Davenport and Root, 1959, Chapter 11]. In equation form, it is assumed that $h^{-1}(t)$ and h(t) exist, where both are impulse responses of realizable systems (i.e., $h^{-1}(t) = h(t) = 0$ for $t<0$), and

$$h^{-1}(t) \circledast h(t) = u_0(t) \tag{56}$$

$$h(t) \circledast h(-t) = R_n(t) \tag{57}$$

or, equivalently,

$$|H(f)|^2 = S_n(f) \tag{58}$$

Continuing with these assumptions, it is useful to consider the following hypothetical system.

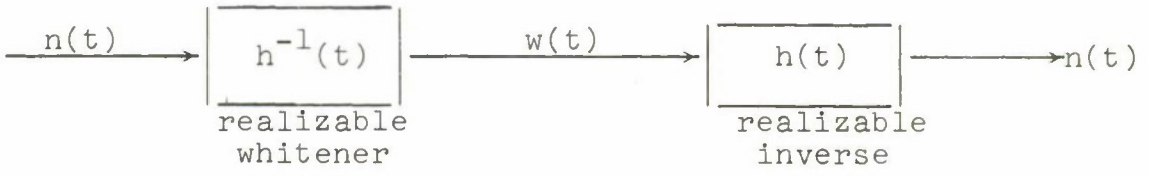


FIGURE 6

The output of this hypothetical system is n(t) because of Equation (56). w(t) is white noise, with

$$\overline{w(t)w(t+\tau)} = u_0(\tau) \tag{59}$$

Since n(t) can be recovered immediately from w(t) with the realizable filter h(t), nothing is lost by basing the estimate, $\hat{n}(t)$, on w(t) for t<0 rather than on n(t) for t<0, and some mathematical convenience is gained.

By inspection of Figure 6

$$w(\alpha) = \int_{-\infty}^{(\alpha)} n(\beta)h^{-1}(\alpha-\beta)\,d\beta \tag{60}$$

$$n(t) = \int_{-\infty}^{(t)} w(\alpha)h(t-\alpha)\,d\alpha \tag{61}$$

where the limits in parentheses are redundant due to the realizability of h(t) and $h^{-1}(t)$.

Rewriting Equation (61)

$$n(t) = \int_{-\infty}^{0} w(\alpha)h(t-\alpha)d\alpha + \int_{0}^{(t)} w(\alpha)h(t-\alpha)d\alpha \tag{62}$$

Because w(t) is white, and the two integrals involve nonoverlapping sections of w(t), these two integrals define independent (zero mean, Gaussian) random processes. The first integral represents the contribution to n(t) of the white noise preceding t=0. The second integral is that of the white noise following t=0. To minimize $(n(t)-\hat{n}(t))^2$, the first integral should be chosen as $\hat{n}(t)$.

$$\hat{n}(t) = \int_{-\infty}^{0} w(\alpha)h(t-\alpha)d\alpha$$

$$= \int_{-\infty}^{0} d\alpha h(t-\alpha) \int_{-\infty}^{(\alpha)} d\beta h^{-1}(\alpha-\beta)n(\beta)$$

$$= \int_{-\infty}^{0} d\beta \int_{(\beta)}^{0} d\alpha h(t-\alpha)h^{-1}(\alpha-\beta)n(\beta) \qquad (63)$$

With this choice, the error, $n_e(t)$, will be

$$n_e(t) = n(t)-\hat{n}(t) = \int_{0}^{(t)} w(\alpha)\, h(t-\alpha)d\alpha \qquad (64)$$

and, as already observed, $n_e(t)$ and $\hat{n}(t)$ will be independent, zero-mean, nonstationary Gaussian processes.

The autocorrelation function of $n_e(t)$ will be needed in the next section. In Appendix A it is shown that

$$b(t,\tau) \triangleq \overline{n_e(t)n_e(\tau)} = \int_{(0)}^{\min(t,\tau)} h(v)h(v+|t-\tau|)dv \qquad (65)$$

whereas

$$R_n(t-\tau) = \overline{n(t)n(t+\tau)} = \int_{(0)}^{\infty} h(v)h(v+|t-\tau|)dv \qquad (66)$$

2. Choosing $\hat{n}(t)$ to minimize $E[|Z_n(f)|^2]$

As indicated earlier, another criterion of some interest would be choosing $\hat{n}(t)$ to minimize $E[|Z_n(f)|^2]$. Because $n(t)$ is assumed Gaussian, the optimum (in the sense of minimizing $|Z_n(f)|^2$) predictor may be written as a linear operation on $n(t)$, $t<0$.

$$\hat{n}(t) = \int_{-\infty}^{0} g(t,\alpha)n(\alpha)\, d\alpha \qquad (67)$$

The problem is to choose $g(t,\alpha)$ to minimize $\overline{|Z_n(f)|^2}$. It is not at first obvious whether or not this can be done independently for all f, so for the present, f is regarded as a single fixed number. The details of the derivation appear in Appendix A, and only an outline of the important steps appears below.

Since,

$$n_e(t) = n(t) - \int_{-\infty}^{0} g(t,\alpha)n(\alpha)\, d\alpha \tag{68}$$

$$\overline{|Z_{n_e}(f)|^2} = \int\int dt\, d\tau\; a(t)a(\tau)\; e^{-j2\pi f(t-\tau)}\overline{n_e(t)n_e(\tau)} \tag{69}$$

is a functional of $g(t,\alpha)$. Considering

$$g(t,\alpha) = g_0(t,\alpha) + \varepsilon g_\varepsilon(t,\alpha) \tag{70}$$

it is desired to choose $g_0(t,\alpha)$ in such a way that $\overline{|Z_{n_e}(f)|^2}$ will be a minimum at $\varepsilon=0$, for any choice of $g_\varepsilon(t,\alpha)$. Choosing $g_0(t,\alpha)$ to satisfy

$$\int_{-\infty}^{0} d\alpha g_0(t,\alpha)R_n(\alpha-\beta) = R_n(t-\beta) \quad \text{for } \beta<0 \tag{71}$$

does insure that

$$\frac{\partial}{\partial\varepsilon}\,\overline{|Z_{n_e}(f)|^2}\Big|_{\varepsilon=0} = 0 \tag{72}$$

and

$$\frac{\partial^2}{\partial\varepsilon^2}\,\overline{|Z_{n_e}(f)|^2}\Big|_{\varepsilon=0} > 0 \tag{73}$$

The integral equation, (71), is solved by

$$g_0(t,\alpha) = \int_{(\alpha)}^{0} d\beta h^{-1}(\beta-\alpha)\; h(t-\beta) \tag{74}$$

Inspection of Equation (63) indicates that $g_0(t,\alpha)$, as specified by Equation (74), leads to the same $\hat{n}(t)$ as discussed in the previous section.

With this choice of $g_0(t,\alpha)$ and hence of $\hat{n}(t)$ it is possible to write expressions for $|Z_{n_e}(f)|^2$ and $\mathrm{var}[|Z_{n_e}(f)|^2]$, but unfortunately, since $n_e(t)$ is a nonstationary process, these equations are not easy to simplify.

$$\overline{|Z_n(f)|^2} = F(f,-f) \tag{75}$$

$$\mathrm{var}[|Z_n(f)|^2] = |F(f,-f)|^2 + |F(f,+f)|^2 \qquad (76)$$

where

$$F(f_1,f_2) = \iint dt\,d\tau\, a(t)a(\tau) \int_0^{\min(t,\tau)} dv\, h(v) \cdot \cdot h(v+|t-\tau|)e^{-j2\pi(f_1 t+f_2\tau)} \qquad (77)$$

This can be simplified somewhat to yield

$$F(f,-f) = 2\int_{(0)}^{\infty} d\alpha \int_0^{\infty} d\beta \int_{(0)}^{\alpha} dv\, a(\alpha)a(\alpha+\beta)h(v)h(v+\beta)\cos 2\pi f\beta \qquad (78)$$

$$F(f,f) = 2\int_{(0)}^{\infty} d\alpha \int_0^{\infty} d\beta \int_{(0)}^{\alpha} dv\, a(\alpha)a(\alpha+\beta)h(v)h(v+\beta)e^{-j2\pi f(\beta+2\alpha)} \qquad (79)$$

In both cases, the corresponding expressions for the stationary case, ($|Z_n(f)|^2$ without using $\hat{n}(t)$) may be obtained by changing the upper limit on the v integration from α to $+\infty$. In the stationary case these expressions can be simplified in terms of the Fourier Transforms of a(t) and h(t). Unfortunately, comparable simplification does not appear possible in the nonstationary case considered here, unless specific assumptions about a(t) and h(t) are made, and thus far even this has not helped much.

In Appendix A expressions are given for F(f,-f), assuming a(t) and h(t) are rectangular pulses of arbitrary length. In the case where they have the same length, these expressions reduce to

$$\overline{|Z_{n_e}(f)|^2} = 1/2\,\overline{|Z_n(f)|^2} \qquad (80)$$

In the stationary case, discussed earlier, it is usually possible to neglect $|F(f,+f)|^2$, the second term in the expression for $\mathrm{var}[|Z_n(f)|^2]$. It is not yet clear whether or not this is also true of the nonstationary case. Even for the special case of rectangular a(t) and h(t), we have as yet been unable to obtain a simple expression for $|F(f,+f)|^2$.

Physically, the reduction in the variance of the spectral estimate would be expected to depend on the ratio of the durations of a(t) and h(t), with the variance being quite small (i.e., $\hat{n}(t)$, a very good estimate) when a(t) is short compared to h(t) and relatively large when the reverse is true. (If h(t) = 0 for t>B, $\hat{n}(t)$ = 0 for t>B).

In summary, it appears possible to reduce the noise contribution to the spectral estimate by subtracting an estimate of the noise waveform that is based on the noise preceding the signal. The analytical predictions of the resulting variability in the estimate have yet to be worked out, and will probably require some simplifying assumptions and approximations. The details of implementing some approximation to this calculation on a digital computer also have to be worked out. In general, the output of the whitener depends on the infinite past, and this dependence would have to be truncated in any implementation. Actually the finite-discrete version of the problem is essentially the same as the linear mean square regression problem [see, for example, Cramér, 1954, pp. 302-305] and should be straightforward to program.

G. Some Comments on the Papers of McIvor and by Larrowe and Crabtree

Both of these papers attempt to apply to seismic waveforms a concept of a time-varying power spectrum. The basic concept can be described in the notation of this chapter by allowing the location of the time window, a(t), to vary. In this notation, then,

$$z_{t_0}(t) = a(t-t_0)y(t) \tag{81}$$

and

$$|Z_{t_0}(f)|^2 = |\int a(t-t_0)y(t)e^{-j2\pi ft}df|^2 \tag{82}$$

y(t) is used here for the seismic waveform to avoid specifying whether it is a transient (as x(t) was) or a continuing process (as n(t) was). $|Z_{t_0}(f)|^2$ would be labelled by words such as instantaneous power spectrum.

With this formulation, several questions immediately arise, most of which stem from the nontrivial question of what is the meaning of $|Z_{t_0}(f)|^2$? Is it an estimate of a power (or energy) density spectrum or is it an interesting functional of y(t) in its own right. Whereas the power density spectrum of a stationary random process and the energy density spectrum of a transient both have appealing "physical" interpretations,* it is not at all clear what interpretation is to be given to $|Z_{t_0}(f)|^2$, unless assumptions are made about how slowly it varies with t_0. In our opinion there is a burden on the user of a "time-varying spectrum" to demonstrate its physical relevance, its utility, or, preferably, both. we do not believe this is satisfactorily done by either of the papers considered.

McIvor completely neglects the problem of variability, which can, of course, be important. If, for example, y(t) is a stationary Gaussian process, then, no matter what a(t) is used

$$\mathrm{var}[|Z_{t_0}(f)|^2] \geq (E[|Z_{t_0}(f)|^2])^2 \qquad (83)$$

Furthermore, he is fundamentally in error when he asserts that Blackman and Tukey's lag window is just another time window and claims that their indirect method is simply a special case of his (essentially direct) method.

Larrowe and Crabtree briefly mention, and then ignore, the problem of variability by making assumptions that imply it does not exist. They divide the observation interval into

*The average power (or the total energy) of the output of a narrow band filter at frequency f_0 is proportional to the value of the power (or energy) density spectrum evaluated at f_0.

several subintervals, and, for each, they apply the equation (their Equation 36)

$$p(\omega) = \frac{1}{2\pi(t_n - t_{n-1})} \left| \int_{t_{n-1}}^{t_n} f(t)\, e^{-j\omega t}\, dt \right|^2 \qquad (84)$$

to estimate the PDS. In this equation, f(t) is the seismic waveform. They assume (page 16) "each [subinterval] long enough to permit a reasonably accurate computation of p(ω)" [using this equation]. As discussed in Section C of this chapter (and elsewhere), no matter how long the interval, if f(t) is a stationary Gaussian process,

$$\mathrm{var}[p(\omega)] \geq \overline{(p(\omega))}^2 \qquad (85)$$

Two pages later, in a derivation based on this assumption, they let $\Delta t = t_n - t_{n-1}$ go to zero!

It turns out that the result of this derivation is essentially correct and is of some interest. This result is that an equivalence exists between the time window and a suitably defined frequency window. This equivalence is also demonstrated by McIvor, but his derivation is a little difficult to follow because it involves several normalizations that are needed later in his paper. As this equivalence is fairly easy to demonstrate in the notation of this chapter, a brief derivation follows.

$$Z_{t_0}(f_0) = \int_{-\infty}^{\infty} a(t - t_0) y(t) e^{-j2\pi f_0 t}\, dt \qquad (86)$$

Considering a filter with impulse response

$$h_{f_0}(t) = a(-t) \cos 2\pi f_0 t \qquad (87)$$

and output w(t), yields

$$
\begin{aligned}
w(t) &= y(t) \otimes h_{f_0}(t) \\
&= \int y(\tau)a(\tau-t)\cos 2\pi f_0(t-\tau)\, d\tau \\
&= \mathrm{Re}\left[\int y(\tau)a(\tau-t)e^{j2\pi f_0 t}\, e^{-j2\pi f_0 \tau}\, d\tau\right] \\
&= \mathrm{Re}\; e^{j2\pi f_0 t} Z_t(f_0) \\
&= |Z_t(f_0)|\cos(2\pi f_0 t + \angle Z_t(f_0)) \qquad (88)
\end{aligned}
$$

Thus the envelope of the output of the filter at time t_0 is equal to $|Z_{t_0}(f_0)|$, and therefore using the time window $a(t-t_0)$ located at t_0, to estimate the PDS at f_0 is in a sense equivalent to using the frequency window $\frac{1}{2}[A^*(f-f_0)+A^*(f+f_0)]$, located at f_0, and examining the envelope at time t_0. Of course, in practical situations, calculating the envelope may be difficult, but in the case where $|Z_{t_0}(f_0)|$ varies only slowly with t_0, there should be no problem. In principle, another filter $h_2(t)$ with

$$h_2(t) = a(-t)\sin 2\pi f_0 t \qquad (89)$$

could be used, in which case the sum of the squares of the outputs of the two filters at time t_0 would be proportional to $|Z_{t_0}(f_0)|^2$.

SECTION IV

Automatic Test for pP in Shallow Earthquakes

The use of the reflected phase pP to determine the depth of shallow earthquakes is often rendered ineffective by the masking of pP by the coda of P. For earthquakes above the Moho, pP is usually buried in the first 13 seconds of coda following P onset where it is often intractably unrecognizable even by the skilled observer. Consequently, a test has been devised to extract pP from the overlying coda by computer.

The method is based on signal enhancement of pP through phased sums of a seismic array of continental dimensions (3000-7000 kilometers). This processing scheme is similar to one previously reported*, however, several important additions have been made to facilitate comparison of shallow quakes and surface events. (As previously given, the test was unable to distinguish between the two classes.)

Assume a set of stations i=1,...,n and epicentral distances Δ_i. Determine the angle of incidence i_i for each Δ_i (Richter, 1958, pp. 664-6). Assume a set of test depths, h_j in the range 10 to 41 kilometers. (Increments of 1.6 km were used). Calculate a set of time delays τ_{ij} of P-pP by:

$$\tau_{ij} = (2h_i/c) \cos i_i$$

For each test depth h_j align the seismograms for proper relative displacement according to τ_{ij}. Form a weighted sum of the seismograms, by multiplying each by a factor

---

*GAC Report 1456-2026-9, January 1966.

$$\frac{S_i}{N_i^2} = \frac{\left\{(1/T_1)\int^{T_1} S_i^2(t)dt\right\}^{\frac{1}{2}}}{(1/T_2)\int^{T_2} n_i^2(t)dt}$$

where S_i and N_i are the mean energies in the P and noise regions respectively. For this computation a P interval of one second was chosen. This interval was centered about the first definite zero crossing of P. The noise interval chosen was from 13 to 3 seconds before P onset. The weighting factor is derived in Appendix B, using the constraint that the signal-to-noise ratio of the phase-sum seismogram be a maximum with respect to all other possible weighting factors.

After alignment and weighting, sum seismograms $s_j(t)$ are formed for each test depth h_j. A window of duration τ is placed about the location $\tau_j=(2h_i/c)$ in each $s_j(t)$, and the average energy in that window measured:

$$S_j = \frac{1}{T} \int_{\tau_j-T/2}^{\tau_j+T/2} S_j^2(t)dt$$

The corresponding average energy C_j ofthe P-coda of $S_j(t)$ is measured outside the window about τ_j, and a test statistic formed:

$$\epsilon_j = (S_j/C_j)$$

A plot of this statistic v.s. h_j should show peaks at certain values of h_j. One reason, in the case of a shallow earthquake, is the presence of pP in the τ_j interval. A second reason for peaks observed with surface and deep events is the purely random occurrence of more energy in one interval than in the others.

The ambiguity is easily removed by formation of a new test statistic, η_j, which is the product of ϵ_j and the average cross correlation of the n seismograms on the interval of duration τ in which ϵ_j is a maximum:

$$\eta_j = \epsilon_j \bar{\rho}_{peak}$$

When η_j is plotted, the shallow earthquakes are emphasized more than the events with no pP in the coda interval. Typically, $\bar{\rho}_{peak}$ for shallow earthquakes is .5, whereas for surface and deep events it is close to zero. (See Table 1).

The test has been applied to twelve earthquakes and six surface events. Plots of η_j for earthquakes are shown in Figures 7 and 8. Plots for the surface events are not shown, as η_j for all six was found to be negligible. Maximum values of η_j for surface events and earthquakes are given in Table 1, along with the h_j's at which they occurred. For surface events, the mean η_{max} was .11 while for shallow earthquakes it was 5.1. Values of depth, h_j, agree very well with the nominal depths in nine of the ten cases, standard deviation being 2.8 kilometers. The earthquake of 15 July 1963, while nominally at 60 km depth, registered a distinct peak at 19 km on the 7 diagram. (This anomaly will be investigated more closely.) The 25 August 1963 earthquake is peculiar in that it's η-h diagram has three distinct peaks, but these may have been caused by water-air and earth-water interfacial reflections under the Sea of Okhotsk. This problem has not been resolved conclusively yet.

Further research will be reported on this topic as more events are tested. Also, the use of $\bar{\rho}_{peak}$ as a possible discriminant between shallow earthquakes and shots bears further statistical investigation.

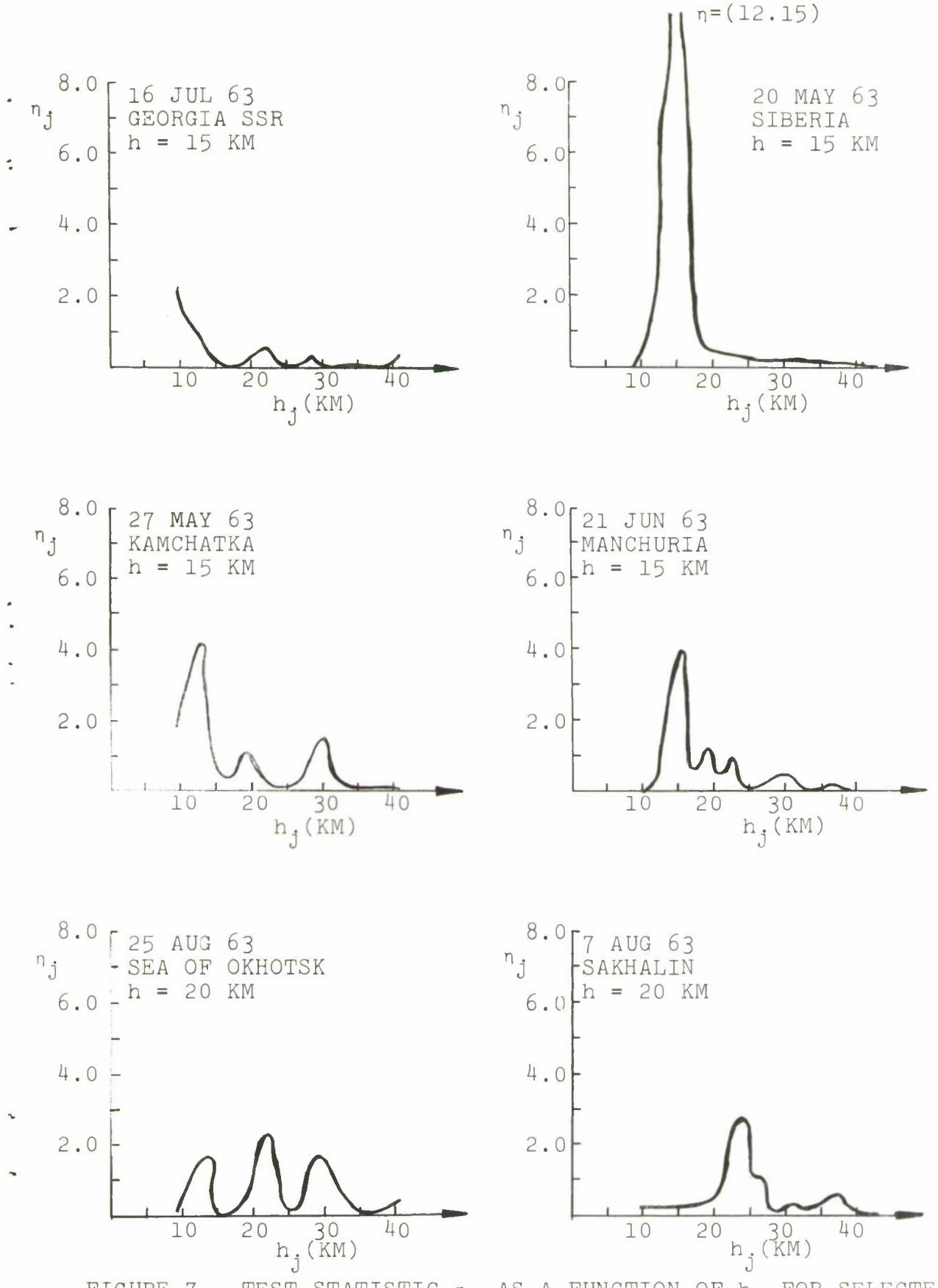


FIGURE 7 - TEST STATISTIC η_j AS A FUNCTION OF h_j FOR SELECTED EARTHQUAKES

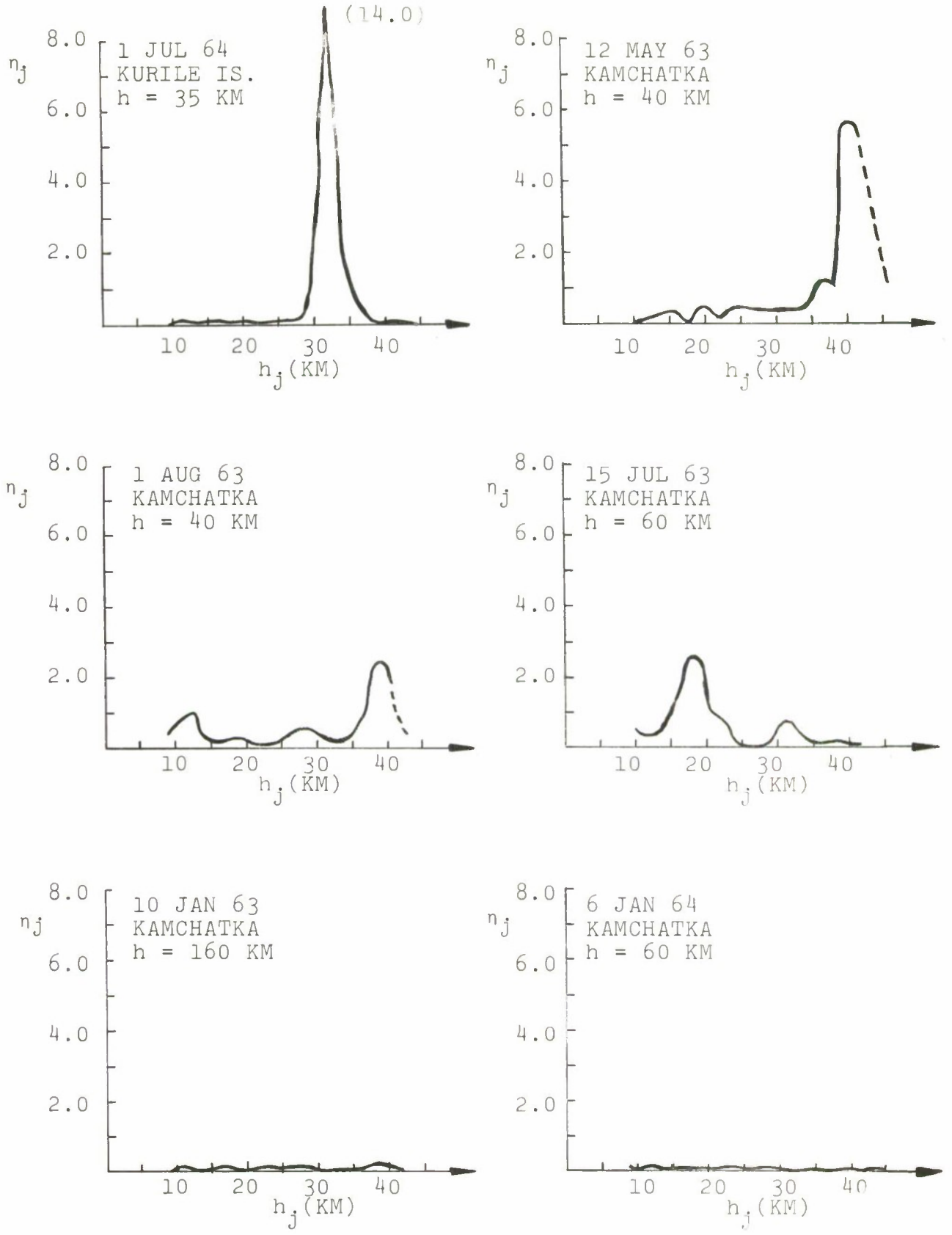


FIGURE 8 - TEST STATISTIC η_j AS A FUNCTION OF h_j FOR SELECTED EARTHQUAKES

Event	Mag.	Depth (PDE)	Depth (ours) h_j	$\epsilon_{Max.}$	$\bar{\rho}_{peak}$	$\epsilon_{Max}\ \bar{\rho}_{peak} = \eta_{Max}$	
7/16/63	6.0	15	10	5.79	.39	2.26	SHALLOW EARTHQUAKES
5/20/63	4.9	15	13	16.66	.73	12.15	
5/27/63	5.7	15	13	5.47	.70	3.83	
6/21/63	4.8	15	15	7.21	.56	4.04	
8/25/63	4.7	20	22	4.48	.52	2.31	
8/7/63	5.1	20	24	5.09	.54	2.75	
7/1/64	5.2	35	32	25.00	.56	14.0	
5/12/63	5.0	40	40	6.85	.82	5.61	
8/1/63	5.1	40	40	3.69	.64	2.36	
7/15/63	4.9	60	19	3.96	.54	2.56	
1/10/63	-	160	-	1.05	.08	.084	DEEP EARTHQUAKES
1/6/64	6.0	60	-	2.93	.03	.088	
#2		-		2.83	.019	.054	SHOTS
4		-		2.48	.14	.35	
5		-		1.81	.046	.083	
7		-		4.86	.086	.427	
10		-		2.77	-.050	-.14	
27		-		5.81	-.014	-.081	

TABLE 1

REFERENCES

1. E. A. Flinn, R. A. Hartenberger, D. W. McCowan, "Two Examples of Maximum - Likelihood Filtering of LASA Seismograms," Report from Earth Sciences Division, Teledyne Industries. June 8, 1966.

2. Semiannual Technical Summary Report to the Advanced Projects Agency on Seismic Discrimination, Lincoln Laboratory, M.I.T. December 31, 1964.

3. J. Capon, R. J. Greenfield, R. T. Lacoss, "Off-Line Signal Processing Results for the Large Aperture Seismic Array," M.I.T., Lincoln Laboratory Technical Note 1966-37. July 11, 1966.

4. J. Capon, et al, "A Frequency-Domain Synthesis Procedure for Multidimensional Maximum-Likelihood Processing of Seismic Arrays," M.I.T., Lincoln Laboratory Technical Note 1966-29, May 6, 1966.

5. D. G. Brennan, "Linear Diversity Combining Techniques," Proc. IEEE. June, 1959.

6. B. D. Steinberg, "Spatial Correlations Observed across LASA from Longshot," General Atronics Technical Note 1512-2033-2 prepared for the Advanced Research Projects Agency. May 18, 1966.

7. P. E. Green, R. A. Frosch, C. F. Romney, "Principles of an Experimental Large Aperture Seismic Array," Proc. IEEE. December, 1965.

8. I. K. McIvor, 1964, "Methods of Spectral Analysis of Seismic Data," Bull. of Seis. Soc. of Am., 54, 1213-1232.

REFERENCES CONT.

9. V. L. Larrowe and R. E. Crabtree, 1963, "Analog Computation of Time-varying Power Spectra of Seismic Waves," Report No. 3798-15-T/5178-8-T, Institute of Science and Technology, University of Michigan.

10. R. B. Blackman and J. W. Tukey, 1959, The Measurement of Power Spectra, Dover Publications, Inc., New York.

11. W. B. Davenport, Jr. and W. L. Root, 1958, Random Signals and Noise, McGraw Hill Book Co., Inc., New York.

12. E. Parzen, 1961, "Mathematical Consideration in the Estimation of Spectra," TECHNOMETRICS, 3, 167-190.

13. H. W. Bode and C. E. Shannon, 1950, A Simplified Derivation of Linear Least-square Smoothing and Prediction Theory, PROC. IRE, 38, 417-426.

14. H. Cramer Mathematical Methods of Statistics, Princeton University Press, Princeton, 1954.

APPENDIX A

1. General Nonstationary Case

Definitions and assumptions:

$n(t)$: a zero mean, not necessarily stationary, Gaussian process, with $\overline{n(t)n(\tau)} = R_n(t,\tau)$

$x(t)$: a pulse "signal" for which the Fourier Transform, $X(f) = \int x(t)e^{-j2\pi ft}dt$, exists

$a(t)$: the "time window"; it is assumed that $A(f)$ exists

$Z(t) = a(t)[x(t)+n(t)]$

With these definitions, the following equations may be written by inspection.

$$Z(f) = A(f) \otimes X(f) + \int a(t)n(t)e^{-j2\pi ft}\,dt \qquad (A\text{-}1)$$

$$= Z_x(f) + Z_n(f) \qquad (A\text{-}2)$$

$$E[|Z(f)|^2] = \iint dt\,d\tau\, a(t)a(\tau)\overline{[x(t)+n(t)][x(\tau)+n(\tau)]}e^{-j2\pi f(t-\tau)}$$

$$= |A(f)\otimes X(f)|^2 + \iint dt\,d\tau\, a(t)a(\tau)R_n(t,\tau)e^{-j2\pi f(t-\tau)}$$

$$= |Z_x(f)|^2 + \overline{|Z_n(f)|^2} \qquad (A\text{-}3)$$

The expected value of $|Z(f)|^2$ is simply the sum of the signal and noise contributions. It should be noted for future reference that the integral defining $|Z_n(f)|^2$ must be real and nonnegative for any legitimate $R_n(t,\tau)$.

$$\mathrm{cov}[|Z(f_1)|^2,|Z(f_2)|^2] = E[|Z(f_1)|^2|Z(f_2)|^2] - E[|Z(f_1)|^2]E[|Z(f_2)|^2] \qquad (A\text{-}4)$$

$$E[|Z(f_1)|^2|Z(f_2)|^2] = \iiint\!\!\int dt\,d\tau\,d\alpha\,d\beta\, a(t)a(\tau)a(\alpha)a(\beta)\cdot e^{-j2\pi f_1(t-\tau)}e^{-j2\pi f_2(\alpha-\beta)}\cdot \overline{[x(t)+n(t)][x(\tau)+n(\tau)][x(\alpha)+n(\alpha)][x(\beta)+n(\beta)]} \qquad (A\text{-}5)$$

When the integrand is multiplied out and expected values are taken, only terms with 0, 2 or 4 factors of $n(\cdot)$ will be non-zero. To simplify the expression involving all four of the $n(\cdot)$, the following property of zero mean Gauss processes is employed.

$$\overline{n(t)n(\tau)n(\alpha)n(\beta)} = R_n(t,\tau)R_n(\alpha,\beta)+R_n(t,\alpha)R_n(\tau,\beta) + R_n(t,\beta)R_n(\tau,\alpha) \qquad \text{(A-6)}$$

Define

$$F(f_1,f_2) = \iint dt\,d\alpha\, a(t)a(\alpha)R_n(t,\alpha)e^{-j2\pi(f_1t+f_2\alpha)} \qquad \text{(A-7)}$$

and note that $F(f_1,f_2) = F(f_2,f_1) = F^*(-f_1,-f_2)$ and that $|Z_n(f)|^2 = F(f,-f)$. With this notation, the various contributions to the integral of Equation (A-5) may be tabulated as follows.

Term in Integrand	Contribution to Integral
1) $\overline{n(t)n(\tau)n(\alpha)n(\beta)}$	$F(f_1,-f_1)F(f_2,-f_2)+\|F(f_1,f_2)\|^2+\|F(f_1,-f_2)\|^2$
2) $x(t)x(\tau)x(\alpha)x(\beta)$	$\|Z_x(f_1)\|^2\|Z_x(f_2)\|^2$
3) $x(t)x(\tau)\overline{n(\alpha)n(\beta)}$	$\|Z_x(f_1)\|^2F(f_2,-f_2)$
4) $\overline{n(t)n(\tau)}x(\alpha)x(\beta)$	$\|Z_x(f_2)\|^2F(f_1,-f_1)$
5) $x(t)x(\alpha)\overline{n(\tau)n(\beta)}$	$Z_x(f_1)Z_x(f_2)F(-f_1,-f_2)$
6) $x(t)x(\beta)\overline{n(\tau)n(\alpha)}$	$Z_x(f_1)Z_x^*(f_2)F(-f_1,f_2)$
7) $x(\tau)x(\alpha)\overline{n(t)n(\beta)}$	$Z_x^*(f_1)Z_x(f_2)F(f_1,-f_2)$
8) $x(\tau)x(\beta)\overline{n(t)n(\alpha)}$	$Z_x^*(f_1)Z_x^*(f_2)F(f_1,f_2)$

Since

$$\overline{|Z(f_1)|^2} = |Z_x(f_1)|^2 + F(f_1,-f_1), \qquad \text{(A-8)}$$

$E[|Z(f_1)|^2]E[|Z(f_2)|^2]$ will have four terms, which are identical to the terms appearing in lines 2-4 and the first entry in line 1 of the above table. Thus the covariance (Equation (A-4)) is given by the second and third entries of line 1 plus lines 5-8.

$$\text{cov}[|Z(f_1)|^2,|Z(f_2)|^2] = |F(f_1,f_2)|^2+|F(f_1,-f_2)|^2$$

$$+ 2\text{Re}[Z_x(f_1)[Z_x(f_2)F(-f_1,-f_2)+Z_x^*(f_2)F(-f_1,f_2)] \quad \text{(A-9)}$$

This specializes to

$$\text{var}[|Z(f)|^2] = |F(f,f)|^2 + |F(f,-f)|^2$$

$$+ 2\text{Re}Z_x^2(f)F(-f,-f)+2|Z_x(f)|^2F(f,-f) \quad \text{(A-10)}$$

where the fact that $F(f,-f)$ is real has been used in dropping the real part operator from the last term.

2. Stationary Case, $R_n(t,\tau) = R_n(t-\tau)$

In the case where

$$R_n(t,\tau) = R_n(t-\tau) = \int S_n(f)e^{j2\pi f(t-\tau)}df \quad \text{(A-11)}$$

the expressions of the previous section may be simplified. Specifically,

$$F(f_1,f_2) = \int\int dtd\alpha a(t)a(\alpha)R_n(t-\alpha)e^{-j2\pi f_1 t}e^{-j2\pi f_2 \alpha}$$

$$= \int dta(t)e^{-j2\pi f_1 t}[a(t)e^{-j2\pi f_2 t} \circledast R_n(t)]$$

The term in the brackets has a Fourier Transform

$$A(f+f_2)S_n(f)$$

Thus the total integral is

$$F(f_1,f_2) = A(f) \circledast A(f+f_2)S_n(f)\Big|_{f=f_1}$$

$$= \int duA(f_1-u)A(u+f_2)S_n(u) = \int duS_n(u)A^*(u-f_1)A(u+f_2) \quad \text{(A-12)}$$

This gives

$$\overline{|Z(f)|^2} = |Z_x(f)|^2 + \int duS_n(u)|A(u-f)|^2 \quad \text{(A-13)}$$

$$\text{var}[|Z(f)|^2] = |\int du S_n(u)|A(u-f)|^2|^2 + |\int du S_n(u)A^*(u-f)A(u+f)|^2 + 2|Z_x(f)|^2 \int du S_n(u)|A(u-f)|^2 + 2\text{Re}Z_x^2(f)\int du S_n(u)A^*(u+f)A(u-f) \quad \text{(A-14)}$$

$\text{Cov}[|Z(f_1)|^2, |Z(f_2)|^2]$ may be obtained by substituting (A-12) into (A-9).

A special case of some interest occurs when

$$A(f) = 0 \quad \text{for } |f| > W \quad \text{(A-15)}$$

If (A-15) holds, then for any f_1, f_2 such that

$$||f_1|-|f_2|| > W$$

$F(f_1, \pm f_2)$ will be zero and thus

$$\text{cov}[|Z(f_1)|^2, |Z(f_2)|^2] = 0 \quad \text{(A-16)}$$

Furthermore, if (A-15) holds

$$A(u+f)A^*(u-f) = 0 \quad \text{for } |f| > W \quad \text{(A-17)}$$

and, therefore, for $|f| > W$

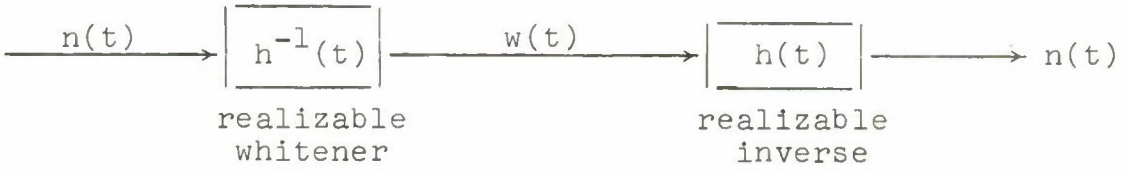
$$\text{var}[|Z(f)|^2] = |\int du S_n(u)|A(u-f)|^2|^2 + 2|Z_x(f)|^2 \int du S_n(u)|A(u-f)|^2 = (\overline{|Z_n(f)|^2})^2 + 2|Z_x(f)|^2 \overline{|Z_n(f)|^2} \quad \text{(A-18)}$$

3. Nonstationary Case Using $n_e(t) = n(t)-\hat{n}(t)$ as Derived in Section F

In Section F, an estimate of $n(t)$, $\hat{n}(t)$, that is optimal in the sense of minimizing $n_e^2(t)$, where

$$n_e(t) = n(t)-\hat{n}(t) \quad \text{(A-19)}$$

and $\hat{n}(t)$ is based only on values of $n(t)$ for $t<0$, was derived by considering the hypothetical system shown below.



This estimate was given by

$$\hat{n}(t) = \int_{-\infty}^{0} w(\alpha)h(t-\alpha)d\alpha \qquad \text{(A-20)}$$

which gives

$$n_e(t) = \int_{0}^{(t)} w(\alpha)h(t-\alpha)d\alpha \qquad \text{(A-21)}$$

where the upper limit in parentheses is redundant since h(t) is the impulse response of a realizable system.

For convenience, define

$$b(t,\tau) = \overline{n_e(t)n_e(\tau)} = \int_0^{(t)} d\alpha \int_0^{(\tau)} d\beta \overline{w(\alpha)w(\beta)}h(t-\alpha)h(t-\beta) \qquad \text{(A-22)}$$

Since

$$\overline{w(\alpha)w(\beta)} = u_0(\alpha-\beta) \qquad \text{(A-23)}$$

Equation (A-22) may be simplified to

$$b(t,\tau) = \begin{cases} \int_0^{(\min(t,\tau))} d\alpha h(t-\alpha)h(\tau-\alpha) & \min(t,\tau)>0 \\ 0 & \min(t,\tau)<0 \end{cases} \qquad \text{(A-24)}$$

Letting

$$v = \min(t,\tau) - \alpha \qquad \text{(A-25)}$$

this may be rewritten as

$$b(t,\tau) = \begin{cases} \int_0^{\min(t,\tau)} dv h(v)h(v+|t-\tau|) & \min(t,\tau)>0 \\ 0 & \min(t,\tau)<0 \end{cases} \qquad \text{(A-26)}$$

To contrast this with $R_n(t-\tau)$, note that

$$R_n(\tau) = h(\tau) \circledast h(-\tau) \qquad \text{(A-27)}$$

and therefore

$$R_n(t-\tau) = \int_0^\infty dv h(v)h(v+|t-\tau|) \qquad \text{(A-28)}$$

Now defining

$$z(t) = a(t)[x(t)+n_e(t)]$$

the moments of $|Z(f)|^2$ may be obtained by substituting $b(t,\tau)$ for $R_n(t,\tau)$ in the expressions of Section 1 of this appendix. This leads to

$$F(f_1,f_2) = \iint dt d\tau a(t)a(\tau)\int^{\min(t,\tau)} dv h(v)h(v+|t-\tau|)\cdot \cdot e^{-j2\pi(f_1 t+f_2\tau)} \qquad \text{(A-29)}$$

Separating this integral into the two regions, $t<\tau$ and $t>\tau$ and substituting

$$\alpha = \min(t,\tau), \qquad \beta = |t-\tau| \qquad \text{(A-30)}$$

leads to

$$F(f_1,f_2) = \int_{-\infty}^\infty dt\int_{-\infty}^t d\tau a(\alpha)a(\alpha+\beta)\int_0^\alpha dv h(v)h(v+\beta)e^{-j2\pi[f_1(\alpha+\beta)+f_2\alpha]} + \int_{-\infty}^\infty dt\int_t^\infty d\tau a(\alpha)a(\alpha+\beta)\int_0^\alpha dv h(v)h(v+\beta)e^{-j2\pi[f_1\alpha+f_2(\alpha+\beta)]} \qquad \text{(A-31)}$$

but

$$\int_{-\infty}^\infty dt\int_{-\infty}^t d\tau = \int_{-\infty}^\infty d\tau\int_\tau^\infty dt = \int_{-\infty}^\infty d\alpha\int_0^\infty d\beta \qquad \text{(A-32)}$$

and

$$\int_{-\infty}^\infty dt\int_t^\infty d\tau = \int_{-\infty}^\infty d\alpha\int_0^\infty d\beta \qquad \text{(A-33)}$$

Therefore,

$$F(f_1,f_2) = \int_\infty^\infty d\alpha\int_0^\infty d\beta\int_0^\alpha dv a(\alpha)a(\alpha+\beta)h(v)h(v+\beta)e^{-j2\pi(f_1+f_2)\alpha}\cdot \cdot (e^{-j2\pi f_1\beta} + e^{-j2\pi f_2\beta}) \qquad \text{(A-34)}$$

and, in particular,

$$F(f,-f) = 2\int_{-\infty}^\infty d\alpha\int_0^\infty d\beta\int_0^\alpha dv a(\alpha)a(\alpha+\beta)h(v)h(v+\beta)\cos 2\pi f\beta \qquad \text{(A-35)}$$

$$F(f,+f) = 2\int_{-\infty}^{\infty} d\alpha \int_{0}^{\infty} d\beta \int_{0}^{\alpha} dv\, a(\alpha)a(\alpha+\beta)h(v)h(v+\beta)e^{-j2\pi f(2\alpha+\beta)} \qquad (A\text{-}36)$$

To convert Equations (A-34) to (A-36) to the stationary case (for which the autocorrelation function is given by Equation (A-28) rather than (A-26)) it is only necessary to change the upper limit on the integration from α to $+\infty$. With this change, Equations (A-34)-(A-36) can be simplified as has already been demonstrated by Equation (A-12). Without this change, it does not appear possible to simplify these equations further without making specific assumptions about a(t) and h(t). One set of assumptions, which leads to some simplification, is considered in a later section of this appendix.

4. Choice of $\hat{n}(t)$ to Minimize $\overline{|Z_{n_e}(f)|^2}$

Considering now an estimate, $\hat{n}(t)$, of the form

$$\hat{n}(t) = \int_{-\infty}^{0} g(t,\alpha)n(\alpha)d\alpha \qquad (A\text{-}37)$$

the problem is to choose $g(t,\alpha)$ to minimize $\overline{|Z_{n_e}(f)|^2}$. [Note, by inspection of Equation (63), it can be seen that the choice of $\hat{n}(t)$ considered in the previous section corresponds to

$$g(t,\alpha) = \int_{(\alpha)}^{\infty} du\, h(t-u)h^{-1}(u-\alpha) \;] \qquad (A\text{-}38)$$

Substituting

$$n_e(t) = n(t) - \int_{\infty}^{0} g(t,\alpha)n(\alpha)d\alpha \qquad (A\text{-}39)$$

into the integral defining $|Z_{n_e}(f)|^2$ and taking expectations yields

$$\overline{|Z_{n_e}(f)|^2} = \iint dt\,d\tau\, a(t)a(\tau)e^{-j2\pi f(t-\tau)}\{R_n(t-\tau) - \int_{-\infty}^{0} d\beta\, g(\tau,\beta)\cdot$$

$$\cdot R_n(t-\beta) - \int_{-\infty}^{0} d\alpha\, g(t,\alpha)R_n(\tau-\alpha) + \iint_{-\infty}^{0} d\alpha\, d\beta\, g(t,\alpha)g(\tau,\beta)$$

$$\cdot R_n(\alpha-\beta)\} \qquad (A\text{-}40)$$

Letting

$$g(t,\alpha) = g_0(t,\alpha) + \varepsilon g_\varepsilon(t,\alpha) \qquad \text{(A-41)}$$

the problem is to choose $g_0(t,\alpha)$ in such a way that $\overline{|Z_{n_e}(f)|^2}$ is a minimum at $\varepsilon=0$ for any $g_\varepsilon(t,\alpha)$.

$$\frac{\partial}{\partial\varepsilon}\overline{|Z_{n_e}(f)|^2}\Big|_{\varepsilon=0} = \iint dt\,d\tau\, a(t)a(\tau)e^{-j2\pi f(t-\tau)}\{0-\int_{-\infty}^{0} d\beta\, g_\varepsilon(\tau,\beta)\cdot$$

$$\cdot R_n(t-\beta)-\int_{-\infty}^{0} d\alpha\, g_\varepsilon(t,\alpha)R_n(\tau-\alpha)+\iint_{-\infty}^{0} d\alpha\, d\beta[g_0(t,\alpha)g_\varepsilon(\tau,\beta) +$$

$$+g_0(\tau,\beta)g_\varepsilon(t,\alpha)]R_n(\alpha-\beta)\}$$

$$= \iint dt\,d\tau\, a(t)a(\tau)e^{-j2\pi f(t-\tau)}\{\int_{-\infty}^{0} d\beta\, g_\varepsilon(\tau,\beta)[\int_{-\infty}^{0} d\alpha\, g_0(t,\alpha)\cdot$$

$$\cdot R_n(\alpha-\beta)-R_n(t-\beta)] + \int_{-\infty}^{0} d\alpha\, g_\varepsilon(t,\alpha)[\int_{-\infty}^{0} d\beta\, g_0(\tau,\beta)R_n(\beta-\alpha) -$$

$$-R_n(\tau-\alpha)]\} \qquad \text{(A-42)}$$

A sufficient condition for Equation (A-42) to be zero, for any choice of $g_\varepsilon(t,\alpha)$ is

$$\int_{-\infty}^{0} d\alpha\, g_0(t,\alpha)R_n(\alpha-\beta) = R_n(t-\beta) \qquad \text{for } \beta<0 \qquad \text{(A-43)}$$

Guessing that the solution to the integral equation is given by Equation (A-38), which corresponds to the earlier derivation of $\hat{n}(t)$, yields

$$g_0(t,\alpha) = \int du\, h(t-u)h^{-1}(u-\alpha) \qquad \text{(A-44)}$$

$$\int_{-\infty}^{0} d\alpha\, g_0(t,\alpha)R_n(\alpha-\beta) = \int_{-\infty}^{0} d\alpha \int du\, h(t-u)h^{-1}(u-\alpha)R_n(\alpha-\beta)$$

$$= \int_{-\infty}^{0} du\, h(t-u)\int d\alpha\, h^{-1}(u-\alpha)R_n(\alpha-\beta)$$

$$= \int_{-\infty}^{0} du\, h(t-u)[h^{-1}(u) \circledast R_n(u-\beta)]$$

[equation continued]

$$= \int_{-\infty}^{0} du h(t-u)h[-(u-\beta)]$$

$$= \begin{cases} R_n(t-\beta) & \text{for } \beta<0 \\ R_n(t-\beta)-\int_0^{\min(t,\beta)} h(t-u)h(\beta-u)du & \text{for } \beta>0 \end{cases}$$

$$= R_n(t-\beta)-b(t,\beta) \qquad \text{(A-45)}$$

where, of course, $b(t,\beta)$ is zero for $\beta<0$. Thus $g_0(t,\alpha)$, as given by Equation (A-44) is a solution to Equation (A-43).

To show that this does indeed give a minimum of $\overline{|Z_{n_e}(f)|^2}$, it is necessary to show that the second derivative is positive at $\varepsilon = 0$.

$$\frac{\partial^2}{\partial \varepsilon^2}\overline{|Z_{n_e}(f)|^2} = \iint dt d\tau a(t)a(\tau)e^{-j2\pi f(t-\tau)}\int_{-\infty}^{0} d\alpha \int_{-\infty}^{0} d\beta g_0(t,\alpha)$$

$$\cdot g_0(\tau,\beta)R_n(\alpha-\beta)$$

$$= \iint dt d\tau a(t)a(\tau)e^{-j2\pi f(t-\tau)}\int_{-\infty}^{0} d\beta g_0(\tau,\beta)R_n(t-\beta)$$

$$= \iint dt d\tau a(t)a(\tau)e^{-j2\pi f(t-\tau)}[R_n(\tau-t)-b(\tau,t)] \qquad \text{(A-46)}$$

where Equation (A-43) is used for the first integration and Equation (A-45) for the second.

Since $\hat{n}(t)$ and $n_e(t)$ are independent, zero mean processes and

$$n(t) = \hat{n}(t) + n_e(t) \qquad \text{(A-47)}$$

it follows that

$$R_n(t-\tau) = \overline{\hat{n}(t)\hat{n}(\tau)} + b(t,\tau) \qquad \text{(A-48)}$$

and therefore the quantity in the brackets in Equation (A-46) is the autocorrelation function of $\hat{n}(t)$. Therefore Equation (A-46) is simply $\overline{|Z_{\hat{n}}(f)|^2}$, which cannot be negative.

5. Nonstationary Case with Rectangular a(t) and h(t)

In an effort to obtain some measure of the improvement attainable by using $\hat{n}(t)$, specific assumptions about a(t) and h(t) are now considered. Unfortunatley, even with these assumptions, only rather limited results have been obtained.

Specifically, it is assumed that

$$a(t) = \begin{cases} 1 & 0<t<T \\ 0 & \text{elsewhere} \end{cases} \tag{A-49}$$

$$h(t) = \begin{cases} 1 & 0<t<B \\ 0 & \text{elsewhere} \end{cases} \tag{A-50}$$

From this it is easy to show

$$A(f) = e^{-j2\pi f(T/2)} \frac{\sin \pi fT}{\pi f} \tag{A-51}$$

$$S_n(f) = |H(f)|^2 = \left(\frac{\sin \pi fB}{\pi f}\right)^2 \tag{A-52}$$

Pursuing these assumptions,

$$F(f,-f) = 2\int_0^\infty d\beta \underbrace{\int_{-\infty}^{\infty} d\alpha a(\alpha)a(\alpha+\beta)}_{T-\beta \text{ for } \beta<T} \underbrace{\int_{-\infty}^{\alpha} dv h(v)h(v+\beta)\cos 2\pi f\beta}_{\min(\alpha,B-\beta)}$$

If B>T

$$= 2\int_0^T d\beta \cos 2\pi f\beta \int_0^{T-\beta} d\alpha \cdot \alpha = \int_0^T d\beta \cos 2\pi f\beta (T-\beta)^2$$

$$= \frac{1}{2}\left(\frac{\sin \pi fT}{\pi f}\right)^2 \otimes \left(\frac{\sin \pi fT}{\pi f}\right)^2 \tag{A-53}$$

If B<T

$$F(f,-f) = 2\int_0^B d\beta \cos 2\pi f\beta \left[\int_0^{B-\beta} d\alpha \cdot \alpha + \int_{B-\beta}^{T-\beta} d\alpha (B-\beta)\right]$$

[equation continued]

$$= \int_0^B d\beta \cos 2\pi f\beta [(B-\beta)^2 + 2(B-\beta)(T-B)]$$

$$= \frac{1}{2}\left(\frac{\sin \pi fB}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fB}{\pi f}\right)^2 + (T-B)\left(\frac{\sin \pi fB}{\pi f}\right)^2 \qquad \text{(A-54)}$$

In summary

$$\overline{|Z_{n_e}(f)|^2} = F(f,-f) = \begin{cases} \frac{1}{2}\left(\frac{\sin \pi fT}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fT}{\pi f}\right)^2 & B>T \\ \frac{1}{2}\left(\frac{\sin \pi fB}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fB}{\pi f}\right)^2 + (T-B)\left(\frac{\sin \pi fB}{\pi f}\right)^2 & B<T \end{cases} \qquad \text{(A-55)}$$

These should be compared to the corresponding expression for the same a(t) and h(t), but without $\hat{n}(t)$. This is

$$\overline{|Z_n(f)|^2} = S_n(f) \circledast |A(f)|^2 = \left(\frac{\sin \pi fT}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fB}{\pi f}\right)^2 \qquad \text{(A-56)}$$

Examining a few special cases,

i) for B=T $\overline{|Z_{n_e}(f)|^2} = \frac{1}{2}\overline{|Z_n(f)|^2} = \frac{1}{2}\left(\frac{\sin \pi fT}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fT}{\pi f}\right)^2$ (A-57)

ii) for B>>T $B\left(\frac{\sin \pi fT}{\pi f}\right)^2 \simeq \overline{|Z_n(f)|^2} >> \overline{|Z_{n_e}(f)|^2}$

$$= \frac{1}{2}\left(\frac{\sin \pi fT}{\pi f}\right)^2 \circledast \left(\frac{\sin \pi fT}{\pi f}\right)^2 \qquad \text{(A-58)}$$

iii) for B<<T $\overline{|Z_n(f)|^2} \simeq \overline{|Z_{n_e}(f)|^2} \simeq T\left(\frac{\sin \pi fB}{\pi f}\right)^2$ (A-59)

The equations indicate an "improvement" of 1/2 when the maximum correlation time of the noise equals the length of the observation interval, a large improvement when the interval is much shorter than the maximum correlation time of the noise, and essentially no improvement if the correlation time of the noise is short compared to the total observation interval.

APPENDIX B

Weighting Factors

Assume that $s_1(t) + n_1(t)$ and $s_2(t) + n_2(t)$ are two seismograms of different signal-to-noise ratio which are to be added together coherently. A weight or bias a is to be given to the second so that the signal-to-noise ratio of the sum seismogram is a maximum:

$$\frac{\int\{s_1(t) + as_2(t)\}^2 dt}{\int\{n_1(t) + an_2(t)\}^2 dt} \overset{*}{=} \text{Maximum}$$

It is appropriate in this case to assume that the noise-cross-noise term

$$\int n_1(t)n_2(t)dt = 0,$$

as stations 1 and 2 are chosen sufficiently far apart for two noises to be independent, and the noise is assumed to be Gaussian.

$$\frac{\int\{s_1(t) + as_2(t)\}^2 dt}{\int n_1^2(t)\,dt + a^2\int n_2^2(t)dt}$$

Set $\int n_1^2(t)dt = n_1^2 \quad \int n_2^2(t)dt = n_2^2$

$\int s_1^2(t)dt = s_1^2 \quad \int s_2^2(t)dt = s_2^2$

Then our expression becomes:

---

* All integrals are assumed to be over unit time for expressional clarity.

$$\frac{s_1^2 + s_2^2 a^2 + 2a\int s_1(t)s_2(t)dt}{n_1^2 + a^2 n_2^2}$$

Since the signals $s_1(t)$ and $s_2(t)$ are coherent, i.e. mutually proportional, the integral may be written:

$$\int s_1(t)s_2(t)dt = \left(\int s_1^2(t)dt \int s_2^2(t)dt\right)^{\frac{1}{2}}$$

$$= s_1 s_2$$

(because the Schwartz inequality reduces to an equality in this case.) The ratio to be maximized then becomes

$$\frac{(s_1 + as_2)^2}{n_1^2 + a^2 n_2^2}$$

For ease in differentiation, we make the substitution

$$(s_1/s_2) = k_1 \quad (n_1/n_2) = k_2$$

$$\frac{(s_1 + as_2)^2}{n_1^2 + a^2 n_2} = \frac{(k_1 + a)^2}{k_2^2 + a^2} \quad \frac{s_2^2}{n_2^2}$$

To find the maximum with respect to a, set

$$\frac{d}{da}\left(\frac{(k_1+a)^2}{k_2+a^2}\right)\frac{s_2^2}{n_2^2} = 0$$

or

$$k_2^2 + a^2 - a(k_1 + a) = 0$$

$$k_2^2 - ak_1 = 0$$

$$a = \frac{k_2^2}{k_1} = \frac{n_1^2}{n_2^2} \frac{s_2}{s_1}$$

The sum seismogram is therefore

$$S(t) = s_1(t) + n_1(t) + \frac{n_1^2}{s_1} \frac{s_2}{n_2^2} \left(s_2(t)+n_2(t)\right)$$

Multiplying S(t) + N(t) by a constant does not change the signal-to-noise ratio, so we may write an equivalent sum as

$$\frac{s_1}{n_1^2} \left[s_1(t) + n_1(t)\right] + \frac{s_2}{n_2^2} \left[s_2(t) + n_2(t)\right]$$

Generalizing this to an arbitrary number of seismograms, we find that the proper weighting factor (for seismogram i) to maximize total signal-to-total-noise ratio is

$$\frac{s_i}{n_i^2} = \frac{\left\{\int s_i^2(t)dt\right\}^{\frac{1}{2}}}{\int n_i^2(t)dt}$$

Note: In practice, signal-to-noise ratio is much greater than one, so that we can approximate the true signal power $\int s_i^2(t)dt$ by $\int (s_i(t) + n_i(t))^2 \, dt$.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) General Atronics Corporation 1200 E. Mermaid Lane Philadelphia, Penna. 19118	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

LARGE APERTURE SEISMIC ARRAYS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Second Quarterly Technical Report June - September 1966

5. AUTHOR(S) (Last name, first name, initial)

Not Applicable

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June - September 1966	81	14

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 19(628)-5981 b. PROJECT NO.	1597-2033-8
c. ARPA Order No. 800 d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) Not Applicable

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this report is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Advanced Research Projects Agency Washington, D.C.

13. ABSTRACT

This report describes three topics considered during the second quarter of Contract AF19(628)-5981.

Three topics reported on are:

a) the estimation of the spectra of transient signals with additive noise,
b) the theoretical aspects of clustering seismometers in a large aperture seismic array,
c) the results of tests of a revised automatic pP phase detection technique.

A method has been devised to estimate the energy density spectra of transient events in such a way that the variance of the spectral estimates produced by additive noise is minimized.

The study of clusters shows:

1) dependence of array S/N performance on cluster spacing,
2) a way in which to determine the amount of clustering possible without S/N degradation depending on noise conditions at the receiving site.

Tests with the revised automatic pP detector have shown 100% correct results with surface focus events (6), 83% correct results for earthquakes (12), with the remaining earthquakes producing anomalies.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Signal Processing Seismic LASA Array Theory Spectral Analysis						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

Unclassified

Security Classification

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS ELECTRONIC SYSTEMS DIVISION (AFSC)
LAURENCE G. HANSCOM FIELD, BEDFORD, MASSACHUSETTS 01731

REPLY TO ATTN OF: ESE(Mr. O'Brien/5322) 7 Apr 67

SUBJECT: Release of Technical Document

TO: ESTIP

Identification: ESD-TR-67-220

ESE has no objection to release of the above-cited document to the Clearinghouse for Scientific and Technical Information.

JOHN T. O'BRIEN
Chief, Public Information Division
Information Office

1 Atch
n/c