An Analysis of Nonlinear Behavior in Delta-Sigma Modulators

by
S.H. Ardalan
and
J.J. Paulos

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University

December 1986

CCSP-TR-86/15

This paper is a major extention of TR-86/6, "Stability Analysis of High-Order Sigma-Delta Modulators," to sinusoidal excitation and the calculation of signal-to-noise ratios. It also supercedes TR-86/6.

An Analysis of Nonlinear Behavior in Delta-Sigma Modulators

Sasan H. Ardalan and John J. Paulos

Center for Communications and Signal Processing
Dept. of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695-7914
Tel (919)737-2336

Abstract

This paper introduces a new method of analysis for Delta-Sigma Modulators based on modeling the nonlinear quantizer with a linerized gain, obtained by minimizing a mean-square-error criterion[7], followed by an additive noise source representing distortion components. In the paper, input-signal amplitude dependencies of Delta-Sigma Modulator stability and signal to noise ratio are analyzed. It is shown that due to the nonlinearity of the quantizer, the signal-to-noise ratio of the modulator may decrease as the input amplitude increases prior to saturation. Also, a stable third-order Delta-Sigma Modulator may become unstable by increasing the input amplitude beyond a certain threshold. Both of these phenomenon are explained by the nonlinear analysis of this paper. The analysis is carried out for both DC and sinusoidal excitations.

Table of Contents

1.1 Introduction

Recently, Delta-Sigma Modulation has been receiving increased attention as an alternative to conventional A/D converters [1,2,3,9]. Delta-Sigma is one of a class of systems which use oversampling and 1-bit quantization to achieve high resolution A/D conversion at a lower rate. Oversampling is attractive for many systems in that the analog anti-alias filtering requirements are relaxed. In addition, Delta-Sigma Modulators can generally be implemented with few precision circuits.

A scanning of the literature on Delta-Sigma Modulation reveals a lack of comprehensive analyses of the modulator which fully account for the nonlinearity of the system. An exception is the fundamental work presented in [4] which clearly illustrates the influence of nonlinear phenomenon on circuit behavior for first order modulators and DC inputs. Thus, very few analytical design aids, which predict nonlinear behavior, can be found for Delta-Sigma Modulation. Many researchers have observed experimental phenomenon for which, to date, no analytical explanations exists. Some examples include: the instability of high-order modulators as the input signal amplitude is increased, the actual decline of signal-to-noise ratio as the input amplitude is increased prior to saturation (see the experimental results of [5] and [12]), and the observation reported in [5] in which very high internal signal variances are encountered for large input amplitudes.

In this paper, higher-order Delta-Sigma Modulators are analyzed. These modulators are preferred because of their superior noise performance and relative freedom from harmonic quantization effects such as noise thresholding [4,5]. These higher-order systems can be unstable for certain values of the loop parameters and, in the case of third-order Delta-Sigma Modulators, for certain ranges of the input signal.

This paper introduces a new method of analysis for Delta-Sigma modulators based on modeling the nonlinear quantizer with a linearized gain obtained by minimizing a mean-square-error criterion[7]. This method has been used to derive regions of stability for higher-order modulators, including both parameter and signal dependencies. In addition, the method can be used to obtain more accurate solutions for quantization noise spectra and signal-to-noise ratios for several classes of input signals. In this paper, we first consider the case where the input signal is a DC amplitude. The approach is then extended to sinusoidal signals.

1.2 Background on Methodology

The quasilinear method for modeling the nonlinearity in nonlinear feedback systems was first introduced by Booton [13]. This technique, known as the describing function method, has been widely used and analyzed [7]. This method, however, is based on neglecting distortion components produced by the nonlinearity. The assumption here is that these components are filtered as they are fed back to the nonlinearity input. While this assumption yields adequate results in many nonlinear feedback systems, it will be shown that it is inappropriate for the analysis of Delta-Sigma Modulators.

The method for modeling the nonlinearity presented in this paper is thus distinguished from the describing function technique in that the distortion components produced by the nonlinearity are not neglected. This paper will show that the distortion components affect the dynamic behavior of the modulator dramatically. This includes instability of higher-order modulators and the increase in baseband noise as the input signal amplitude is increased.

In [8], Smith reviewed the different approaches for including the distortion components in the modeling of the nonlinearity. One method consisted of modeling the nonlinearity by a quasilinear gain followed by an additive noise source representing the distortion components produced by the nonlinear element. For example, using this model West et. al. [6] were able to predict

baseband distortion in a nonlinear feedback system. Since the modulator response is now seen to consist of both a signal related to the input to the modulator and random distortion components, the response of the nonlinearity to multiple inputs must be considered. To this end, the method of multiple linearized gains for multiple inputs has been used. This technique has been extensively studied in [7]. Using this method, a nonlinear feedback system can be represented as two interlocked linear systems.

In this paper, unlike the work in [7] where the multiple inputs are applied externally, the random component is produced internally by the nonlinearity in the modulator. Smith has presented experimental results which show that under certain conditions the random distortion components appearing at the nonlinearity input may have a Gaussian probability density function (pdf). In the present work, this observation is exploited for the case of higher-order modulators. Thus, the nonlinearity is modeled as two linearized gains, one for the input signal to the system and one for the random distortion components which are assumed to have a Gaussian pdf. The nonlinear system is then analyzed as two interlocked linear systems.

In this paper, with the nonlinear modeling approach outlined above as a basis, a novel approach for analyzing Delta-Sigma Modulators is introduced. Using this approach, closed form analytical expressions are obtained which relate, for example, the additive quantization noise to the input-signal amplitude. Similarly, expressions are derived which relate the ratio of the amplitude of the signal component to the random noise component at the nonlinearity input to the input signal amplitude. Such expressions have not been derived before. Using these expressions, it is then possible to compute the value of the linearized gains as a function of input amplitude by solving a set of simultaneous nonlinear and integral equations. From the gains, the signal-to-noise ratio of the modulator and other system parameters including the variance and magnitude of various signals within the modulator can be calculated.

2.Delta-Sigma Modulation

A simple representation of a first-order Delta-Sigma Modulator is shown in Fig. 1. This circuit can be implemented with a differential integrator, a comparator, and a flip-flop or sample-and-hold amplifier. The output of this system is a bit stream whose pulse density is proportional to the applied input signal amplitude. (A more conventional digital representation of the input signal can be obtained by decimation and baseband filtering of the pulse stream.) In previous work [4,5,3], this system has been modeled by replacing the nonlinear element with a unity-gain linear element followed by an additive noise process. This simple model has proven to be inadequate for the accurate analysis of higher-order modulator stability since it does not reflect the dependence of the nonlinear system on the input signal to the nonlinearity.

A second-order double-loop modulator is shown in Fig. 2. In order to analyze the modulator, fictitious sample-and-hold circuits are inserted into the circuit as shown in Fig. 3. It is possible now to obtain a z-domain block diagram of the second order modulator as shown in Fig. 4. Since the sample-and-hold circuit and the integrator in the inner loop are cascaded with no sampler in between, they must be combined prior to obtaining the z-Transform. Thus,

$$H_2(z) = Z\left[\frac{1-e^{-T_s s}}{s}\frac{1}{s}\right] = (1-z^{-1})Z\left[\frac{1}{s^2}\right] = \frac{T_s}{z-1}. \tag{2.1}$$

The z-Transform of a simple integrator becomes,

$$H_1(z) = \frac{z}{z-1}. \tag{2.2}$$

It can be shown that any continuous-time or multi-loop system can be represented in the z-domain

as shown in Fig. 5, with the appropriate choice of the loop filters $H_i(z)$ and $H(Z)$. The z-domain representation, however, is an approximation to the actual behavior of the continuous system. For a second-order loop we have,

$$H_i(x) = \frac{\alpha_1 H_1(z)}{1 + \alpha_1 H_1(z)} \tag{2.3}$$

and,

$$H(z) = [1 + \alpha_1 H_1(z)]\alpha_2 H_2(z). \tag{2.4}$$

3.Modeling of the Nonlinear Quantizer

3.1 DC Input

Consider the random signal z(t) with mean value m_x and zero mean random component y(t).

$$z(t) = m_x + y(t) \tag{3.1}$$

Consider the nonlinear mapping of $z(t)$ into $u(t)$,

$$u(t) = N(y(t), m_x) \tag{3.2}$$

With reference to Fig. 6, we associate the linearized gain K_y with the zero mean random component y(t) and K_x with the DC mean value or offset of the signal z(t). In Fig. 6, x(t) is a DC signal equal to m_x in this case. The identification problem is to determine K_x and K_y such that the mean square error in modeling the nonlinear element using the linearized gains is minimized. Thus we must minimize,

$$\sigma_e^2 = E\{e^2(t)\} = E\{[u(t) - K_y y(t) - K_x m_x]^2\} \tag{3.3}$$

Taking the partial derivatives and setting them to zero we obtain,

$$\frac{\partial \sigma_e^2}{\partial K_y} = -2E\{u(t)y(t)\} + 2K_y E\{y^2(t)\} = 0 \tag{3.4}$$

$$\frac{\partial \sigma_e^2}{\partial K_x} = 2K_x m_x^2 - 2m_x E\{u(t)\} = 0 \tag{3.5}$$

which yield,

$$K_y = \frac{E\{u(t)y(t)\}}{E\{y^2(t)\}} \tag{3.6}$$

$$K_x = \frac{E\{u(t)\}}{m_x} \tag{3.7}$$

In terms of the probability density function of y(t) the results are

$$K_y = \frac{1}{\sigma_y^2} \int_{-\infty}^{\infty} yN(y + m_x)p(y)dy \tag{3.8}$$

$$K_x = \frac{1}{m_x}\int_{-\infty}^{\infty} N(y+m_x)p(y)dy \tag{3.9}$$

An important consequence of using the above linear gains is that the error, e(t), becomes uncorrelated with y(t).

$$E\{y(t)e(t)\} = E\{y(t)u(t)\} - K_y E\{y^2(t)\} - K_x m_x E\{y(t)\} = 0 \tag{3.10}$$

In the application of the above modeling technique to the field of nonlinear control, the error $e(t)$ is usually neglected. This is based on the assumption that the error is filtered by the plant after feedback and forms a negligible part of the input signal to the nonlinearity [7]. For this reason, y(t) is usually assumed to be dependent only on the input to the control system. In Delta-Sigma Modulation, however, the nonlinearity introduces spectral components which cover a wide bandwidth, including the baseband. In this case, the error, e(t), represents the noise due to quantization which forms a major component of the modulator pulse stream. Furthermore, in many cases in nonlinear control, the output of the nonlinearity is the input to the plant. Hence, it is substantially filtered before feedback to the nonlinearity input. In contrast, the output of the nonlinear quantizer is the desired modulated pulse stream in Delta-Sigma Modulation which is *is directly fed back* and subtracted from the input signal. For this reason, we must include the error term e(t) in the modeling of the nonlinearity. We will show that the quantization noise has a major impact on stability for high order modulators.

In this paper we will assume that e(t) has a white spectrum. The nonlinearity, in particular a one-bit quantizer, produces harmonics of the input signal across the spectrum of interest including base-band. Therefore, this assumption seems partly justified [8]. Thus we replace the nonlinear quantizer in the modulator by the two linearized gains followed by an additive noise source representing the error. This is illustrated in Fig. 7 and Fig. 8, where we have separated the response due to the zero mean random component and the DC response. We are assuming that the input to the modulator is a DC level equal to m_x. From the Figures, we have for the steady state DC response,

$$m_e = \frac{H_i(1)H(1)}{1+H(1)K_x}m_x \tag{3.11}$$

and for the response to the random noise component,

$$E_n(z) = N(z)\frac{H(z)}{1+H(z)K_n} \tag{3.12}$$

If we assume that n(k) is white with variance σ_n^2 then,

$$\sigma_e^2 = \frac{\sigma_n^2}{2\pi}\int_{-\pi}^{\pi} \frac{|H(e^{j\omega})|^2}{1 + K_n[H(e^{j\omega}) + H^*(e^{j\omega})] + K_n^2|H(e^{j\omega})|^2}d\omega \tag{3.13}$$

In most applications of Delta-Sigma Modulation the forward path includes an integrator. Therefore, $H(1)\to\infty$ and,

$$m_e \to \frac{H_i(1)}{K_x}m_x \tag{3.14}$$

Although we make no assumptions on the probability density distribution of the noise $n(k)$, the signal into the nonlinearity due to the noise, $e_n(k)$, is a twice or triple integrated version of the noise $n(k)$ for second and third-order loops. Since integration of a random variable tends to

make its distribution approach a Gaussian distribution, we will assume a Gaussian distribution for $e_n(t)$. Substantial errors may result if this assumption is not true [8]. The linearized gains can be calculated based on (3.7) and (3.8) with a Gaussian distribution assumed for p(y).

$$K_n = \frac{2\Delta}{\sigma_e\sqrt{2\pi}} e^{-\frac{m_e^2}{2\sigma_e^2}} \tag{3.15}$$

$$K_x = \frac{\Delta}{m_e} erf\left(\frac{m_e}{\sigma_e\sqrt{2}}\right) \tag{3.16}$$

In the above expressions, the nonlinearity is assumed to be a one-bit quantizer with an output of +/- Δ. Substituting for K_x in (3.13) we obtain,

$$m_e = m_x \frac{m_e}{\Delta} erf\left(\frac{m_e}{\sigma_e\sqrt{2}}\right) \tag{3.17}$$

where we have assumed $H_i(1) \rightarrow 1$ if an integrator is used. Thus,

$$erf\left(\frac{m_e}{\sigma_e\sqrt{2}}\right) = \frac{m_x}{\Delta} \tag{3.18}$$

Define,

$$\rho = \frac{m_e}{\sigma_e\sqrt{2}}$$

Then,

$$\rho = erf^{-1}\left(\frac{m_x}{\Delta}\right) \tag{3.19}$$

The above results show that ρ is independent of the loop gains and is directly related to the DC input amplitude m_x. From Fig. 7 and Fig. 8, the Delta-Sigma pulse stream can be written as,

$$p(k) = e(k)K_n + n(k) + m_e K_x \tag{3.19}$$

Now, since p(k) fluctuates between the two levels $-\Delta$ and Δ, its power is constant and equal to Δ^2. Hence,

$$E\{p^2(k)\} = E\{e^2(k)\}K_n^2 + \sigma_n^2 + m_e^2 K_x^2 = \Delta^2 \tag{3.20}$$

where we have used the fact that the crosscorrelation between e(k) and n(k) is zero since we are using the linearized minimum mean square error gains. From (3.20) we can derive the expression for the variance of the modeled additive noise as a function of input DC amplitude,

$$\sigma_n^2 = \Delta^2\left[1 - \frac{m_x^2}{\Delta^2} - \frac{2}{\pi} e^{-2[erf^{-1}(\frac{m_x}{\Delta})]^2}\right] \tag{3.21}$$

This expression shows that the noise variance depends only on the input DC level and is independent of the loop gains. When $m_x \rightarrow \Delta$, we have $\sigma_n^2 \rightarrow 0$.

We now examine the dependence of the gain K_n on the input DC amplitude. From (3.18) we observe that as m_x approaches Δ, ρ becomes very large. Based on (3.14), K_n decreases exponentially with ρ^2. Hence, K_n will decrease with increasing amplitude (although $\sigma_e \rightarrow 0$ as

$m_x \to \Delta$, the exponential term in (3.14) decreases at a faster rate). As the gain K_n decreases with amplitude the noise shaping of the Delta-Sigma Modulator changes. The implications of this non-linear phenomenon on stability and signal to noise ratio will be examined later.

In order to test the theoretical results derived above, a second-order double-loop digital Delta-Sigma Modulator was simulated with a DC input. The mean value m_e and the variance σ_{en}^2 of the zero-mean random component $e_n(k)$ at the input to the nonlinearity were obtained after 1024 iterations. Also, the linear gains K_x and K_n were calculated from the simulation using time averages based on (3.6) and (3.7). From the computed gains, the additive noise, n(k), was obtained from the simulation by subtracting the quantizer output from $e_n(k)K_n + m_e K_x$. The loop gains were, $\alpha_1 = 1.0$ and $\alpha_2 = 1.0$. The quantizer step size was $\Delta = 1.0$. In Fig. 9, the simulated and calculated value of ρ, based on evaluating (3.18) as a function of DC amplitude m_x , are plotted. In Fig. 10, the simulated and calculated value of the additive noise variance, σ_n^2, based on expression (3.25), are plotted. In both cases the theoretical results agree closely with the simulation results.

3.2 Sinusoidal Input and Nonlinearity Modeling

Consider the case where the input to the nonlinear quantizer consists of a sinusoid, x(t), and a Gaussian signal, y(t). Based on the previous discussion, we can associate the linear gains K_x and K_y with the sinusoidal and Gaussian inputs. See Fig. 6. Minimizing the mean square error between the linearized system and the actual nonlinearity, the following linearized gains are obtained.

$$K_x = \frac{1}{\sigma_x^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xN(x+y)p(y)q(x)dxdy \tag{3.22}$$

$$K_y = \frac{1}{\sigma_y^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yN(x+y)p(y)q(x)dxdy \tag{3.23}$$

where,

$$q(x) = \frac{1}{\pi}(a_x^2 - x^2)^{\frac{1}{2}} \tag{3.24}$$

is the probability density function (pdf) of the sinusoid with amplitude a_x, and

$$p(y) = \frac{2}{\sigma_y \sqrt{\pi}} e^{-\frac{y^2}{2\sigma_y^2}} \tag{3.25}$$

is the pdf of the Gaussian input y(t). The linear gains described by the equations above have been solved for the case of an ideal relay, which is equivalent to a one-bit quantizer, by Atherton [7]. The results follow,

$$K_y = (\frac{2}{\pi})^{\frac{1}{2}} (\frac{\Delta}{\sigma_y}) M(½, 1, -\rho^2) \tag{3.26}$$

$$K_x = (\frac{2}{\pi})^{\frac{1}{2}}(\frac{\Delta}{\sigma_y})M(½,2,-\rho^2) \tag{3.27}$$

where,

$$\rho = \frac{a_x}{\sqrt{2}\sigma_y} \tag{3.28}$$

is the square root of the ratio of the variance of the sinusoidal component to the Gaussian component at the nonlinearity input. The functions $M(\alpha,\gamma,x)$ are the Confluent Hypergeometric Functions [10]:

$$\frac{\Gamma(\gamma-\alpha)\Gamma(\alpha)}{\Gamma(\gamma)}M(\alpha,\gamma,x) = \int_0^1 e^{xt}t^{\alpha-1}(1-t)^{\gamma-\alpha-1}dt \tag{3.29}$$

By replacing the nonlinear quantizer by the linearized gains followed by an additive noise source, n(t), the Delta-Sigma Modulator can be separated into two interlocked linear systems as illustrated in Fig. 7 and 8. In one system, the input forcing function is the sinusoid x(k). In the other system, the forcing function is the additive noise source n(k) produced by quantization. From Fig. 7, we have,

$$\frac{E_x(z)}{X(z)} = \frac{H_i(z)H(z)}{1+K_xH(z)} \tag{3.30}$$

If integrators are used in the modulator, then in the base-band region and assuming that the frequency of x(k) is small,

$$\frac{E_x(z)}{X(z)} \approx \frac{1}{K_x} \tag{3.31}$$

It follows that,

$$\sigma_{ex}^2 = \frac{1}{K_x^2}\sigma_x^2 \tag{3.32}$$

where, for a sinusoid, $\sigma_x^2 = \frac{a_x^2}{2}$.

As in the previous section, we assume that n(k) is white. We also assume that the input to the nonlinearity due to the noise source n(k) , $e_n(k)$, is Gaussian. From Fig. 8,

$$\sigma_{en}^2 = \frac{\sigma_n^2}{2\pi}\int_{-\pi}^{\pi}\frac{|H(e^{j\omega})|^2}{|1+K_nH(e^{j\omega})|^2}d\omega \tag{3.33}$$

The power at the output of the Delta-Sigma Modulator is constant for a single bit quantizer. Hence,

$$E\{p^2(k)\} = \sigma_n^2 + K_x^2\sigma_{ex}^2 + K_n^2\sigma_{en}^2 = \Delta^2 \tag{3.34}$$

Substituting for the linearized gains and solving for the additive noise variance,

$$\sigma_n^2 = \Delta^2[1 - \frac{2}{\pi}\rho^2M^2(½,2,-\rho^2) - \frac{2}{\pi}M^2(½,1,-\rho^2)] \tag{3.35}$$

Now, from (3.32) and (3.27),

$$K_x^2 = \frac{\sigma_x^2}{\sigma_{ex}^2} = \frac{2}{\pi}\frac{\Delta^2}{\sigma_{en}^2} M^2(½,2,-\rho^2) \tag{3.36}$$

Hence, the ratio ρ can be obtained as a function of the input sinusoid amplitude, a_x. This is done by solving the following nonlinear equation for ρ,

$$\rho^2 M^2(½,2,-\rho^2) = \frac{\pi}{4}\frac{a_x^2}{\Delta^2} \tag{3.37}$$

Using this expression, the additive noise variance can be found as a function of input-sinusoid amplitude from (3.35),

$$\sigma_n^2 = \Delta^2[1 - \frac{a_x^2}{2\Delta^2} - \frac{2}{\pi} M^2(½,1,-\rho^2)] \tag{3.39}$$

An analysis of the above results shows that the gains K_n and K_x decrease as the input amplitude of the sinusoid increases.

Fig. 11 shows the calculated value of ρ obtained by solving (3.37) using the Newton-Raphson technique plotted as a function of input sinusoid amplitude a_x with $\Delta = 1$. In the same Figure, the additive noise variance σ_n^2 is also plotted. Note that the additive noise variance remains almost constant, decreasing only slightly, and its value is close to that obtained if we assume that the noise is unformally distributed (i.e. $\frac{\Delta^2}{3}$). However, the magnitude of ρ is seen to increase as a_x approaches Δ. This implies that the variance of the input-signal-related component at the quantizer input becomes very large. This will be examined in the next section.

A few words are in order about the Confluent Hypergeometric Functions. First of all, we note that

$$M(½,1,-x^2) = e^{-\frac{x^2}{2}} I_0(\frac{x^2}{2}) \tag{3.40}$$

where $I_0(x)$ is the modified Bessel function of zero order. Furthermore,

$$M(½,1½,-x^2) = \frac{\pi^{½}}{2x} erf(x) \tag{3.41}$$

This function is encountered for DC inputs. However, such nice closed form solutions do not exist for the case $M(½,2,-x^2)$. We note, however, that [10]

$$M(\alpha,\gamma,z) = \frac{\Gamma(\gamma)}{\Gamma(\gamma-\alpha)}(-z)^{-\alpha}[1 + O(|z|^{-1})] \tag{3.42}$$

when Real (z) <0. $\Gamma(n)$ is the well known Gamma Function. Using the above expression, we obtain the following approximations as x becomes large,

$$M(½,1,-x^2) \approx \frac{1}{\pi^{½}x} \tag{3.43}$$

$$M(½,2,-x^2) \approx \frac{2}{\pi^{½}x} \tag{3.44}$$

$$M(½,1½,-x^2) \approx \frac{\pi^{½}}{2x} \tag{3.45}$$

Equation (3.44) is consistent with the exact relationship (3.41) as x becomes large.

4. Noise Spectra and Signal to Noise Ratio

In this section we will analyze the effects of input-signal amplitude on the shaping of the noise spectra and the signal-to-noise ratio of the Delta-Sigma Modulator. The noise transfer function can be written from the block diagram in Fig. 8 as,

$$NTF(z) = \frac{P_n(z)}{N(z)} = \frac{1}{1 + K_n H(z)} \tag{4.1}$$

For the second order loop in Fig. 4.,

$$P_n(z) = N(z)\frac{(1-z^{-1})^2}{(1-z^{-1})^2 + \alpha_2 K_n[(1-z^{-1})+\alpha_1]z^{-1}} \tag{4.2}$$

The spectra of the noise can be obtained from (4.2) by noting that we have assumed n(k) to be a white noise process. Thus,

$$N(e^{j\omega T_s})N^*(e^{j\omega T_s}) = \frac{\sigma_n^2}{f_s} \tag{4.3}$$

Hence, the expression for the noise spectra becomes,

$$S_{nn}(f) = P_n(e^{j\omega T_s})P_n^*(e^{j\omega T_s}) \tag{4.4}$$

$$= \frac{16\sigma_n^2 \sin^4(\pi\frac{f}{f_s})}{16\sin^4(\pi\frac{f}{f_s})[1-\alpha_2 K_n] + 4\alpha_2 K_n \sin^2(\pi\frac{f}{f_s})[\alpha_2 K_n - 2\alpha_2\alpha_1 K_n - 2\alpha_1] + (\alpha_1\alpha_2 K_n)^2}$$

Clearly, the shaping of the noise spectra depends on the gain K_n. As the input signal amplitude increases , K_n decreases and the noise moves in-band. This is shown in Fig. 12 and Fig. 13 where the noise spectra based on (4.4) are plotted for K_n large and small. This nonlinear phenomenon, predicted theoretically above, is also observed in actual simulations. In Fig. 14, the amplitude spectrum of the Delta-Sigma Modulator pulse stream, obtained by digital simulation of a second-order system, which high lights the baseband region is shown. The sampling rate was 1.024 MHz. The spectrum was obtained using a 4096 point FFT and only the first 160 points are shown. The input signal consisted of a 5 KHz tone with an amplitude of 0.13. Again, $\Delta = 1.0$. Notice that the in-band noise is very small. However, in Fig. 15, where the amplitude is increased to 0.93, the in-band noise increases due to the shaping of the noise spectra as the gain K_n decreases.

The above result implies that the signal-to-noise ratio of the Delta-Sigma Modulator is not a linear function of the variance of the input signal. This result cannot be obtained by the simple model used for the quantization process in [5] and [9].

To obtain an approximate expression for the signal-to-noise ratio, we note that within the base-band $f \ll f_s$. Thus,

$$S_{nn}(f) \approx \frac{16\sigma_n^2 \sin^4(\pi \frac{f}{f_s})}{\alpha_1^2 \alpha_2^2 K_n^2} \tag{4.5}$$

The in-band noise is calculated by integrating (4.5) over the base-band,

$$\sigma_{nB}^2 = \int_0^{f_B} S_n(f)df \approx \frac{16\pi^4 \sigma_n^2}{[5\alpha_1\alpha_2 K_n]^2}(\frac{f_B}{f_s})^5 \tag{4.6}$$

Therefore, the signal-to-noise ratio becomes,

$$SNR = \frac{\sigma_x^2}{\sigma_{nB}^2} \approx \frac{5\sigma_x^2 \alpha_1^2 [\alpha_2 K_n(\sigma_x)]^2}{16\pi^4 \sigma_n^2(\sigma_x)}(\frac{f_s}{f_B})^5 \tag{4.10}$$

In the above expression the dependence of the gain K_n and the additive noise variance σ_n^2 on the input signal variance σ_x^2 has been indicated.

Exact Numerical Calculation of Signal to Noise Ratio

Based on the analytical results presented in section 3, we can compute the signal-to-noise ratio as a function of input-signal amplitude for both DC and sinusoidal inputs. For DC inputs, equations (3.13), (3.15), (3.19) and (3.21) represent coupled, nonlinear and integral equations. The numerical solution of these simultaneous equations, given a DC input amplitude m_x, yield the linearized gain, K_n, and σ_n^2. Once K_n and σ_n^2 are known, then the noise spectra $S_{nn}(f)$ can be numerically integrated over the baseband to yield the in-band noise component σ_{nB}^2. The SNR is then defined as the ratio,

$$SNR_{DC} = \frac{m_x^2}{\sigma_{nB}^2} \tag{4.12}$$

For sinusoidal inputs, the coupled, nonlinear and integral equations, (3.37), (3.38), (3.33) and (3.26) must be solved. In this case the SNR is defined as,

$$SNR = \frac{a_x^2}{2\sigma_{nB}^2} \tag{4.13}$$

Figure 16 shows the calculated SNR of second and third-order modulators plotted against the DC amplitude. The sampling rate was 1.024 Mhz and the baseband bandwidth was 4 kHz. Notice that the SNR increases and then decreases as the input amplitude approaches to within 10 dB of the saturation point ($m_x = \Delta$). This theoretical result, which has been observed experimentally in [5] and [12], is caused by the decrease of the gain K_n as the amplitude is increased. The decrease in K_n in turn modifies the spectral noise shaping of the modulator causing the noise to move in-band. The SNR decreases although based on (3.21) the additive noise variance decreases with increasing amplitude!

Another observation concerning Fig. 16 is that the gain in SNR by using a third-order loop is not substantial in this case. However, if the sampling rate is doubled a significant advantage can be gained by using a third-order loop over a second-order loop as demonstrated in Fig. 17. The reason for this is that for a third-order loop although the baseband noise is smaller than for a second-order loop the noise increases at a larger rate as the baseband bandwidth is increased. In Fig. 18 the calculated signal-to-noise ratio of a second and third-order modulator, with a fixed input amplitude, are shown as a function of the oversampling ratio $\frac{f_s}{f_b}$. From the Figure, we

observe that a third-order modulator loses its SNR advantage over a second order modulator as the oversampling ratio is decreased.

The SNR of a second order modulator with a sinusoidal input is shown in Fig.19 as a function of input amplitude a_x. The sampling rate was $f_s = 1.024MHz$ and the baseband bandwidth was $f_b = 4\ kHz$. Again the SNR reaches a maximum and then declines as the sinusoid amplitude increases.

As was pointed out in section 3, the linear gain K_n decreases as the input amplitude increases. This is shown in Fig. 20 where the calculated gain is plotted against input amplitude for a sinusoidal signal. Interestingly, the variance of the random noise at the quantizer input actually increases as the amplitude increases. The calculated value of this variance, σ_e^2, is plotted in Fig. 21. Note that σ_e^2 increases well beyond its value at low amplitudes where it constitutes the major component of the quantizer input. Also, the variance of the sinusoidal component at the quantizer input also increases, surpassing the noise variance σ_e^2. This is also shown in Fig. 21. This analytical observation, which has been observed in experimental circuits in [5], has important consequences on the design of actual circuits where the dynamic range of signals is limited. An important contribution of the present analytical results is that proper design considerations, accounting for the large variances, can be made when actual circuits are implemented.

5. Stability Analysis

One of the major problems associated with higher-order Delta-Sigma Modulators is their stability. In this section we present a stability analysis of the second and third-order modulators. Expressions are derived which give bounds on the loop gains for stable operation. Furthermore, we show that a stable third-order modulator will become unstable as the input-signal amplitude is increased beyond a certain threshold.

5.1 Double Loop System

The transfer function of the double loop Delta-Sigma Modulator, between the input $x(k)$ and output $p_x(k)$ is,

$$T(z) = \frac{K\alpha_2\alpha_1 H_1(z)H_2(z)}{1+\alpha_2 K H_2(z)[1+\alpha_1 H_1(z)]} \tag{5.1}$$

where K is the linearized gain. The denominator of (5.1) is common for the input signal and noise transfer functions except that the appropriate gain must be substituted. As the amplitude of the input signal increases, the linearized gains decrease. In this section we derive stability conditions as a function of loop parameters, α_1 and α_2, and as a function of the gain K. To obtain the frequency response we substitute $z = e^{j\omega T_s}$, in (5.1). The denominator can then be written as,

$$D(j\omega T_s) = 1 + KGH(e^{j\omega T_s}) \tag{5.2}$$

where, if ideal integrators are used,

$$GH(e^{j\omega T_s}) = \frac{-\alpha_2}{4\sin^2(\frac{\omega T_s}{2})}[1 + \alpha_1 - \cos(\omega T_s) + j\sin(\omega T_s)]$$

In order to determine the relationship between α_1 , α_2, and K for a stable system, we set the

imaginary part of $GH(e^{\omega T_s})$ to zero and its real part equal to $\frac{-1}{K}$. Thus,

$$\sin(\omega T_s) = 0 \tag{5.3}$$

Hence, the frequency of oscillation of the limit cycle just at the point of instability is,

$$f = f_s/2 \tag{5.4}$$

Also, $\cos(\omega T_s) = -1$ for $f = f_s/2$. Hence,

$$\frac{\alpha_2 K}{4}(\alpha_1 + 2) = 1 \tag{5.5}$$

Therefore, when the above condition is met the circuit will produce a sustained oscillation at a frequency half of the sampling rate. For stability,

$$\frac{\alpha_2 K}{4}(\alpha_1 + 2) < 1 \tag{5.6}$$

From this expression, it is clear that if α_1 is chosen such that the system is stable, decreasing K does not cause instability. Thus the system will remain stable for increasing input-signal amplitude. This is verified by the Nyquist plot for the second-order system in Fig. 22 ($\alpha_1=0.5$ $\alpha_2=1.0$). This is not the case for the third-order system as will be shown below. Since increasing α_2 directly increases the variance of the signal to the quantizer, K will decrease accordingly. Therefore, the only degree of freedom is the choice of α_1. To increase the signal-to-noise ratio, α_1 must be increased. This leads to instability based on the expression above, where $\alpha_1<2$ must be satisfied assuming that $\alpha_2 K \approx 1$.

5.2 Third Order System

The transfer function for the third-order system is,

$$T(z) = \frac{K\alpha_3\alpha_2\alpha_1 H_1 H_2 H_3(z)}{1 + K\alpha_3 H_3(z)\left[1 + \alpha_2 H_2(z)[1 + \alpha_1 H_1(z)]\right]} \tag{5.7}$$

Proceeding as in the second order case we obtain two solutions for the third-order system. The first solution is given by,

$$\frac{K\alpha_3}{8}[4 + 2\alpha_2 + \alpha_1\alpha_2] \leq 1 \tag{5.8}$$

with a corresponding frequency of oscillation of $f = f_s/2$. From this expression, a stable third-order system can be designed by proper choice of α_1 and α_2. There are two degrees of freedom since changing α_3 causes K to adjust accordingly. For stability, α_2 must be small.

The frequency of oscillation for the second solution is determined by the following equation,

$$\sin^2\left(\frac{\omega T_s}{2}\right) = \frac{\alpha_1\alpha_2}{4} \tag{5.9}$$

Setting the real part of $GH(e^{\omega T_s})$ equal to $\frac{-1}{K}$ we obtain,

$$\frac{K\alpha_3}{8\sin^3(\frac{\omega T_s}{2})}[4\sin^3(\frac{\omega T_s}{2})+2\alpha_2\sin(\frac{\omega T_s}{2}) + \alpha_1\alpha_2\sin(\frac{\omega T_s}{2})]\leq 1 \tag{5.10}$$

Substituting for $\sin(\frac{\omega T_s}{2})$ from (5.9) we obtain,

$$\alpha_3 K[1 + \frac{1}{\alpha_1}]\leq 1 \tag{5.11}$$

A Nyquist plot for the third-order system is shown in Fig. 23 ($\alpha_1=0.1,\alpha_2=0.1,\alpha_3=1.0$). We observe that, as the linearized gain K decreases, we approach the second solution given by (5.9) and (5.10). If K decreases further, the third-order system will become unstable. In other words, increasing the amplitude of the input signal to a stable third-order Delta-Sigma modulator will cause the modulator to become unstable. There is an amplitude region, however, for which the modulator is stable.

From (5.11) notice that α_1 determines the range for which decreasing K causes instability. If a small α_1 is chosen, then the allowable amplitude range is increased at the expense of signal-to-noise ratio. Surprisingly, α_2 only effects the frequency of oscillation. Hence, a large value can be used for α_2 to increase SNR provided that (5.8) is satisfied for stability.

At this point an interesting question arises. Which linearized gains, K_x or K_n cause instability in a third-order system? To answer this question we evaluate the ratio $\frac{K_n}{K_x}$ as $m_x\rightarrow\Delta$. Thus for a DC signal,

$$\lim_{m_x\rightarrow\Delta}\frac{K_n}{K_x} = \lim_{m_x\rightarrow\Delta}\frac{2}{\pi^{1/2}}\rho\frac{e^{-\rho^2}}{erf(\rho)} = 0 \tag{5.12}$$

since $\rho\rightarrow\infty$ as $m_x\rightarrow\Delta$.

Hence, K_n is smaller than K_x. Thus the distortion components will cause the modulator to go unstable. This supports the notion that the distortion components due to the quantizer must be included in the analysis of the modulator using linearized gains.

6. Conclusions

A new method of analysis for Delta-Sigma Modulators based on modeling the nonlinear quantizer with minimum-mean-square error linearized gains followed by an additive noise source representing distortion components is described. Closed form expressions have been derived which relate the quantizer additive noise variance to the input-signal amplitude. The effects of increasing input-signal amplitude on the shaping of the noise spectra and signal-to-noise ratio have been presented. The signal-to-noise ratio of the modulator is calculated directly as a function of input-signal amplitude using analytical methods. The analysis is carried out for both DC and sinusoidal excitations.

Moreover, regions of stability for second and third-order loops have been obtained, including bounds on the loop gains for stable operation. It is shown that the stability of third-order modulators is input-signal amplitude dependent.

7.References

(1) M.W. Hauser, P.J. Hurst, and R.W. Brodersen, "MOS ADC-filter combination that does not require precision analog components," *IEEE ISSCC Dig. of Tech. Papers*, pp. 80-82, Feb. 1985.

(2) B.P. Agrawal and K. Shenoi, "Design Methodology for Delta Sigma Modulation,"*IEEE Trans. on Communications*, vol. COM-31, no. 3, March 1983.

(3) T. Misawa, J.E. Iwersen, L.J. Loporcaro, and J.G. Ruch, "Single-chip per channel CODEC with filters utilizing Delta- Sigma Modulation," *IEEE J. Solid-State Cir.*, vol. SC-16, pp. 333-341, Aug. 1981.

(4) J.C. Candy and O.J. Benjamin, "The structure of quantization noise from Sigma-Delta Modulation," *IEEE Tr. on Comm.*, vol. COM-29, pp. 1316-1323, Sept. 1981.

(5) J.C. Candy, "A use of double integration in Sigma Delta Modulation," *IEEE Tr. on Comm.*, vol. COM-33, pp. 249-258, Mar. 1985.

(6) J.C. West, J.C. Douce, and B.G. Leary, "Frequency spectrum distortion of random signals in non-linear feedback systems," *IEE Monograph*, No. 419M, Nov. 1960, pp. 259-264.

(7) D.P. Atherton, Nonlinear Control Engineering, Van Nostrand, London, 1975.

(8) H. W. Smith, "Approximate analysis of randomly excited nonlinear controls,", Research Monograph No. 34, MIT Press, Cambridge, Massachusetts.

(9) Hanafy Meleis, Pierre Le Fur, "A novel architecture design for VLSI implementation of an FIR decimation filter," *Int. Conf. on Acoustics, Speech and Sig. Proc.* Orlando, Florida, 1985.

(10) M. Abramowitz, I. A. Stegun, "Handbook of Mathematical Functions," Dover Publications, New York, 1970.

(11) S. H. Ardalan, John Paulos, "Stability analysis of high-order sigma-delta modulators,"*1986 International Symposium on Circuits and Systems*,San Jose, Cal., May 1986.

(12) Max W. Hauser, Robert W. Brodersen, "Circuit and technology considerations for MOS Delta-Sigma A/D converters," *1986 International Symposium on Circuits and Systems*, San Jose, Cal,. May 1986.

(13) R. C. Booton Jr., "Nonlinear control systems with statistical inputs," M.I.T. Dynamic Analysis and Control Laboratory, Report No. 61, March 1952.

Figure Captions

Fig. 1. First Order Delta-Sigma Modulator

Fig. 2. Second Order Delta-Sigma Modulator

Fig. 3. Second Order Modulator with Fictitious Sample and Holds

Fig. 4. Z-domain Second Order Modulator

Fig. 5. Block Diagram of Z-Domain Delta-Sigma Modulator

Fig. 6. Minimum Mean Square Error Linearized Gain Modeling of Nonlinearity

Fig. 7. Equivalent Linearized System for DC and Sinusoidal Input

Fig. 8. Equivalent System for Quantization Noise

Fig. 9. Calculated and Simulated Values of $\rho = \frac{m_x}{\sqrt{2\sigma_e}}$

Fig. 10. Calculated and Simulated Additive Noise Variance σ_n^2

Fig. 11. Calculated Additive Noise Variance, σ_n^2, and ρ Sinusoidal Input

Fig. 12. Calculated Noise Spectrum, $K_n = 1.0$

Fig. 13. Calculated Noise Spectrum, $K_n = 0.1$

Fig. 14. Simulated Amplitude Spectrum, $a_x = 0.13$

Fig. 15. Simulated Amplitude Spectrum $a_x = 0.93$

Fig. 16. Calculated SNR for 2nd and 3rd Order Modulator, $\frac{f_s}{f_b} = 256$

Fig. 17. Calculated SNR for 2nd and 3rd Order Modulator, $\frac{f_s}{f_2} = 512$

Fig. 18. SNR Dependence on $\frac{f_s}{f_b}$ 2nd and 3rd Order Modulator

Fig. 19. Calculated SNR, Sinusoidal Input 2nd Order Modulator

Fig. 20. Calculated Linearized Gain, K_n, Sinusoidal Input 2nd Order Modulator

Fig. 21. Calculated Variances at Quantizer Input, 2nd Order Sinusoidal Input

Fig. 22 Nyquist Plot 2nd Order Modulator ($\alpha_1 = 0.5$, $\alpha_2 = 1.0$)

Fig. 23 Nyquist Plot 3rd Order Modulator ($\alpha_1 = 0.1$, $\alpha_2 = 0.1$, $\alpha_3 = 1.0$)

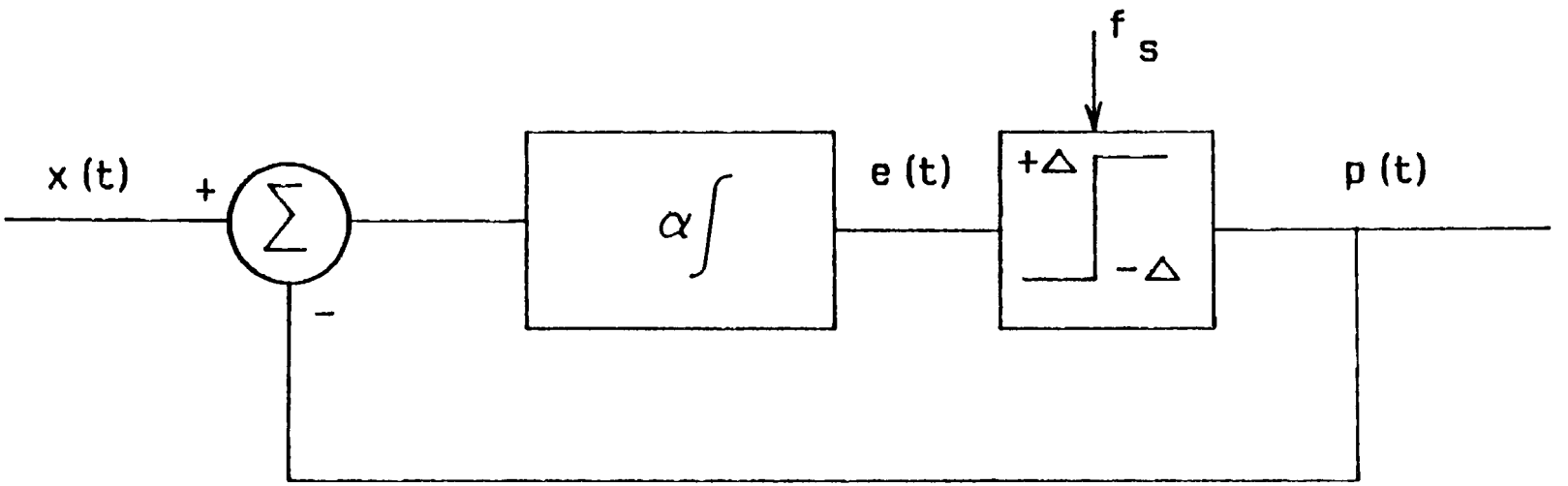

Fig. 1. First Order Delta-Sigma Modulator

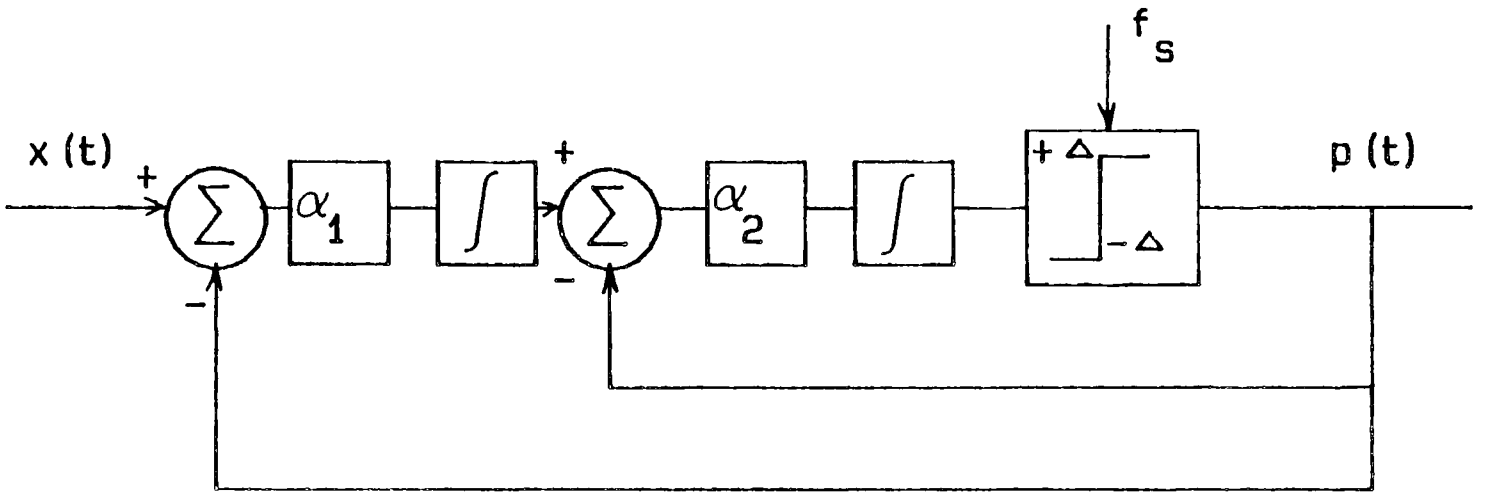

Fig. 2. Second Order Delta-Sigma Modulator

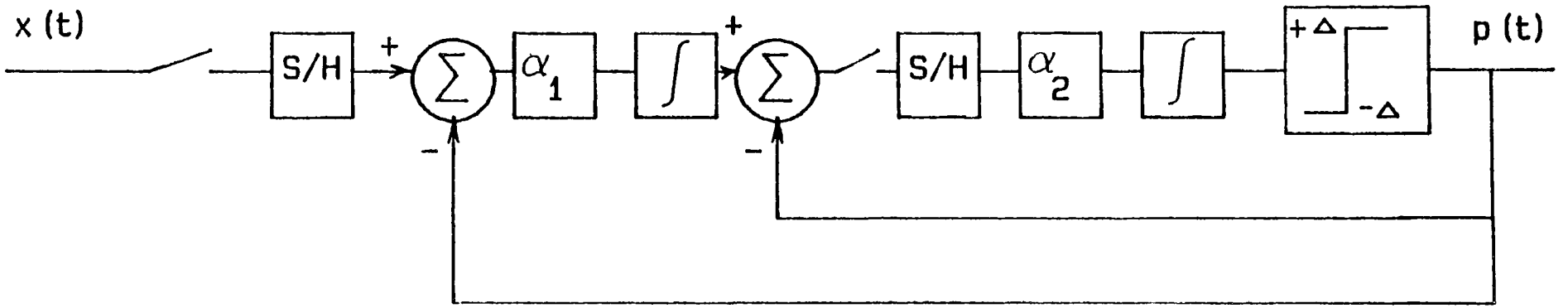

Fig. 3. Second Order Modulator with Fictitious Sample and Holds

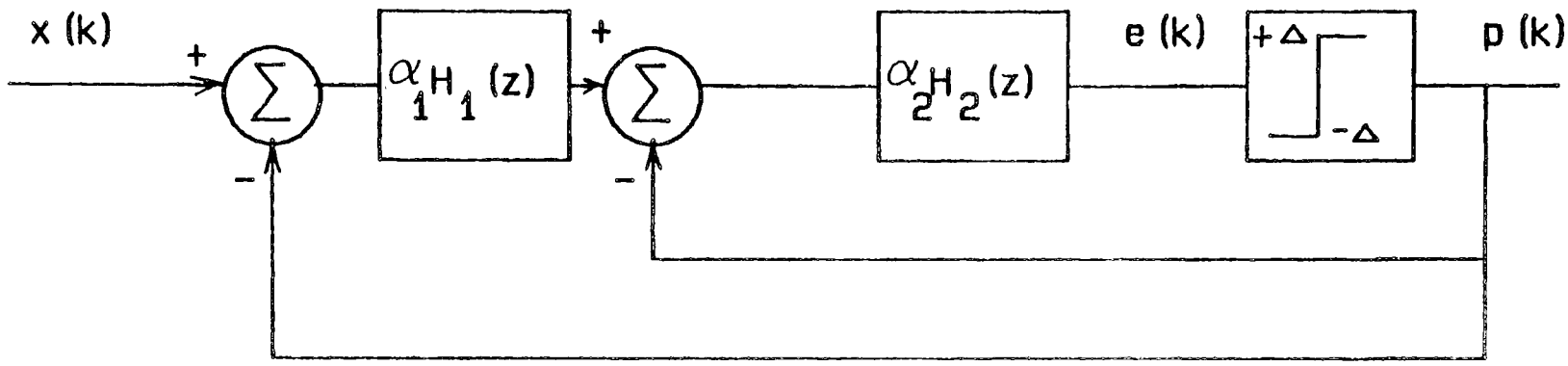

Fig. 4. Z-domain Second Order Modulator

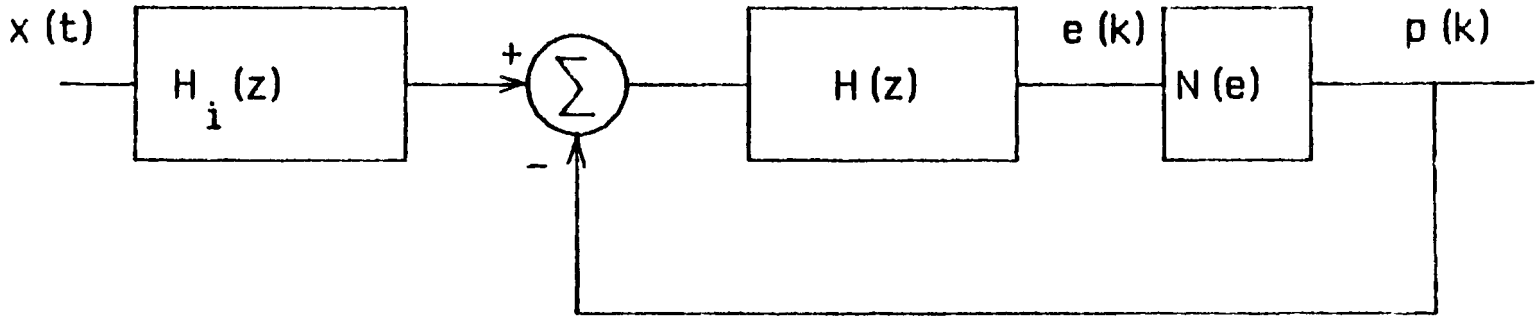

Fig. 5. Block Diagram of Z-Domain Delta-Sigma Modulator

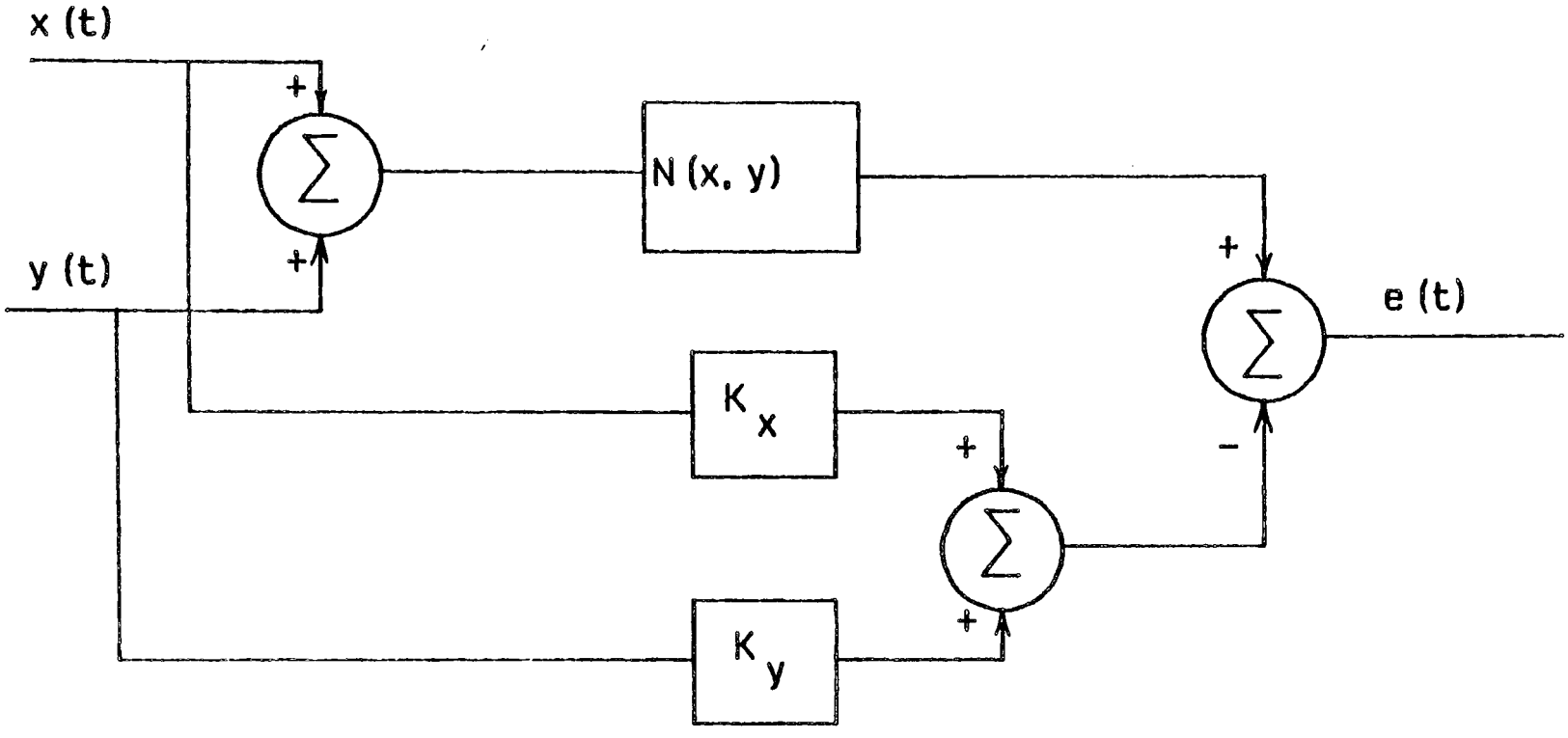

Fig. 6. Minimum Mean Square Error Linearized Gain Modeling of Nonlinearity

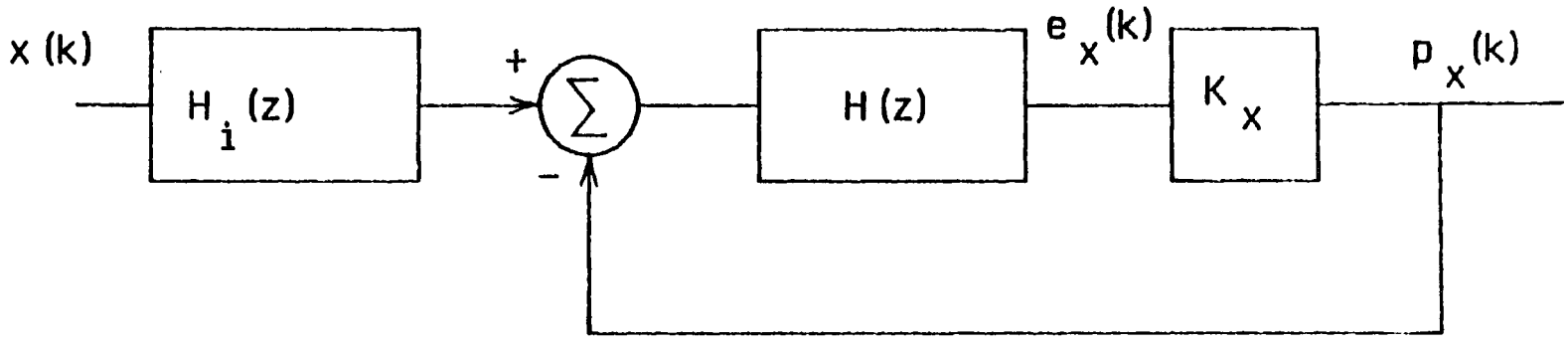

Fig. 7. Equivalent Linearized System for DC and Sinusoidal Input

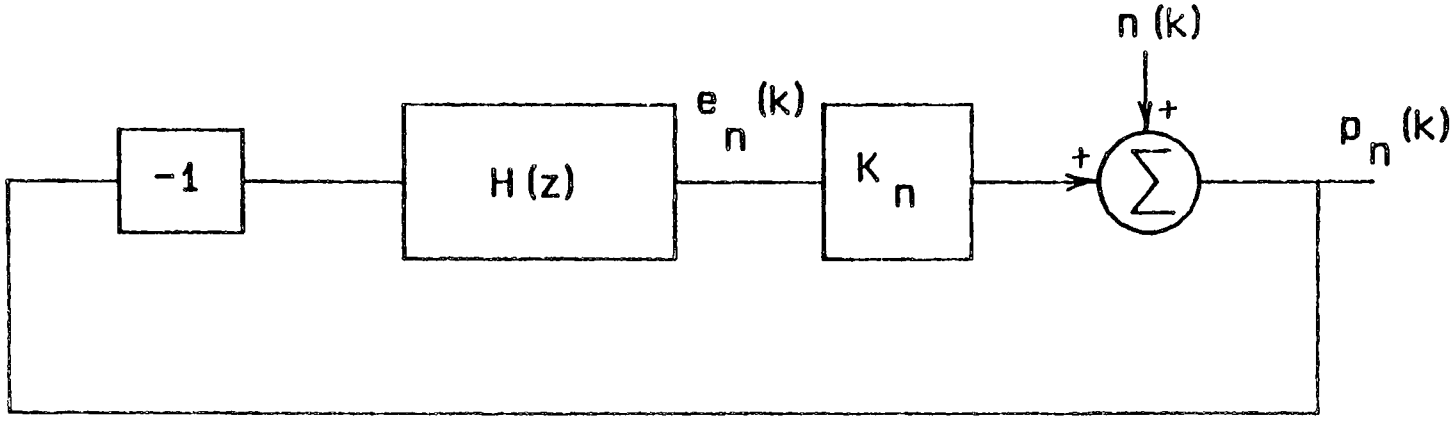

Fig. 8. Equivalent System for Quantization Noise

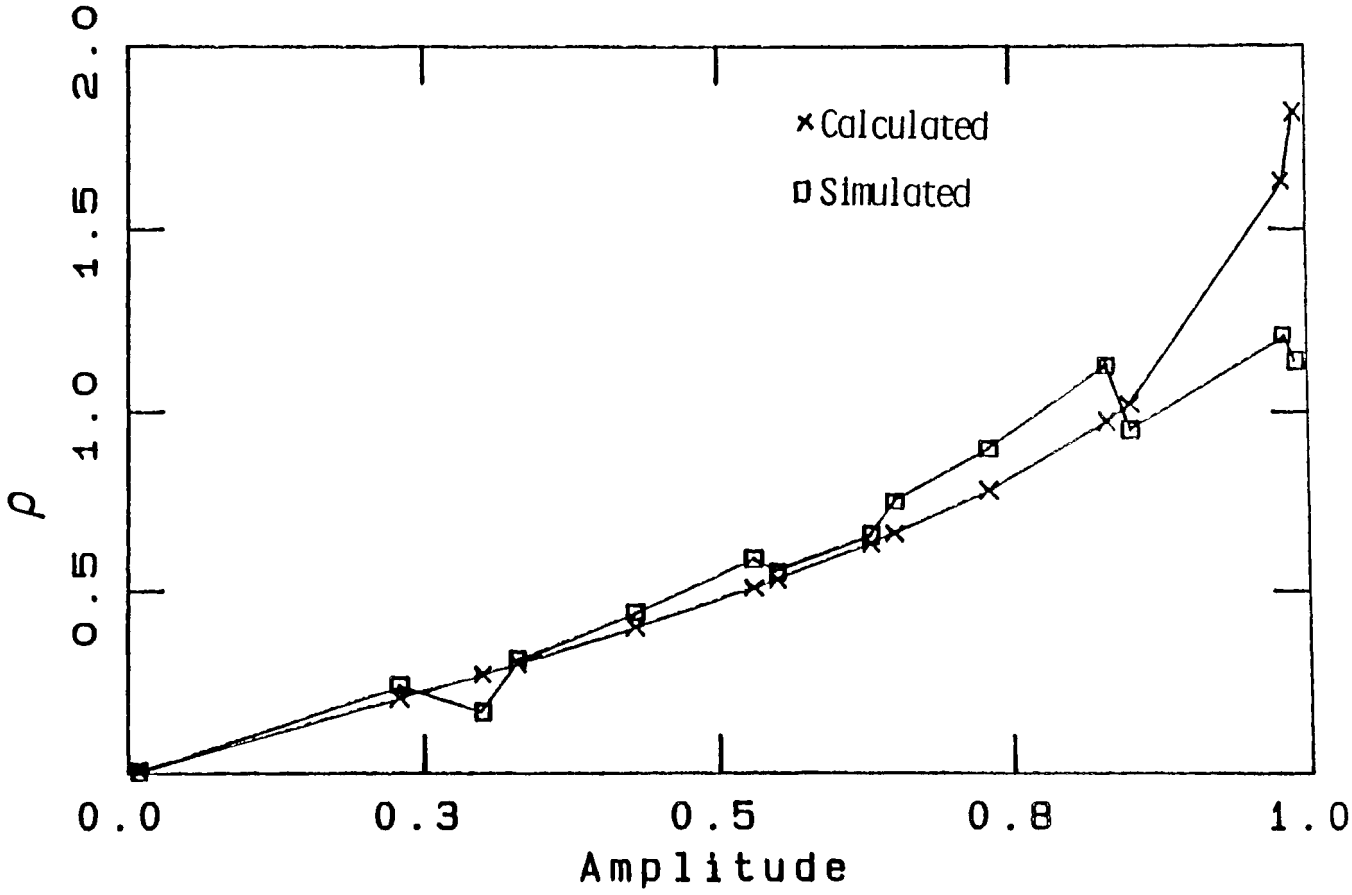

Fig. 9. Calculated and Simulated Values of $\rho = \frac{m_x}{\sqrt{2\sigma_e}}$

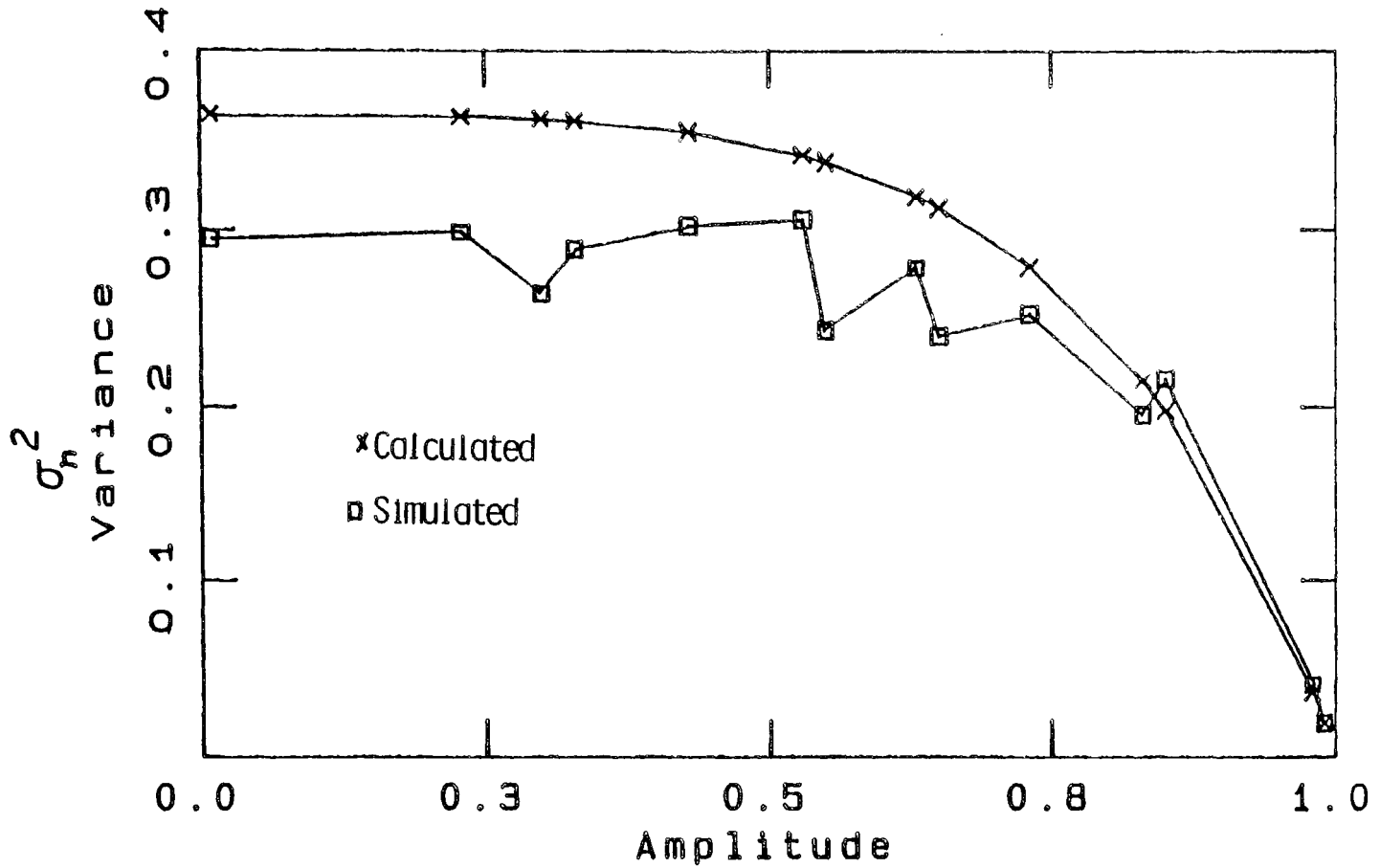

Fig. 10. Calculated and Simulated Additive Noise Variance σ_n^2

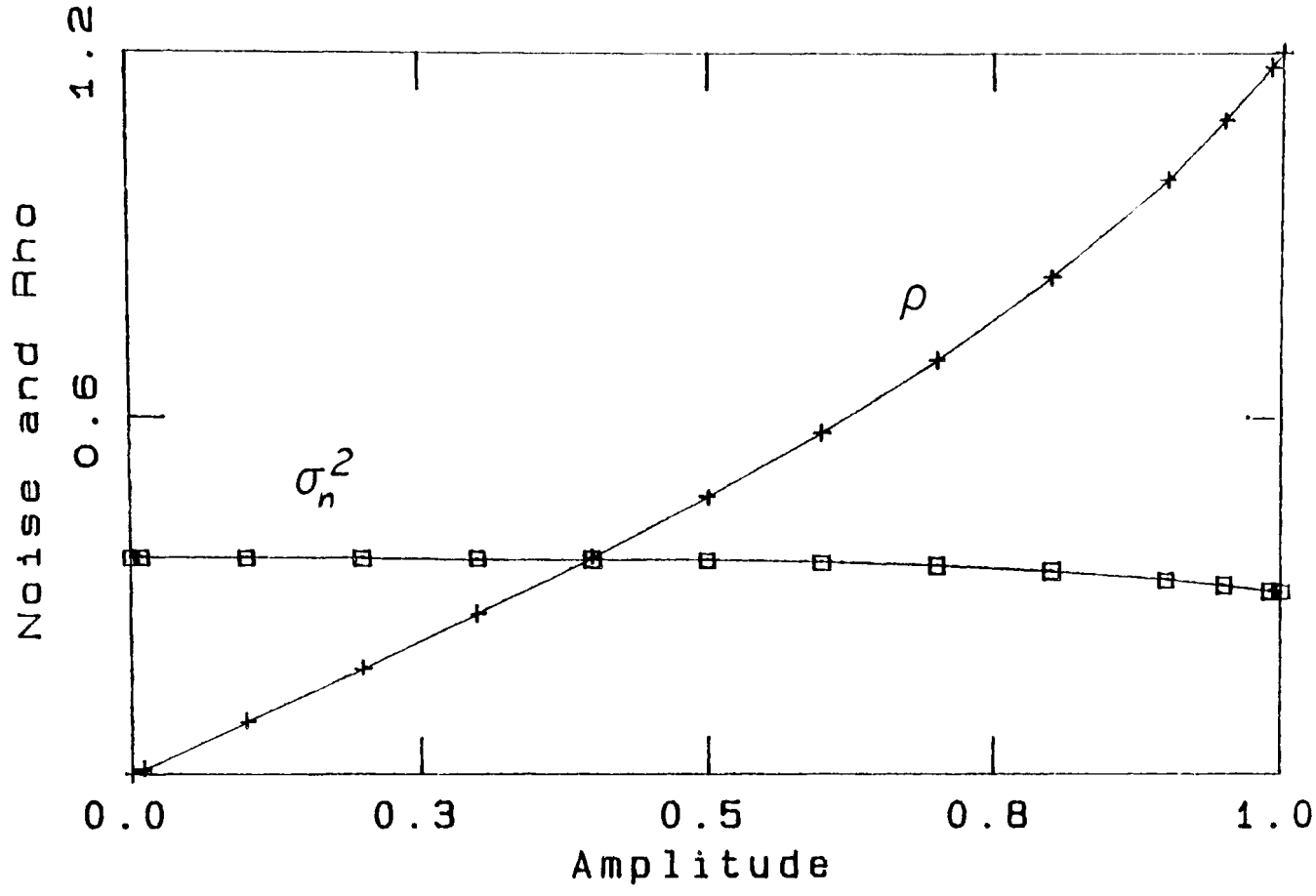

Fig. 11. Calculated Additive Noise Variance, σ_n^2, and ρ Sinusoidal Input

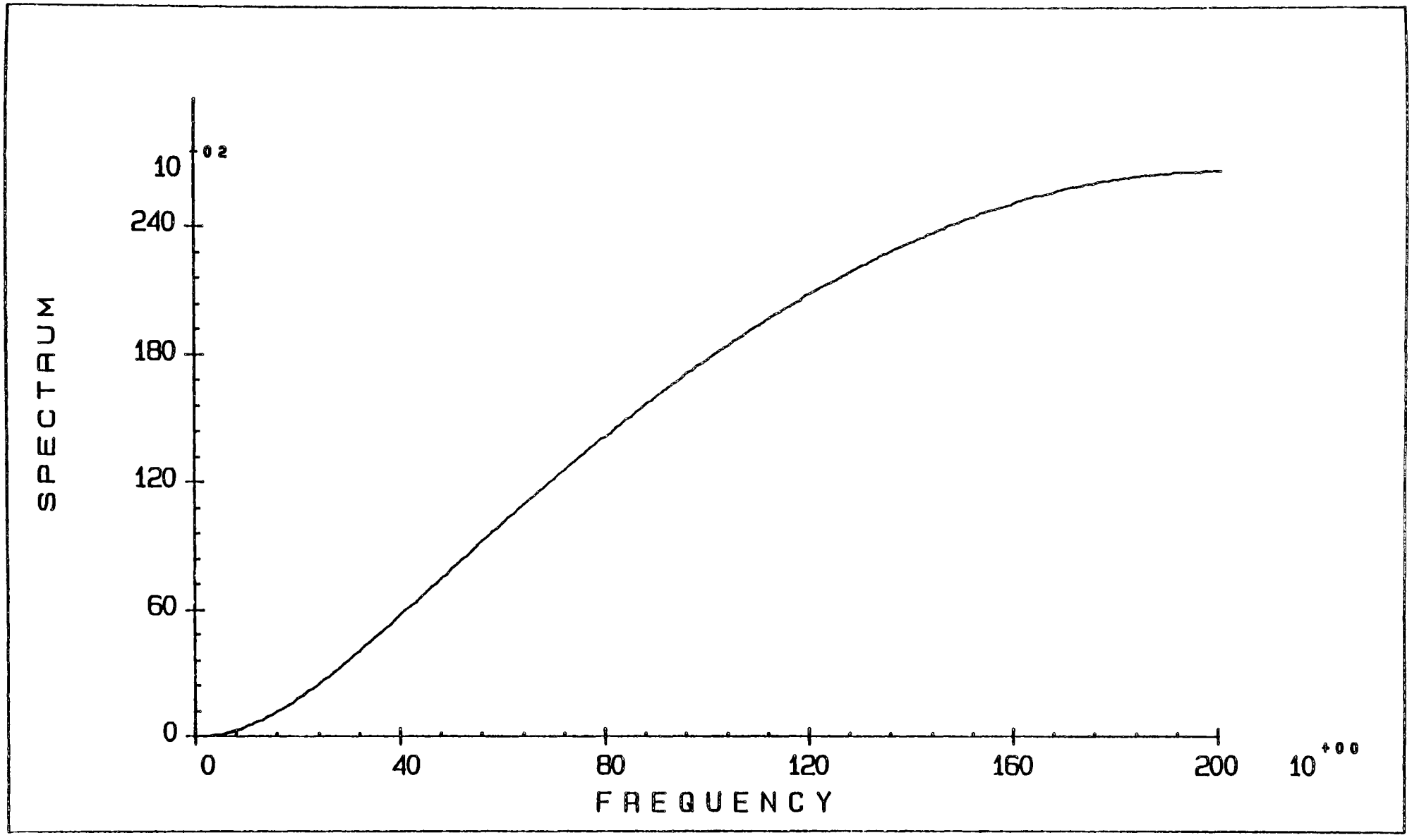

Fig. 12. Calculated Noise Spectrum, $K_n = 1.0$

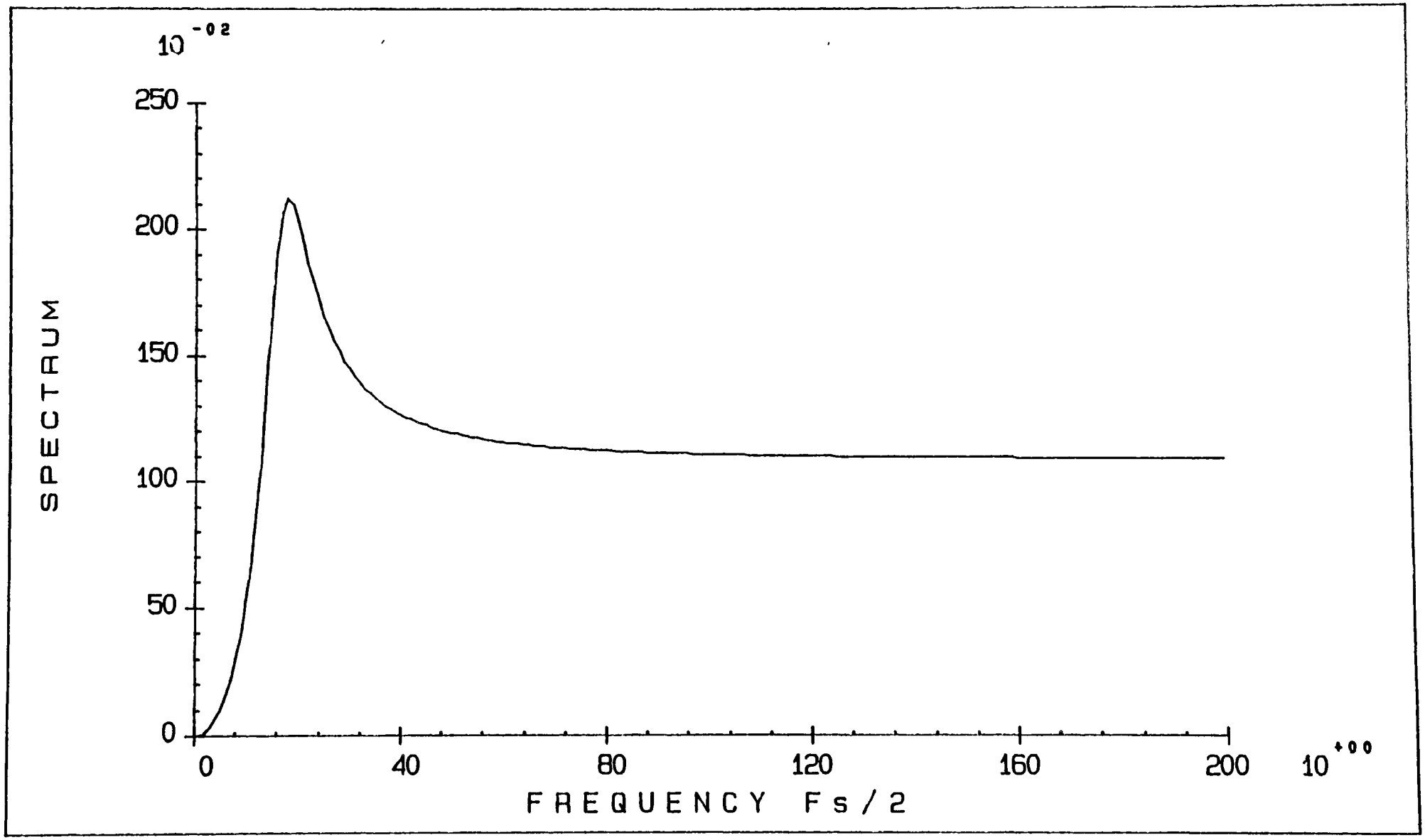

Fig. 13. Calculated Noise Spectrum, $K_n = 0.1$

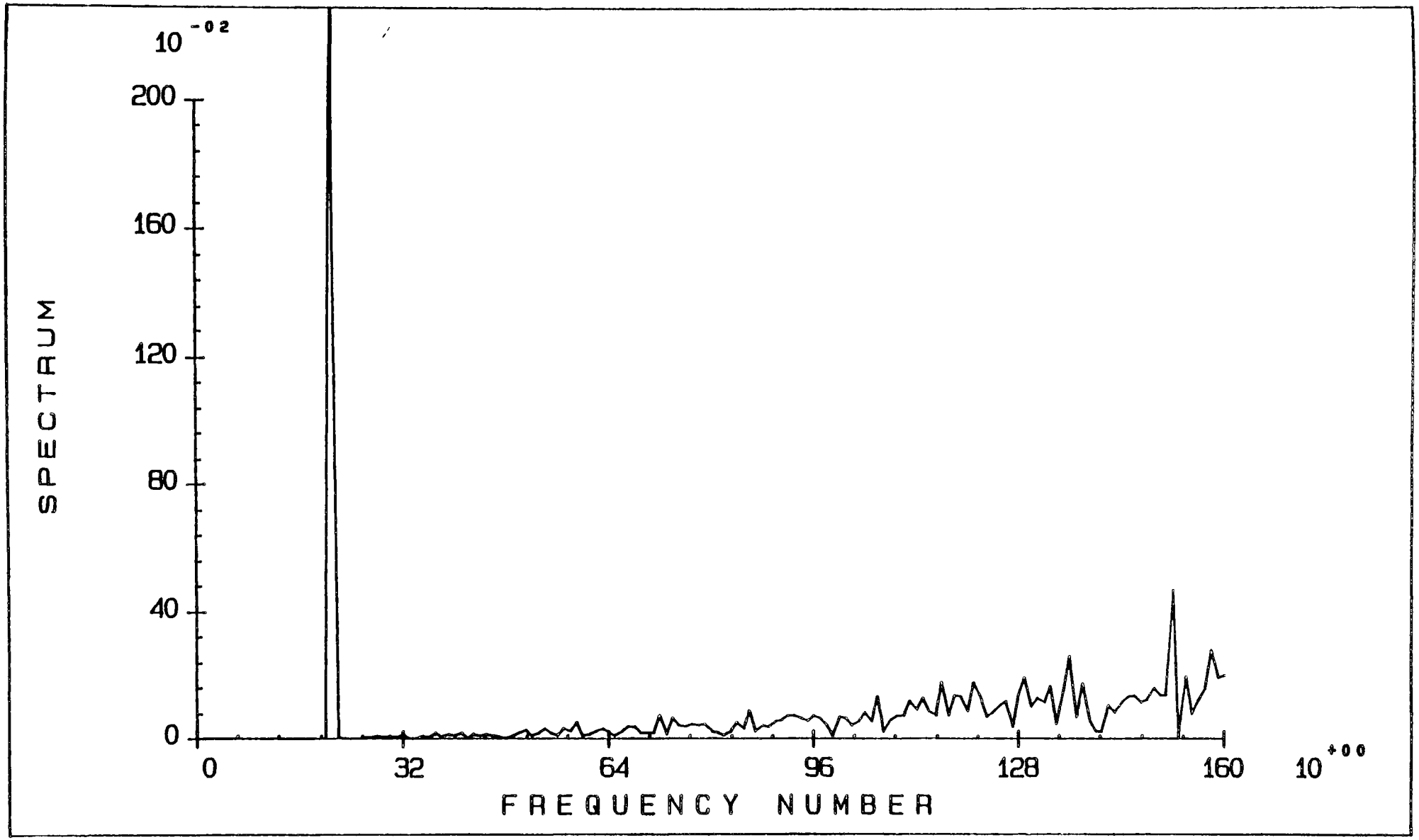

Fig. 14. Simulated Amplitude Spectrum, $a_x = 0.13$

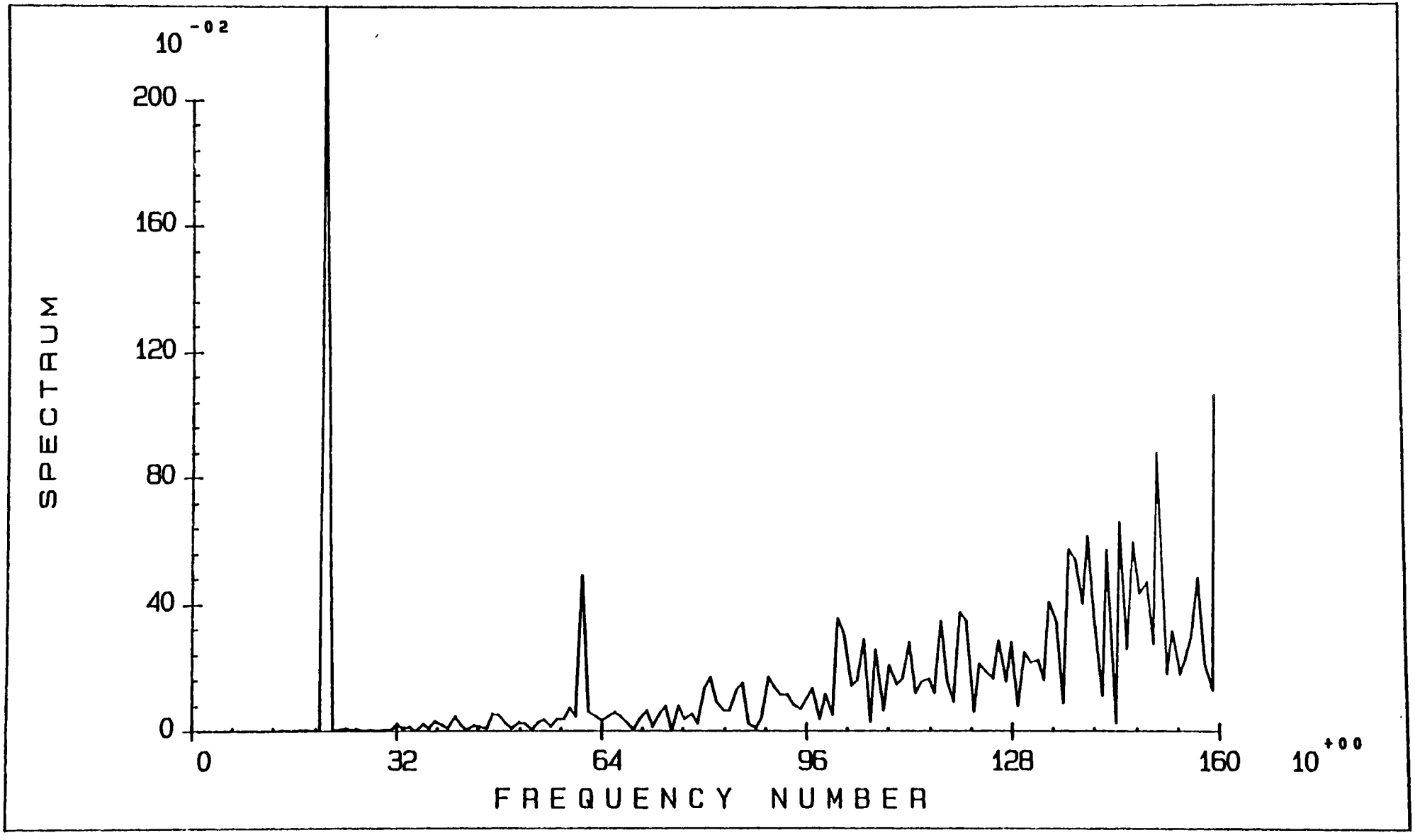

Fig. 15. Simulated Amplitude Spectrum $a_x = 0.93$

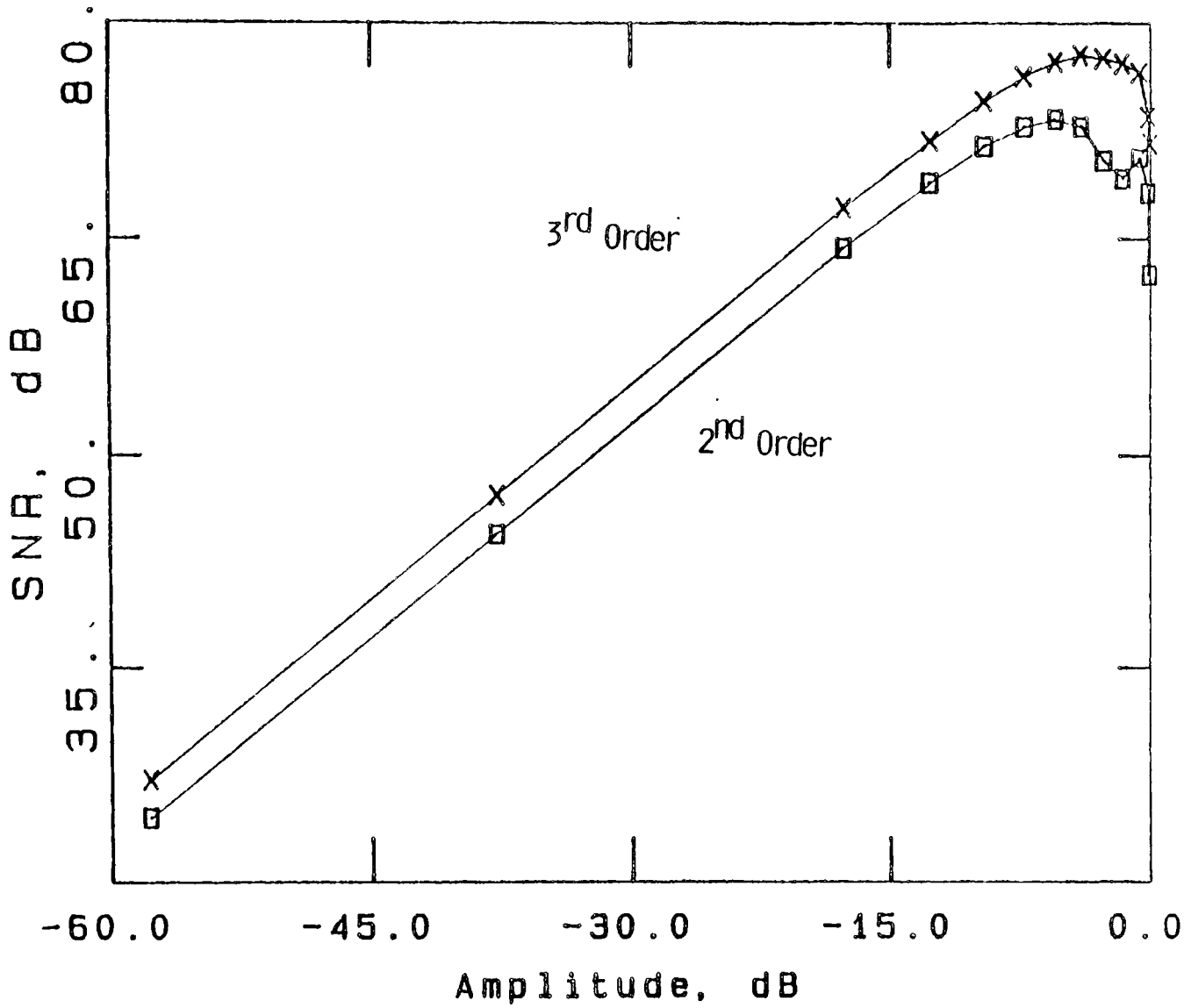

Fig. 16. Calculated SNR for 2nd and 3rd Order Modulator, $\frac{f_s}{f_b} = 256$

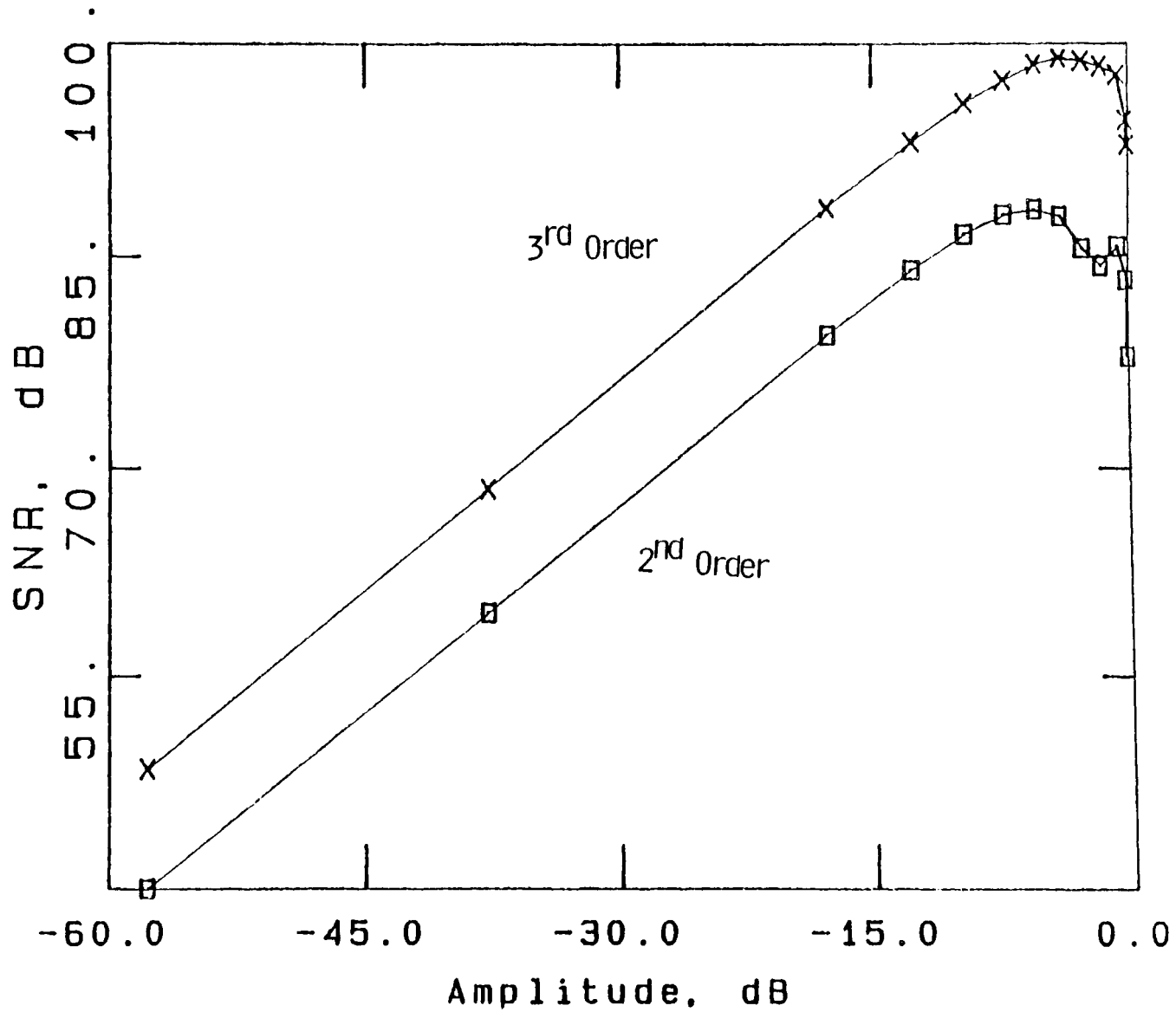

Fig. 17. Calculated SNR for 2nd and 3rd Order Modulator, $\frac{f_s}{f_2}=512$

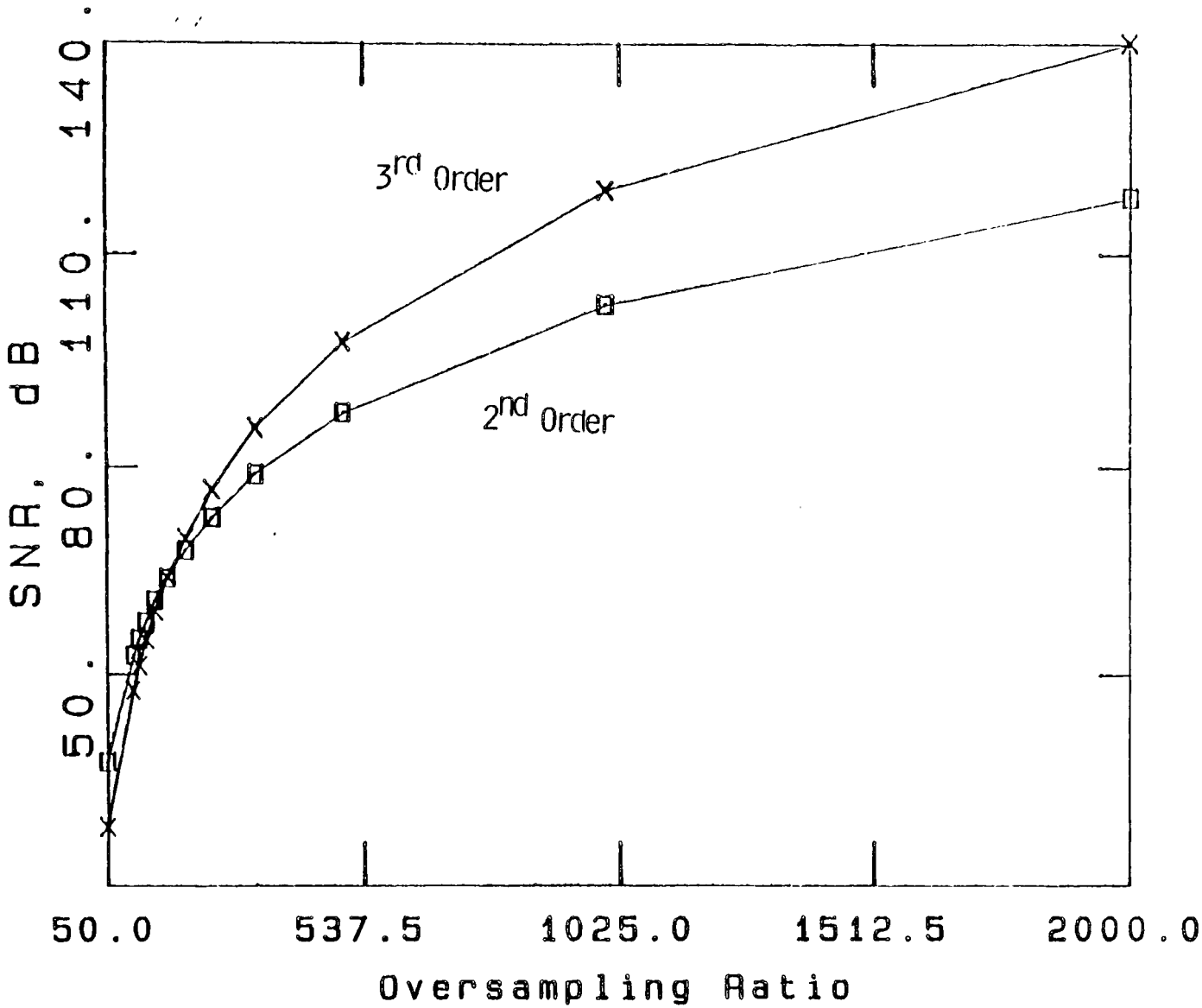

Fig. 18. SNR Dependence on $\frac{f_s}{f_b}$ 2nd and 3rd Order Modulator

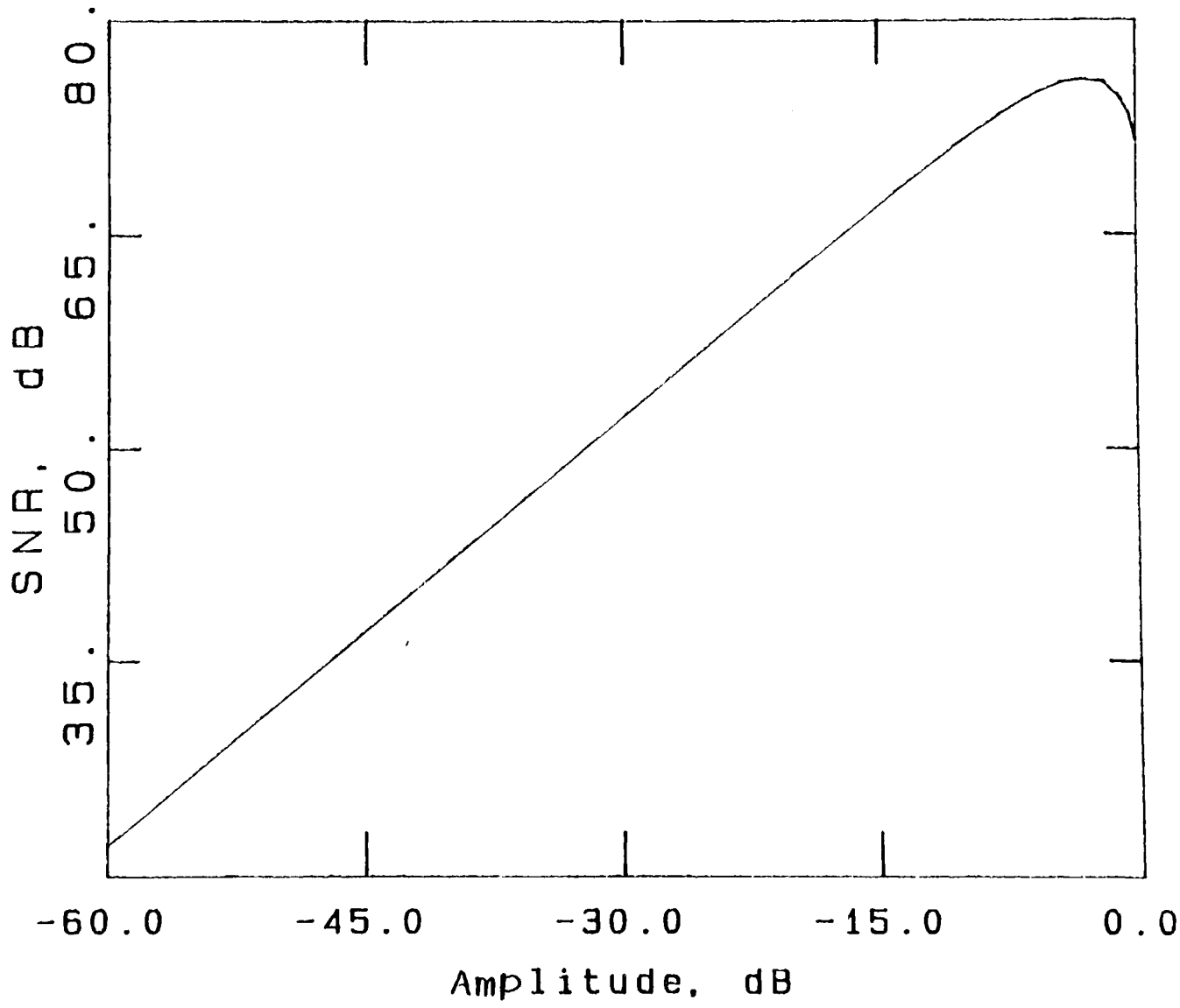

Fig. 19. Calculated SNR, Sinusoidal Input 2nd Order Modulator

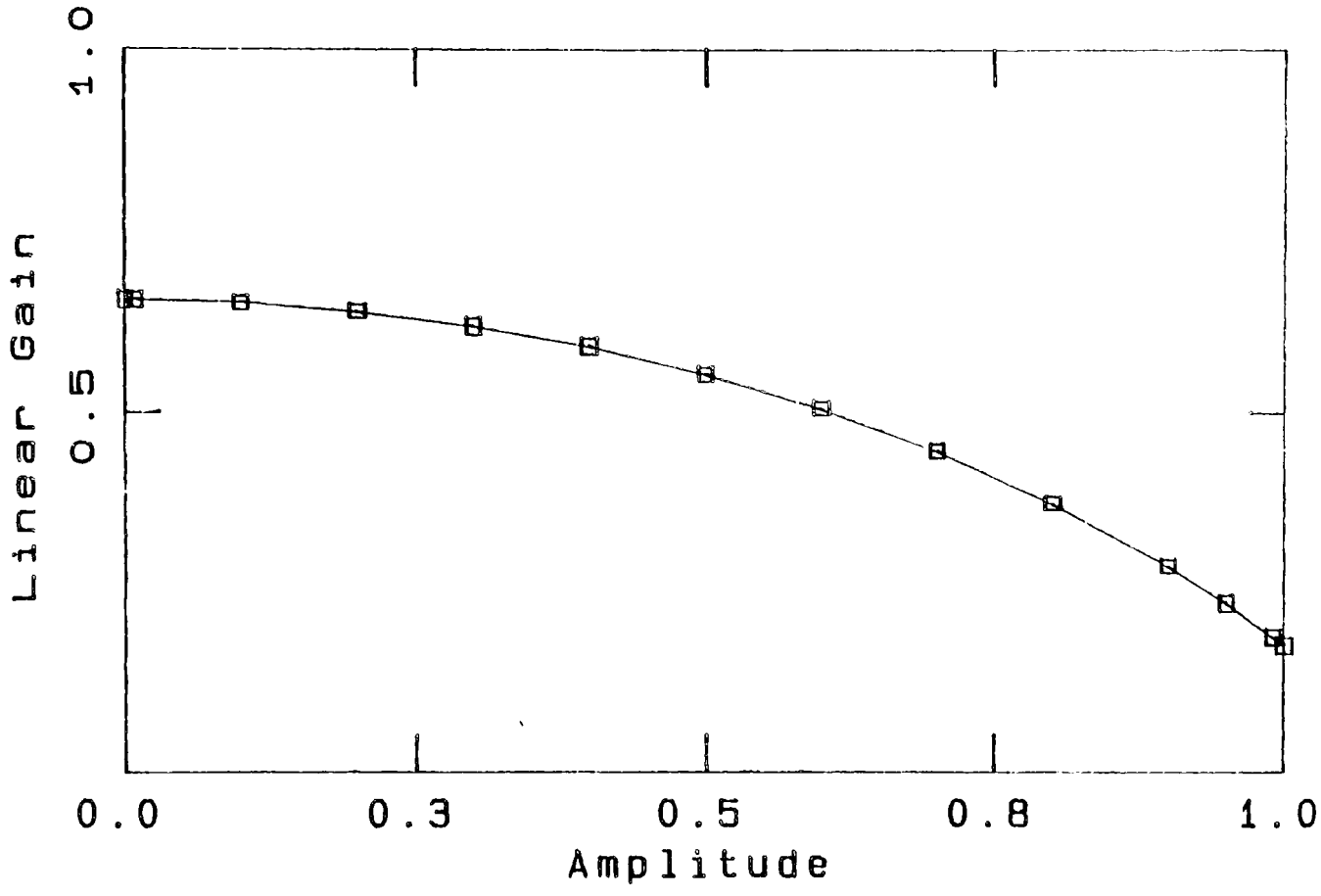

Fig. 20. Calculated Linearized Gain, K_n, Sinusoidal Input 2nd Order Modulator

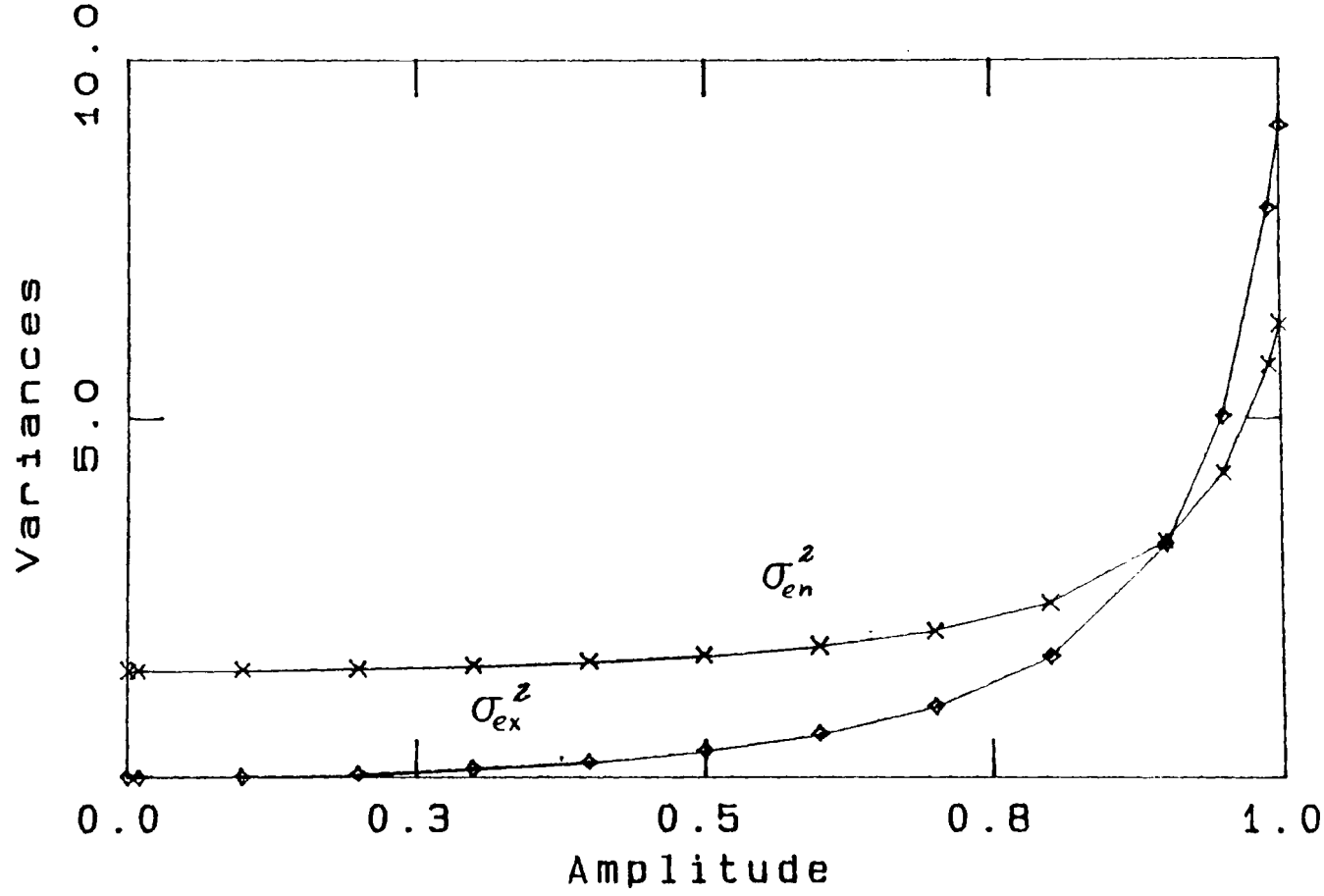

Fig. 21. Calculated Variances at Quantizer Input, 2nd Order Sinusoidal Input

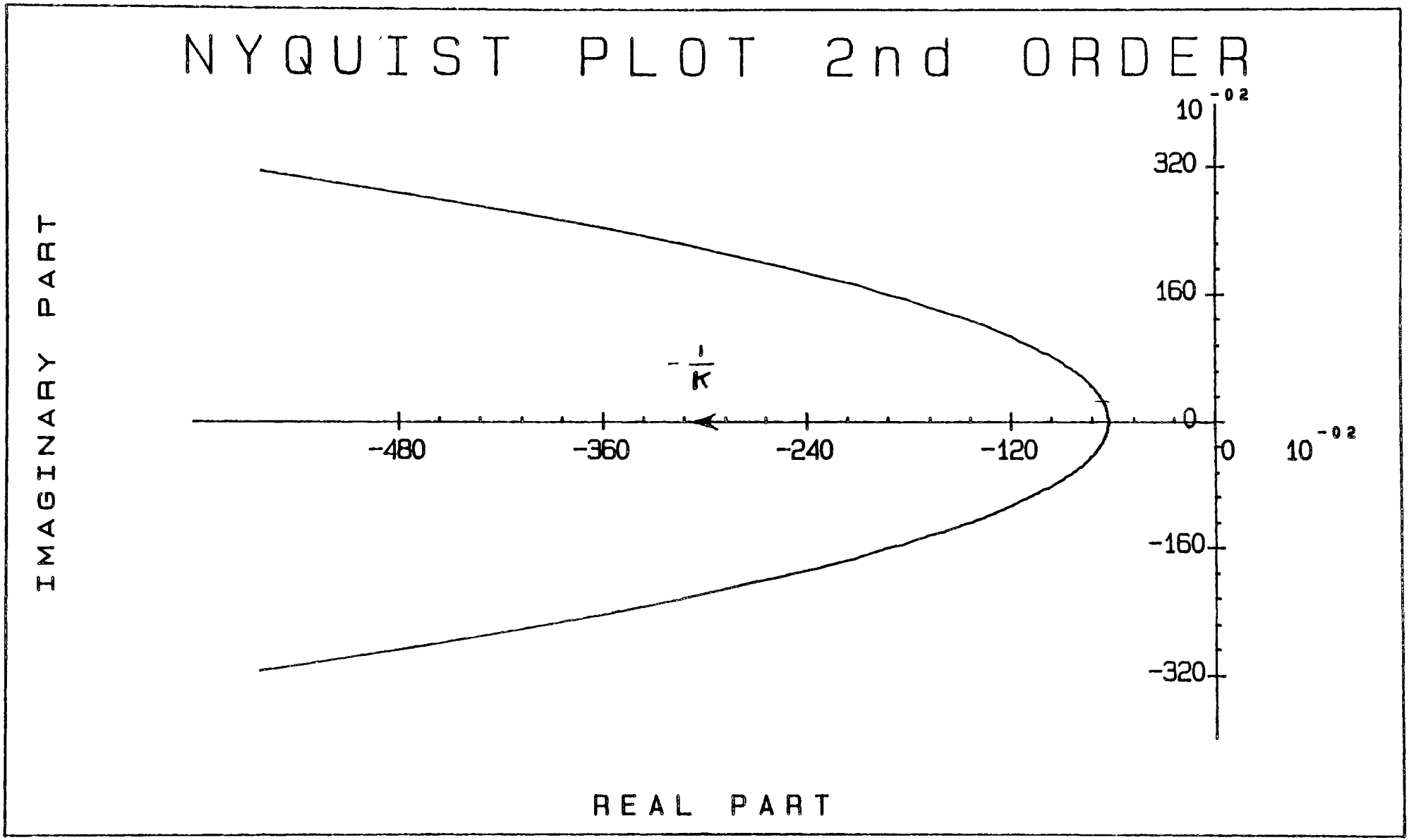

Fig. 22 Nyquist Plot 2nd Order Modulator ($\alpha_1 = 0.5$, $\alpha_2 = 1.0$)

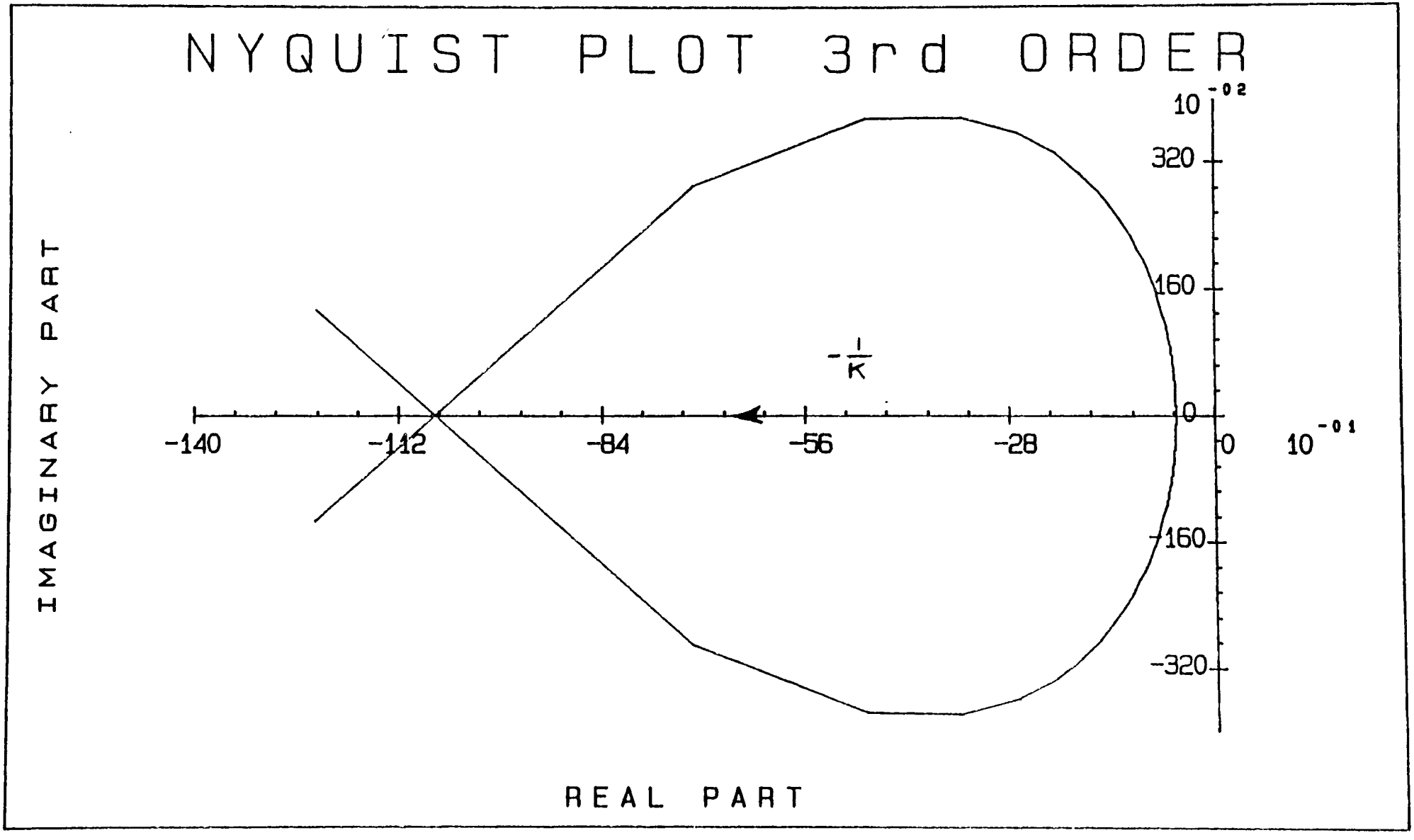

Fig. 23 Nyquist Plot 3rd Order Modulator ($\alpha_1=0.1$, $\alpha_2=0.1$, $\alpha_3=1.0$)