

VLBA Fringe Phasing Error

John Granlund

February 20, 1987

VLBA Correlator Memo No. 82

(870225)

1. Introduction

The phases of a correlation function and of the corresponding correlator count, at the same lag, generally disagree. The difference is between phase after quantization but before quantization-correction on the one hand, and phase either before quantization or after correction on the other.

A preliminary analysis considered the effect of quantization on the correlation coefficient

$$\rho = \cos \omega\tau$$

of a strong, narrow spectral line. If ρ is the result of performing Van Vleck correction on the correlator count r of a two-level correlator according to

$$\rho = \sin\left(\frac{\pi r}{2}\right),$$

the two-level correlator count in this case must be

$$r = \frac{2}{\pi} \sin^{-1} \rho = \frac{2}{\pi} \sin^{-1}(\cos \omega\tau).$$

This is a sawtooth waveform having equal rates of rise and fall. Its peaks, troughs, and zeros agree with those of $\cos \omega\tau$, but its phase alternately leads and lags the phase of $\cos \omega\tau$. At most, when $\omega\tau = 26$ degrees, the sawtooth leads the cosine in phase by 6.65 degrees. This phase error is signal-dependent: As

$$R_{\text{hmax}} = \max_{\tau} \{|\rho(\tau)|\}$$

is reduced, making the portion of the correction curve that is actually used more nearly a straight line, the phase error is also reduced toward zero.

To correct for source position, the earth's rotation, and local-oscillator drifts, it has been VLBI practice to phase shift the quantized signals by amounts

that would properly complete the phase corrections if these phase shifts were, instead, applied to the signals before quantization. At this time, VLBA plans call for a continued use of this erroneous VLBI fringe phasing practice, and, to make matters worse, there is much interest in observing water masers and other strong, large-Rhmax sources. The purpose of this memo is to investigate the effects of adopting the current VLBA phasing plans.

Current VLBA plans call for the use of Nobeyama's FX correlator, but it should be emphasized that the phasing error under consideration is neither caused nor influenced by this choice; the same error would be produced by a lag correlator. The trouble is that the right phase shifts are being applied in the wrong places. These erroneous applications are (1) fringe phasing before FFT and (2) fractional bit-shift phasing after FFT; phasing error would be eliminated entirely if these two were, somehow instead, done either before quantization or after quantization-correction. The view that the "twiddle-factor" phasing within the FFT is also incorrect can, I suppose, also be justified, but I prefer the view that the FFT of an incorrect quantity is being correctly obtained. I shall confine attention in this memo to the effects of (1) fringe phasing before the FFT.

2. Analytical Methods

It will be noted that the preliminary analysis of the spectral line was completed by dealing with correlations and correlator counts -- and perhaps also their Fourier transforms -- without the need to treat the corresponding random signals using Monte Carlo techniques. This has saved much time, and it has produced results that are good to the full machine accuracy of my desktop computer, but it has prevented me from simulating anything but full machine accuracy phase shifting, FFT computation, etc. The investigations of this memo will be completed in the same way.

2.1 Phase and Phase-Shifting

What is the phase of the sawtooth-shaped correlator count, and, while we're at it, what is the amplitude? And how do phase and amplitude vary with lag? In the case of $\rho(\tau) = \cos \omega\tau$, we insist that we "know" that the phase is $\omega\tau$ and that the amplitude is always 1, but if we weren't so damn sure of ourselves, we could calculate phase and amplitude in a way that applies equally well to any real waveform, including the sawtooth-shaped count:

- * Use the Fourier transform to calculate the spectrum of $\rho(\tau)$.
- * Double the positive-frequency spectral components and discard the negative-frequency components.
- * Transform back to obtain the complex lag function $\hat{\rho}(\tau)$.

Because it has no negative-frequency spectral components, this complex function of lag is analytic in the upper half of the complex τ -plane. Following customary signal processing usage, I'll call it the analytic signal, or, in the case of a correlator count, the analytic correlator count. In the case $\rho(\tau) = \cos \omega\tau$, the

analytic correlation coefficient comes out $\hat{\rho}(\tau) = e^{j\omega\tau}$. The amplitude of $\rho(\tau)$ is $|\hat{\rho}(\tau)|$, the phase is $\text{Arg}[\hat{\rho}(\tau)]$, and these statements apply to any real Fourier-transformable waveform $\rho(\tau)$. It may be noted that the so-called complex cross-correlation measured by the VLA is exactly the analytic cross-correlation in the sense of this memo.

Although the three-step computational procedure outlined above is numerically by far the most efficient when the Fourier transform can be approximated by an FFT, the procedure can be easily simplified -- taking special care with limits and limit-taking -- to read

$$\hat{\rho}(\tau) = \rho(\tau) + j H\{\rho(\tau)\} ,$$

in which $H\{\rho(\tau)\}$ is the Hilbert transform of $\rho(\tau)$ in the sense of Schwartz [1966], or the negative of the Hilbert transform of $\rho(\tau)$ in the sense of Bateman, Titchmarsh, and, I suppose, Hilbert, himself. From this it is evident that

$$\text{Re}\{\hat{\rho}(\tau)\} = \rho(\tau) \quad \text{and} \quad \text{Im}\{\hat{\rho}(\tau)\} = H\{\rho(\tau)\} .$$

From the foregoing, it is clear that a version of $\rho(\tau)$ that has been advanced in phase by θ can be expressed as

$$\text{Re}\{\hat{\rho}(\tau) e^{j\theta}\} ,$$

and this observation is the key to the phase-shifting method that will be used in this investigation. There are, however, several peculiarities that should be understood before the method is used.

One would like to be sure that the result of advancing the phase of a waveform by θ , followed by retarding the phase by θ , is the original waveform. But the Hilbert transform of a constant -- the result of retarding the phase of a constant waveform by 90 degrees -- is zero. So all of, or part of, a constant waveform may be lost by first advancing and then retarding its phase by θ . Since phase shifting is a linear operation, we must conclude that phase advancing a waveform by θ and then phase retarding the result by θ will remove some of the average value of the waveform, and this is not acceptable. But the correlations, correlation coefficients, and correlator counts to be dealt with here all diminish rapidly enough with lag so that

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \rho(\tau) d\tau = 0 ;$$

their average values are all zero before any phase shifting, so there is no problem.

This happy conclusion must be changed considerably when the Fourier transforms used to evaluate the phase-shifted waveforms are replaced by FFT's. $\rho(\tau)$ is then taken to be periodic with a finite period T , so its new average value may no longer be zero. Since the average value of a waveform is proportional to its zero-frequency spectral component, we conclude that phase shift calculations using the FFT will not be accurate if the waveform has a d-c spectral component. A similar difficulty accompanies the highest-frequency spectral component, whose waveform is sampled just twice during one of its periods; phase shifting this waveform can make it zero at sample time. Thus, in using the FFT to calculate a phase-shifted version of a waveform, full accuracy can only be obtained if its spectral components at zero frequency and at half of the sampling rate, w , are absent.

If spectral components at the frequencies 0 and $w/2$ are to be removed, this must be done before quantization; it would not do to sample and quantize, FFT, and then to set these spectral components to zero. In fact, it should be remembered that spectral components offset from the two "target" frequencies by $\pm (\text{integer}) \cdot w$ will be aliased by the sampling process to the target frequencies, so these offset components should also be removed. In total, spectral components at d-c, $\pm w/2$, and all harmonics of $\pm w/2$ should be removed before quantization. However, my results show that when the spectral components at the target frequencies after aliasing are only moderately weak, their effects on the rest of the spectrum after processing are too small to be noticed.

Finally, in doubling the positive-frequency spectral components and discarding the negative-frequency ones, when using the FFT to determine the analytic signal, it should be noted that spectral components at the target frequencies are shared between the positive and negative frequencies. If these specific spectral components are not to be removed, they should be retained but should not be doubled at this step of the process.

2.2 The Simulation

An unquantized test correlation is given a phase lag θ , such as might result from propagation or local-oscillator drift. This phase-shifted correlation is normalized to produce a correlation coefficient and quantized, as in the preliminary analysis of the introduction, to produce a correlator count. The phase-shifted correlator count is next advanced in phase by θ to produce a fringe-rotated correlator count. We would like to believe that, after quantization correction and un-normalization, this function is exactly the original test correlation, but, of course, it is not.

In accordance with current VLBA plans, the fringe-rotated correlator count will be averaged over a time interval involving many revolutions of θ to reduce the variance of the result before the last steps of quantization correction and un-normalization are applied. Although the first results of this investigation omit this average over θ , the more pertinent results include it. The computations for these latter results are block diagrammed in Figure 1.

The first line of Figure 1 is an initialization that produces the analytic correlation, used in the second line to calculate the phase-retarded correlation. The second line completes the average over θ of the fringe-rotated correlator count. It is followed repeatedly, once for each point used in calculating the

average. The values of θ used are uniformly distributed over one revolution, the last point taken at $\theta = 2\pi$; the number of points is noted. The signs in the FFT blocks are the signs of the exponents used in the FFT kernels; thus FFT^- calculates a power spectrum from a correlation, and FFT^+ calculates a correlation from a power spectrum. In all cases, a 1024-point FFT has been used. On lines 1 and 3, "Spect" means a power spectrum, but on line 2 it means the spectrum of a correlator count. "Spect⁺" and "Spect⁻" mean the sets of positive- and negative-frequency spectral components, respectively.

The normalization and quantization blocks of Figure 1 deserve consideration. For all tests, a correlation has been chosen that reaches its maximum absolute value at zero lag, and this value is positive. But in the introduction, it was suggested that the parameter that best describes signal strength is R_{hmax} . Evidently normalizing the zero-lag correlation $C(0)$ produces

$$R_{\text{hmax}} = \frac{C(0)}{A},$$

which can be used to determine the normalizing power A for a given R_{hmax} . For the two-level correlator, I have used the arc sine and the sine for quantization and quantization-correction, as in the introduction. For the four-level correlator, I used a pair of very tight rational-function fits to the correction curve that were obtained by Fred Schwab for the equal-threshold case $n = 3$, $v_0 = 0.995687$. (See VLBA Correlator Memo No. 75 for notation.) The quantization curves came from reversion -- solving for r vs. ρ -- of the correction curves. For the low-range fit, $0 \leq |\rho| \leq 0.9$, an odd polynomial in $\frac{2}{\pi} \sin^{-1} \rho$ that produced an rms error of fit to the quantization curve of 1.7×10^{-8} was used. Newton's method produced full machine accuracy for the high range.

3. Results and Discussion

Two different test correlation functions have been used, the second suggested by results from the first. The reasons for this and several other changes are incorporated in the discussion.

3.1 The Notched Spectrum

Figure 2 shows the aliased spectrum of the truncated autocorrelation function of white noise filtered by an appropriate IF filter. Also, shown dotted, is the real and positive power spectrum of this noise further filtered by a notch filter that has been contrived to make spectral differences as evident as possible. The Fourier transform of the notch filter's power gain has been windowed to round -- very slightly -- its sharp corners. The autocorrelation function corresponding to this notched spectrum is chosen as the first test correlation. It is expected that distortion products, caused by the fringe phasing error under investigation, will partially fill the notch and thus provide some estimate of the dynamic range of spectral determination when current VLBA phasing plans are employed.

Figures 3 through 6 show results for two-level quantization; Figures 7 through 10 show results for four-level quantization.

With Rhomax = 0.4, Figure 3a shows the spectra resulting from the choice of six different values of θ ; the average over θ has been omitted. These spectra are no longer real; their magnitudes -- in decibels -- have been plotted. One of the choices is not $\theta = 0$; the spectrum for that choice overlies the dotted test spectrum. Spectral values $S(0)$ and $S(16)$ at the target frequencies before processing are noted. Sampling rate $w = 32$ MHz is being used, and for the bandpass IF filter chosen, $S(0)$ is aliased from ± 32 MHz.

Three different averages over θ have been completed in collecting ordinates for the solid curves of Figure 3b, which addresses the question of how many points -- different values of θ -- must be used to get a suitable approximation to the true average over infinitely many points. In this case, since there are three curves that fall on top of each other, eight points seem to be enough, but it will be noted from further results that more points are sometimes needed. When enough points have been used, the final spectra are, except for roundoff error, pure real, but they are not everywhere positive. The final spectrum of Figure 3b, for example, is negative in the ranges centered on 16, 22, and 32 MHz and bounded by the sharp spectral dips in the figure.

Parameter Q is discussed in connection with Figures 5. For the time being, $Q = 0$ means that the plotted spectra have been calculated exactly as is shown by the block diagram of Figure 1.

Spectral components at the target frequencies of the test spectrum were set to zero before the three solid curves of Figure 3b were reworked and plotted in Figure 3c. Infinitely deep dips in the dotted test spectrum are an evident result of this change, but the final solid curves seem to agree with those of the previous figure, except, perhaps, at the target frequencies. From here on, the correctness of the pre-quantization phase shifting will be ensured by setting the target-frequency spectral components of the test spectrum permanently to zero.

Figures 4a through 4j constitute a run in which Rhomax is reduced from 1 in steps of about 2 dB. It will be noted that more than 16 points for the average over θ are appropriate with Rhomax = 1, and that Rhomax must be reduced considerably if the dotted test spectrum is to be reproduced.

In Figure 5a, with Rhomax = 1, quantization correction, from the bottom line of Figure 1, has been entirely skipped: In place of the correction curve

$$\rho = \sin\left(\frac{\pi r}{2}\right),$$

the straight line through the origin

$$\rho = \frac{\pi r}{2}$$

has been used. The final spectrum is now entirely positive. This has suggested that some combination of the quantization-corrected result of Figure 4a and the uncorrected result of Figure 5a might show less distortion than either of its components. Accordingly, the quantization-correct and un-normalize block of Figure 1 was modified to deliver the "correlation"

$$C = A \left[Q \frac{\pi r}{2} + (1 - Q) \sin\left(\frac{\pi r}{2}\right) \right] ,$$

in which r is the average over θ of the fringe-rotated correlator count, and A is the normalizing power. With R_{hmax} maintained at unity, parameter Q is adjusted in Figures 5b through 5f to reach what appears to be its best value of 0.35 in Figure 5e.

The run over values of Q displayed in Figures 5 is repeated in Figures 6, but with $R_{\text{hmax}} = 0.4$. This time the best choice seems to be $Q = 0.55$, with the final spectrum shown in Figure 6g.

Four-level quantization produces notably less distortion than two-level quantization. Four-level results are arranged in the same order that was used for the two-level results. Figures 7 show spectra for (the same) selected values of θ , averages over θ using 8 and 16 points, and the effect of removing spectral components at the target frequencies. R_{hmax} is varied (over the same values) in the run of Figures 8. Q is varied in the runs of Figures 9 and Figures 10, using the same values of R_{hmax} that were chosen for the corresponding two-level runs. The modified "correlation" used for this purpose was

$$C = A \left[Q r \frac{dp(0)}{dr} + (1 - Q) \rho(r) \right] ,$$

in which $\rho = \rho(r)$ is the four-level quantization-correction curve.

From the runs of Figures 4 -- two-level quantization -- and Figures 8 -- four-level quantization -- in which R_{hmax} is varied, plots of the dynamic range of spectral determination vs. R_{hmax} can be taken. I have been rather sloppy in this, using the ad hoc definition

Dynamic range \equiv Attenuation of 22 MHz component of final spectrum.

These curves are so signal-dependent that a more careful determination did not seem justifiable. I doubt the applicability of the curves to spectral shapes much different from that of the notched spectrum that produced them.

The curves are plotted in Figure 11. Near $R_{\text{hmax}} = 0.1$, four-level quantization offers almost 7 dB more dynamic range than two-level quantization and, with R = dynamic range expressed as a power ratio,

$$R = \frac{36.8}{(R_{\text{hmax}})^2} \quad \text{-- 2-level quantization}$$

$$R \approx \frac{179}{(\text{Rhomax})^2} \quad \text{-- 4-level quantization .}$$

With the notch only 60 dB deep, it seems strange that the curves hook to the right to report a dynamic range in excess of 60 dB at small Rhomax. The reason is that over the range of Rhomax plotted, the 22 MHz components of the spectra are negative. To reach the positive, 60 dB down values expected when Rhomax is really made small, these spectral components must be driven through zero, at which point the ad hoc definition would give a dynamic range of infinitely many decibels.

Two-Level Quantization

	Rhomax = 1			Rhomax = 0.4		
Q	0	.35	1	0	.55	1
Dynamic Range (dB)	12	19	7	20	28	15

Four-Level Quantization

	Rhomax = 1			Rhomax = 0.4		
Q	0	.55	1	0	.65	1
Dynamic Range (dB)	17	27	15	26	37	24

Table 1. Summary of Runs on Q.

Table 1 is a summary of the four runs in which Q was varied to find the best combinations of corrected and uncorrected (fringe-rotated and averaged) correlator counts. For each method of quantization and for each of two values of Rhomax, the dynamic range is given for complete correction, for the best combination, and for no correction. Some 10 dB of dynamic range can be gained by using the best combination, but the Q for that best combination depends on Rhomax and the method of quantization.

The dynamic ranges of Table 1 should not be compared with the dynamic ranges plotted in Figure 11. I have been somewhat more careful, and certainly more conservative, in preparing Table 1. For it, dynamic range means the greatest attenuation above which each spectral component differs from the corresponding spectral component of the dotted test spectrum by less than 1 dB. I have read the runs on Q with a magnifying glass, certainly to no better than 1 dB of dynamic range by this definition.

3.2 Search for an Optimum Correction Curve

The runs on Q have produced "quantization correction" curves that show less distortion than the curves that apply with the fringe phasing error eliminated.

This has motivated the search for an optimum two-level quantization-correction curve to be described here.

The overall plan for the search is to modify the quantization-correct and un-normalize block of Figure 1 to use a rational function of r -- with enough parameters to be quite flexible -- for "quantization correction". The form finally selected was

$$\rho(r) = r \frac{p_0 + p_1 r^2 + p_2 r^4 + p_3 r^6 + p_4 r^8}{1 + p_5 r^2 + p_6 r^4} ;$$

the flexibility of this form is demonstrated by the results. Finally, the parameters of the form are adjusted to minimize the sum of squares of the decibel differences between the final spectrum and the test spectrum.

Some changes are needed first. Even with $R_{\text{hmax}} = 1$, most of the correlation coefficients of the previously used notched test spectrum are very small -- too small, in fact, for it to be determined whether they have been quantized or not. What is needed for this search is a test correlation coefficient whose values are more-or-less uniformly distributed over $-1 \leq \rho \leq 1$. The sawtooth from -1 to 1 and back to -1 was rejected because its spectrum would be concentrated too near the target frequency $f = 0$; the test correlation coefficient consisting of many cycles of this sawtooth was also rejected because its higher harmonics, which decay painfully slowly, would fall out of band, to be aliased back in band by the sampling process. Instead, something like the sinusoidal correlation coefficient $\rho(\tau) = \cos(2\pi f_0 \tau)$, with $f_0 = w/4$, was used. But the distribution of a sinusoid is accentuated near its peak and trough values. The test correlation coefficient

$$\rho(\tau) = \sqrt{1 - \left(\frac{2w\tau}{N}\right)^2} \cos(2\pi f_0 \tau) , \quad -\frac{N}{2w} < \tau \leq \frac{N}{2w}$$

was finally chosen. If infinitely many samples of a cosine are taken, uniformly distributed over its period, and these are weighted at random by samples of the radical, uniformly distributed over its domain, the product $\rho(\tau)$ is uniformly distributed over $-1 \leq \rho \leq 1$. But, of course, with only 1024 samples, the distribution deviates -- in a seemingly random way -- from uniformity. The range $0 \leq |\rho| \leq 1$ was subdivided into 128 levels and the samples in each level were counted as f_0 was adjusted. The most uniform distribution resulted with $f_0 = \frac{131}{512} \cdot w$, and this value was selected for the search.

With normalizing power A , the test correlation, itself, is $A\rho(\tau)$, and its power spectrum is

$$S(f) = \frac{\pi N A}{2w} \left\{ \frac{J_1 \left[\frac{\pi N}{w} (f - f_0) \right]}{\frac{\pi N}{w} (f - f_0)} + \frac{J_1 \left[\frac{\pi N}{w} (f + f_0) \right]}{\frac{\pi N}{w} (f + f_0)} \right\},$$

in which $J_1(x)$ is the Bessel function of first kind and order one. The function $\frac{J_1(x)}{x}$ is similar to $\frac{\sin x}{x}$, but it decays slightly more rapidly as x gets large, and

$$\lim_{x \rightarrow 0} \frac{J_1(x)}{x} = 1/2.$$

Then, neglecting the contribution from the spectral term centered at $-f_0$, the normalizing power must be

$$A = \frac{4w}{\pi N} = \frac{1}{8\pi} \text{ watts}$$

to achieve a peak power density of 1 watt/MHz at f_0 with a sampling rate $w = 32$ MHz and an $N = 1024$ point FFT in use.

Over the range of positive lags dealt with by the FFT, the test correlation goes through 131 cycles of decaying oscillation, and the power spectrum goes through some 256 cycles of oscillation in the range of positive in-band frequencies. I'll not try to plot either. It may also be noted that, since the power spectrum takes on both positive and negative values, this test correlation is a cross-correlation.

The objective function to be minimized in determining the parameters of the optimum correction curve was modified to speed its computation, to eliminate unwanted terms, and to include appropriate penalties for phase errors between the final spectrum and the test spectrum, even though these errors are mere roundoff errors when enough points have been used in calculating the average over θ . With an array of output spectral values T_i -- having small imaginary parts -- and the corresponding array of input spectral values, S_i , consider

$$|dB| = 10 \left| \log \frac{T_i}{S_i} \right| = \frac{10}{\ln 10} \left| \ln \frac{T_i}{S_i} \right| = \frac{10}{\ln 10} \left| \ln \left(1 + \frac{T_i - S_i}{S_i} \right) \right|.$$

If the decibel error is small, as is to be expected, then

$$|dB| \approx \frac{10}{\ln 10} \left| \frac{T_i - S_i}{S_i} \right|,$$

and an appropriate objective function is

$$F = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{10}{\ln 10} |dB| \right)^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left| \frac{T_i - S_i}{S_i} \right|^2.$$

With minimization completed,

$$\text{rms } |dB| \text{ error} \approx \frac{10 \sqrt{F}}{\ln 10}.$$

It may be noted that the objective function includes phase errors, in quadrature with magnitude errors, at the rate

$$\frac{\ln 10}{10} \text{ radians} = 13.19 \text{ degrees per decibel.}$$

Since the input spectrum has been scaled to make its largest spectral value unity, there will be no point in including in the objective function terms for which $|S_i|$ is much less than 10^{-6} ; in fact, spectral values at the target frequencies, which will be set to zero, must be excluded. To achieve these purposes, the objective function is further modified to read

$$F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{|T_i - S_i|^2}{|S_i|^2 + \delta}.$$

$ S_i $ (dB)	-30	-32	-34	-36	-38	-40	-42	-44	-46	-48	-50
Weight	.990	.975	.941	.863	.715	.500	.285	.137	.059	.025	.010

Table 2. Weights Produced by the Modified Objective Function when $\delta = 10^{-8}$.

To exclude objective function terms involving input spectral values less than 10^{-n} in magnitude, δ is set at 10^{-2n} . This produces a smooth cut-off of terms involving the smaller input spectral values, as is illustrated by Table 2.

Before the most recent modification, it was possible to calculate the rms $|dB|$ error directly from the final value of the objective function, but now the objective function is a weighted sum of errors. To make such a sum interpretable as the mean

square $|dB|$ error, it must be divided by the sum of weights. The final arrangement is

$$W = \left(\frac{\ln 10}{10}\right)^2 \sum_{i=0}^{N-1} \frac{|S_i|^2}{|S_i|^2 + \delta}$$

and

$$F = \frac{1}{W} \sum_{i=0}^{N-1} \frac{|T_i - S_i|^2}{|S_i|^2 + \delta},$$

from which

$$\text{rms } |dB| \text{ error} = \sqrt{F}.$$

A 1024-point average over θ of the fringe-rotated correlator count has been used in computing the various output spectra for the minimizations, but a 512-point average was also accumulated for check purposes. After each minimization, the objective function was re-calculated with $|T_i - S_i|^2$ replaced by the squared magnitude of the difference between the output spectrum calculated from the 1024-point average over θ and the output spectrum calculated from the 512-point average. In all cases, the square roots of such check "objective functions" were more than two orders of magnitude less than the corresponding rms $|dB|$ errors. This has shown that the 1024-point average over θ was more than adequate.

The objective function was minimized for a variety of values of δ , both with correction curve parameters p_1 through p_6 set to zero -- no correction -- and with all seven parameters free to be adjusted -- full correction. p_0 is, of course, $\frac{dp(0)}{dr}$. The results are shown in Table 3.

$-5 \log \delta$ (dB)	No correction		Full correction	
	p_0	rms $ dB $ error	p_0	rms $ dB $ error
15	1.286 879	0.737 748		
20	1.391 845	0.666 376	1.352 854	0.654 019
25	1.470 347	0.512 023		
30	1.519 546	0.363 525	1.520 166	0.363 223
35	1.546 539	0.246 543		
40	1.557 017	0.190 254	1.557 207	0.190 100
45	1.557 365	0.225 719		
50	1.542 320	0.484 311	1.547 041	0.476 945
55	1.405 006	1.375 888		
60	0.743 229	3.144 086	1.183 805	1.397 876
65	0.130 159	4.158 034		

Table 3. Results of the Search for an Optimum Correction Curve.

Rather than listing all seven correction curve parameters for each value of δ , I believe it will be more informative to describe the curves. The largest average over θ of the fringe-rotated correlator count occurs at zero lag and is $r = 0.81$. Except for the curve with 60 dB cut-off -- i.e., $\delta = 10^{-12}$ -- it is not possible to distinguish the plotted full-correction curves, over the range $0 \leq |r| \leq 0.81$, from straight lines through the origin, even with the aid of a straight

edge. The near agreements between values of $p_0 = \frac{dp(0)}{dr}$ and rms |dB| error for the no-correction lines through the origin and the full-correction curves suggest this. The curve with 40 dB cut-off is plotted in Figure 12a. The rms |dB| error reaches a minimum with a 40 dB cut-off and rises as δ is increased above or reduced below this value. This suggests that major errors are made with large spectral values and small spectral values, with lesser error in between. As the rms |dB| error diminishes, the slope, p_0 , increases, but it never reaches

$\pi/2 = 1.570796$, the slope of the Van Vleck correction curve at $r = 0$. I suppose this is because it is more accurate to reduce the output when the computed result is less certain.

The curve with 60 dB cut-off is plotted in Figure 12b. The fact that the rms |dB| error produced by this curve is less than half the rms |dB| error produced by the best straight line shows that there must be something extraordinary about the curve! The nick in the curve above the word "curve" is not a plotter error, but one of two very special ingredients provided by the form: The form produces real zeros at $r = \pm 0.811870$ and real poles at $r = \pm 0.811902$; the violence caused by these pole-zero pairs was not plotted because the 111 points used for the plot were much too widely spaced. The effects of these pole-zero pairs on the correction of the largest counts are illustrated in Table 4: The top three counts are reduced considerably, relative to the next smaller counts.

Lag Number	r	$p(r) - r p(1)$
0	0.810 371	- 0.024 709
1022 } 2 }	0.807 519	- 0.012 318
1020 } 4 }	0.801 473	- 0.008 436
981 } 43 }	0.800 498	- 0.008 191
983 } 41 }	0.799 914	- 0.008 063
979 } 45 }	0.796 942	- 0.007 561
39 } 985 }	0.794 736	- 0.007 296
6 } 1018 }	0.789 836	- 0.006 881

Table 4. Effects of the Pole-Zero Pairs on the Top Fifteen Counts.

The other special ingredient of the form is a pair of imaginary poles at $r = \pm j 0.077\ 059$ and a pair of imaginary zeros at $r = \pm j 0.078\ 899$. These increase the slope of the correction curve in the neighborhood of $r = 0$, as can be seen with the aid of a straight edge.

Having thus demonstrated the flexibility of the quantization correction form, it is time to get practical. The 60 dB cut-off correction curve of Figure 12b is clearly far too signal-dependent to be of any use. In fact, the most practical conclusion of this search for an optimum correction curve is that, for the test correlation employed, no correction should be used: Correlation coefficients should be calculated from two-level correlator counts according to

$$\rho = \frac{\pi r}{2}.$$

3.3 Harmonic Analysis

Why were the distortions produced by the test correlation used in the search for an optimum correction curve so much less than the distortions produced by the notched spectrum? For that matter, since Alan Rogers (VLBA Correlator Memo No. 68) promises no harmonics, why are we finding any distortion at all? In this section, I'll reproduce Rogers' results analytically in the simpler, continuous -- not sampled -- environment. There will be differences between the continuous and sampled environments, but I do not expect them to be significant.

Consider the correlation coefficient

$$\rho(\tau) = \sum_n a_n \cos(\omega_n \tau + \alpha_n)$$

and the effect of processing it using current VLBA fringe phasing plans. After the phase of $\rho(\tau)$ has been retarded by θ , the phase-shifted version is

$$\rho_s(\tau) = \sum_n a_n \cos(\omega_n \tau + \alpha_n - \theta) \equiv \sum_n a_n \cos x_n.$$

The right-hand abbreviation will be used to save space, but it must be remembered that each x_n contains the phase lag $(-\theta)$. $\rho_s(\tau)$ is next quantized using either $r = \frac{2}{\pi} \sin^{-1} \rho$ for a two-level correlator or the corresponding four-level quantization curve. In either case, the curve can be represented by the odd power series

$$r = \sum_{\substack{m=1 \\ m\text{-odd}}}^{\infty} b_m \rho^m,$$

so that the quantized version of $\rho_s(\tau)$ is the correlator count

$$r_s(\tau) = \sum_{\substack{m=1 \\ m\text{-odd}}}^{\infty} b_m \left(\sum_n a_n \cos x_n \right)^m$$

$$= \sum_{\substack{m=1 \\ m\text{-odd}}}^{\infty} b_m \sum_{n_1} a_{n_1} \sum_{n_2} a_{n_2} \dots \sum_{n_m} a_{n_m} \cos x_{n_1} \cos x_{n_2} \dots \cos x_{n_m}.$$

The products of cosines can be broken down by using

$$\cos u \cos v = 1/2[\cos(u + v) + \cos(u - v)].$$

For example, the cubic term of the series, involving ρ_s^3 , reads

$$1/4 b_3 \sum_{n_1} a_{n_1} \sum_{n_2} a_{n_2} \sum_{n_3} a_{n_3} \left[\cos(x_{n_1} + x_{n_2} + x_{n_3}) + \cos(x_{n_1} + x_{n_2} - x_{n_3}) \right. \\ \left. + \cos(x_{n_1} - x_{n_2} + x_{n_3}) + \cos(x_{n_1} - x_{n_2} - x_{n_3}) \right].$$

It must be clear at this point that I'll not come close to expressing $r_s(\tau)$ in closed form, but let's continue, watching the effects on this cubic term. $r_s(\tau)$ is next advanced in phase by θ and averaged over θ . The cubic term of this fringe-rotated count, to be averaged over θ , reads

$$1/4 b_3 \sum_{n_1} a_{n_1} \sum_{n_2} a_{n_2} \sum_{n_3} a_{n_3} \left[\cos(x_{n_1} + x_{n_2} + x_{n_3} + \theta) + \cos(x_{n_1} + x_{n_2} - x_{n_3} + \theta) \right. \\ \left. + \cos(x_{n_1} - x_{n_2} + x_{n_3} + \theta) + \cos(x_{n_1} - x_{n_2} - x_{n_3} + \theta) \right]$$

Since each x_n contains the phase lag $(-\theta)$, only the second and third terms in square brackets survive the average. The surviving cosine terms are the ones whose arguments contain no net multiple of θ ; typical of these are $\cos[(\omega_5 + \omega_6 - \omega_7)\tau + \alpha_5 + \alpha_6 - \alpha_7]$, $\cos[(2\omega_4 - \omega_3)\tau + 2\alpha_4 - \alpha_3]$, and $\cos(\omega_2\tau + \alpha_2)$, itself. The terms $\cos[(\omega_5 + \omega_6 + \omega_7)\tau + \alpha_5 + \alpha_6 + \alpha_7]$ and in particular $\cos(3\omega_1\tau + 3\alpha_1)$ cannot emerge from the average over θ . As Rogers has reported, harmonics are not produced, but many other cross-product terms remain as sources of distortion.

The case in which only one (zero-width) spectral line is present deserves special consideration. The average over θ of the fringe-rotated count in this case is again a single sinusoid at the frequency of the original spectral line, but its amplitude is not what would be obtained from

$$\overline{r(\tau)} = \rho(\tau) \frac{dr(0)}{d\rho} .$$

Rogers complains of an amplitude "scaling", but the amplitude is really in error because the average contains distortion products at the frequency of the spectral line, generated by all of the higher-order terms of the power series for the quantization curve. For the case of two-level quantization with $R_{\text{hmax}} = 1$, I find that

$$\overline{r(\tau)} = \frac{8}{\pi^2} \rho(\tau) .$$

In any case, quantization correction of $\overline{r(\tau)}$ from the single spectral line would produce odd harmonics of all orders.

In spite of the difficulties that have been uncovered so far, it should be noted that the average over θ does bring about the improvements of phase switching, even though it prevents retrieval of the correct spectrum: If the thresholds are not precisely set, so that there are even terms in the power series for the quantization curve, all of these even terms average to zero, leaving a net odd quantization curve that is truly zero when the correlation coefficient vanishes.

The cross-power spectrum used in the search of Section 3.2 for an optimum correction curve was not a zero-width spectral line with non-zero positive-frequency spectral values confined to a single FFT bin, but the line was relatively narrow. Perhaps the small distortions produced by the search can be explained by the approximation of the test spectrum to a single, pure-line spectrum. Rogers' amplitude "scaling" was, after all, corrected by adjusting the slope parameter p_0 . Centered at $f_0 = w/4$, this test spectrum had the unfortunate property that if odd harmonics were produced by the erroneous fringe phasing, they would all be aliased to center again on $w/4$. To make the search a rough experimental check for the presence of odd harmonics, it was reworked with $f_0 = \frac{65}{512} \cdot w = w/8$. With this choice, third and fifth harmonics, if present, would be aliased to $\frac{3}{8} w$, seventh and ninth harmonics to $\frac{1}{8} w$, eleventh and thirteenth harmonics to $\frac{3}{8} w$, etc., etc. The reworked search did not exactly reproduce the results of Table 3, and the extraordinary difference between the "correction" curve and the best straight line through the origin with a 60 dB cut-off did not reappear. But through a cut-off of 45 dB, the rms |dB| errors agreed to better than 0.1 dB. Surely the agreement would have been much worse if harmonics had been produced.

Consider next the effects of current VLBA fringe phasing plans on the processing of a white noise. The correlation function is proportional to a delta function, but what about the correlation coefficient, which must not exceed unity in magnitude? It is necessary to deal with the ultimate delta function as the limit of a pulse whose duration vanishes as the width of the white, low-frequency portion of the spectrum increases without bound. Even the shape of this pulse must be chosen with some care: Suppose that the correlation coefficient, as a function of lag, is represented by a unit-height rectangular pulse of duration ϵ , and that it is planned to allow ϵ to approach zero after the average over θ of the fringe-rotated correlator count has been determined. But the Hilbert transform of the rectangular pulse approaches infinity logarithmically at the lags corresponding to the beginning and the end of the rectangular pulse. So how does one generate a phase-shifted version of this correlation coefficient that can be quantized?

I have chosen

$$S_{\rho}(f) = \rho_m \pi \epsilon e^{-2\pi\epsilon|f|}, \quad \epsilon > 0 \quad \text{and} \quad 0 \leq \rho_m \leq 1$$

as the spectrum of the test correlation coefficient, so that the analytic test correlation coefficient is

$$\hat{\rho}(\tau) = 2 \int_0^{\infty} S_{\rho}(f) e^{j2\pi f\tau} df = \frac{\rho_m \epsilon}{\epsilon - j\tau} = \rho_m \epsilon \frac{\epsilon + j\tau}{\epsilon^2 + \tau^2},$$

and the test correlation coefficient, itself, is

$$\rho(\tau) = \text{Re}\{\hat{\rho}(\tau)\} = \frac{\rho_m \epsilon^2}{\epsilon^2 + \tau^2}.$$

Note that the amplitude of this test correlation coefficient is

$$|\hat{\rho}(\tau)| = \frac{\rho_m \epsilon}{\sqrt{\epsilon^2 + \tau^2}} \leq \rho_m,$$

so that with $\rho_{\text{max}} = \rho_m \leq 1$, no phase shifting of the test correlation coefficient can make its absolute value exceed unity. The width or duration of the test

correlation

$$\frac{\int_{-\infty}^{\infty} \rho(\tau) d\tau}{\max_{\tau} \{\rho(\tau)\}} = \frac{\rho_m \pi \epsilon}{\rho_m} = \pi \epsilon$$

is of the order of ϵ , so allowing ϵ to approach zero at the end will ensure that this analysis has dealt with a white input spectrum.

I shall spare the reader by not including the necessarily heavy mathematics of this investigation in the memo, but I am prepared to deliver copies of the analysis to interested readers. I find that, after the test correlation coefficient has been retarded in phase by θ , quantized using an odd power series, and advanced in phase by θ , the average over θ of each power series term can be calculated first. And I find that each such term-by-term average of the fringe-rotated correlator count has a width that vanishes with ϵ . Thus when $\epsilon \rightarrow 0$, the average over θ of the fringe-rotated correlator count is white. This confirms Rogers' experimental result.

In contrast to the case of the zero-width spectral line, quantization correction of this white correlator count would produce a white correlation.

I must close this Section by offering sympathy to Alan Rogers, who made experimental measurements on what, I suppose, he expected to be the worst two cases of fringe-phasing error, but which have turned out to be, instead, probably the only two cases in which the only effects of the error are "scaling" errors.

4.0 Conclusion

I have not found the worst case test spectrum -- or correlation -- either, but from the results that have been presented, I expect that the worst case spectrum will be white, with a narrow-band section completely notched out. The center frequency chosen for this rectangular notch should have no effect in the continuous environment and very little effect in the sampled environment. Such a notched spectrum is just what telephone engineers have used to measure the performance of multi-channel transmission systems. If this guess is correct, the results from the V-notched spectrum of Section 3.1 are nearly worst case, and the curves of Figure 11 provide an estimate of worst case dynamic range, at least out to 50 dB, when quantization correction is employed.

Do we really insist that the VLBA be able to measure, with a comfortable dynamic range, sources producing ρ_{\max} near unity, or are we willing to quit above $\rho_{\max} = 0.1$? The first choice will, at least, be more expensive.

REFERENCE:

M. Schwartz, W. R. Bennett, and S. Stein, Communication Systems and Techniques, McGraw-Hill, 1966, pp. 29-45.

Figure Captions

Figure 1. Computational Block Diagram.

Figure 2. IF Filter Gain and Notched Test Spectrum.

Figure 3. Two-Level Quantization.

3(a). Spectra for Selected θ .

3(b). Averages over θ with
Various Numbers of Points.

3(c). Setting Target Spectral
Components to Zero.

Figure 4. A Run with Rhomax Varied; Two-Level.

Figure 5. A Run with Q Varied; Rhomax = 1; Two-Level.

Figure 6. A Run with Q Varied; Rhomax = 0.4; Two-Level.

Figure 7. Four-Level Quantization.

7(a). Spectra for Selected θ .

7(b). Averages over θ with
Various Numbers of Points.

7(c). Setting Target Spectral
Components to Zero.

Figure 8. A Run with Rhomax Varied; Four-Level.

Figure 9. A Run with Q Varied; Rhomax = 1; Four-Level.

Figure 10. A Run with Q Varied; Rhomax = 0.4; Four-Level.

Figure 11. Dynamic Range vs. Rhomax.

Figure 12. Optimum Correction Curves.

12(a). $\delta = 10^{-8}$

12(b). $\delta = 10^{-12}$.

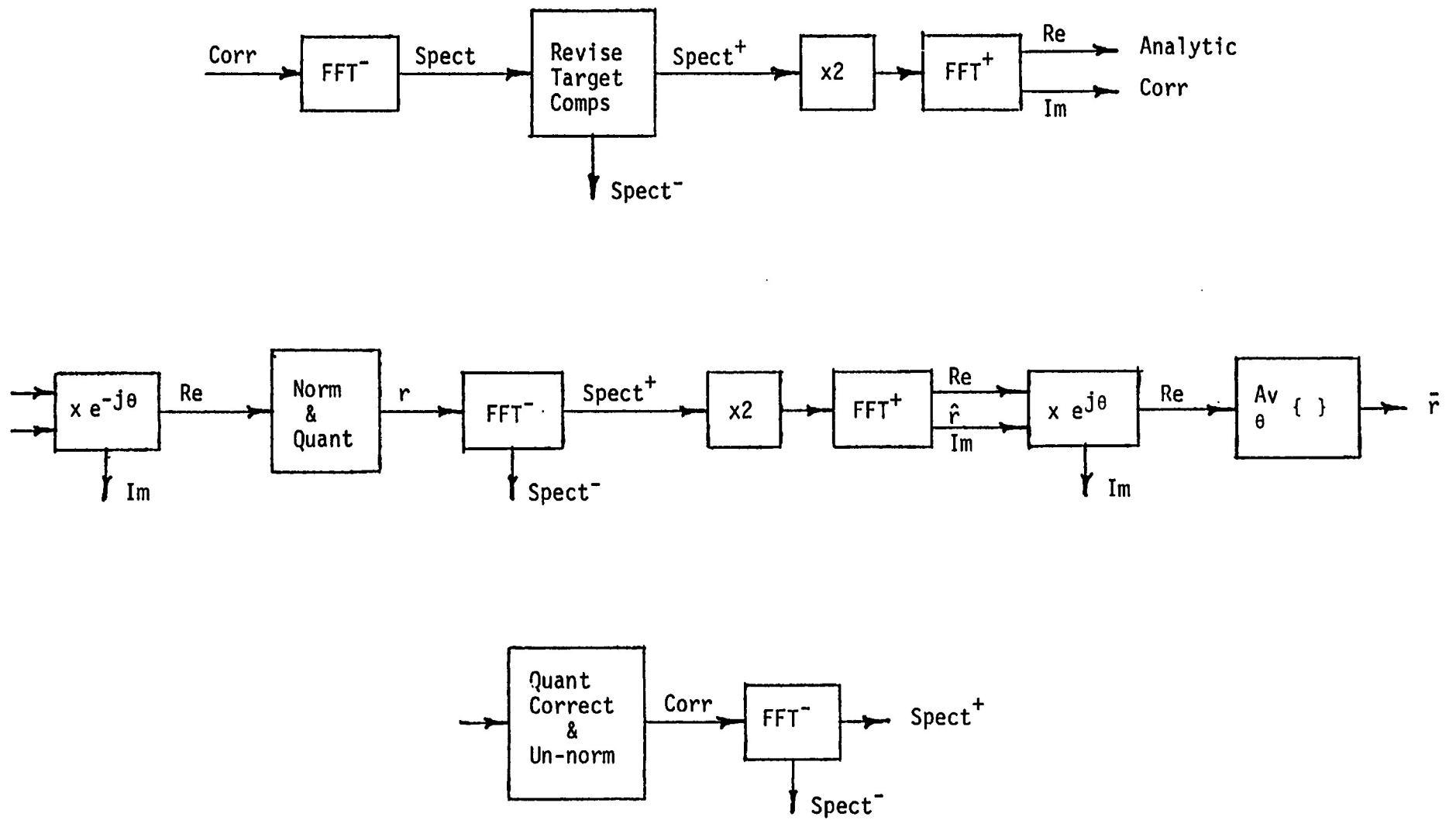


Fig. 1. Computational Block Diagram

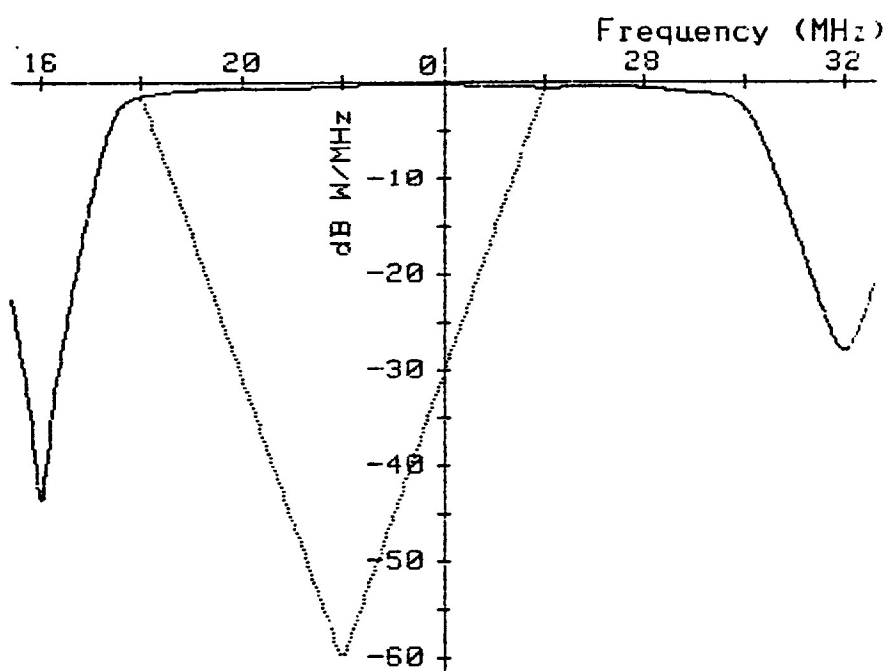
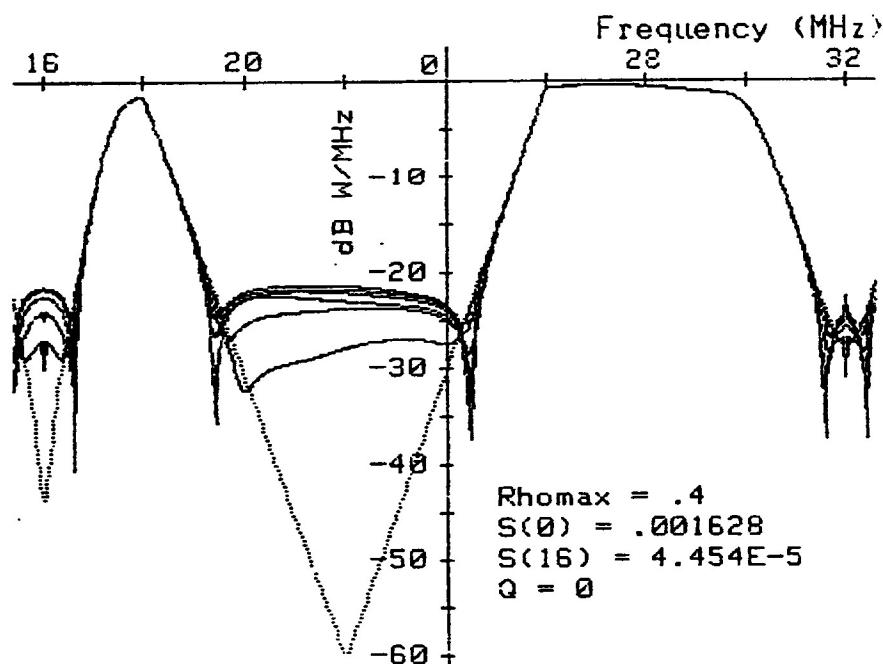
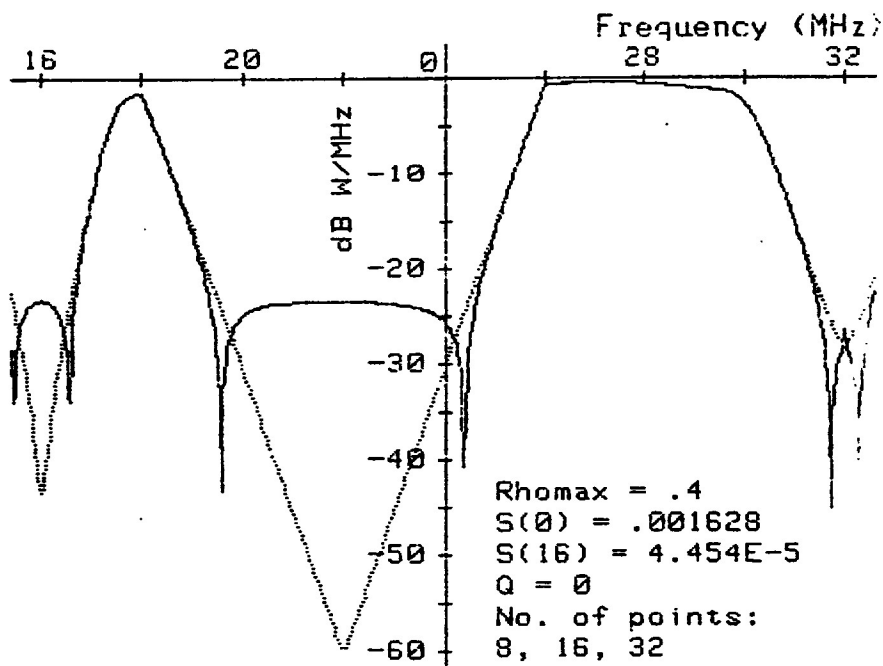


Figure 2. IF Filter Gain and Notched Test Spectrum.

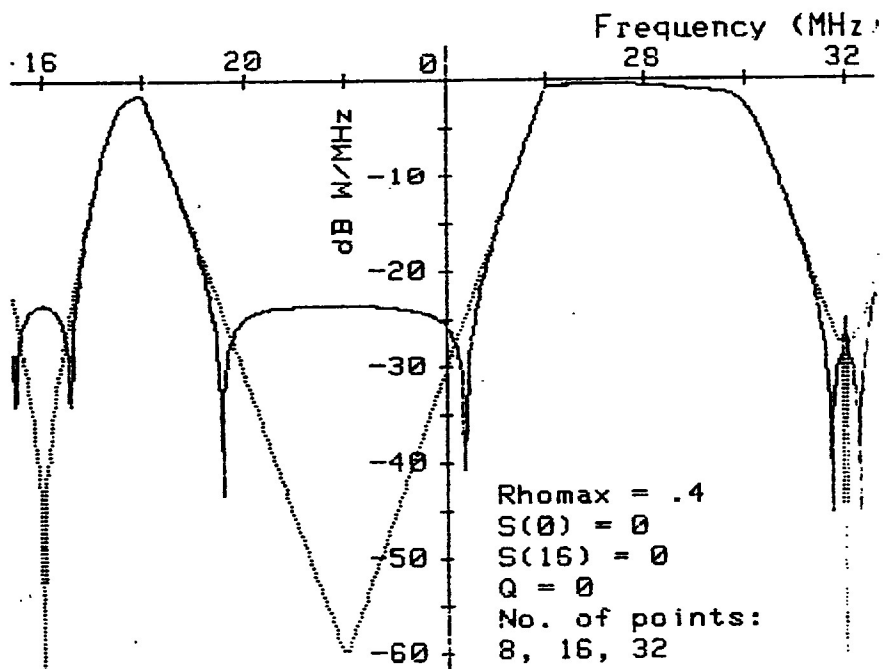


3(a). Spectra for Selected θ .

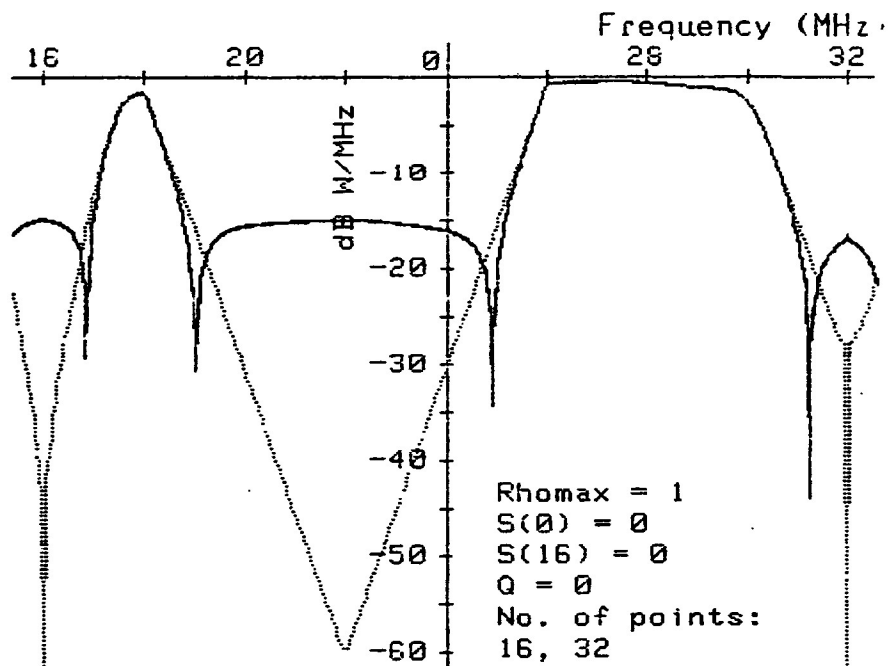
Figure 3. Two-Level Quantization.



3(b). Averages over θ with Various Numbers of Points.

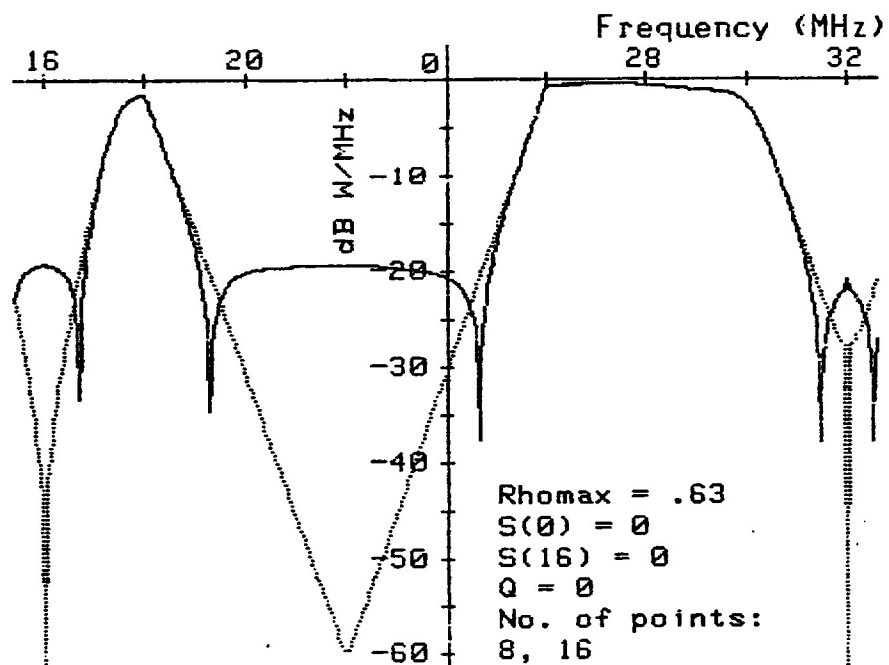


3(c). Setting Target Spectral Components to Zero.

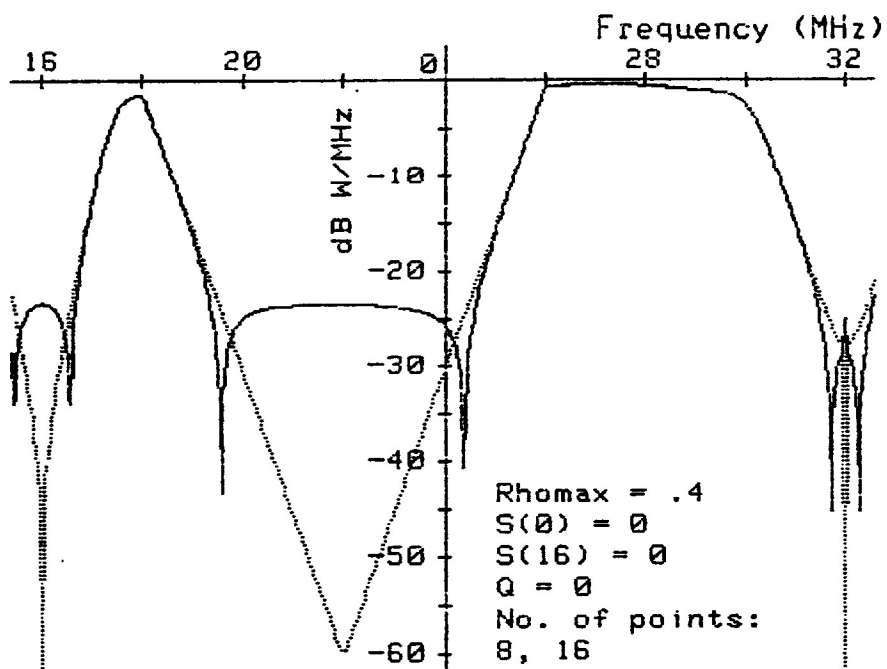


4(a).

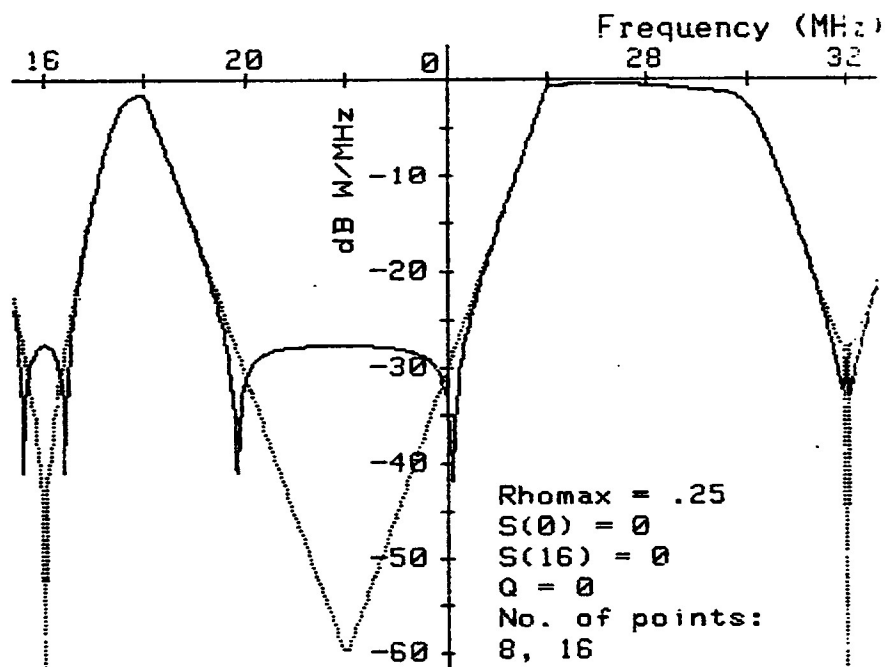
Figure 4. A Run with Rhomax Varied; Two-Level.



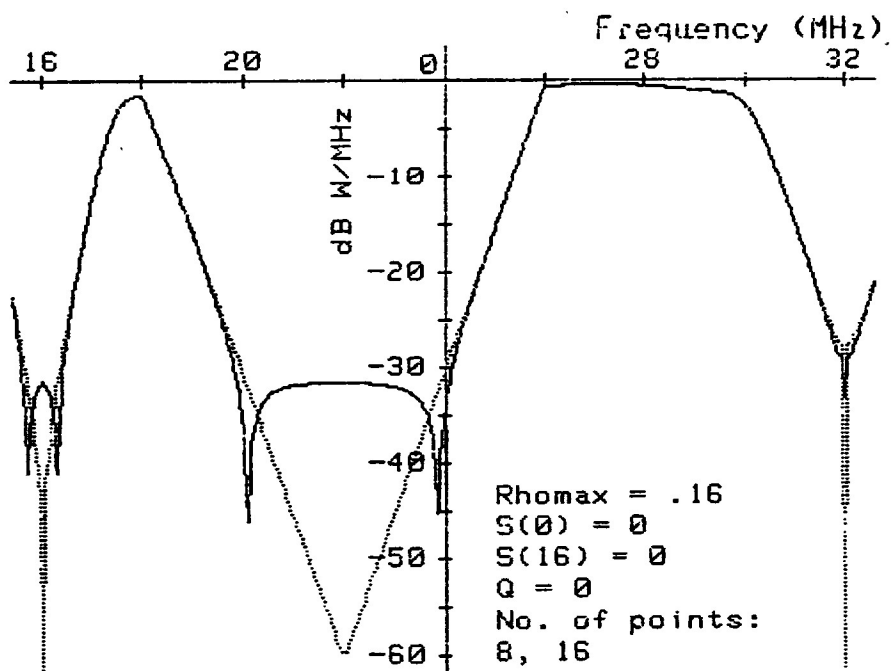
4(b).



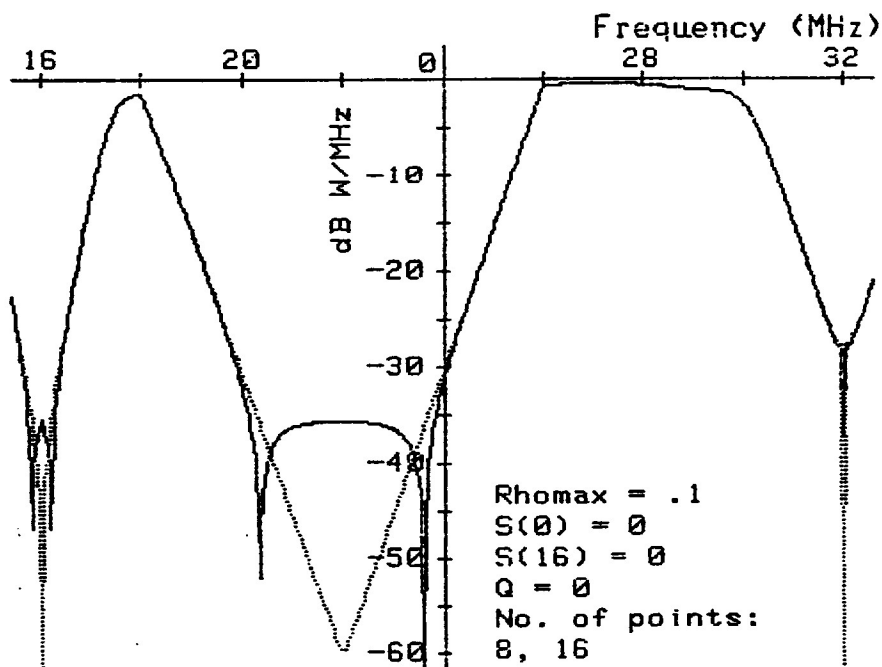
4(c).



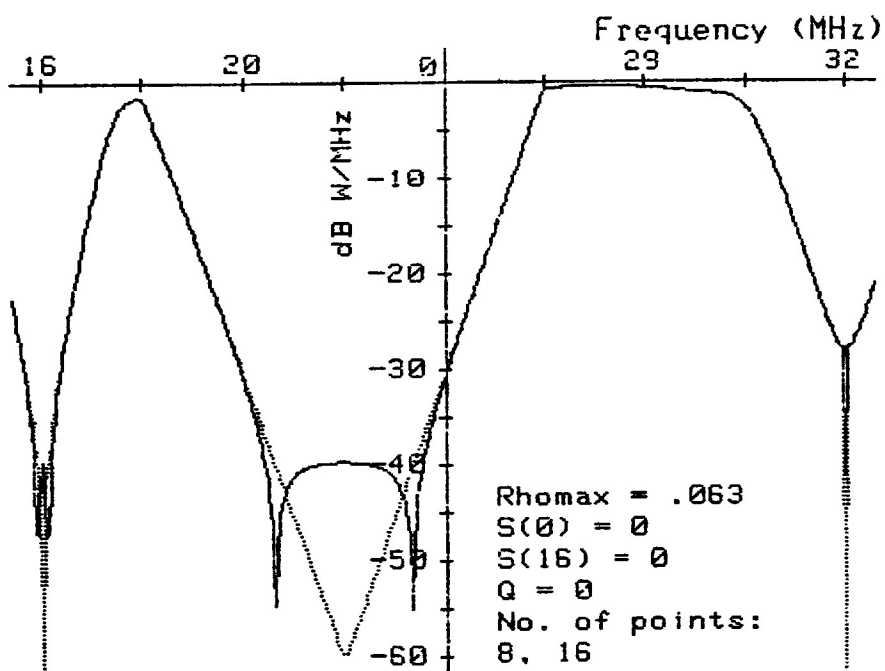
4(d).



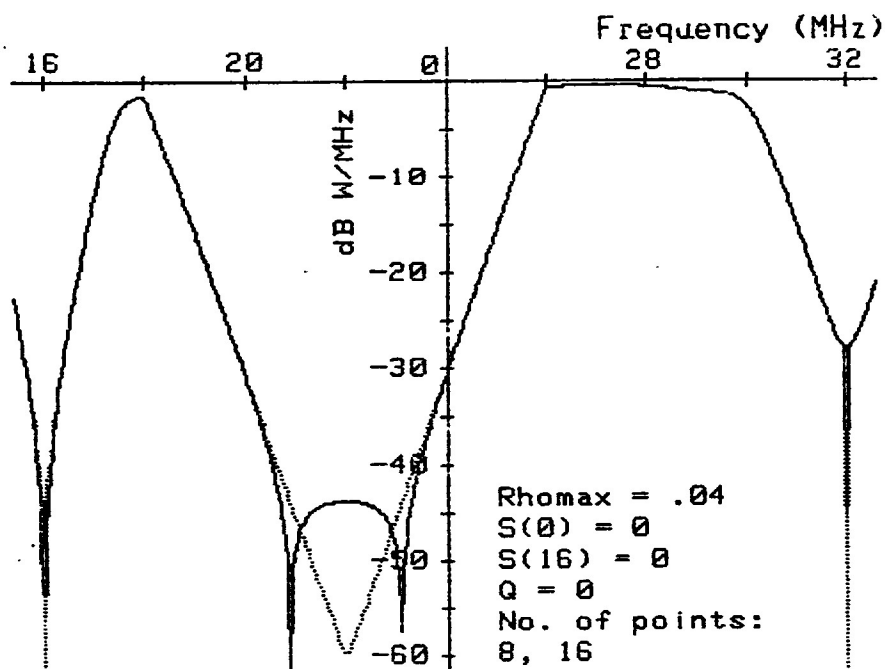
4(e).



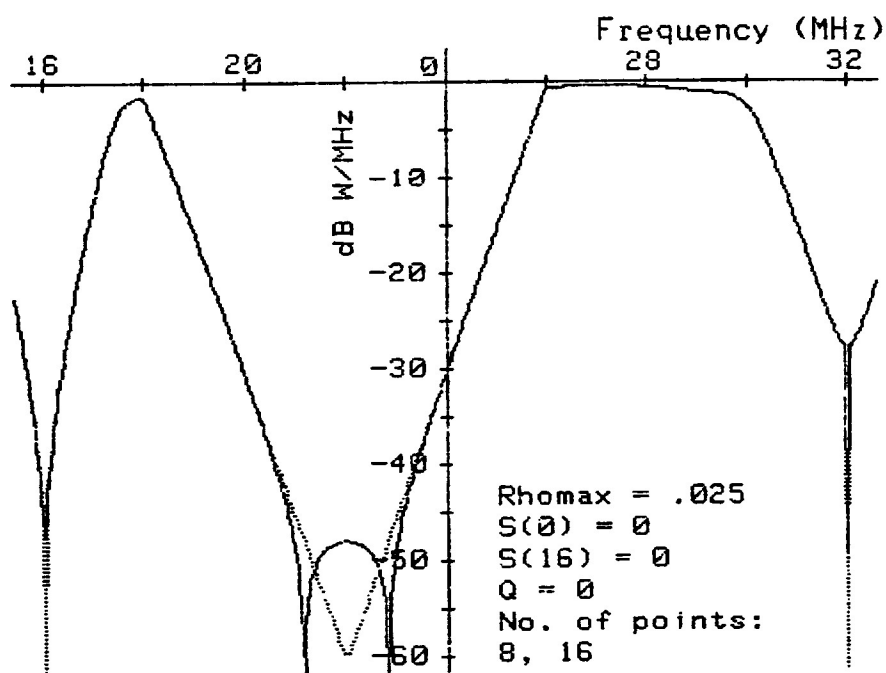
4(f).



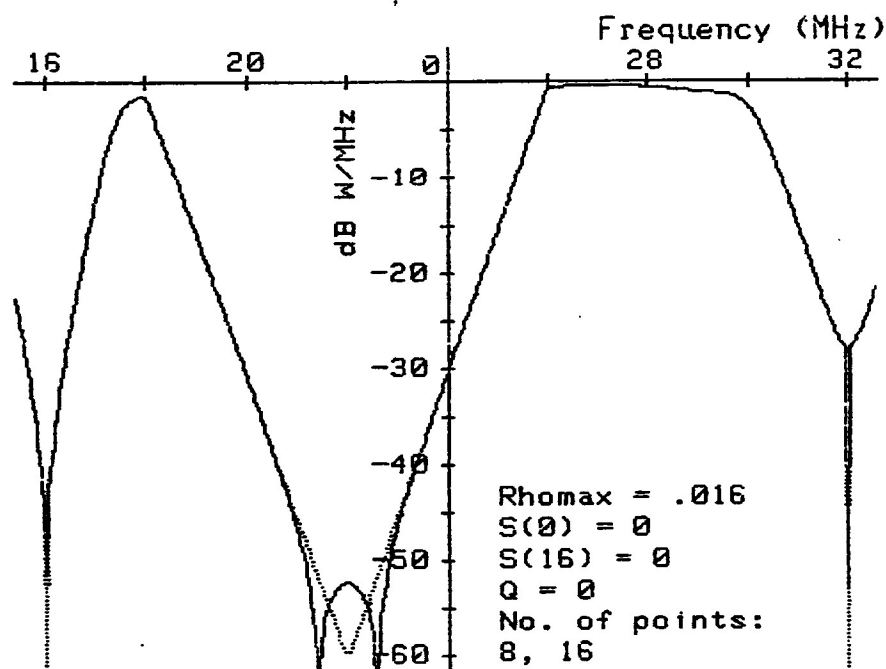
4(g).



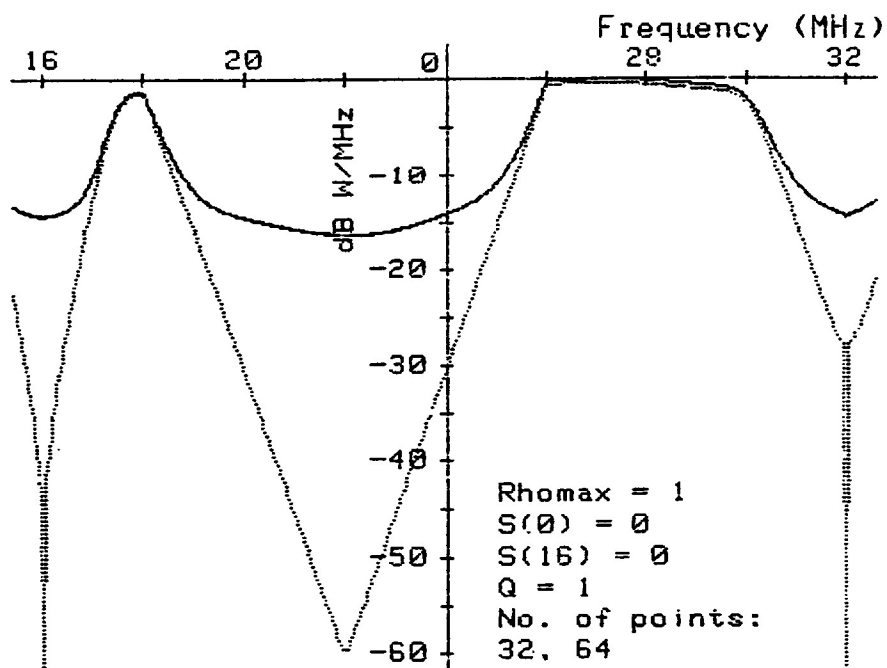
4(h).



4(i).

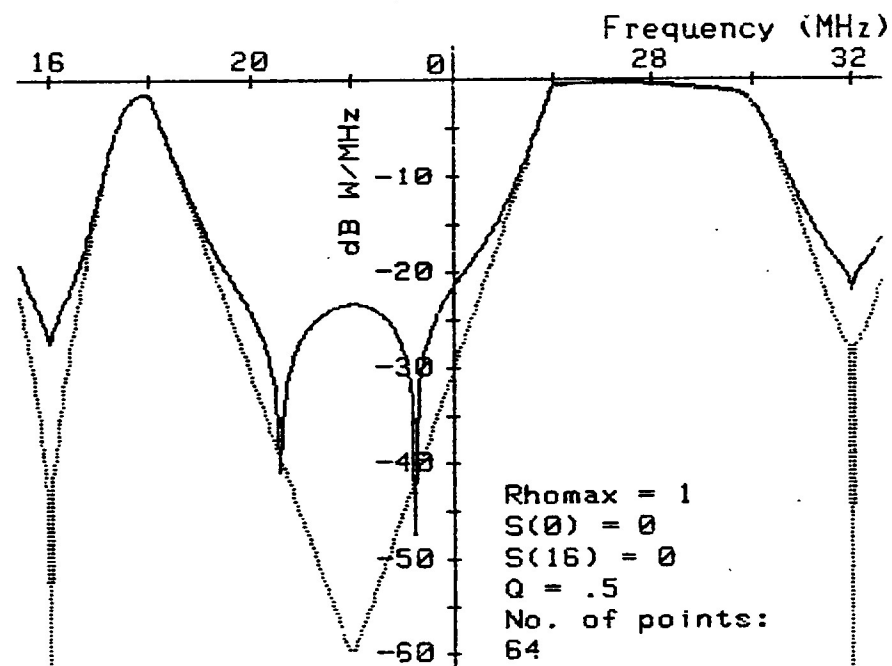


4(j).

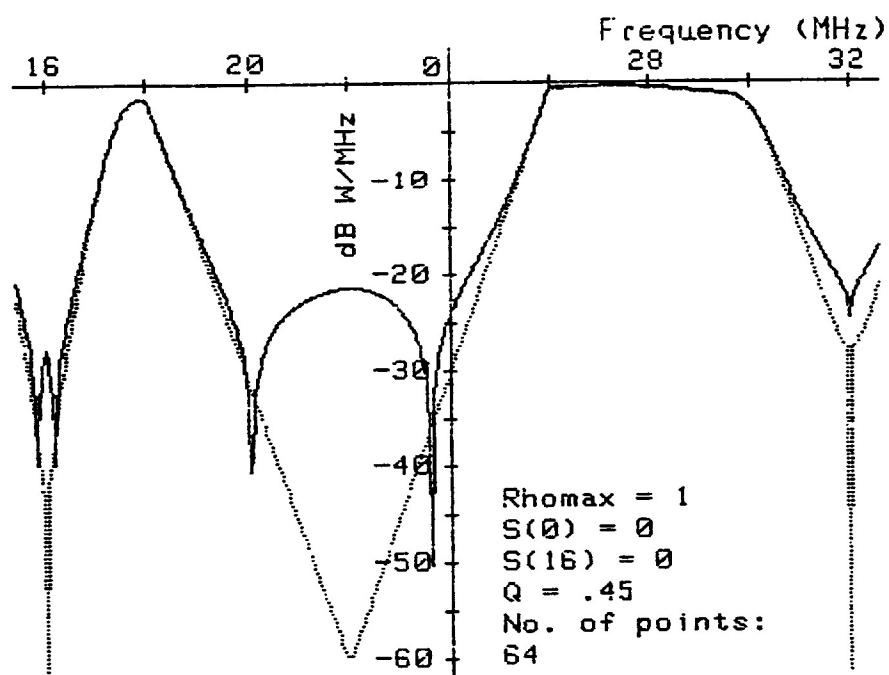


5(a).

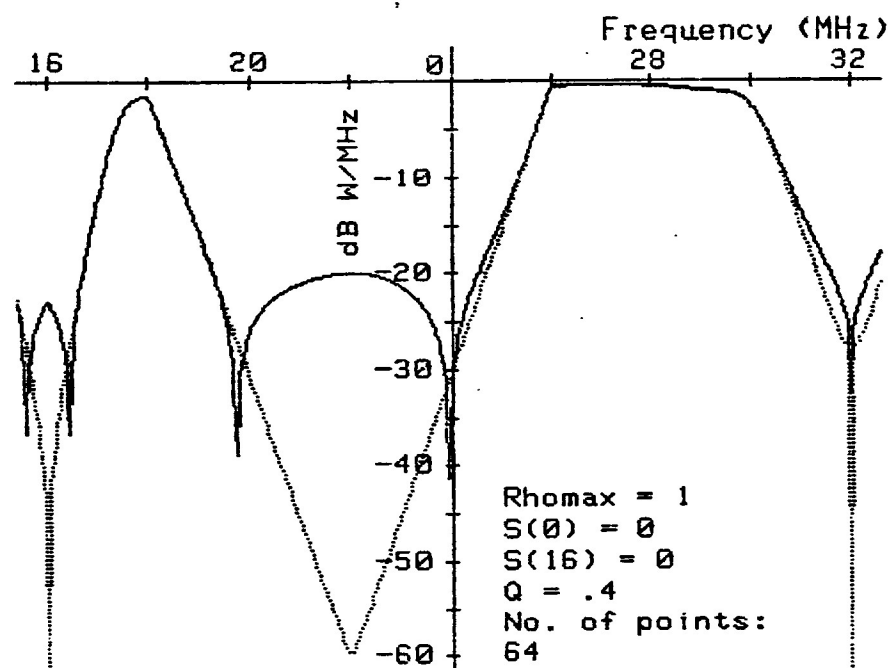
Figure 5. A Run with Q Varied; Rhomax = 1; Two-Level.



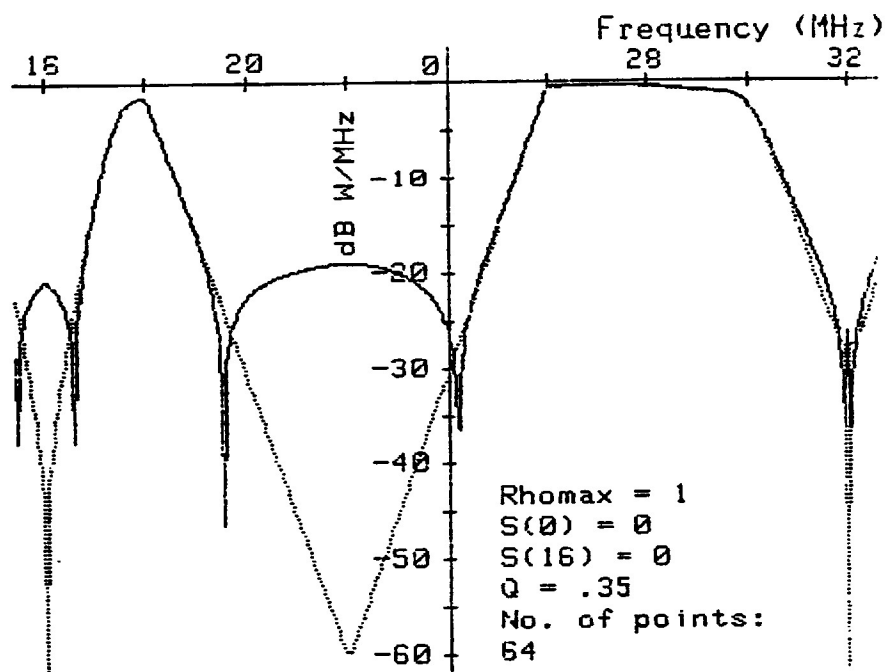
5(b).



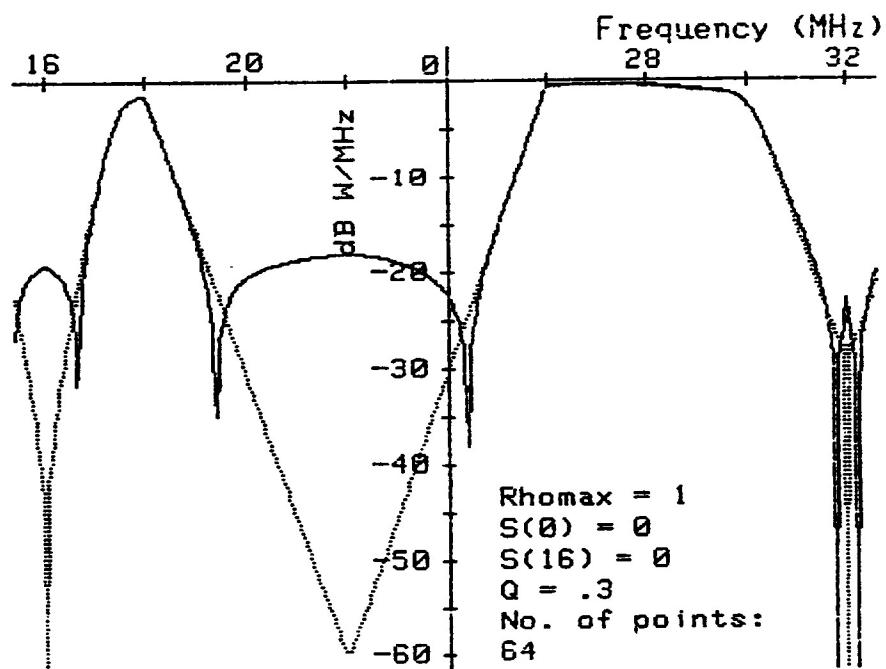
5(c).



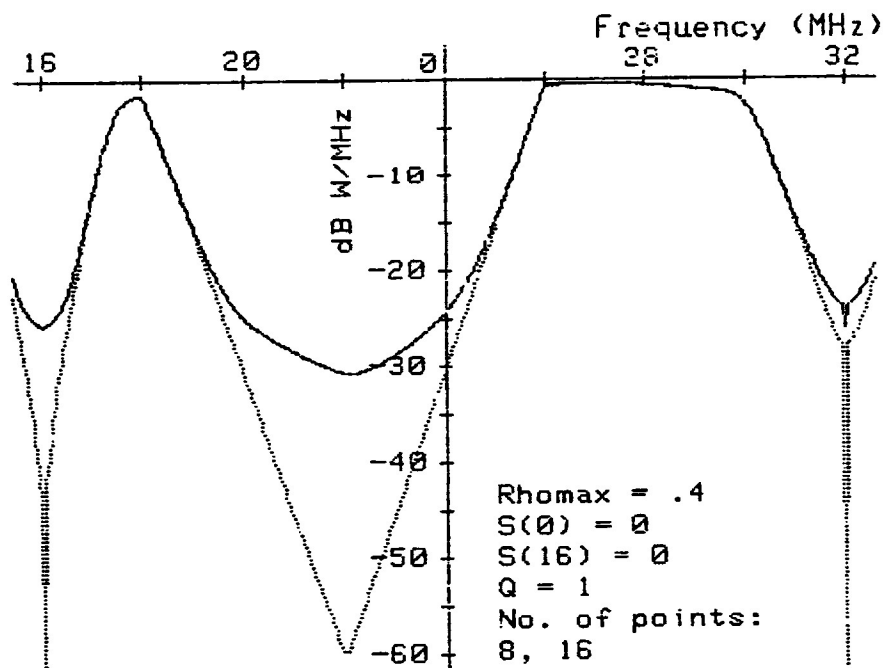
5(d).



5(e).

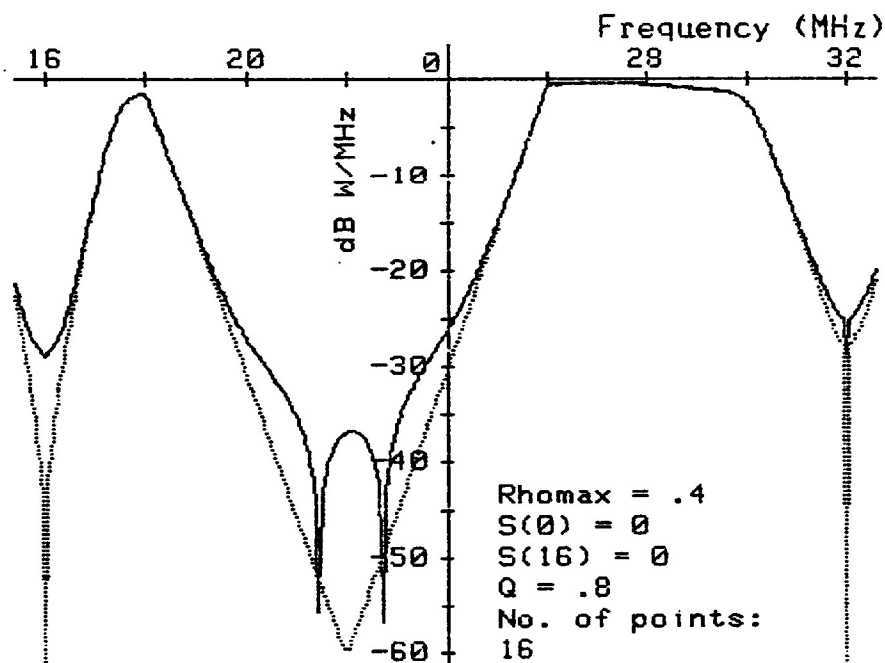


5(f).

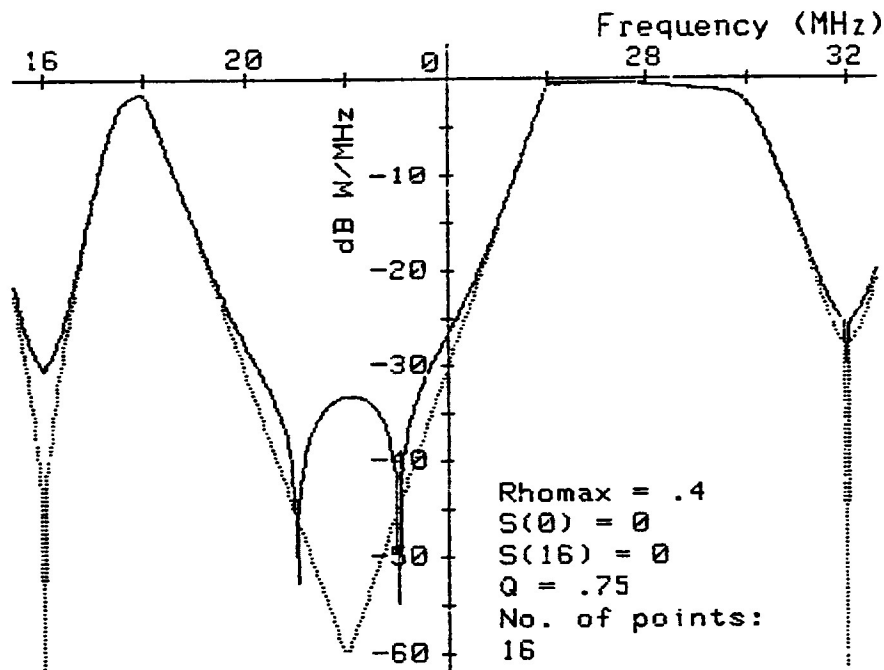


6(a).

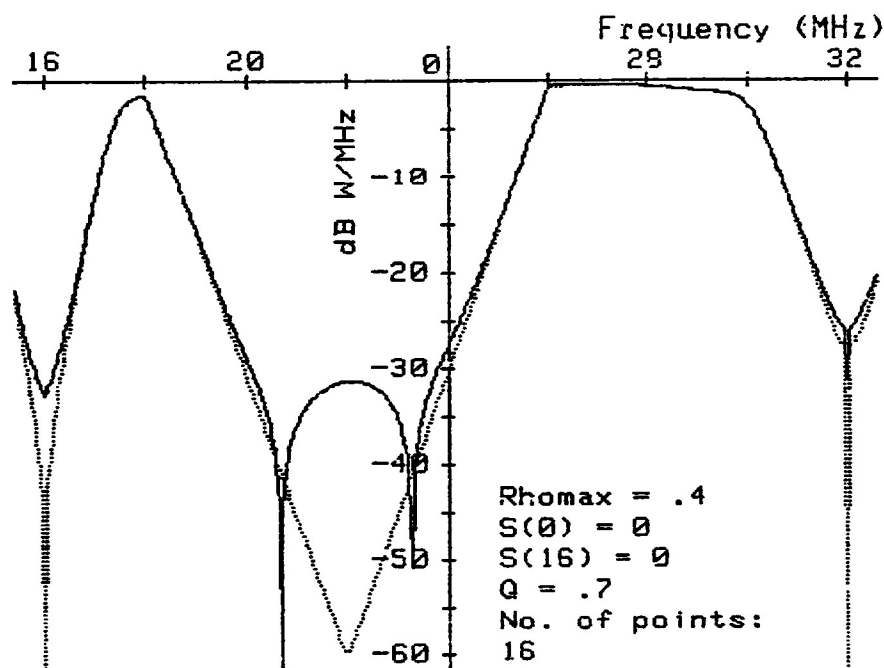
Figure 6. A Run with Q Varied; Rhomax = 0.4; Two-Level.



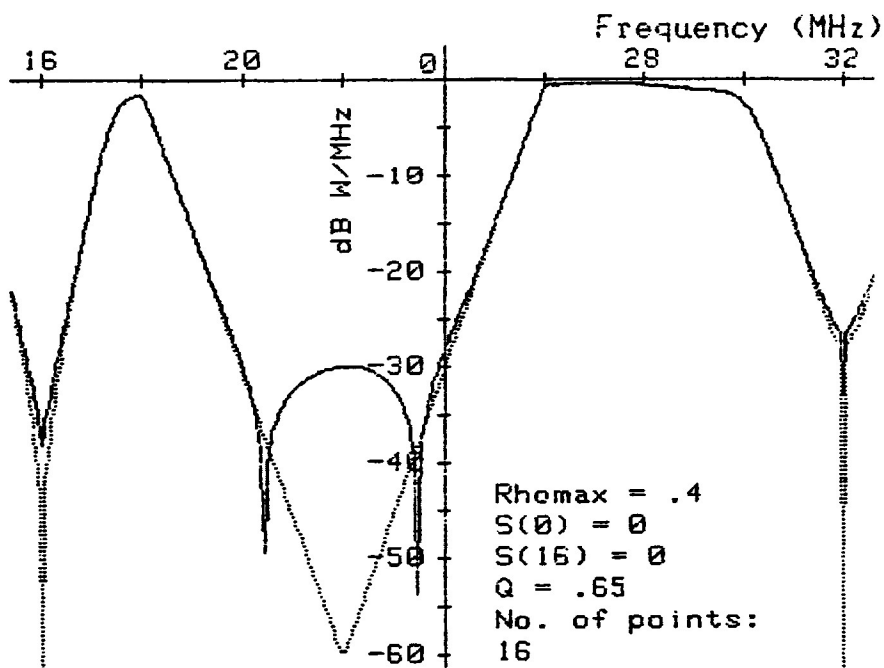
6(b).



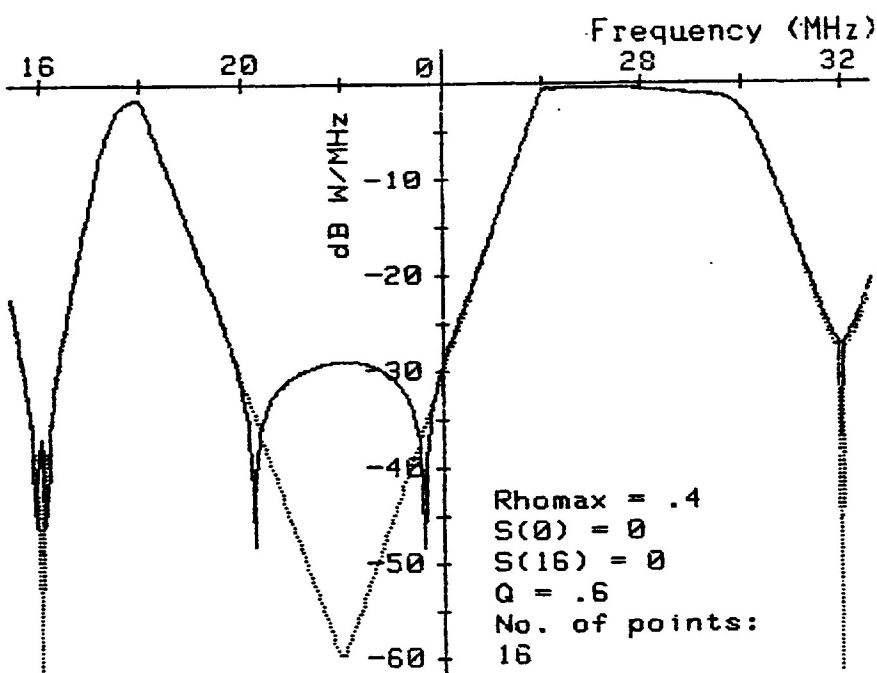
6(c).



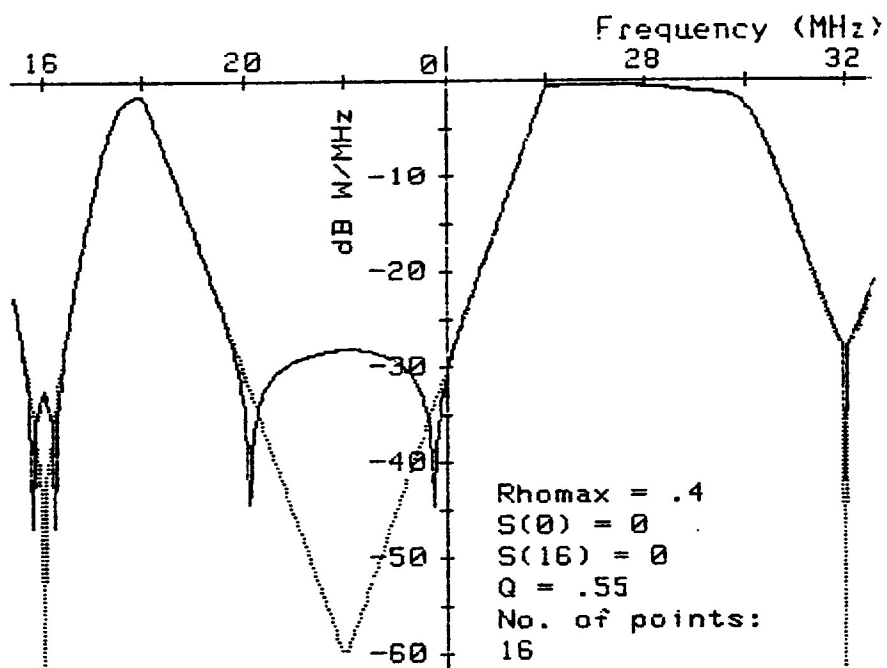
6(d).



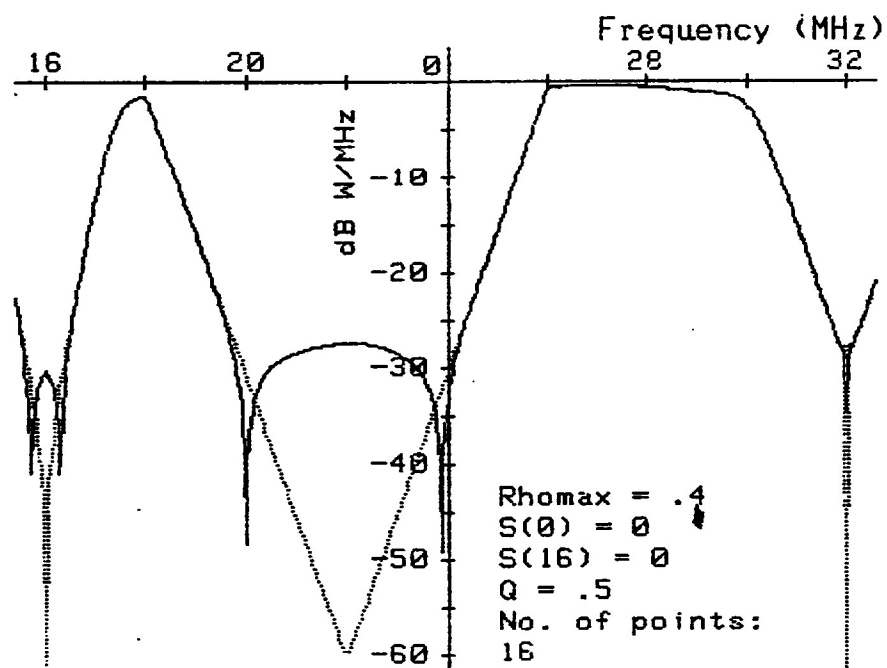
6(e).



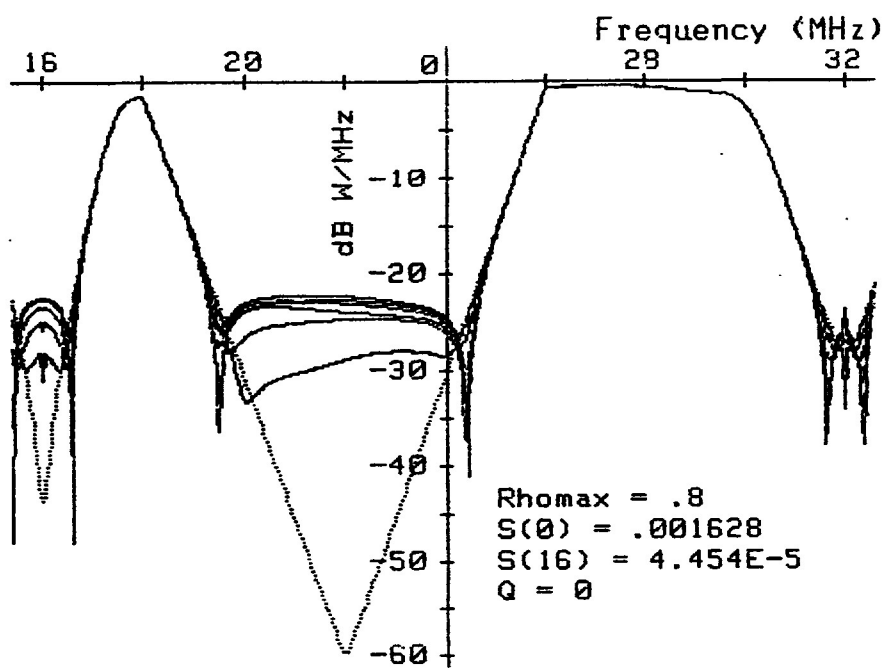
6(f).



6(g).

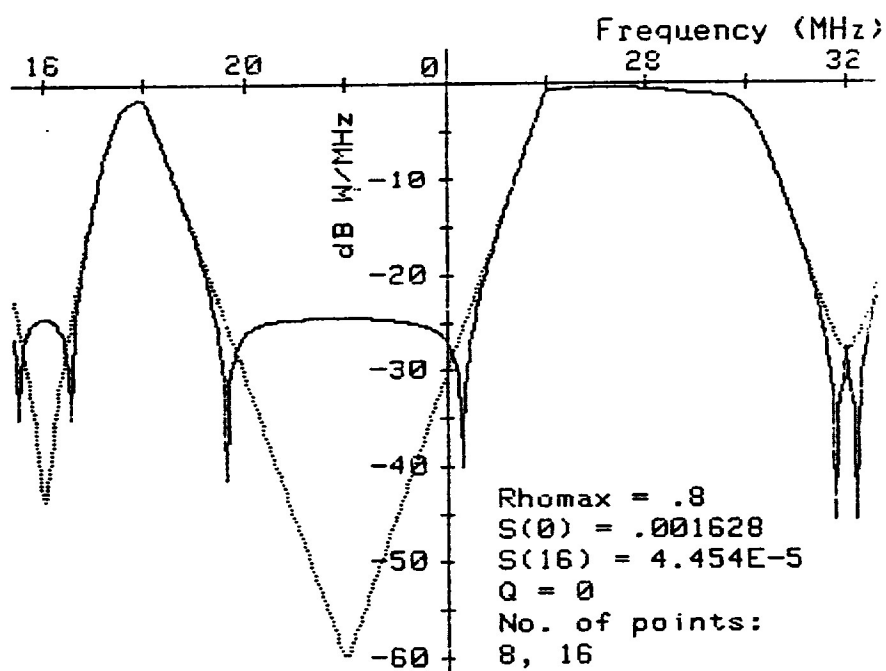


6(h).

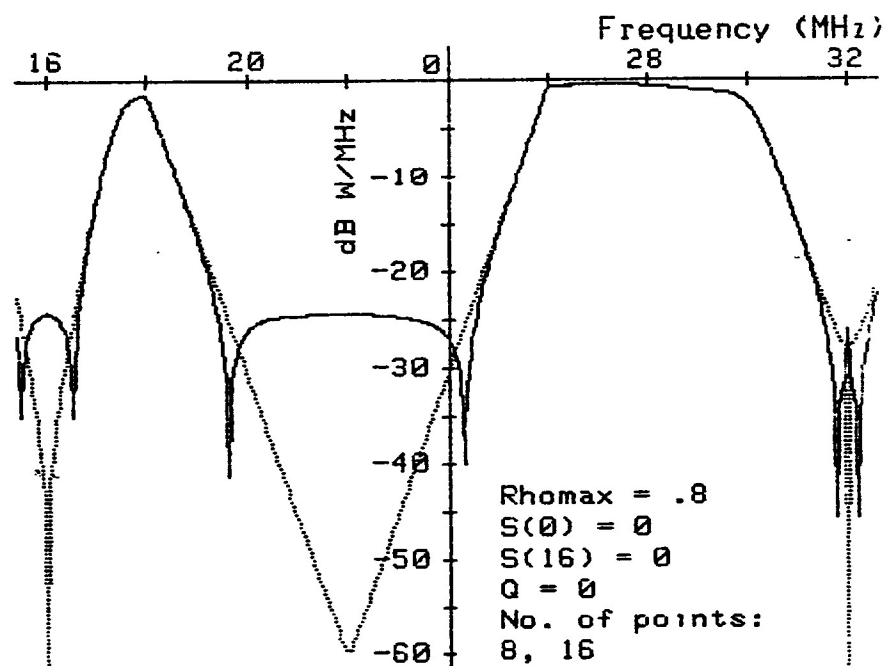


7(a). Spectra for Selected θ .

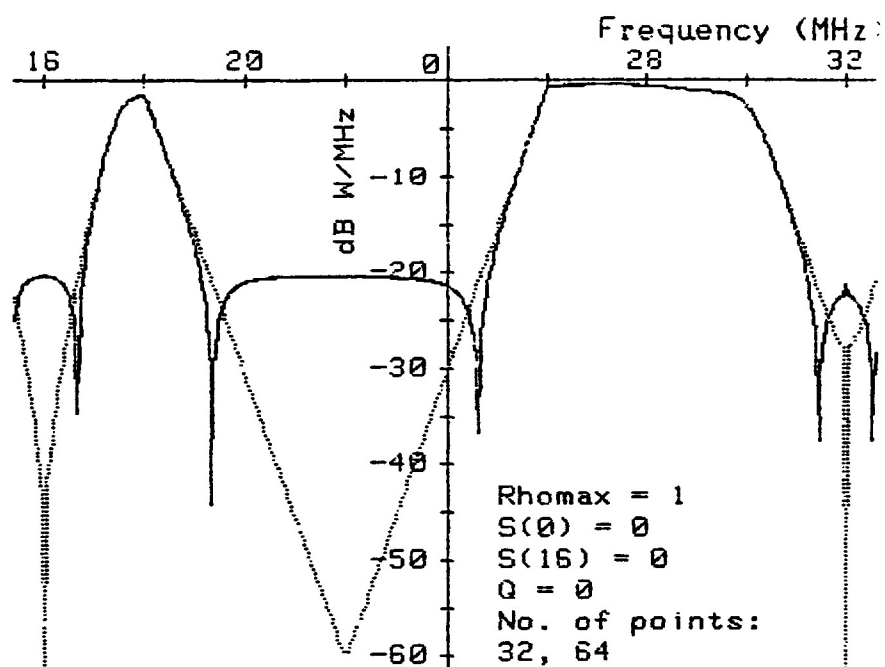
Figure 7. Four-Level Quantization.



7(b). Averages over θ with Various Numbers of Points.

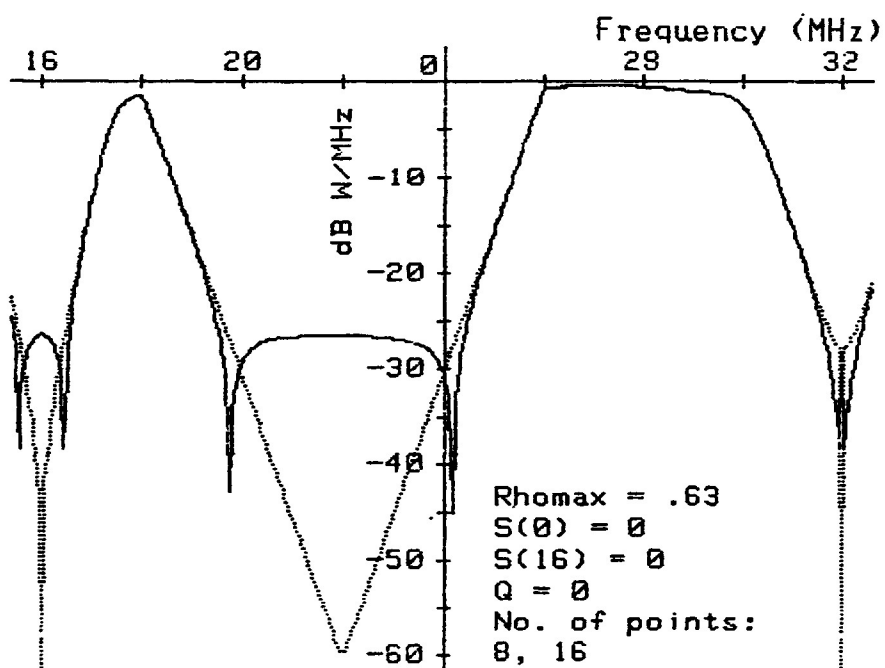


7(c). Setting Target Spectral Components to Zero.

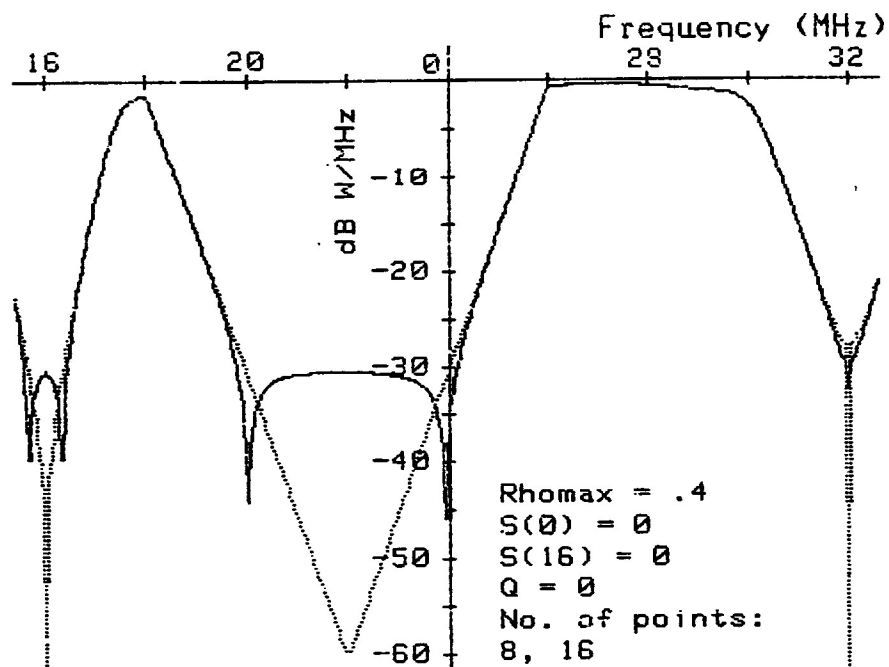


8(a).

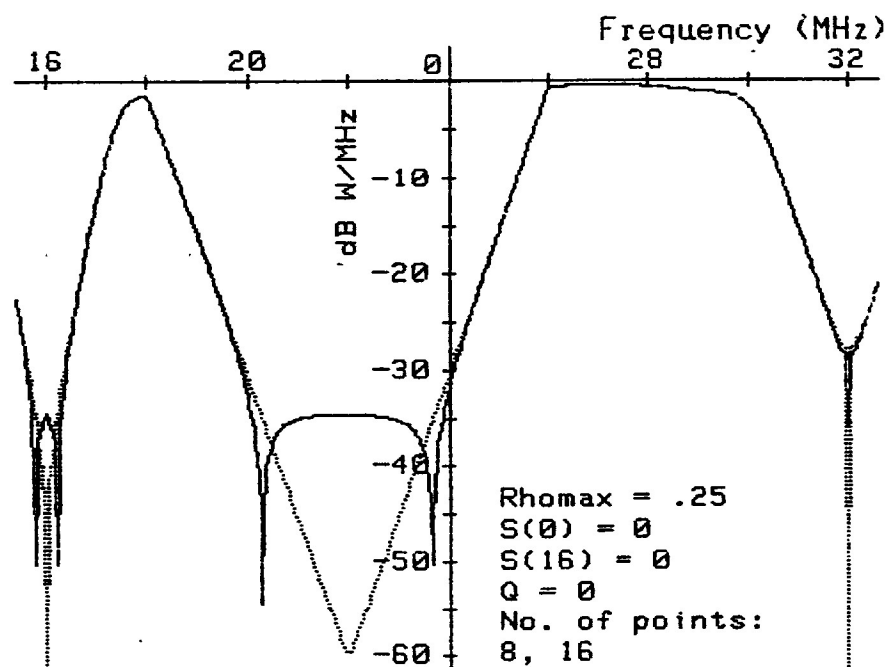
Figure 8. A Run with Rhomax Varied; Four-Level.



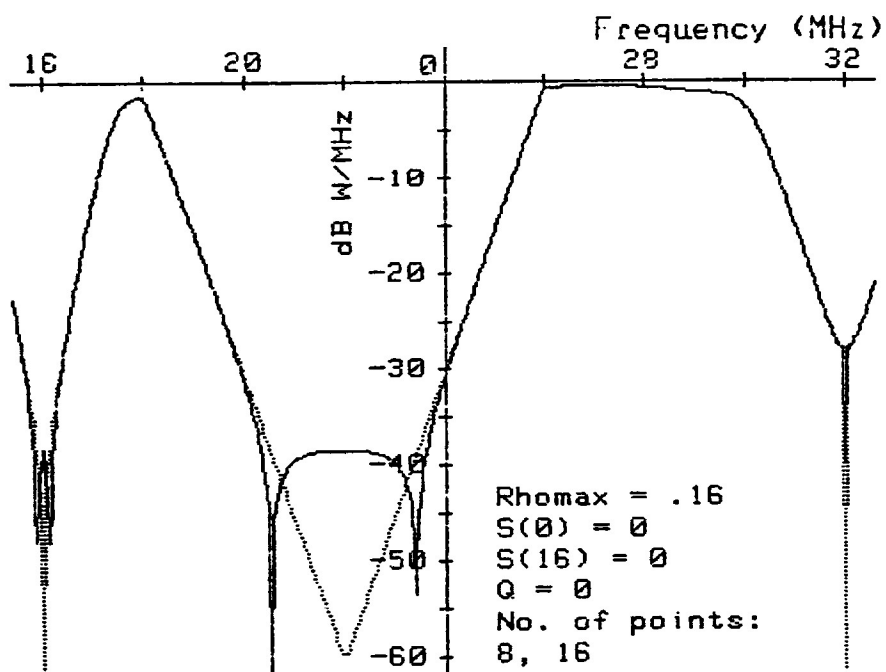
8(b).



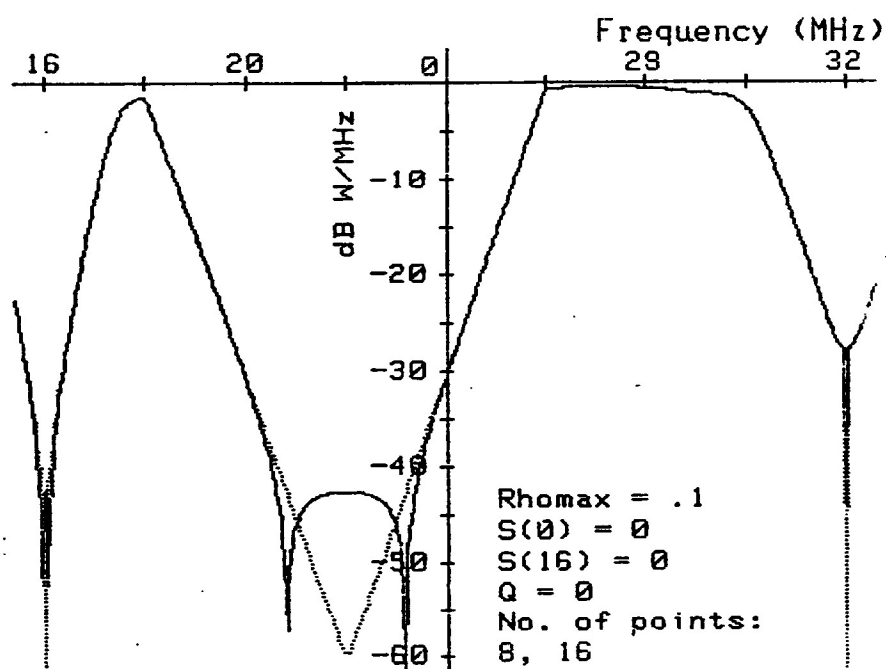
8(c).



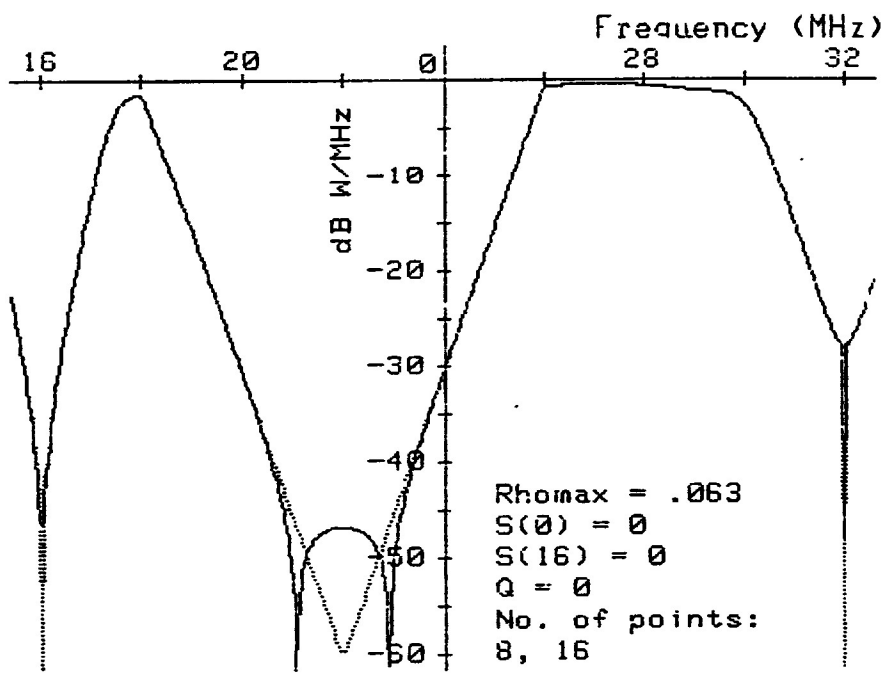
8(d).



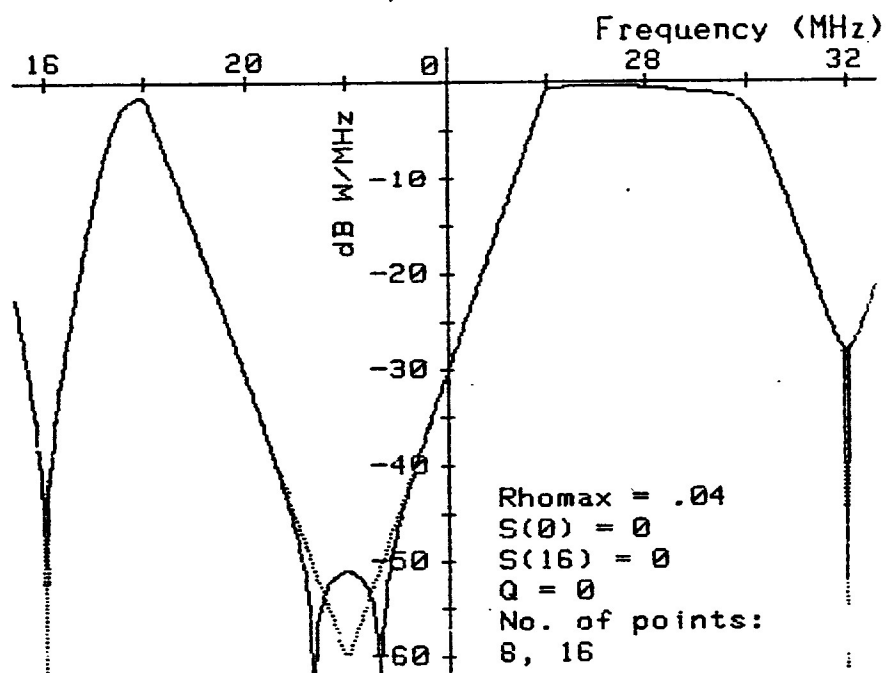
8(e).



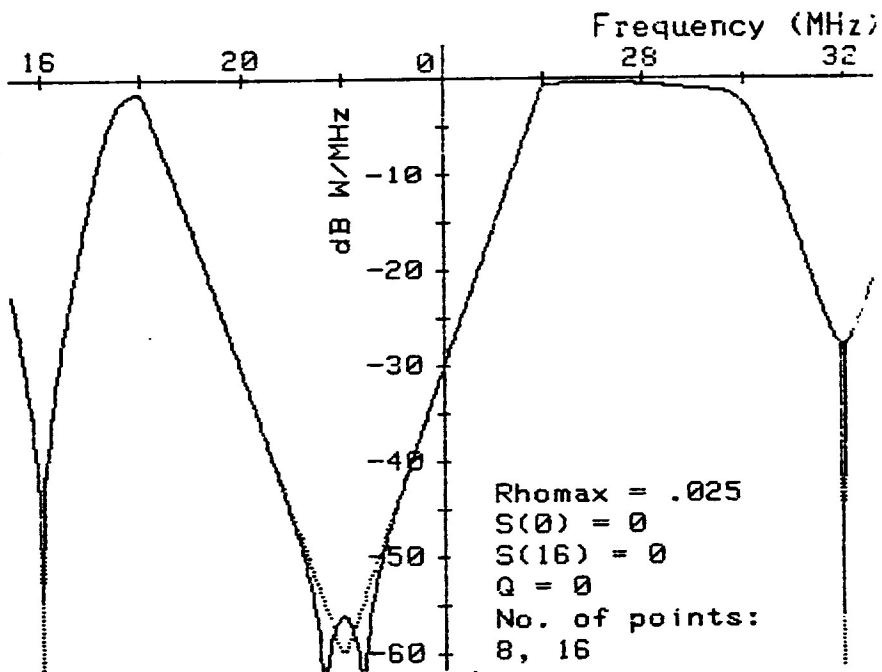
8(f).



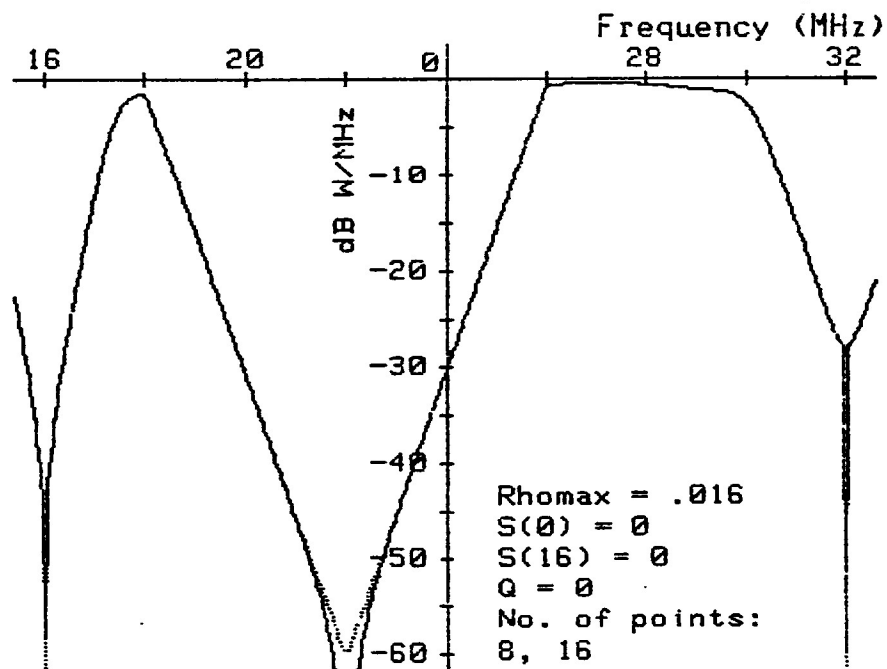
8(g).



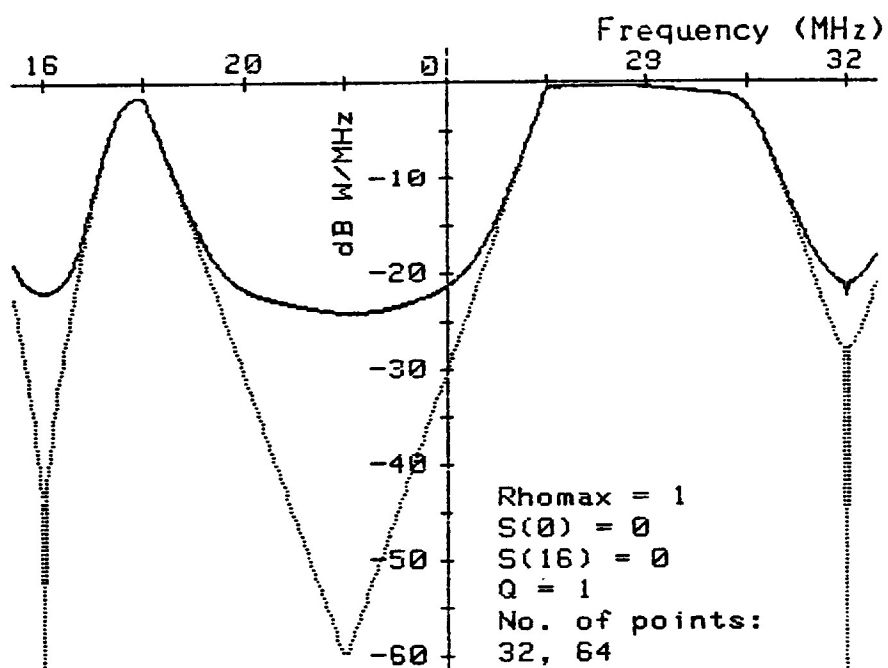
8(h).



8(i).

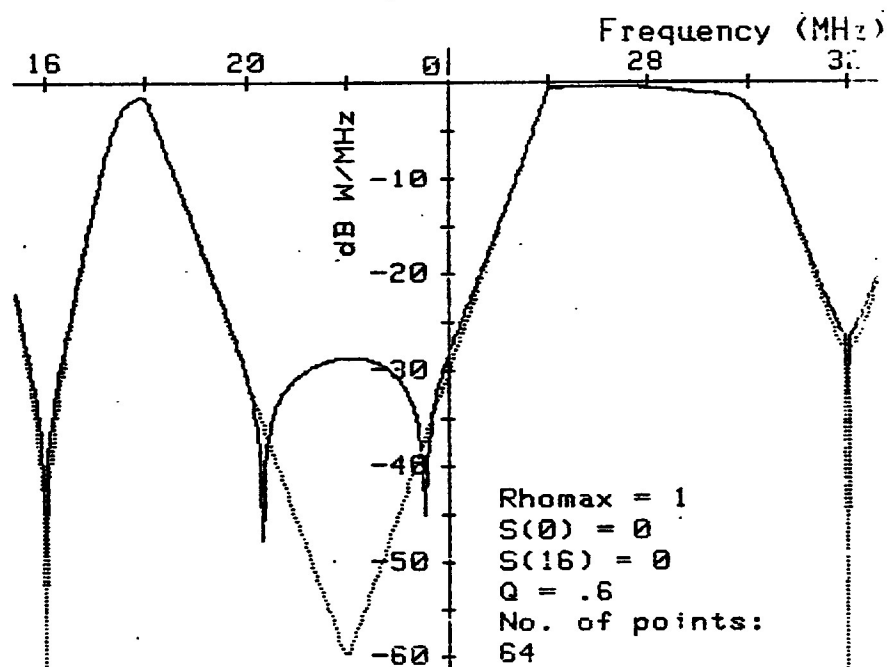


8(j).

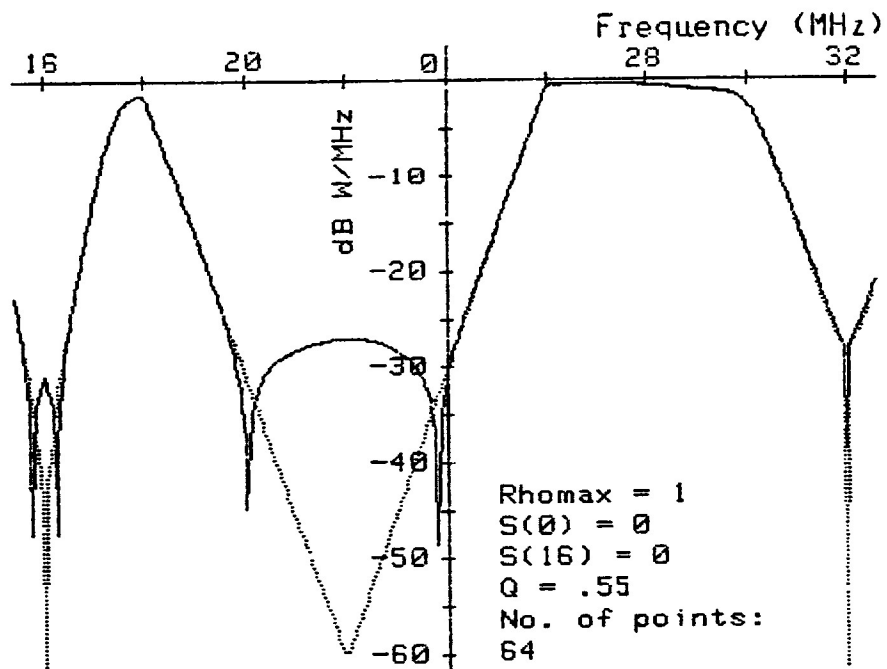


9(a).

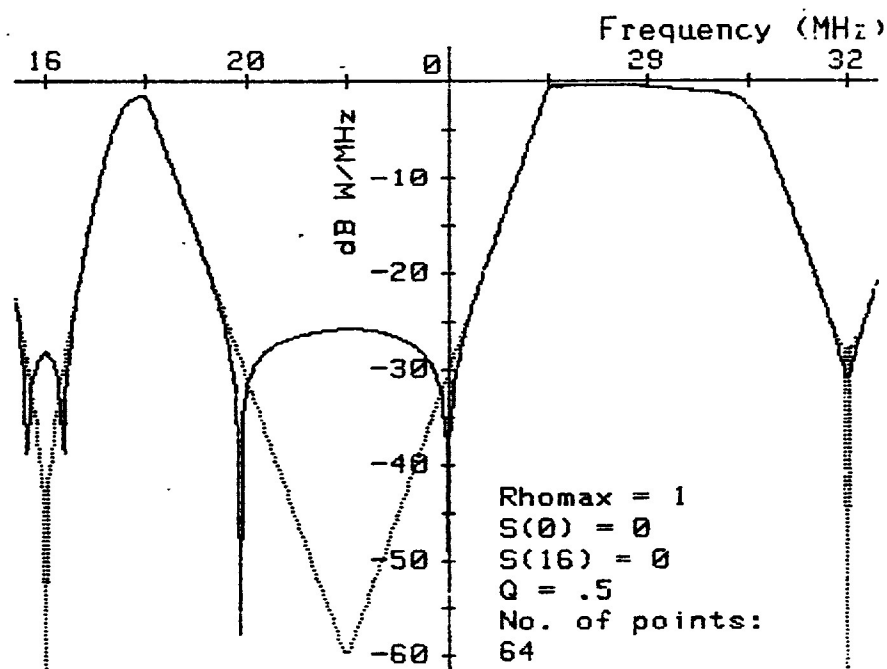
Figure 9. A Run with Q Varied; Rhomax = 1; Four-Level.



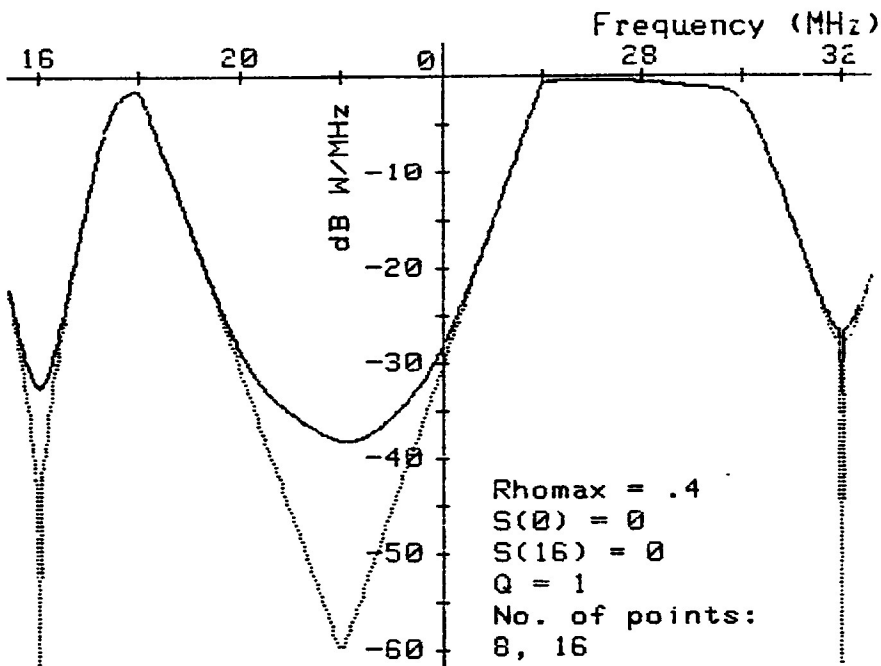
9(b).



9(c).

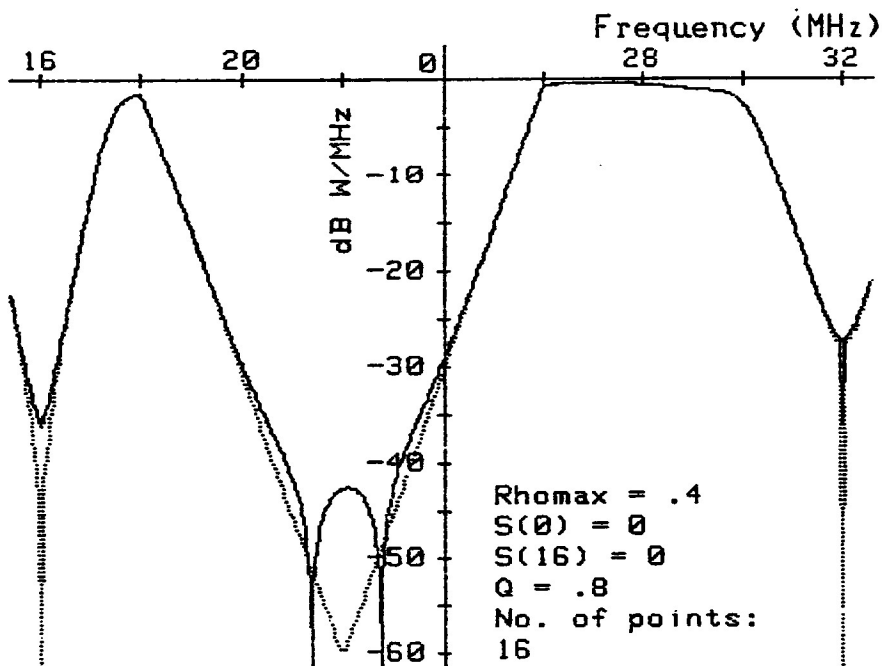


9(d).

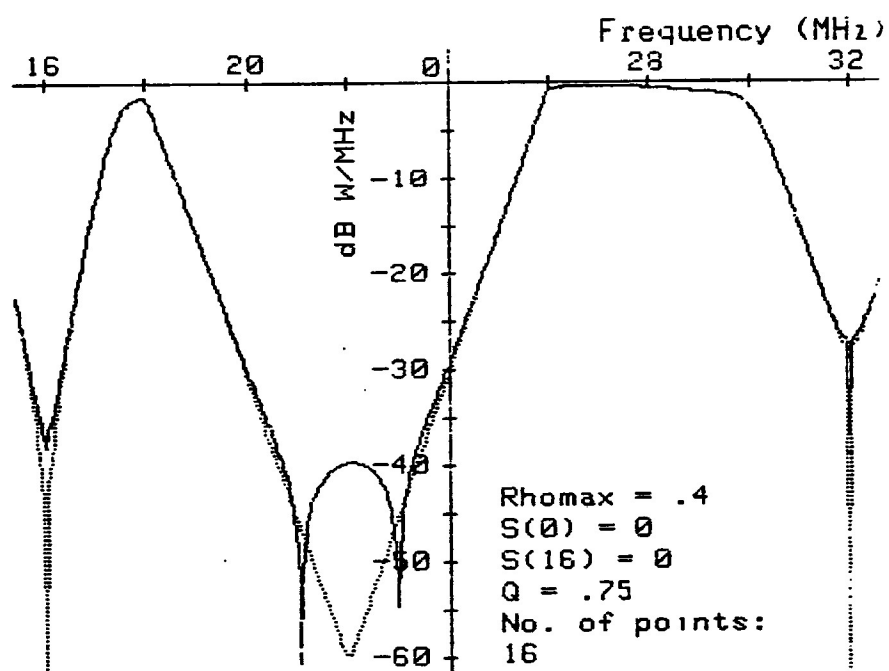


10(a).

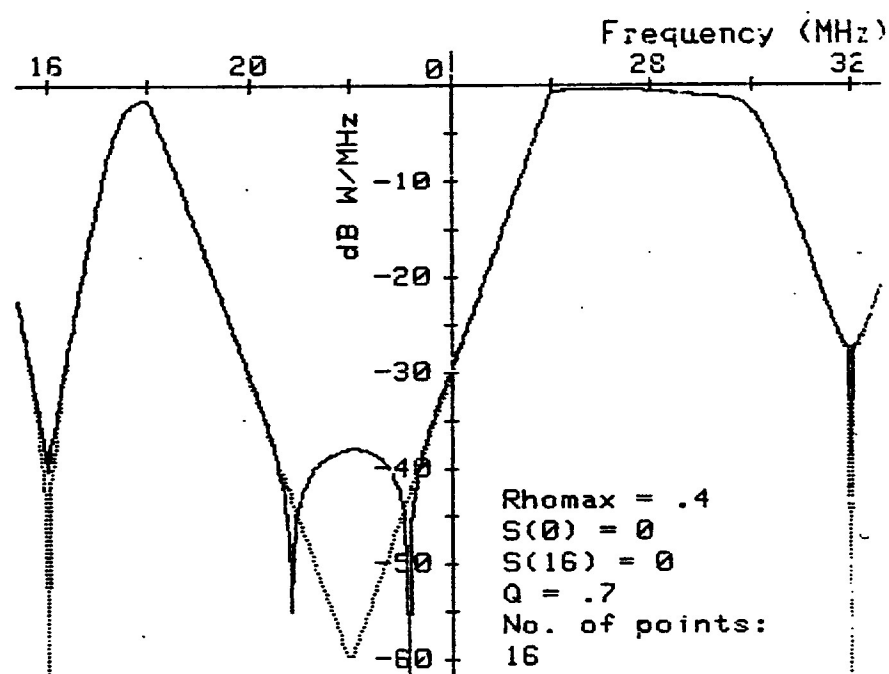
Figure 10. A Run with Q Varied; Rhomax = 0.4; Four-Level.



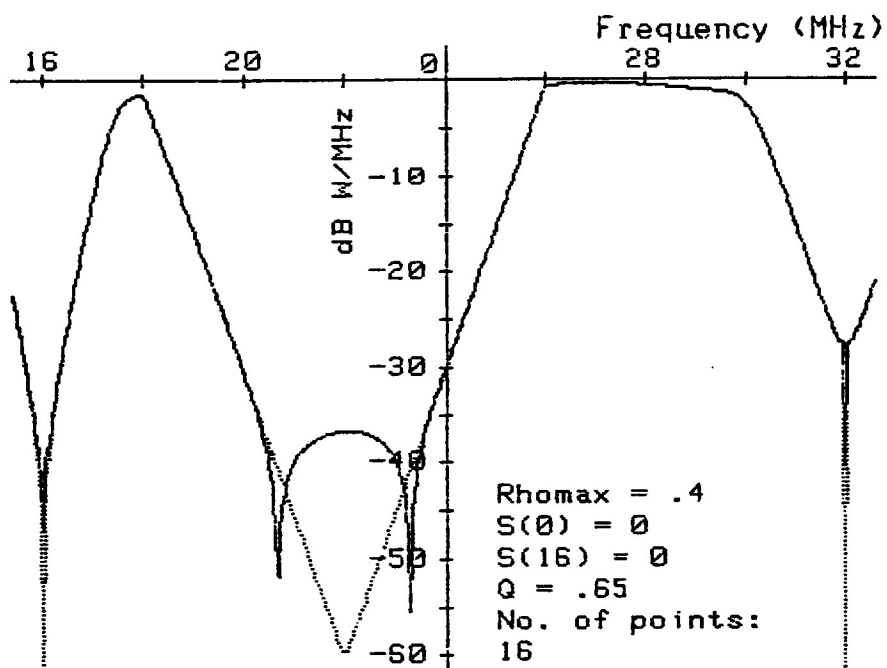
10(b).



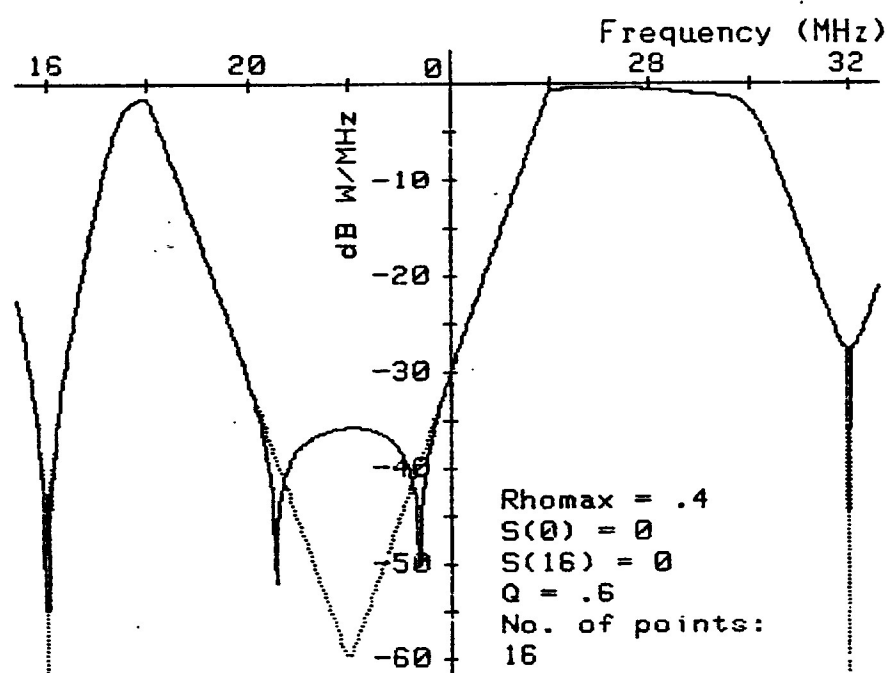
10(c).



10(d).



10(e).



10(f).

40 4970

SEMPER PARATUS
KEUFFEL & ESSER CO. MADE IN U.S.A.

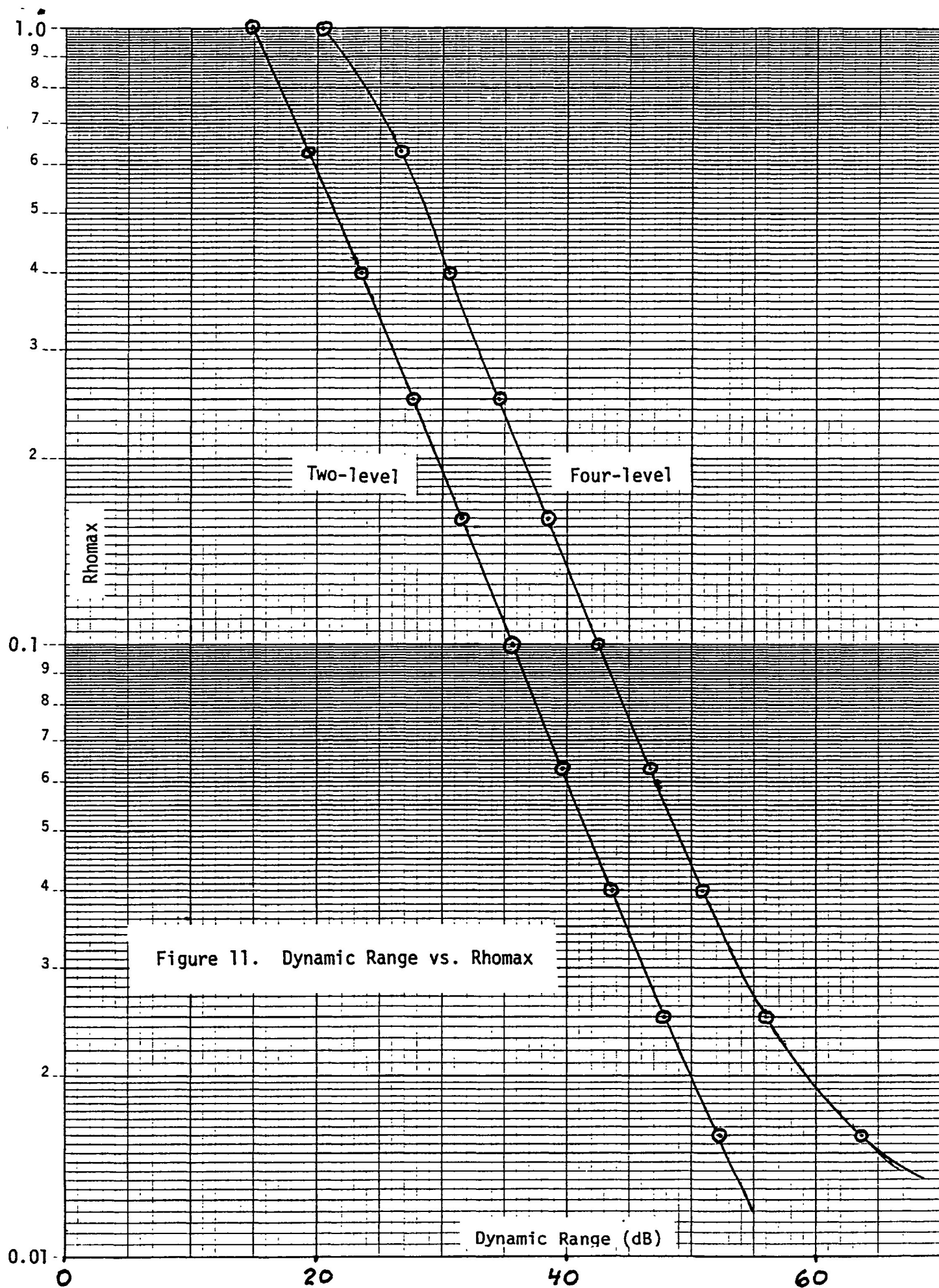
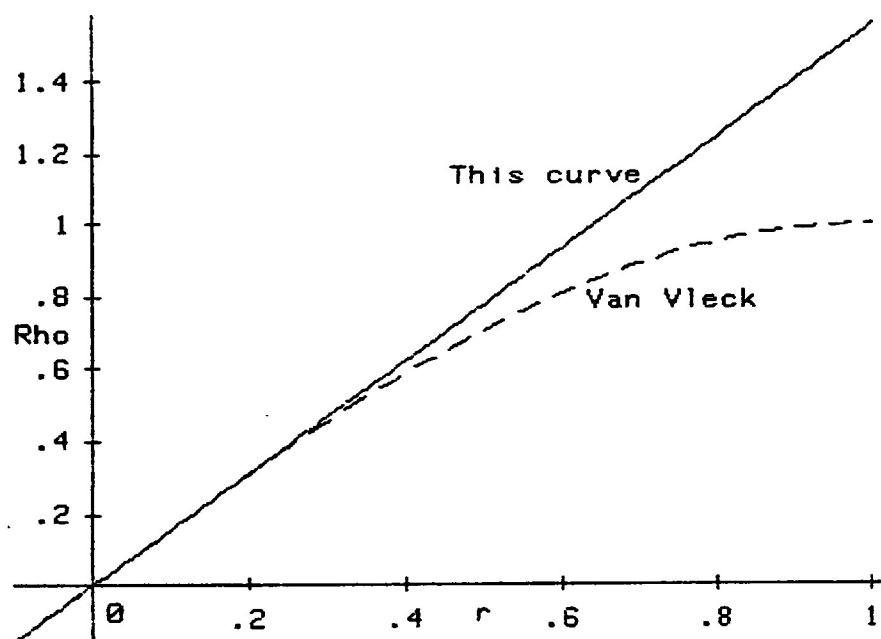
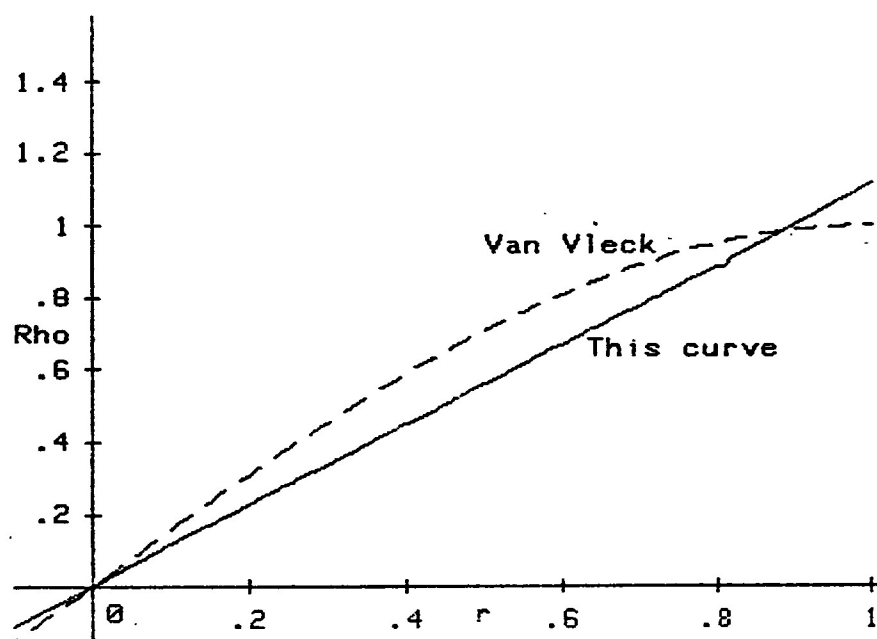


Figure 11. Dynamic Range vs. Rhomax



12(a). $\delta = 10^{-8}$

Figure 12. Optimum Correction Curves.



12(b). $\delta = 10^{-12}$