NOISE IN DELTA MODULATION

by

F.R. D'Emden B.E.(Hons.)

Submitted in partial fulfilment of the
requirements for the degree of
Master of Engineering Science

University of Tasmania

Hobart

March 1973

ACKNOWLEDGEMENTS

This work was carried out in the Electrical Engineering Department of the University of Tasmania. The author wishes to thank all members of the engineering staff and research students with whom he was associated for their assistance and encouragement. In particular the author wishes to thank Professor C.H. Miller, Professor of Electrical Engineering, and Mr. P. Watt, his supervisor, for their help and encouragement. The author also thanks the Public Service Board for its financial assistance and Miss. C. Byard for her perseverance with the typing.

I hereby declare that, except as stated herein, this thesis contains no material which has been accepted for the award of any other degree or diploma in any University or College, and that, to the best of my knowledge this thesis contians no copy or paraphrase of material previously published or written by another person except where due reference is made in the text of this thesis.

SUMMARY

An ivestigation of the noise which is an inherent part of delta modulation due to the nature of the modulation process is presented. A large part of the examination of the noise consists of an investigation of the noise performance of delta modulation. In order to limit unwanted paramenters, the study has been mainly confined to an ideal delta modulator with a perfect, single integration feedback network. The investigation presented is restricted to the modulation - demodulation process in the main, with emphasis on voice communication usage. The feasability of a computer simuation of delta modulation and the design and construction of an experimental modulator is presented. Analysis and comment on the qualitative and quantitative aspects of noise in delta modulation is included with reference to measurements taken, and existing analysis being made.

CONTENTS

INTRODUCTION

In the current state of voice and video telecommunication the use of analogue methods for new communication channels is still dominant. In particular, amplitude modulation, frequency modulation, and direct analogue transmission are most frequently used. With current trends in technology and service requirements, the use of digital communication systems is showing itself to be preferable to analogue methods in many circumstances.

The most developed, and at present the preferred type of digital communication system is pulse code modulation (PCM) which is in commercial use in many technologically advanced countries. Digital modulation systems have the advantages over analogue systems of: the capacity for regeneration with amplification (thus preventing the accumulation of noise over long routes); compatability with digital data transmission; use with time division multiplexing (TDM); and lower power consumption by virtue of the two state nature of the transmitted signal.

The other basic form of digital modulation, which uses a simpler principle of operation, is delta modulation. By virtue of the quantization in time and amplitude of the signal to be modulated by a digital method, some information is discarded in the modulation process. This gives rise to a type of noise intrinsic to the modulation process, and it is an investigation of this type of noise for delta modulation which forms the basis of this thesis.

In the first chapter a brief review is made of continuous carrier modulation and analogue pulse modulation, leading up to digital pulse modulation. This provides a basis for the introduction of delta modulation with discussion of its origin and operating principles. A review of the modifications, improvements and current state of development of delta modulation is contained and a comparison with pulse code modulation is made.

The second chapter consists of an investigation into the nature and magnitude of the noise which is inherent in the delta modulation process for simple, non-companded systems. A review of existing noise analysis with investigation into the factors and definitions involved is made.

In the third chapter the establishment of a computer simulation (using an Elliott 503 computer) for a single perfect integration, simple delta modulator is presented. The noise power results obtained from various simulations are discussed along with the factors influencing the results and the effectiveness of the simulation.

The design, realization and performance of an experimental delta modulator are described in the fourth chapter. Discussion of the method of noise measurement, the accuracy and repeatability of results and limitations of the circuitry is also included.

In the fifth chapter the experimental and computer simulation estimates of the performance of delta modulation are compared and any discrepancy is explained in terms of the different conditions applying to each. The results obtained from the measurements are discussed and analysed relative to the expected performance and the established theoretical descriptions. Where discrepancy between the various performance descriptions is observed it is explained in terms of either; the limitations of the theory, the lack of relevance of the conditions considered to the theory, or inaccuracies in the results due to limitations in the conditions under which the performance results were obtained.

It is shown in this thesis that no theoretical analysis yet determined gives an accurate description of delta modulation performance for all ranges of operating conditions. (ie. for a given set of system parameters; for a full range of input signal values and a representative range of expected input signal types). It has been demonstrated that the most accurate and useful description for a range of input loadings is that of Abate[17], which is empirically derived and gives estimates for the non-overload performance which are close to the theory of O'Neal[35] (derived from van de Weg[34]).

It has been demonstrated in this thesis that simple delta modulation performance can be described by four regions of performance. These regions are clearly distinguishable for sinusoidal inputs but the region of partial slope overload is not readily distinguished from the slope overload region where a broadband input signal is being considered. The performance region of partial slope overload is defined, analysed and confirmed by experimental and computer simulation results for a sinusoidal input. Previously this region of performance has only been referred to descriptively; as for example by O'Neal[35]: "as if the basic granular quantizing noise has been supplemented by furthur noise, due to the onset of the slope overload condition." The performance region of minimum quantizing noise for high clock frequencies has also been described, analysed and confirmed by experimental and computer simulation results. Discussion of the noise power spectrum and its variation with the input signal loading is included along with discussion of the effect of variation in the probability density function of both the error signal and the output-to-line pulse signal. It has been determined that the use of a sinusoid as a test signal is useful in indicating the nature of noise and the qualitative performance of delta modulation but is only of limited relevance quantitatively for specific non-sinusoidal inputs.

CHAPTER 1

THE DEVELOPMENT OF DIGITAL MODULATION SYSTEMS

1.1 Introduction

The era of electrical communication commenced in 1838 when Samuel Morse conveyed information over a distance using the variation of electrical parameters in an electrical transmission medium. This first electrical communication system was the telegraph in which information, in the form of an input message was coded into pulses, using a binary code, which were converted to electrical pulses for transmission by wire to be dectected and decoded at the receiving end.

From this date the demand for electrical communication has soared due to the high speed of information transmission inherent in electrical communication. With the perfection of a transducer for the conversion of sound to a time varying electrical signal directly related to the sound, telephony for direct voice communication resulted. The discovery of the possibility of using electromagnetic radiation as the transmission medium led, with wireless telegraphy, to modulation. A high frequency carrier wave more suited to the transmission medium is modulated by the systematic variation of the carrier in accordance with the input signal. Modulation must be performed in such a way that an acceptable approximation to the input signal can be reconstituted at the receiving end.

Modulation has developed into a vital process in electrical communication systems because of various demands which are satisfied, and benefits which result from modulation. The main benefits of modulation result from the shift in the natural frequency range of the information to a much higher frequency. Among the benefits are:

Ease of electro magnetic radiation at higher frequencies.

The allocation of different signals to different carrier frequencies. This allows many communication paths, each separable from the others, to use a common transmission medium. Thus a single channel with a wide bandwidth can be fully untilized by the simultaneous transmission of many narrow bandwidth signals.

Transmission at higher frequencies allows greater power transmission and the facility to select the frequency of transmission most suited to available equipment and to the transmission medium.

Modulation in general has the effect of suppressing noise interference on the transmitted signal. In particular, certain modulation methods allow a trade-off between bandwidth requirements and noise suppression.

1.2 Types of Modulation

The various methods of modulation can be divided into roughly three categories. Such division is useful in describing the basic principles of any type of modulation. However some commercial communication systems may contain characteristics of more than one type of modulation.

Continuous Wave Modulation

Continuous wave or continuous carrier modulation has a sinusoidal wave as the carrier. Some parameter of this carrier is varied in a continuous fashion by the signal to be transmitted (i.e. the modulating signal). The frequency of the sine wave carrier must be greater than the maximum signal frequency designed to be transmitted, and in general the carrier frequency is many times the maximum signal frequency.

Amplitude modulation (AM) in which the amplitude of the carrier sinusoid is the continuously varied paramenter was the first modulation method used. Amplitude modulation and various modifications of AM remain the dominant methods of modulation in electrical communication.

The other parameters of a sinusoid carrier which can be varied as a function of modulating signal are the phase, which gives phase modulation (PM), and the time derivative of the phase or the instantaneous frequency, which gives frequency modulation (FM). Both these continuous wave modulations are closely allied and are known as angle modulation.

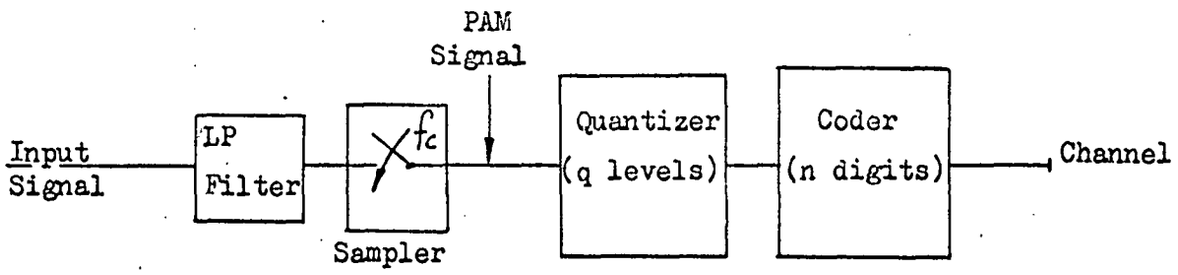
Pulse Modulation

The carrier for pulse modulation is a train of periodic pulses. Some parameter of each pulse is varied by the modulating signal. Clearly any pulse modulation system, since it transmits information only at the instant of the pulse, is a discrete or non-continuous system, For the transmission of non-continuous information the only requirement would be that the pulse rate is matched to the rate of information. The use of pulse modulation for a continuously time varying signal depends on the theory of sampling. By sampling at a rate of at least twice the maximum signal frequency (f_m) all the information of the continuous signal is contained in the samples.

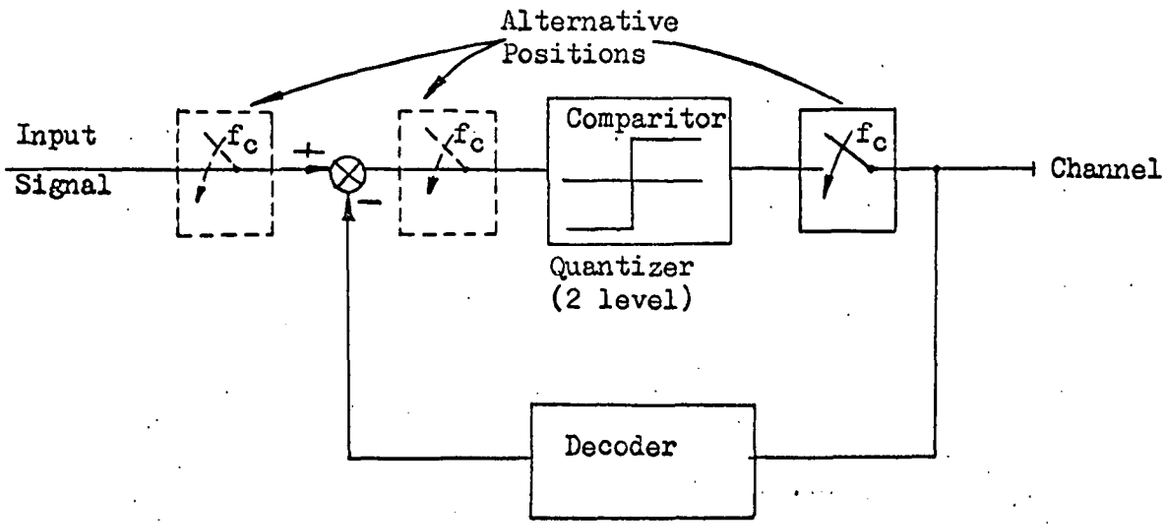
As the pulse rate is determined by the above considerations the parameters of the pulse train carrier suitable for variation are the height, width or position of each pulse. This gives respectively, pulse amplitude modulation (PAM), pulse width modulation (PWM) and pulse position modulation (PPM). Although non-continuous, these methods

FIG 1.1 Schematic of Digital Modulation Methods

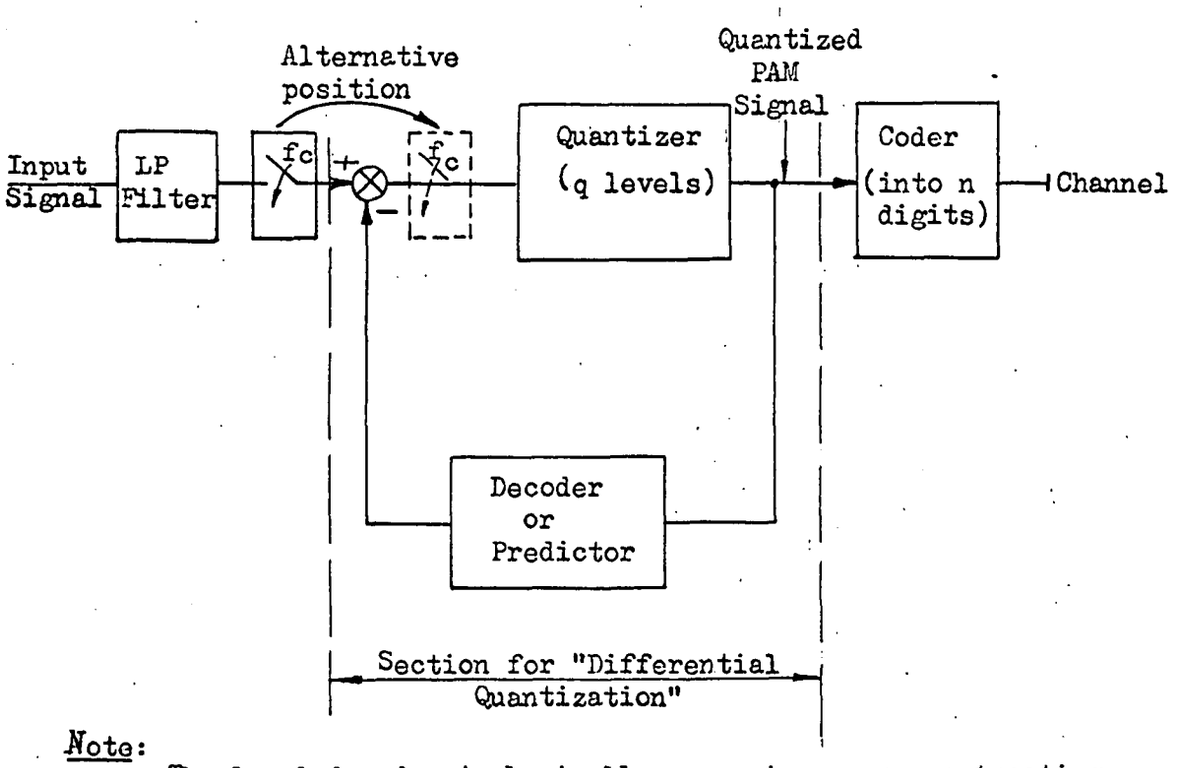
Pulse Code Modulation



Delta Modulation



Differential Pulse Code Modulator



Note:

The local decoder is basically a summing or reconstructing device . ie. a device which predicts the next input signal, sampling instant value , from the previous value.

of pulse modulation are analogue in nature in common with continuous wave modulation. This distinction can be made because the variable paramenter is being changed directly and continuously be the modulating signal.

Digital Pulse Modulation

The feature of digital pulse modulation which makes it markedly different from the other modulation methods is that the periodic pulse train carrier has standard pulses. There is no variation of any pulse parameter with the modulating signal. Instead the information is conveyed by the pulses being assigned one of several states. In general digital pulse modulation systems are binary systems. Hence it is merely the determination of the presence or absence of a pulse which conveys the information.

Because of the discrete nature of the information (either 1 or 0) contained in each pulse, the signal sample values must have discrete values. Thus quantization - the assignment to each sample value of a discrete value - is necessary. In addition, to convey the information of the quantized signal value to one or more pulses some coding is required.

Pulse code modulation (PCM) was the originally conceived method for the digital transmission of continuous time varying signals. Although the concept of PCM was fully defined by A. H. Reeves in 1938 it was not until the last decade that the advances in semiconductor switching elements made PCM links feasable for commercial use. The process of modulation using PCM (see Fig. 1.1) consists of: sampling the analogue signal at the rate required for reconstruction, quantizing the sampled values, and finally coding the quantized amplitudes into a digital format using the standard binary code. For commercial telephony 128 quantizing levels are used that are exactly specified by a group of binary coded pulses.

It is interesting to note that the rapidly growing commercial use of PCM and digital data transmission represents a return to the method used in the original electrical communication - the telegraph. The major advances in electronics however, now permit automatic coding with far more efficient coding systems.

A modification of PCM is to code and transmit only the difference between successive signal sample amplitudes. This method is known as differential pulse code modulation (DPCM) (see Fig. 1.1). The concept of DPCM was defined in the early 1950's and resulted from attempts to increase the efficiency of the digital coding of analogue signals. Because of the correlation between successive sample values of analogue signals such as speech, there is a redundancy in transmitting each sample

value. With DPCM systems there is a feedback path around the quantizer. The feedback network makes a predication of the value of the incoming sample and this is subtracted from the sample value. The feedback network "predicts" by some form of decoding or reconstruction of the differential quantizer output. The simplest system is one which gives the previous sample value as the prediction for the next value.

A special case of DPCM results if the difference value is coded with a one bit cade. This results in each sample's value being represented by only one pulse in the pulse train carrier. The presence or absence of each pulse indicates whether a unit increase or decrease in the reconstructed signal is the closest approximation to the input signal. Such a digital pulse modulation system with unity bit coding of the input sample differential is known as delta modulation (DM). (See Fig. 1.1).

Although delta modulation can be regarded as a special case of DPCM the method was proposed prior to the conception of DPCM. Due to the uniqueness of the unity bit code compared with a code of several digits, delta modulation has many characteristics which make it quite distinct from the other digital pulse modulation methods. A full discussion of the origin and concept of delta modulation is made in Section 1.4. Included is a comparison with PCM and DPCM which points out the fundamental differences and consequences of these differences regarding the practical realization of a digital communication system.

Before proceeding to a comparison of the basic modulation methods it should be pointed out that some actual communication systems are to some degree a hybrid of methods. The most frequently used commination occurs in systems where some pulse modulation method is the initial and fundamental modulation. In such cases it is common in the final stage, for the pulse train (either analogue or digital) to modulate on R.F. or microwave carrier for transmission purposes. The use of this final stage of continuous wave modulation depends on the transmission medium and the distance involved.

1.3 A Comparison of Modulation Types

To compare methods of modulation with the intention of arriving at an optimum method for electrical communication is clearly a meaningless exercise. Due to the large variety of signals, transmission conditions, distances, and quanlity requirements, the priorities of desired characteristics varies depending on the overall communication system under discussion. However in comparing the intrinsic advantages and disadvantages of any particular method of modulation the various factors which are of greatest significance are:

1. The quality of performance.
2. The average transmitted power.
3. Bandwidth required for transmission, B_T.
4. Compatibility with equipment sharing techniques such as multiplexing.
5. Suitability for use with repeaters on long distance or high attenuation transmission paths.
6. Equipment expense factors such as complexity and quantity.

In discussing these factors, terms are defined which will be standard throughout the thesis.

The quality of performance of a communication system depends on a large number of factors many of which are difficult to analyse. However the commonly used measure of quality is the output signal-to-noise ratio (SNR) defined by:

$\frac{S_o}{N_o}$ = Ratio of average signal power to average noise power at the receiver output.

Clearly the output SNR gives no indication of many quality factors such as, the nature of the noise, or the susceptibility of the system to signal distortion or interference. For example, it may be stated that where the noise is basically random white noise uncorrelated to the signal, then acceptable output SNR's for speech are 30 dB or greater. However some particular system or modulation method may give rise to noise whose nature makes a 30 dB SNR totally unacceptable due to the unpleasant sounding nature of the noise. Where such factors are of significance in the comparison of modulation methods they are mentioned separately. In order to make the output SNR a suitable measure of the noise performance inherent with the basic method of modulation alone, it is frequently expressed in terms of the input SNR defined by;

$\frac{S_{in}}{N_{in}}$ = Ratio of the average signal power to average noise power at the input to the detector (or decoder).

A more objective measure would take into account the proportion of the average transmitted power, P_T, at the detector input which is contributed by the noise power, S_{in}. If a system has a significant part of the total transmitted power, P_T, accounted for by, say, a non-information carrying carrier, then the output SNR is reduced compared to a system which devotes all of P_T to the signal power, S_{in}. For this reason it is often more meaningful to express the output SNR in terms of $P_{T/N_{in}}$ or where relevent to indicate what proportion of P_T is accounted for by S_{in}. Perhaps the best measure of relative performance of any modulation method, and the one which is used most frequently throughout this thesis, is the output SNR in terms of Z (see Carlson[2]) where Z is defined by;

	Modulation Method	Transmission Bandwidth, B_T	S_o/N_o (1)		typically	Remarks
	Baseband	f_m (2)	$\frac{S_{in}}{N_{in}}$	Z (3)		
Amplitude Modulation	DSB	$2f_m$	$2\frac{S_{in}}{N_{in}}$	$\frac{m^2}{2+m^2}$Z (4)	Z/3 to Z/10 at best 5dB below Z	Threshold effect for Z < about 13dB.
	DSB-SC	$2f_m$	$2\frac{S_{in}}{N_{in}}$	Z		Requires synchronous detection.
	SSB	f_m	$\frac{S_{in}}{N_{in}}$	Z		
Angle Modulation	NBFM	$D \ll 1, \simeq 2f_m$	$3D^2\frac{S_{in}}{N_{in}}$	$\frac{3}{2}D^2$Z	Very much below Z	Threshold effect for $\frac{S_{in}}{N_{in}}$ < about 13dB. equivalent to Z < 20 $\frac{B_T}{f_m}$
	WBFM	$D \gg 1, \simeq 2f_d = 2Df_m$	$3D^3\frac{S_{in}}{N_{in}}$	$\frac{3}{2}D^2$Z	For D = 5, 16dB above Z	Capacity for S_o/N_o improvement using de-emphasis, typically 12dB.
	PM	$\simeq 2(\phi_d + 1)f_m$		$\frac{1}{2}\phi_d{}^2$Z	at best 7dB above Z, typically Z	for $\phi_d \ll \pi$, have NBPM which is very similar to AM.
Analogue Pulse Modulation	PAM	$\geqslant f_m$	$\geqslant \frac{S_{in}}{N_{in}}$	$\leqslant$ Z		Highly suitable for TDM in which case $B_T = nf_m$ where n = no. of signals multiplexed.
	PPM	B_T	$\leqslant \frac{1}{16}\frac{B_T}{f_m}^3\frac{S_{in}}{N_{in}} \leqslant \frac{1}{16}\frac{B_T}{f_m}^2$Z		of the order of 10dB lower than $(S_o/N_o)_{max}$	Exhibits wide band noise reduction properties.
	PWM	B_T	$\leqslant \frac{1}{4}\frac{B_T}{f_m}^2\frac{S_{in}}{N_{in}} \leqslant \frac{1}{4}\frac{B_T}{f_m}$Z			
	PCM	pf_m typically $7f_m$ for speech		$\frac{3}{2}2^{2p} = \frac{3}{2}2^{2(\frac{B_T}{f_m})}$	for p = 7, S_o/N_o = 44dB	Threshold effect for S_{in}/N_{in} < about 15dB. ie. Z < 30 B_T/f_m = 30p.

FIG 1.2 A Comparison of Various Modulation Methods

Notes:

(1) these expressions are assuming a mean square signal value of $\frac{1}{2}$ the amplitude squared, ie. they are based on a sinusoidal signal.

(2) f_m = max. signal frequency and the signal bandwidth.

(3) Z = Output SNR of baseband transmission under the same transmission conditions.

(4) m = modulation ratio = max. modulating signal amplitude A_{max} / carrier amplitude

(5) f_d = max. frequency deviation.

(6) D = deviation ratio = $\frac{f_d}{f_m}$

(7) ϕ_d = phase modulation index ($\leqslant \pi$)

Z = Output SNR of baseband transmission under the same transmission conditions.

Hence, Z = Ratio of transmitted power to noise power in the message bandwidth.

$= \frac{P_T}{\eta W}$ Where the noise is assumed to be additive white noise of density η and W = message bandwidth.

The performance figures in terms of Z are in terms of the channel quality taking into account both the transmission power and the noisyness of the channel and receiving equipment.

The bandwidth of frequencies required in the transmission channel by a particular system for the transmission of one signal is called the transmission bandwidth and is denoted by B_T. This factor becomes particularly significant when it is desired to transmit many individual signals over a common limited bandwidth transmission facility. The common usage of a transmission channel is achieved by using either frequency division multiplexing (FDM) or time division multiplexing (TDM) [2,4]. TDM has two significant advantages over FDM, namely simpler circuit implementation and relative immunity to interchannel interference or crosstalk. Any pulse modulation system (digital or analogue) can be multiplexed with TDM, whereas FDM is only suitable for use with continuous wave modulation. Where multiplexing is desired (as in long distance telephone communication) the multiplexing method suitable to a particular method of modulation must be considered in any comparison of modulations.

In selecting modulation methods to be used in realizing a communication link there are constraints other than those mentioned already which are not related to the fundamental nature of the method of modulation. For example a new system, although apparently superior for some particular communication task to the currently used system, may well be slow in being implemented due to the well established technology of the existing system. Also the compatibility of a new system to the existing facilities may impede its utilization. However such constraints will not be taken into account in this thesis in the discussion of any modulation system.

1.3.1 Continuous Wave Modulation

Conventional analysis [2,3] gives estimates of the output SNR for c-w modulation methods as shown in Fig. 1.2 along with other relevant characteristics. Such analysis is based on the reasonable assumption that the dominant signal contamination occurs in the channel and early receiver stages and results from white noise, which is independent of the signal. The limitation of the output SNR of AM- DSB modulation results from the signal power being only a small part of the

FIG 1.3 A Comparison of Performance for Various Modulation Methods

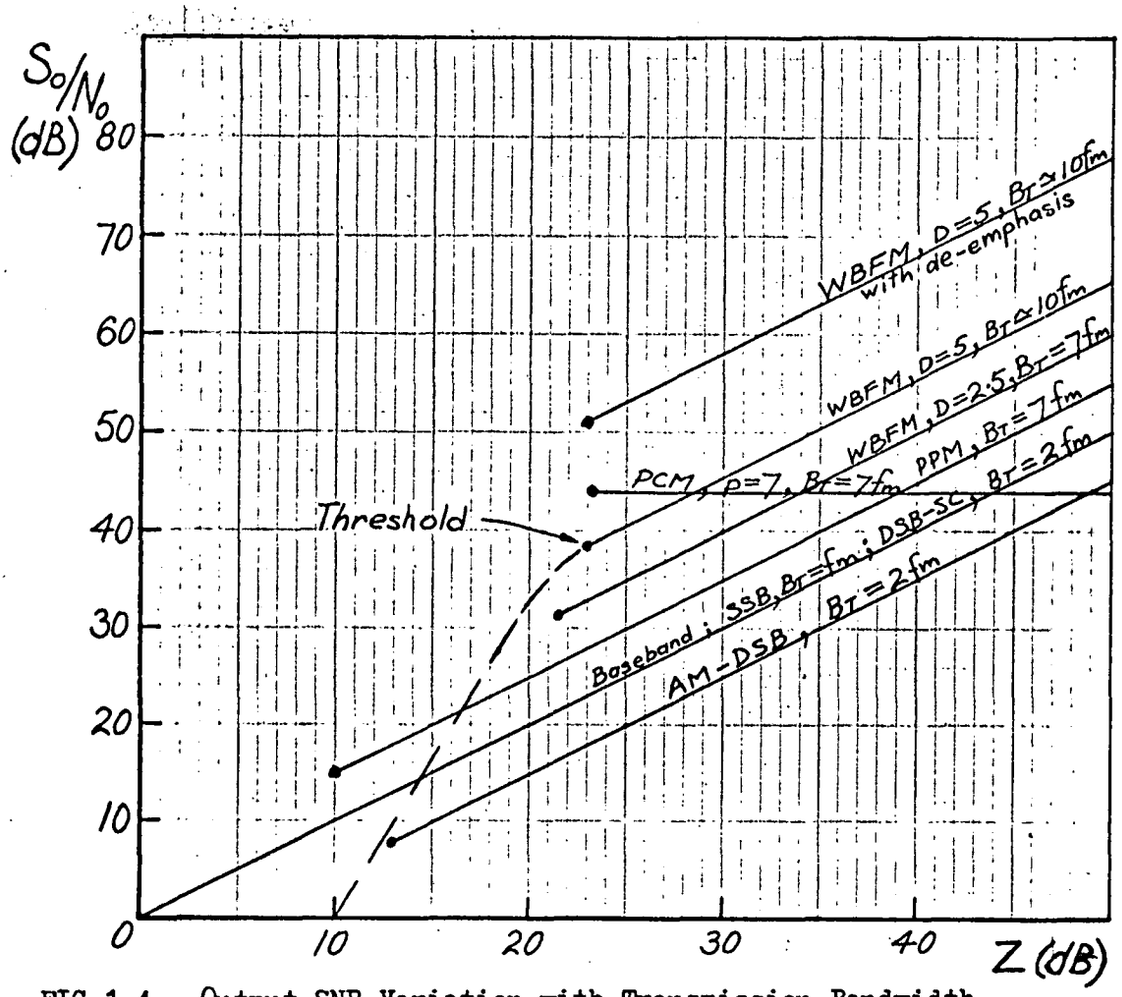
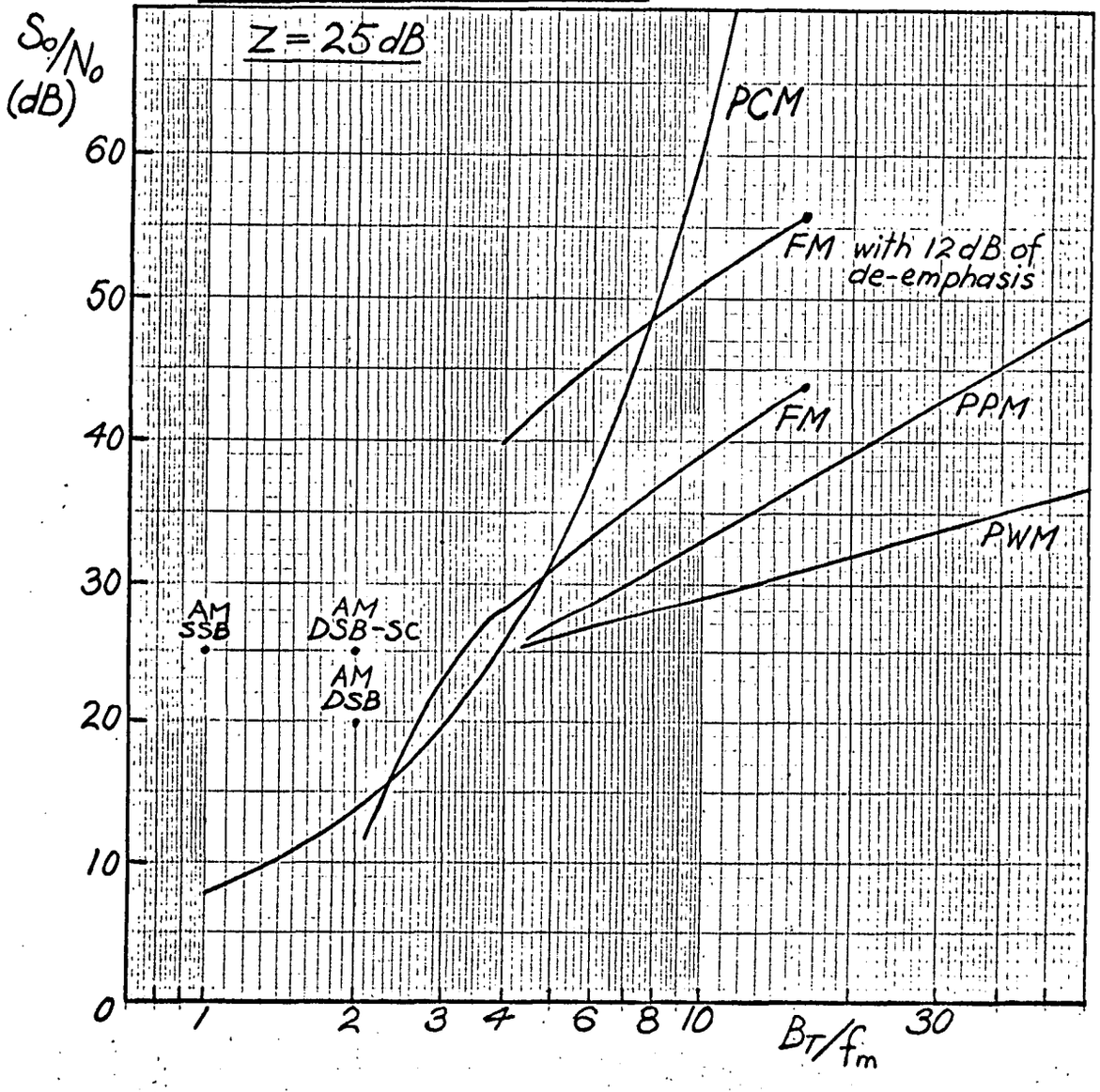


FIG 1.4 Output SNR Variation with Transmission Bandwidth for Various Modulation Methods



transmitted power ($S_{in} = \frac{m^2}{2+m^2} P_T$). As full modulation cannot be maintained in general (i.e. $m < 1$), over 2/3 of the transmitted power will be wasted in the non-information bearing carrier. This, combined with the transmission bandwidth required, makes the performance of AM-DSB compare poorly with SSB modulation. However, the suppressed carrier of SSB or DSB-SC makes the receiver circuitry considerably more complex and thus DSB is used in broadcasting where many receivers of a single channel are involved. The noise performance of SSB degenerates considerably where signal discontinuities are significant and thus DSB or DSB-SC are used for the continuous wave modulation of pulse signals.

Narrow band FM shows no basic improvements over AM but wide band FM has the valuable characteristic, which is frequently required, of wideband noise reduction.

Fig. 1.2 shows the potential WBFM has for large increases in the output SNR relative to baseband transmission. This is achieved at the expense of a greatly increased channel bandwidth requirement. Wideband noise reduction with phase modulation (PM) is not possible as the phase modulation index ϕ_d is constrained to be less than π. Thus PM compares poorly with AM for analogue signals but finds its use in digital pulse train modulation of h.f. carriers (i.e. Phase shift keying or PSK) where it strongly resembles DSB-SC amplitude modulation by digital signals (i.e. amplitude shift keying or ASK). FM is also highly suited to deemphasis filtering by which, typically, an extra 12dB of output SNR can be obtained without any increase in the required transmission bandwidth. Indefinite output SNR improvement for WBFM cannot be achieved by increasing D,(and hence B_T) as a point will be reached where the increased input noise N_{in}, due to increasing B_T, starts to cause severe signal distortion. Therefore WBFM suffers from a threshold effect which severely limits the output SNR for S_{in}/N_{in} of about 13dB or lower (see Fig. 1.3). This represents a channel capacity limit corresponding to Z = about 20 B_{T/f_m}. Clearly this also puts a limit on increasing the transmission bandwidth in favour of reduced transmission power,or the use of extremely noisy transmission paths.

1.3.2. Analogue Pulse Modulation

The benefits of PAM result from its suitability to TDM and it thereby inherits the advantages of TDM over FDM. In fact PAM and also PPM and PWM would almost invariably be used with TDM to give multi-channel use of a single transmission path. The bandwidth and noise performance of PAM are in general inferior to AM but they can tend towards baseband performance.

PPM and PWM are analogous to angle modulation, in that the varying parameter is in the time domain and they exhibit wideband noise reduction.

Fig. 1.2 illustrates the good wideband noise reduction of PPM, with the output SNR increasing with B_T at the same rate as for WBFM (i.e. $S_0/N_0 \propto B_T^2$). The maximum output SNR for PPM falls about 8dB below WBFM for the same conditions. However a practical PPM system could be of the order of 18 dB below WBFM in noise performance. PWM shows wideband noise reduction which is poor compared to FM or PPM. A threshold effect similar to that experienced by FM occurs with PPM and PWM. With the pulse modulations the threshold results from noise peaks which give false pulses if the transmission bandwidth, and hence the noise power, exceeds a threshold level.

As well as its noise reduction property, PPM along with the other analogue pulse modulations has the advantage of having the potential to use an on-off transmitted signal, which gives rise to greater transmitter efficiency.

If, in the final stage of an analogue pulse modulation system, c-w modulation of a high frequency carrier is performed (e.g. ASK or PSK) for transmission purposes, then the transmission bandwidth required (as shown in Fig.1.2) is in general doubled or more than doubled. (The only exception is where a multisignal PAM-TDM pulse train is low pass filtered and transmitted as an analogue signal using AM-SSB.) Overall, analogue pulse modulation systems do not compare favourably with c-w modulation from either the bandwidth or noise performance viewpoints. Their use depends heavily on the desirability of their special merits mentioned previously. The superiority of TDM over FDM however would have guaranteed PAM and PPM a major role in long distance, multi-signal, point to point telephony had not their use been short-circuited by PCM.

1.3.3. Digital Pulse Modulation

The nature of the noise and the noise performance of digital systems is unique, due to the basically different nature of information coding as discussed in Section 1.2. Sampling of a band limited signal involves no contribution to noise but the amplitude quantization of a sampled signal does introduce an error or noise function. This noise results from the loss of information when each sampled value is restricted to take on a discrete value and is known as the quantization noise. In digital systems no attempt is made to transmit the exact signal, but a coded pulse train carries the information of an approximation to the signal, namely the quantized signal. This gives digital systems unique characteristics, listed below, in which the merits of its use lie.

1. The transmitted signal consists of constant size, periodic pulses, the presence or absence of each pulse being the only information required of the transmitter carrier. Thus, provided the noise power is below some threshold level, so that virtually no pulses are lost or false pulses added, there will be no significant loss of information at the receiver and hence no random channel noise at the output.

2. The digital pulse train can be detected and regenerated with the rejection of all accumulated noise and distortion at repeater stations. This is a unique advantage over analogue modulation methods, where the signal connot be distinguished from the signal plus noise and thus repeaters amplify both signal and noise, as well as introducing some further distortion.

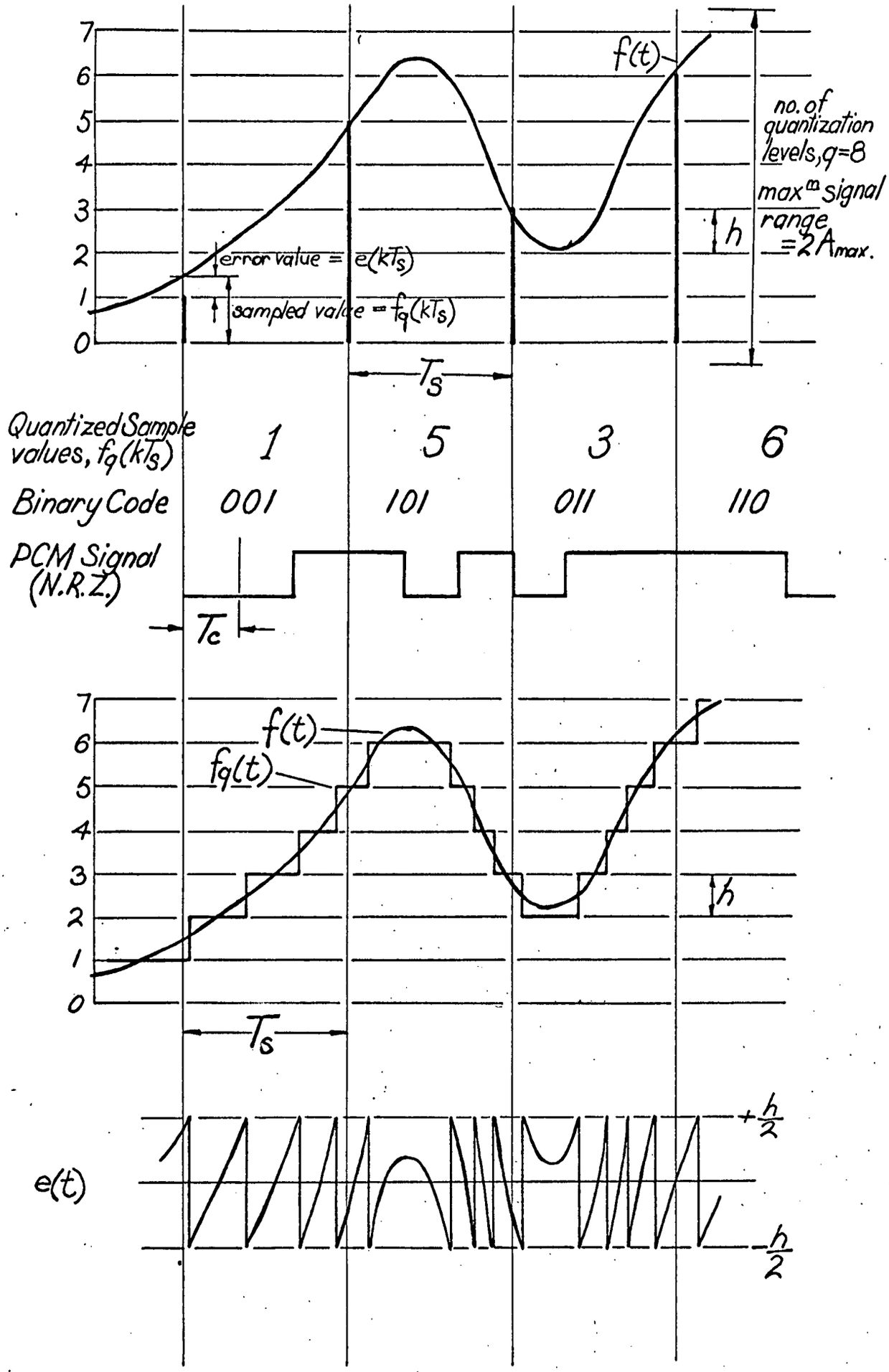
3. Whereas the noise performance of analogue systems is limited by random noise in the channel and by distortion, the noise performance of digital systems is virtually completely determined by the error that results from quantization. Furthermore, provided the channel SNR is above the threshold value, the quantization noise is determined totally by the spacing of the quantizing levels. By spacing the quantizing levels as closely as desired, (with a corresponding increase in the number of quantizing levels), the output SNR can be made as high as desired. As the quantizing level spacing is reduced, and hence the output SNR is increased, there is an increase in the transmission bandwidth required. This indicates that digital systems exhibit wideband noise reduction as will be shown later.

As well as their special advantages, digital pulse systems share the advantages of analogue pulse modulation of being suitable for use with TDM and having an on-off transmission signal. In addition, digital modulation systems, having dominantly digital hardware are highly suited to large scale IC realization, and are to a large degree, compatible with digital data transmission.

Pulse code modulation, having a binary code, has the number of quantizing levels (q) related to the number of pulses (or digits) required to uniquely specify each level (p), by $q = 2^p$. The transmission bandwidth required as shown in Fig. 1.2 is based on that required for the resolution of non-return-to-zero (NRZ) or full width pulses. If half width pulses were used in order to facilitate simpler clock synchronization at the receiver, then the required transmission bandwidth would be doubled.

The analysis leading to the output SNR expressions for PCM, as shown in Fig. 1.2, will be considered in more detail than the previously discussed systems because of its significance in later discussion of delta modulation noise. Some textual discrepency for PCM output SNR estimates occurs, due to varying defintions of the output signal.

FIG 1.5 Signal Quantizing and Coding for PCM



The input signal, f(t), will be considered to be limited by the PCM system, to a maximum value of A_{max} (i.e. $-A_{max} < f(t) < A_{max}$). The peak factor, β, of any signal is defined as the ratio of the maximum allowable signal value to the rms signal value. Thus $\beta = A_{max} / \sqrt{\overline{f^2(t)}}$. Fig. 1.5 shows the quantization and coding of a typical signal using a 3 digit coded PCM system (i.e. p=3, q=8). From this figure it can be seen that for uniformly spaced quantizing levels with spacing h, then the maximum error of any sample is $\pm h/2$. The received quantized sample values can be regarded as the sum of the sampled signal value plus the quantization error value as below:

$$f_q(kT_s) = f(kT_s) + e(kT_s) \qquad \text{where } -h/2 < e(kT_s) < +h/2$$

It is shown (refer Sakrison[1] or Bennett[5]) that for a signal with equal probability of having any value between $\pm A_{max}$ then, since $e(kT_s)$ will have a probability density function uniformly distributed between $\pm h/2$, the quantization noise power, $\overline{e^2}$, is given by

$$\overline{e^2} = h^2/12$$

Sakrison[1] also shows for a signal which is not uniformly distributed that $\overline{e^2} = h^2/12$ is a good approximation, provided h is sufficiently small.

Since the maximum peak to peak signal value, $2A_{max}/q$ is equal to qh then $h = 2A_{max}/q$. As we are considering transmission with the channel SNR sufficiently high to render bit errors insignificant (i.e. above threshold) then the output noise is given by:

$$N_o = \overline{e^2} = \frac{A_{max}^2}{3q^2} = \frac{A_{max}^2}{3x2^{2p}} \quad , \quad \text{since } q = 2^p$$

The output signal power is $S_o = \overline{f^2(t)} = \frac{A_{max}^2}{\beta^2}$

Hence: $\frac{S_o}{N_o} = \frac{3x2^{2p}}{\beta^2}$

For a sinusoidal signal of maximum amplitude; $\beta^2 = 2$ and $S_o/N_o = 3/2\ 2^{2p}$. For a signal with a uniform probability density function over the full range, (such as a triangular wave with amplitude A_{max}), then $\beta^2 = 3$ and $S_o/N_o = 2^{2p}$.

For equispaced quantizing levels, an input signal with uniform distribution over the full range is preferable. Typical input signals, such as speech, are invariably non-uniformly distributed with a predominance of low signal values. (i.e. $\beta^2 \ll 1$). For this reason instantaneous companding, in the form of non-linear compression of the input signal before sampling, with complementary expansion at the receiver output, is usually

employed. Such companding, designed to give a signal with as close to uniform distribution over the full input range as possible, is easier to implement than a PCM system with the quantizing level spacing adjusted to match the signal characteristics. With companding the value of $\beta^2 = 3$ can be approached for typical signals and so the estimate of the output SNR is given as $S_o/N_o = 2^{2p}$. However, to keep a uniform basis for comparison with other systems the slightly higher estimate obtained with a sinusoidal signal of full amplitude is given in Fig. 1.2.

The minimum channel or input SNR before bit errors become significant is the threshold level and is estimated [1,2] at: (S_{in}/N_{in}) Threshold = 30 (15dB). This corresponds to $Z_{th} = 30\ B_T/\ f_m = 30p$.

It is significant to bear in mind that for a typical commercial PCM system for telephony with a 7 digit code (p=7) the companding is such that it reduces the output SNR from 44 dB as estimated by the equation of Fig. 1.2 to about 32 dB. The companding is designed, however, to give a large dynamic range of input signals of about 35 dB.

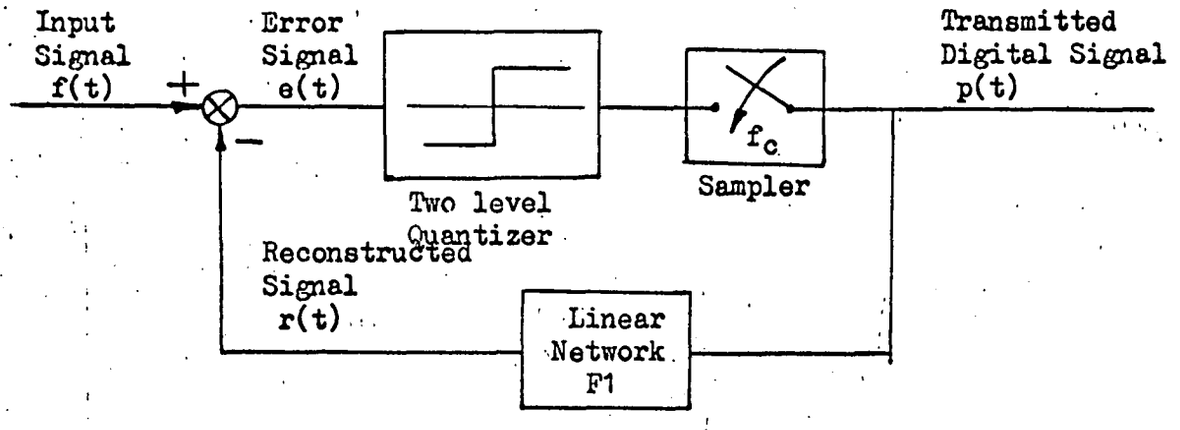
The noise performance analysis for PCM indicates that in addition to the advantages of digital pulse modulation discussed earlier, PCM exhibits a more powerful wideband noise reduction than the other modulation methods (refer Fig. 1.4), with an exponential increase in output for increasing transmission bandwidth. PCM, in common with all digital systems shows no output SNR improvement with improving channel conditions once threshold has been reached. Operating just above the threshold, Fig. 1.3 indicates that for typical speech transmission requirements, and for comparable transmission bandwidth requirements, PCM is considerably superior to the best analogue pulse modulation (PPM) and is only inferior to FM with deemphasis.

Because of all the factors mentioned, and in particular because of its capacity for noiseless regeneration and use with TDM; and because the quite complex quantizing and coding equipment can be shared by all the multiplexed signals; PCM is rapidly growing in use for long-distance telephone transmission.

Differential pulse code modulation (DPCM) has been shown [6,7] to be capable of output SNR improvement if the normalised correlation between adjacent samples is greater than about 0.5. McDonald[7] estimates that for speech signal transmission, DPCM can give from 6 to 10 dB increase in the output SNR relative to PCM for the same conditions. Alternatively DPCM can give performance on a par with commercial PCM with a 6 digit instead of the 7 digit code group for PCM thus giving a reduction in the transmitted pulse rate from 56 to 48 kHz. However DPCM suffers from a severe equipment problem when use with TDM is required; as it frequently would be. This problem arises because a differential quantizer/coder

FIG 1.6 Schematic of Basic Delta Modulation System

Modulator or Coder



Local decoder

Demodulator or Decoder

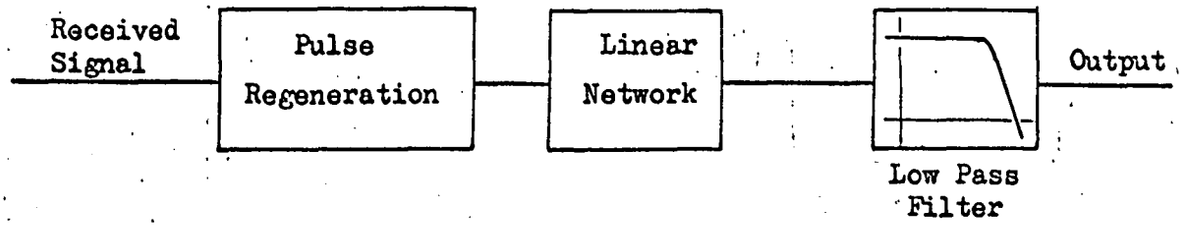
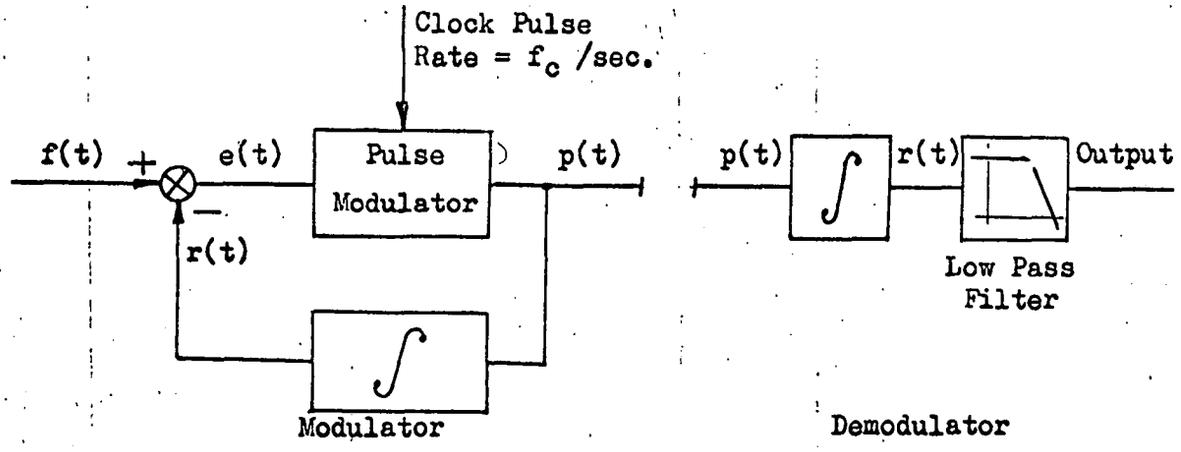
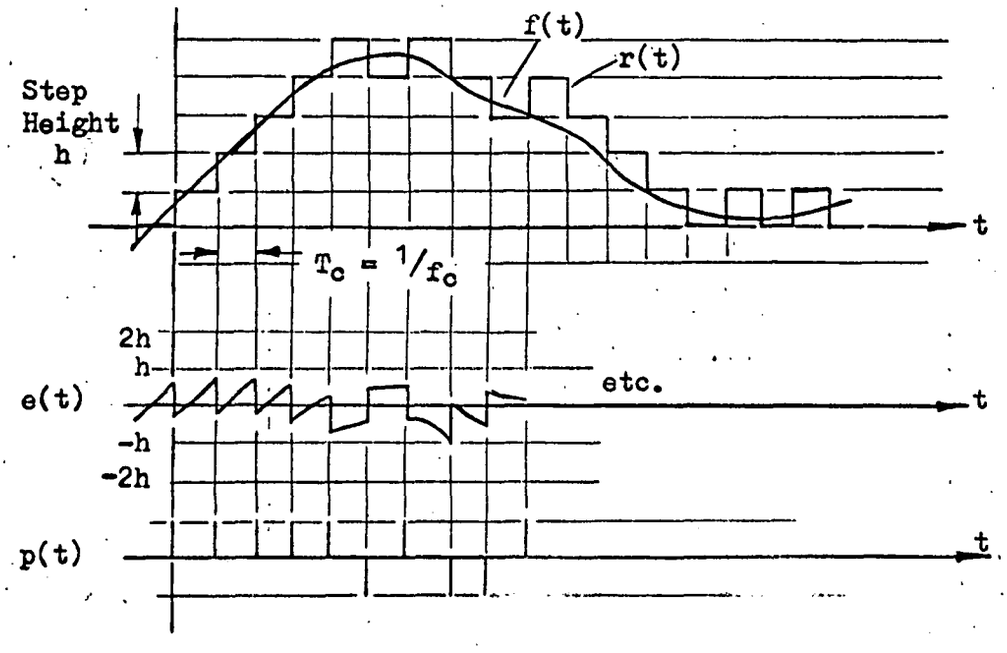


FIG 1.7 Single Integration Delta Modulation





requires consecutive samples to be from the same signal. Therefore the method of sampling and multiplexing many signals prior to their input into a common quantizer/coder as is done with PCM is not so readily available to DPCM.

Some further discussion of this problem as it relates to delta modulation is included in Section 1.4.

1.4 Delta Modulation.

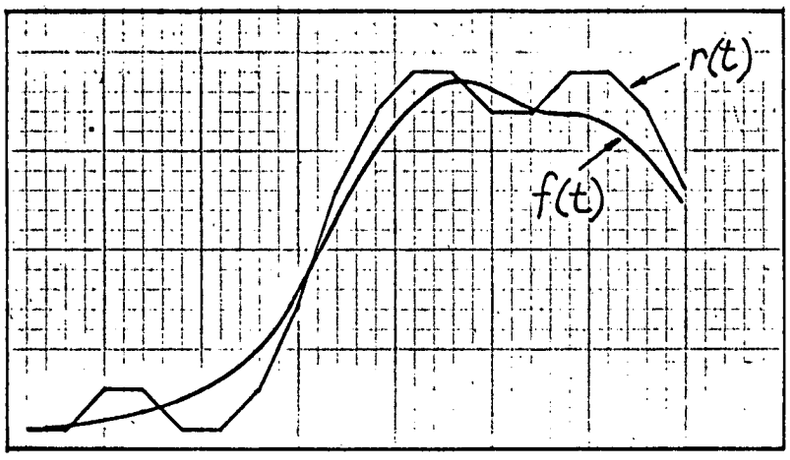
1.4.1 The Basis of Delta Modulation

The concept of delta modulation (DM) as a method for converting analogue signals into digital form was first noted in a French patent in 1949 and was fully described by Schouten, [8] and Jager and Greefkes [8,9] in 1952. The basic method is to code into a one-bit digital code the differenc between the input signal to be transmitted, f(t), and a "predicted" or reconstructed signal, r(t) (See Fig. 1.6). The reconstructed signal is obtained by passing the modulated pulse train through a linear network (Fl). Reconstruction of the signal at the receiving end is achieved by pulse regeneration, (the first step for signal recovery in all digital modulation systems) followed by a linear network, generally the same as Fl, and finally low pass filtering to remove noise components outside the signal bandwidth. Thus the feedback path in the delta modulator can be regarded as a local decoder.

Coding of the difference or error signal, e(t), into a one digit code is achieved using a two-level quantizer and a periodic sampler (or pulse modulator). If the quantizer output is high at a sampling instant, indicating that e(t) is positive, then a pulse (or binary one) is transmitted and r(t) increases in an attempt to form a better approximation or prediction fo the input signal. Likewise if the quantizer output is low, a binary zero (no pulse or a negative pulse) is transmitted and r(t) is decreased. By virtue of the binary nature of the quantizer, the sampling could alternatively be performed directly before the quantizer, without altering the system's performance. The only values which affect the output are those at the clocking instants.

The most basic and first delta system considered, used an integrato in the feedback path with discrete narrow pulses for the output digital signal. Fig. 1.7 shows the system and the wave-forms that result assuming perfect integration and ideal impulses. The receiver output is the same as r(t) after low pass filtering and the output noise signal is very close to e(t) after low pass filtering. The error signal is not identical to the unfiltered noise, as r(t) is delayed behind f(t) by some

FIG 1.8 Waveforms of the Input and Output with a Double Integration Feedback Network



Maximum slope change of r(t) = hf_c in T_c secs.
Therefore ; Maximum rate of change of slope change = hf_c^2

FIG 1.9 Illustrating Output Instability with a Double Integration Feedback Network

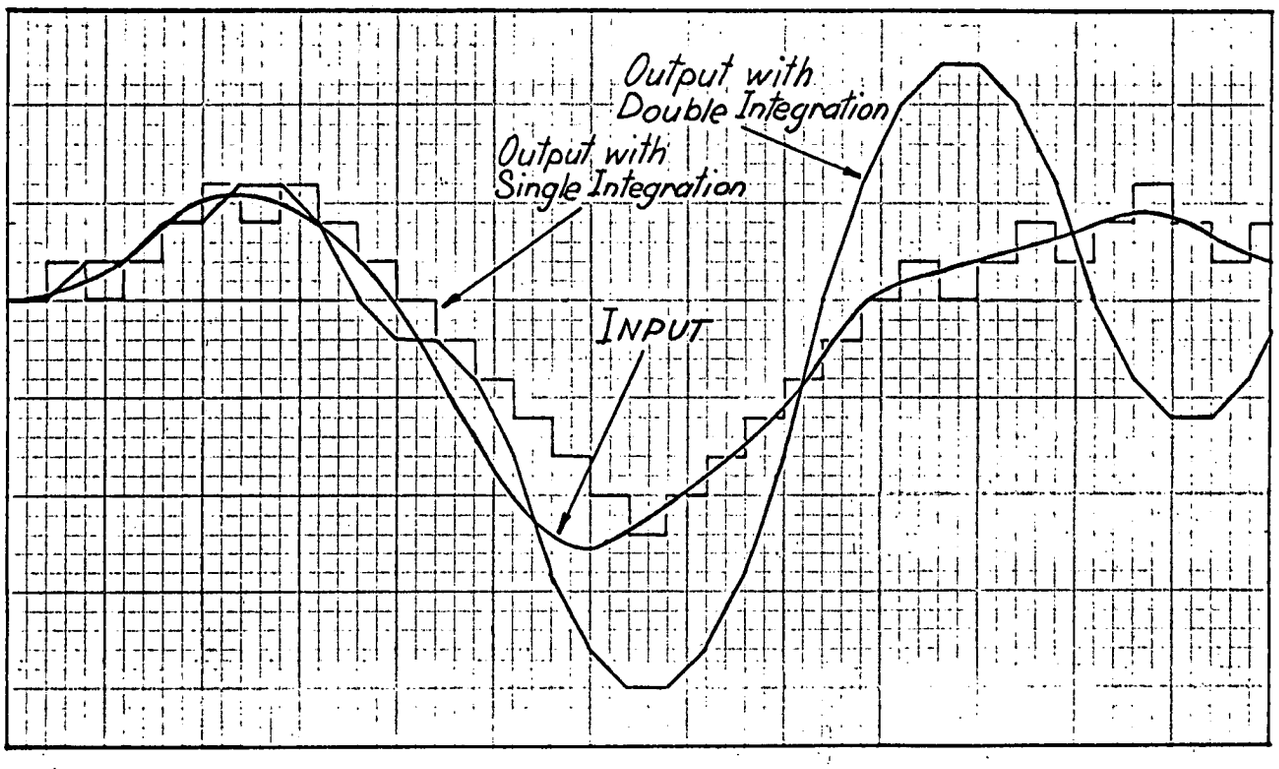
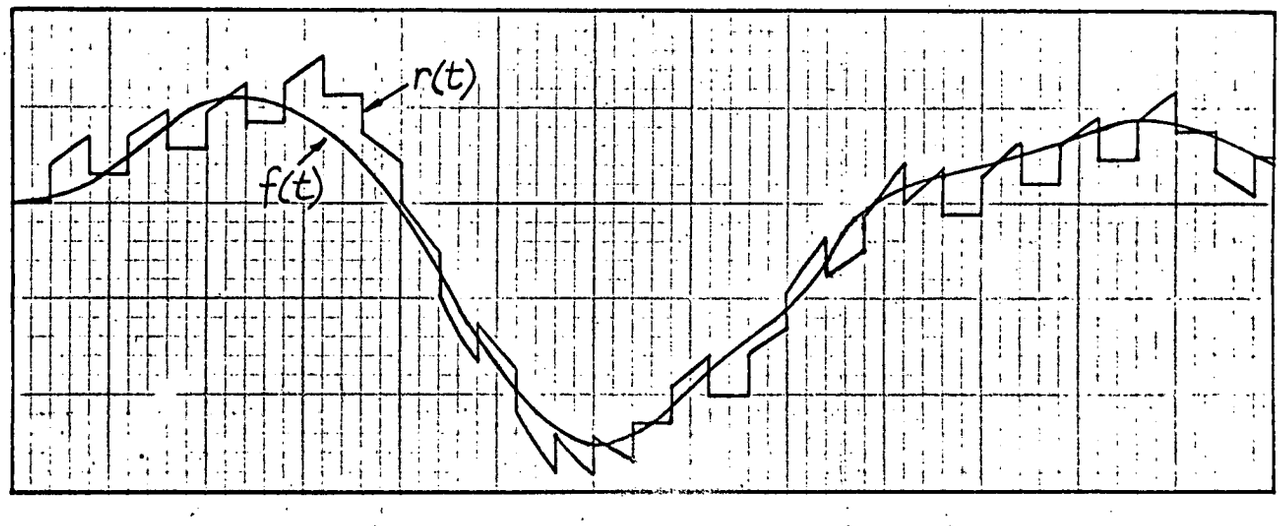


FIG 1.10 Waveform of Output Resulting from a Mixed Integration Feedback Network with a High Frequency Boost to Prevent Instability



fraction of a clock period, giving e(t) some input signal component, as will be shown later. Such a system has an amplitude quantized reconstructed signal, similar to a decoded PCM signal, but certain important differences exist.

1. Each quantized level is constrained to be one unit higher or lower than in the previous sampling period.

2. In such an ideal system there is no limitation on the maximum input signal amplitude as exists with PCM. The limitation on r(t) and hence the input signal, is an amplitude rate of change limitation with the maximum slope given by $f_c h$, where h is the quantizing level spacing or the height of the step response of the integrator.

3. With PCM the quantizing error at any sampling instant is h/2 or less in magnitude as discussed in Section 1.3. For DM, as a consequence of the first point, the error signal at the sampling instant, $e(kT_c) = e(kT_c+)$, can have a magnitude up to h provided the signal slope, f'(t), is within the maximum value

4. To sample at a rate greater than the Nyquist rate with PCM does little to enhance the performance and is wasteful with transmission bandwidth. This contrasts with DM which requires a sampling rate considerably greater than the Nyquist rate in order to have the capability to handle a reasonable range of input signal slopes. Moreover for a given maximum signal slope, increasing the sampling frequency allows the quantizing level spacing to be reduced and thus gives a reduction in the output noise.

1.4.2 Double Integration.

In order to give the reconstructed signal a greater capacity to adapt to the input signal, a double integration feedback network has been proposed[9]. Each binary one output gives a unit slope increase in r(t) and each binary zero output a unit decrease in slope. Typical wave forms for a double integration feedback are shown in Fig. 1.8. From this figure it can be seen that a greater dynamic range of input signal can be handled compared with single integration, giving a correspondingly greater output SNR. The limit on the input signal dynamics, so that r(t) can track the input signal, would be expected to be a limit on the rate of change of slope of f(t) given by $f''(t)_{max} = hf_c^2$. However Fig. 1.9 illustrates that for certain signals whose second derivative does not exceed the limit, instability can arise resulting in oscillations of r(t). Just as with single integration the amplitude of r(t) overcompensates, giving a stepwise approximation of f(t); so with double integration the slope of r(t) overcompensates, but with the possibility of far more serious consequences in the form of instability in r(t).

FIG 1.11 Overload Characteristics Compared with Spectrum of Human Voice

for a sine wave input with f_o = 56kHz.

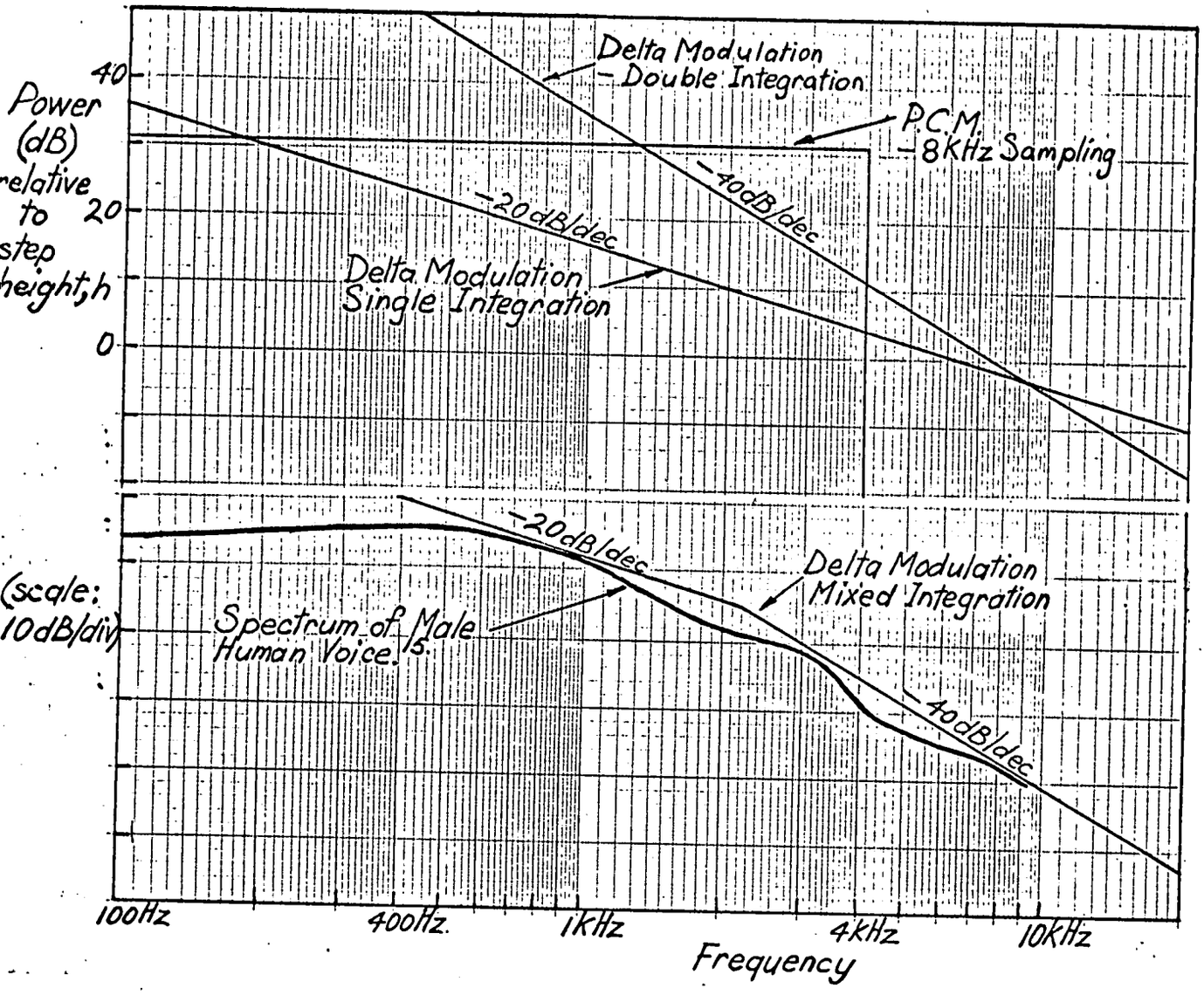
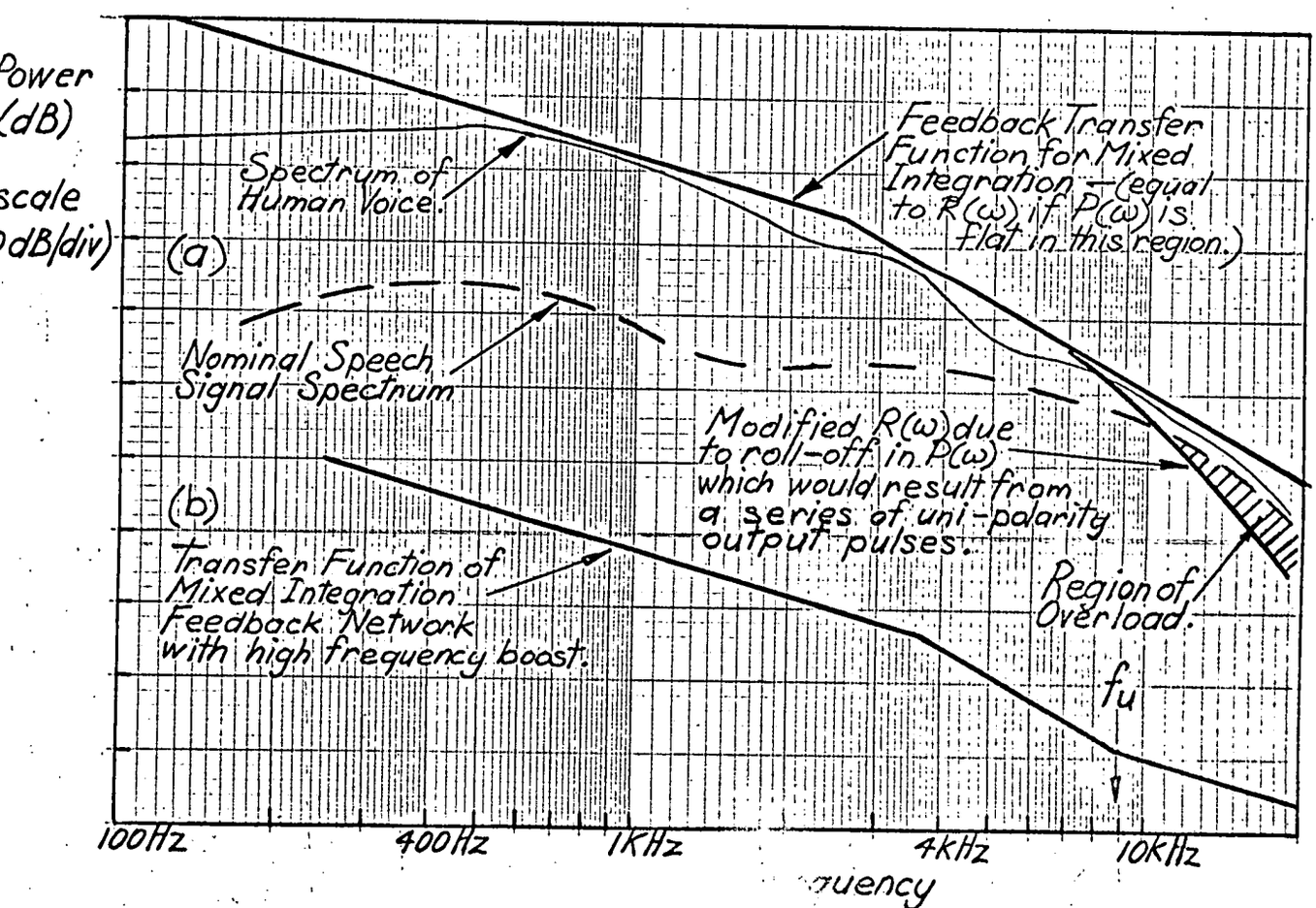


FIG 1.12 Illustrating the Need for High Frequency Boost in the Feedback Transfer Function to Prevent Instability for Double or Mixed Integration Delta Modulation.



Whether oscillations in r(t) will occur depends on the characteristics of the input signal. For a particular type of signal it would seem reasonable that a combination of single and double integration could provide optimum approximation to the input by r(t), while avoiding the instability problem.

1.4.3 Optimum Feedback Network

With direct amplitude quantization as occurs with PCM, there exists the same amplitude capacity, A_{max}, for all signal frequencies. For input signals such as speech when the amplitude spectrum[15] falls with increasing frequency (see Fig. 1.11), this results in redundancy in the systems' amplitude capacity for the higher signal frequencies.

Considering again, for single integration DM, the limit on f(t) to prevent overload. The maximum slope capability of r(t) is $f_c h$ and so for a sinusoidal input signal, $f(t) = A\sin 2\pi ft$, the maximum amplitude, A_{max}, to prevent slope overload is given by $f'(t)_{max} = A_{max} 2\pi f = f_c h$. Hence $A_{max} = \frac{f_c h}{2\pi f}$. Thus, for a given single integration DM system the input amplitude capacity decreases at 20dB per decade for increasing signal frequency. This indicates a good adaption of DM to voice communication. For a double integration feedback network, neglecting the stability problem, the limit on a sinusoidal signal to prevent overload is given by:

$$f''(A)_{max} = (2\pi f)^2 A_{max} = h f_c^2.$$

Hence, $$A_{max} = \left(\frac{f_c}{2\pi f}\right)^2 h$$

Thus for double integration the overload characteristic has the maximum signal amplitude falling at 40dB per decade for increasing signal frequency. This direct relationship between the feedback network transfer function and the overload characteristic means, that by suitably combining single and double integration in the feedback path, the overload characteristic can be made to correspond fairly closely to the signal spectrum. This mixed integration provides a reduction of redundancy in the signal amplitude capacity of the modulator as shown in Fig. 1.11.

An alternative approach to the question of the optimum feedback network is one based on the principle that the maximum information will be carried by each pulse when the removal of inherent signal redundancies by the use of negative feedback is optimized.

It would appear that the maximum removal of signal redundancy will be achieved when r(t), the reconstructed or "predicting" signal, depends on each previous sampling instant value to the same degree as the input signal depends on its previous sampling instant values. Based on this it

would be reasonable to attempt to give r(t) on autocorrelation function which was the same as the average autocorrelation function of the input signal for time displacements greater than or equal to T_c. This is a matter of trying to achieve a power spectrum for r(t) which corresponds to the average power spectrum of the input for frequencies below f_c. A similar approach was briefly used by de Jager[9] to arrive at his "double integration" feedback network. Based on the assumption of a flat (white noise) spectrum for the output pulse signal, p(t) over the frequency range of interest; de Jager concluded that for voice communication, the feedback network's transfer function should be directly related to the average speech spectrum.

Lacking evidence as to the exact nature of the output pulse signal, it will be assumed as a first approximation that p(t) has a random polarity, so that some idea of the power spectrum, $P(\omega)$, of p(t) can be obtained in order to make an estimate of the optimum feedback transfer function. This assumption gives a pulse signal power spectrum which is flat for all frequencies as shown in Appendix A, From this it would be concluded that a feedback transfer function which matched the average input spectrum would give a spectrum for r(t) which also matched the average input spectrum, as desired. (It should be noted that, while being matched, the power spectrum or r(t) should slightly exceed that of the input, particularly at the higher signal frequencies, so that higher order overloads cannot occur.) However, as shown in Appendix A, as the probability of no change in the output pulses' polarity increases beyond one half, so the power spectrum of the pulse train begins to roll-off in the frequency range $0 - f_c/2$. Such a situation would be characterized by series of successive output pulses with the same polarity. The effect would be to reduce the higher frequency (up to $f_c/2$) power of r(t). If the input signal (which will be taken to be speech) contained considerable power at some higher frequencies (as shown in Fig. 1.12) when the higher frequency spectral density of r(t) was reduced as discussed, then the signal power could exceed the capability of r(t) in this frequency region. This would result in a higher order overload, e.g., r(t) being unable to match the derivative of the slope rate of change of f(t) (i.e. $f'''(t)$) or some combination of $f''(t)$ and $f'''(t)$. Such a situation is shown diagramatically in Fig. 1.12. It is hypothesized that such higher order derivative overload results in the onset of an oscillatory wave in r(t).

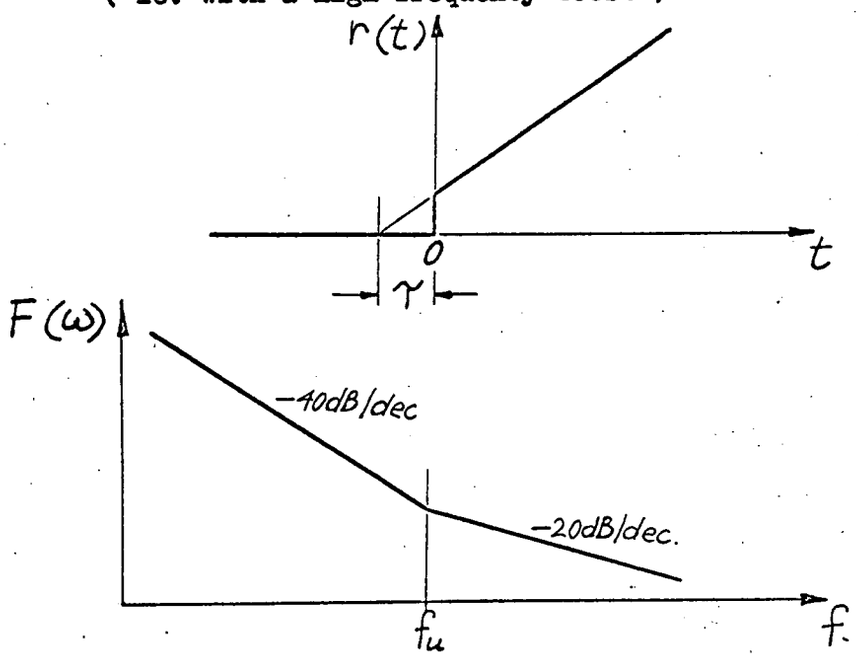
Considering again Fig. 1.9 it is observed that associated with the onset of the oscillation there is a fairly rapid change in the input signal slope which would be associated with a high frequency signal component. Also the preceeding output pulse signal has a series of consecutive pulses of the same polarity. This correlation between the noted conditions at the onset of oscillation and those predicted to be associated with a higher order derivative overload, supports the hypothesis that oscillation results from a higher order derivative overload. It can also be seen from Fig. 1.9 that the greater the number of consecutive pulses

of the same polarity in p(t), the greater the amplitude and lower the frequency of any ensuing oscillation.

A suitable corrective measure to prevent oscillation, based on the above arguement, would be to include a high pass section in the feedback network which takes effect at the higher signal frequencies, in order to compensate for the power spectrum roll-off. The frequency at which the roll-off in $P(\omega)$ becomes significant depends on f_c and P_{nc}, as shown in Appendix A, where P_{nc} is the probability of no change in the output pulse trains' polarity between successive clock intervals. Typically, for $P_{nc} = 3/4$ and $f_c = 56$ kHz, the power of p(t) is down 3dB at 10kHz. To make a reasonable estimate of the desired frequency, f_u as shown in Fig. 1.12 at which the high frequency boost to r(t) should take effect, the maximum P_{nc} for given system and signal parameters should be known.

De Jager[9] consideres the stability problem with double integration as one of a lack of "prediction". He proposes a partial bypass of the second integrator in order to give an immediate step response in r(t) prior to the slope change. This is achieved with a feedback network having an impulse time response as shown in Fig. 1.13, with the corresponding transfer function

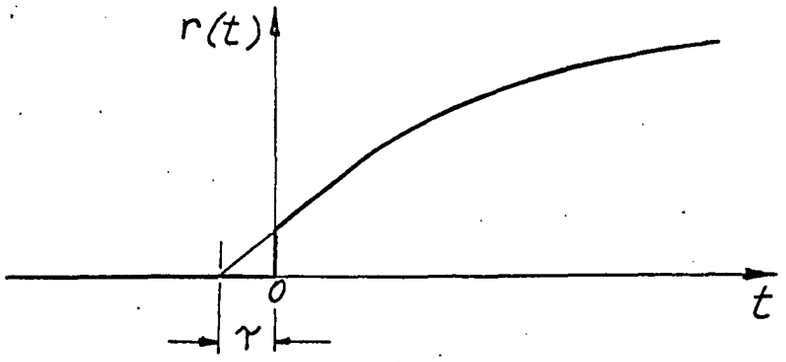
FIG 1.13 Feedback Network Impulse Response and Transfer Function for Double Integration with Partial Bypass (ie. with a high frequency boost)



This modification to the feedback network, suggested by de Jager in order to achieve stability, is of the same form as proposed above. De Jager suggests intuitively that γ should be about equal to T_c and hence that $\omega_u \simeq 1/T_c$. Thus $f_u \simeq f_c/2\pi$. By measurement on a particular system and input signal, de Jager found the $f_u = f_c/2\pi$ was close to optimal. Using a combination of single and double integration for the feedback network, with partial bypass of the second integrator, the time response

to a unit impulse will be of the form shown in Fig. 1.14. The corresponding frequency characteristic is shown in Fig. 1.12 (b), with $f_u = f_c/2\pi$ for f_c = 56kHz.

FIG 1.14 Impulse Response for Mixed Integration Feedback Network with Partial Bypass
(ie. with the high frequency boost)



This gives a typical r(t) signal as shown in Fig. 1.10. Such a mixed integration delta modulation system is frequently referred to in the literature as double integration delta modulation, since pure double integration is seldom considered.

1.4.4 Unique Characteristics of Delta Modulation

Delta modulation shares the advantage mentioned earlier of digital modulation systems such as: noisless regeneration capacity, a two state transmission signal, suitability to digital I.C. instrumentation, and so on. Delta modulation also has several characteristics which make it unique from DPCM and particularly from PCM.

The unity-bit code of DM results in a two-level quantizer (or comparitor) where a complex, many-level quantizer (typically 128 levels) is required for PCM and DPCM. Additionally, the one bit code gives a sampling rate equal to the output pulse rate, resulting in typically a much higher sampling rate for DM. This removes the necessity of low pass filtering the input signal to prevent aliasing error as required with PCM and DPCM. As a result of these features, basic DM exhibits remarkable circuit simplicity, which was indeed the fundamental reason for the early interest in DM.

As discussed in Section 1.4.3 there is a natural adaption of the amplitude capability spectrum of any differential system to the mean speech spectrum. In particular this applies where the feedback network can be modified to give an optimum adaption. In addition to giving a minimum redundancy in capacity, this adaption to speech also provides a smooth roll-off for frequencies outside the range of immediate interest. Thus a more natural sound would be expected compared to PCM or DPCM, where sharp band limiting of the higher frequencies is required in order to give

a finite bandwidth suitable to sampling at the Nyquist rate (typically 8kHz).

The use of a one-bit code however, has a major disadvantage. In order to achieve a reasonable signal amplitude capacity with a one-bit code, a sampling rate much greater than the Nyquist rate is required. Since all the information of a signal is contained in the sampled signal when the Nyquist rate is used, the higher rate would seem redundant from the information transmission viewpoint. In fact, as will be discussed in Chapter 2, delta modulation shows an increase in the output SNR approximately proportional to; from the third to the fifth power of the output pulse rate, depending on the feedback network. This compares poorly with PCM and DPCM where an increased pulse rate is used to increase the number of digits per code group, giving an exponential SNR increase with increasing pulse rates; as discussed in Section 1.3. For typical commercial PCM a maximum output SNR of about 44dB for a pulse rate of 56kHz is obtained (Fig. 1.2). As will be shown in Chapter 2, DM for comparable performance requires an output pulse rate of about 75 kHz.

Another major difficulty with DM, which is encountered with any differential pulse modulation method, relates to its use with TDM. As differential (negative feedback) coders depend on the high correlation between consecutive sampled signal values, the multiplexing of signals cannot be performed before the coder[7,10]. For the coding of a multiplexed signal a much more complex coder would be required, with the facility to store the current predicted value, $r(kT_c)$, for each channel, so as to be available for comparison with the next sample from the same channel. Thus the very significant benefit to PCM, for long haul voice communication; of a common quantizer and coder for many signals, is not shared by DPCM or DM. However the basic simplicity of DM and the current trends in I.C. technology, are making the use of a delta coder/decoder at each transducer continually more feasable. This could result in totally digitized communication links, with the resulting reduction in vulnerability to distortion and interference as an additional benefit.

FIG 1.15 Companded (or Adaptive) Delta Modulation - Schematic

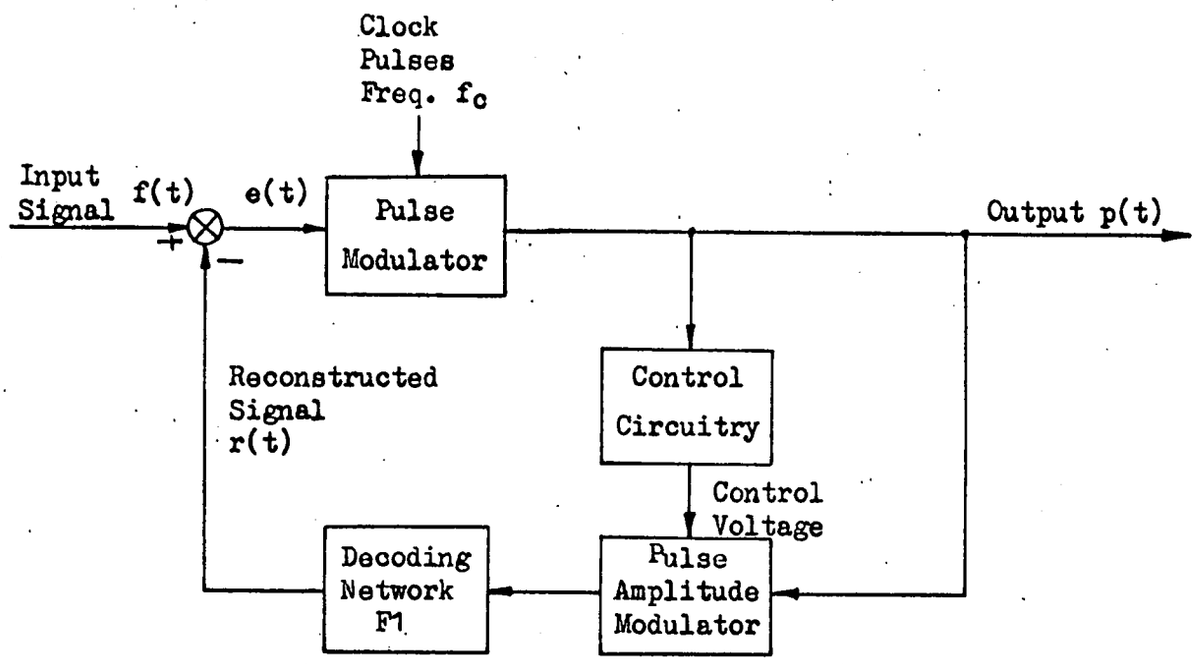


FIG 1.16 Continuous Delta Modulation - Greefkes[11]

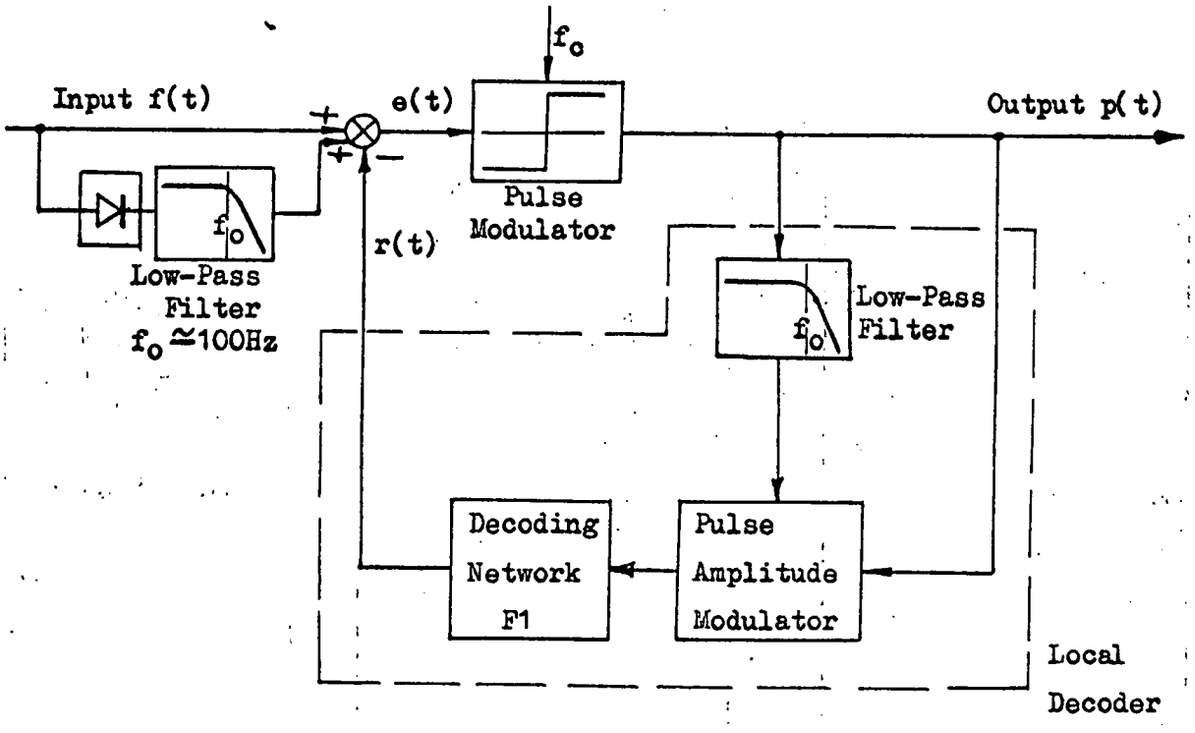
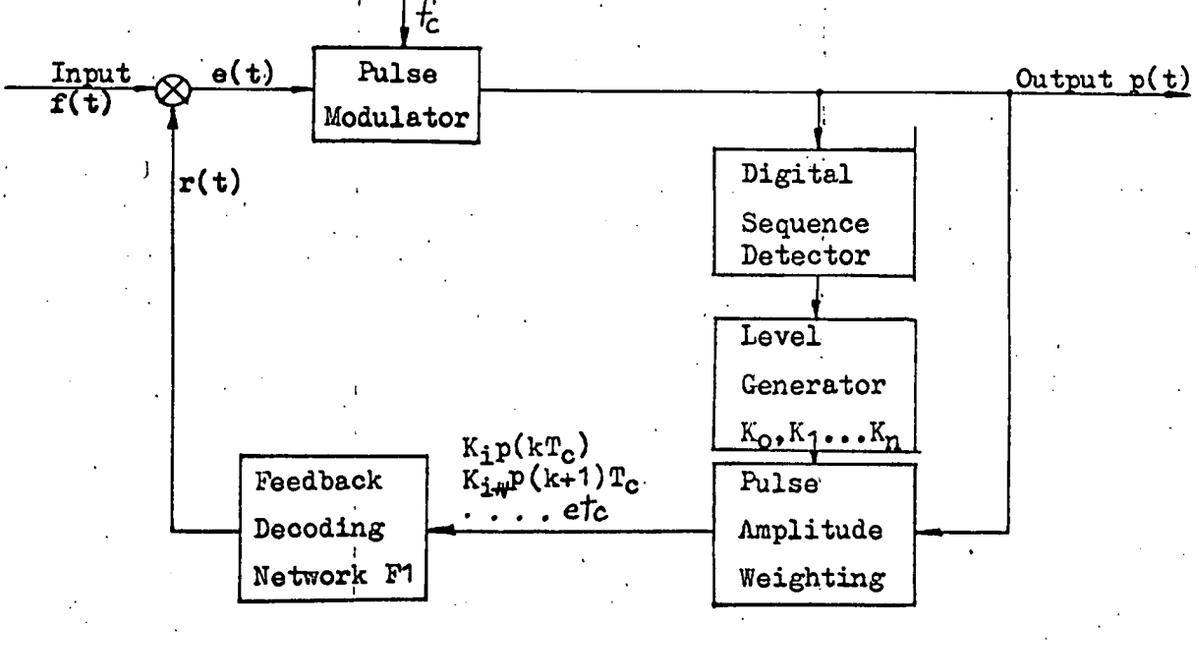


FIG 1.17 Discrete (Instantaneous) Companding - Schematic



1.5 Modified Delta Modulation Systems

The main purpose for modifying, or adding to the basic delta modulation system has been to improve the performance under maximum signal conditions and to increase the dynamic range of input signals which will give satisfactory performance. Ideally it would be desirable to have a constant output SNR for a wide range of input signal powers; (typically a SNR of greater than 30dB for an input dynamic range of about 35dB is required for commercial telephony). This would give a noise power which varied directly with the signal power and was thus "masked" by the signal equally well over the required range of input signals.

As mentioned in Section 1.3.3 the signal dynamic range of PCM is increased using instantaneous or static companding. This provides a good dynamic range, but at the expense of the maximum output SNR. Modifications to DM are generally designed to provide dynamic companding, which is achieved by varying the "step size" of r(t), (or more precisely, by varying the size of the pulses into the feedback network) according to the level of the input signal. Thus dynamic companding is designed to give an additional facility for prediction by r(t) of the input signal. Just as the mixed integration feedback network discussed in Section 1.4.3, is designed, to give long term speech spectrum matching; so companding is designed, through pulse size adaption, to give shorter term signal level matching. From the information theory viewpoint, dynamic companding is introduced in order to increase the information carried by each output pulse, by reducing the correlation between pulses. Reduction in the correlation occurs as the pulse train spectrum tends to become flat. As discussed in Appendix A this corresponds to the pulse train tending towards a random binary signal. Thus companding can be regarded as an attempt to increase the information content of the output pulse train by increasing the probability of a change in pulse polarity when a series of consecutive "ones" or "zeros" is occuring (that is, when the input signal level is tending towards overload).

The basic system for companding DM is shown in Fig. 1.15. The implementation of companding can be divided into roughly two categories; continuous companding and discrete companding. With continuous companding, first proposed by Greefkes and de Jager[11], the step size is varied in a continuous manner by a control voltage derived from the input signal level or the derivative of the input signal. The control voltage varies at a muc slower rate than the pulse rate and the companding is often referred to as "syllabic companding", as the rate of step size variation is frequently made to roughly correspond to the rate of signal level change of speech, or the "syllabic rate".

The continuous delta modulation for speech proposed by Greefkes and de Jager[11] incorporates a level detection circuit at the input as shown in Fig. 1.16. This adds a low frequency signal, below the

minimum speech frequency, thus incorporating the control voltage in the transmitted pulse signal as a variation in the mean number of positive pulses. A similar system is presented by Tomazawa and Kaneko[12] which eliminates the input level detection circuitry. Instead the control circuitry contains a reconstruction filter, the output of which is level detected and used as the control voltage for the pulse amplitude. This provides a square law expansion of the signal content of p(t) and does not involve an increase in the low frequency or dc content of p(t). Brolin and Brown[13] developed a compander in which a level detector from the input feeds a separate delta coder with a low clock rate. The reconstructed signal of the control delta coder determines the step size, and the two digital pulse streams are multiplexed for transmission. Other continuous delta modulation systems have been considered (see Project Mallard [14]) all similarly based on a detection of the density of the input pulses to the feedback decoding network, F1, with the pulse weight being adapted in a continuous manner accordingly. For a clock rate of 56kHz and a test signal of 800Hz, the performance of continuously companded systems are typically of the order of: maximum output SNR of 30dB for a dynamic range of inputs of 35dB or greater[11,12,15].

Discrete companding bases the variation of the step size on the preceding sequence of binary output pulses. The step size is usually changed in discrete increments and at a rate equal to the clocking rate. For this reason such adaption is frequently referred to as instantaneous companding and is equivalent to a non-companded delta modulation system with static (instantaneous) compression of the input signal and expansion of the output. The step height control voltage is derived from the binary output via digital logic circuitry (see Fig. 1.17). One of the first companded delta modulation systems proposed was Winkler's "high information delta modulation" [16] which employed discrete instantaneous companding. After two consecutive output pulses of the same polarity, the local decoder (consisting of a sequence detector and an exponential level generator) doubles the weighting K_1, of the input to the feedback network (F1) for each successive pulse of the same polarity. A change in polarity halves the weighting (or the control voltage) until the minimum level, K_o, is reached. A basically similar system was developed by Schindler[15] which also exhibits a logarithmic companded characteristic. Abate[17] gives a general discussion on discrete companding with DM and concludes from computer simulation that equispaced levels give marginally better performance than exponentially spaced levels for a signal with a uniform spectrum. Wing[18] has proposed a discrete adaptive DM system exhibiting very simple circuitry. His proposed optimum system has the step height depending linearly on the proportion of ones (or zeros) in the preceeding four binary output pulses.

With the wide choice of logic circuitry and step weighting, no method has so far shown itself greatly superior; while many new digital systems based on the output binary sequence are being proposed.[19,20,21] To a large degree the best logic and weighting system will depend on the type of input signal[22]. Instantaneous companding has two disadvantages relative to continuous companding. With the step size being determined by the immediate output pulse sequence, any mismatching between the transmission and receiving pulse weighting circuitry will produce distortion in the received signal [16,24]. On the other hand, mismatching with continuous companding only causes error in the level of the signal at the output. Secondly, for a given maximum step size, continuous companding under full load employs the maximum step size over the entire range of instantaneous signal values. Instantaneous companding however, like static companding for PCM, has varying step sizes even under full load, depending on the instantaneous signal value. For this reason an instantaneous companded system cannot achieve the same maximum output SNR ratio as a non-companded or continuously companded system with the same maximum step size.

These factors have led to the use of syllabic (continuous) companding for the currently more promising delta systems and in particular for the system recently proposed by Greefkes[23] and investigated further by Zarda and Hauser[24]. Although continuous, Greefkes' "digitally controlled delta modulation" uses digital sequence detection of the output pulses. After four consecutive pulses of the same polarity (indicating that a certain loading level has been exceeded) the sequence detector gives a "high" output. By low pass (syllabic) filtering of the sequence detector output, the pulse amplitude control voltage is generated. This gives a control voltage directly related to the mean time (over a 10 msec. period) that the loading level is exceeded. Thus the system is kept fairly close to the overload (optimum) condition for a wider range of input signal levels. The use of digital circuitry allows a greater range of step height variation (companding range) without stability problems. Typically, "digitally controlled delta modulation" has been shown[24] to give an output SNR of 30dB or greater for a dynamic range of input signals of 45dB, with an 800Hz signal and a clock frequency of only 40kHz; the maximum SNR being about 37dB. The best published results for an instantaneous dynamic companded delta system appear to be those given by Schindler's system develped at IBM[15]. For an 800 Hz sinusoidal signal with a 56kHz clocking rate, an output SNR of greater than 30dB over a signal range of 35dB, and a maximum SNR of about 39dB, were measured.

With the many illustrations of companded delta modulations potential performance superiority over companded PCM, for voice communication; it is now only the development of a low cost I.C. delta

coder/decoder that is required for delta modulation to become competitive or superior to PCM for commercial telephony.

Other modifications to the basic delta modulation method include; delta-sigma modulation, in which the input signal undergoes pre-emphasis with an integrating network prior to entering a conventional delta modulator.[25,26] Also, asynchronous delta modulation[27] for which the clock rate is not fixed but varies according to the demands of the signal. Such variations on the basic delta modulation method are of interest for communication situations where their unique characteristics are of benefit. Such situations however, fall outside the mainstream of electrical communications.

CHAPTER 2.

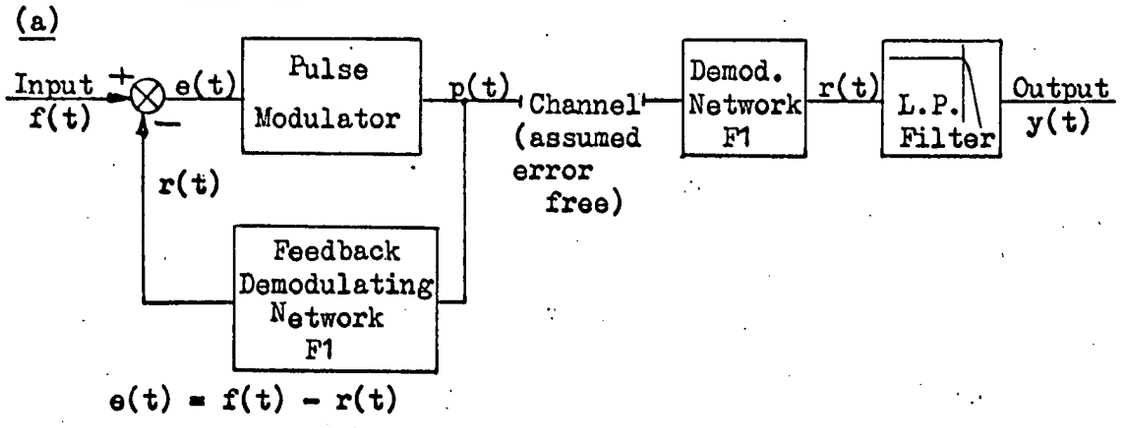
INVESTIGATION OF NOISE INHERENT IN DELTA MODULATION SYSTEMS

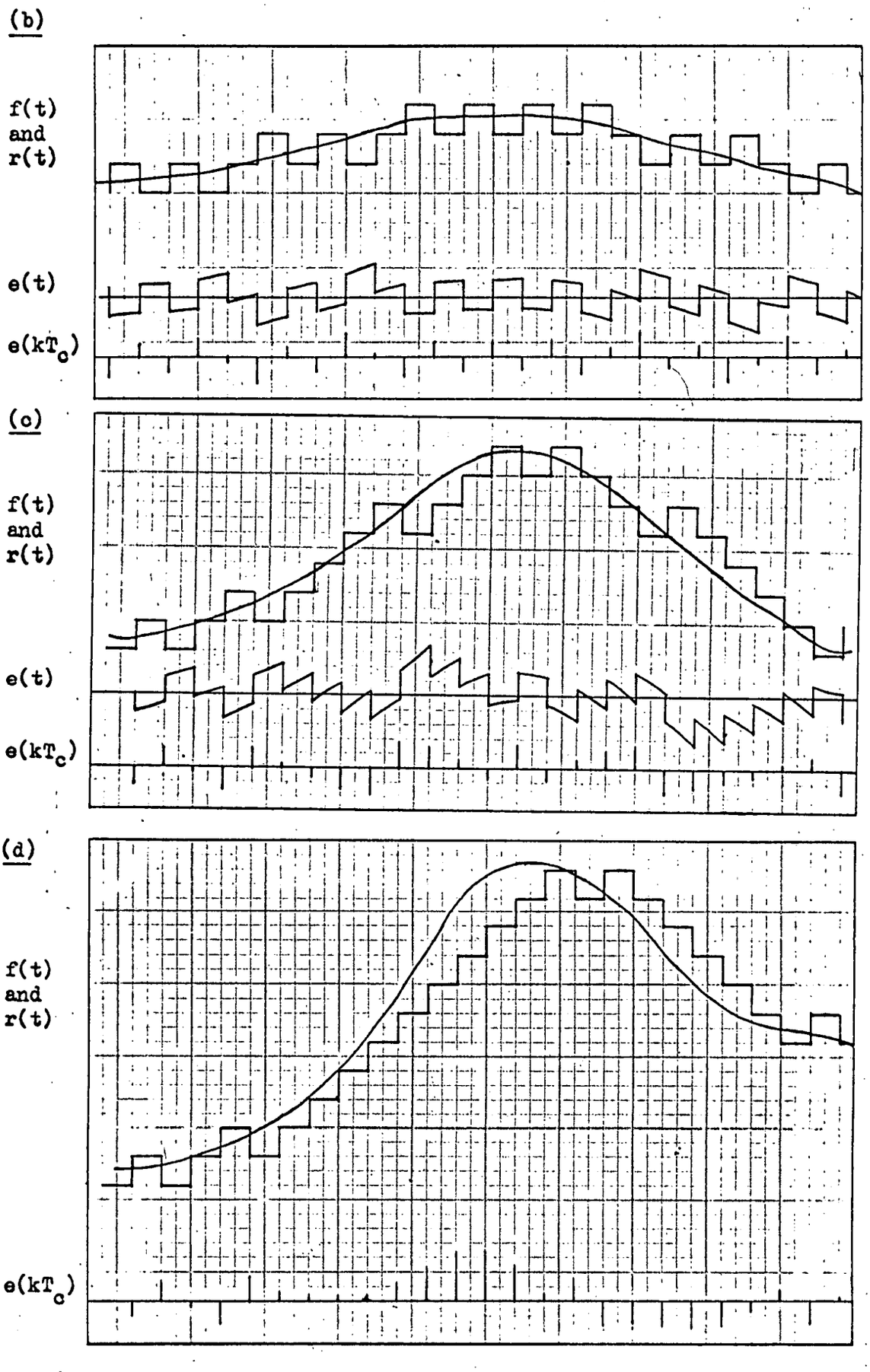
2.1 Introduction

In order to represent a continuous signal by a binary pulse sequence, it is necessary to consider an approximation of the signal. The approximation to discrete amplitude values, is necessary because a finite binary sequence can only specify a discrete amplitude. From the information viewpoint, the approximation is necessary to limit the amplitude information to a finite value, because a binary sequence can contain only a finite amount of information. Thus a binary representation of a continuous signal, necessarily involves the introduction of an error funtion. The error funtion gives rise to noise which is inherent in any digital modulation method. The noise resulting from amplitude quantization, is completely distinct from the usual sources of noise in electrical communication such as distortion, interference, and random noise in the channel and equipment. Because of pulse regeneration with digital modulation methods, these types of noise are totally eliminated at the pulse regeneration stage of a receiver and their contribution from the regeneration stage to the final output would be negligible. Their effect can be to cause errors in the regenerated binary signal which gives rise to output noise. As discussed in Section 1.3.3 for PCM and more fully analysed by Carlson[2], the probability of an error in any particular pulse decreases rapidly for an input SNR increasing beyond some threshold value. For delta modulation, as with PCM, provided some input SNR threshold is exceeded, (as it is under normal operating conditions), there is virtually no contribution to the output noise by pulse errors. (Because the effect of a pulse error is different for different modulation methods, the input SNR threshold for delta modulation will not be the same as that for PCM. In fact it will vary with delta modulation depending on the actual implementation. However the threshold input SNR for PCM, given conservatively at 15dB[2], would be of the same order as that for delta modulation. Discussion of the effect of channel errors can be found in Johnson[20], Wolf[30] and elsewhere.) It can therefore be concluded that for a delta modulation system operating under normal conditions, virtually all noise at the output is due to noise introduced in the modulating process itself, i.e. noise inherent in the modulation.

In this investigation, only non-companded systems will be considered, because the main purpose of companding is to extend the range of input signals for which near optimum, or at least satisfactory performance is achieved and not to effect overall reduction or alteration of the noise. Furthermore the basis for analysis of companded delta modulation would be an analysis of the system with a constant "step size".

FIG 2.1 Typical Waveforms for Variables in a Simple Delta Modulator with Ideal Integration





2.2 Nature of the Noise

The noise produced in a delta modulation process is generally classified as consisting of two types - quantizing or granular noise and overload noise.* The noise resulting solely from the signal approximation, inherent to the modulation, is termed the quantizing noise and it accounts for all the output noise when the magnitude of the input signal is well within the system's capacity. Thus quantizing noise for DM is equivalent to the quantizing noise for PCM which results from the descrete amplitude representation of the continuous input signal. If the input signal magnitude exceeds the capacity of the system in some way, for example amplitude, slope, rate of change of slope and so on, then some type of overload condition ensues giving rise to an overload noise component at the output.

Although these two types of noise are readily distinguishable when one is dominant, the actual transition between types, or a definition of the onset of overloading has proved difficult. Overloading is generally defined in the literature for single integration demodulating networks, (either ideal or lowpass filtering) and is said to occur when the maximum signal slope is greater than the maximum slope the modulator (or r(t)) can achieve. Typical waveforms for the reconstructed signal, r(t), of a DM system with a single, perfect integration feedback network with ideal impulse inputs (referred to hereafter as a simple delta modulator) are shown in Fig. 2.1 for three levels of input loading.

Referring to Fig. 2.1(a), two definition of the total output noise can be made. The first, and the one used in all the early analysis, (de Jager, van de Weg, Zetterberg etc.) defines the noise as being given by error funtion ($e(t) = f(t) - r(t)$) after low pass filtering. Such a definition, does not take into account the delay of r(t) relative to the input, which results from the action of the decoding network, as can be seen in Fig. 2.1. Therefore the error function, e(t), would have a component

---

* This nomenclature for the two types of noise is not used universally throughout the literature on delta modulation. Frequently the term "quantizing noise" is used to refer to all noise resulting from the modulation process itself, and thus includes both granular and overload noise. However the author feels that overload noise is not accurately described as quantizing noise as its nature is totally different from that resulting from the amplitude quantization of a continuous signal.

which was correlated to the input, and for this reason the noise defined from e(t) will be referred to as the correlated noise. The noise which is linearly independent of the input and thus represents more faithfully the unwanted output component will be termed the uncorrelated noise. As such, the uncorrelated noise function is the difference between the demodulator output, y(t), (which is equivalent to r(t) after low pass filtering in the system model of Fig. 2.1) and the input signal with suitable delay and amplitude change required to give a minimum difference. The output signal will be given by this changed but undistorted input which can be written as $kf(t-t_o)$. Thus, the uncorrelated noise function is given by:
$n(t) = y(t) - kf(t-t_o)$ where k and t_o are such that n(t) has minimum power.

One important aspect of the nature of the noise is the subjective effect. Although the output SNR is the accepted measure of the quality of the output, the ultimate measure of output will depend on the effect of the noise power on subjective factors such as intelligibility and "unpleasantness" of the noise. Little subjective evaluation of delta modulation has been made to date as quantization noise is not regarded as being significantly more or less offensive than random noise and it has the advantage over the noise of non-digital systems of not being present during idling periods. Overload is akin to distortion and its subjective effect would be of considerable significance to output quality for a system operating under overload conditions. However as the operation of a delta system under a significant degree of overlaod is not frequently contemplated, the subjective effect has again been given little thought. Subjective evaluation of various digital and linear modulation methods has been performed by Donaldson, Chan and Douville[31,32]. Their evaluation, however, has been directed towards system comparison, rather than correlation between output SNR and the subjective output quality.

Recognition of subjective effects is made by Aaron et al,[33] in their definition of correlated and uncorrelated noise. Their distinction between the two types is different from that proposed above and is more suited to the meaning of "correlated". The correlated noise is defined as that component which could result from passing the signal through a linear network. The uncorrelated noise component is defined as being linearly independent of the signal and is thus additive, uncorrelated noise. Therefore the uncorrelated noise = $y(t) - L[f(t)]$, where L is the linear operator which will give an uncorrelated noise power minimum. Such a definition, while allowing for the totally different subjective nature of correlated and uncorrelated noise, defines an output signal, $L[f(t)]$, which contains distorted input signal components.

2.3 Noise Analysis

Analysis of the noise produced by delta modulation is not suited to the general tools of system analysis because of the presence of both nonlinear elements and a feedback path. As it relates to delta modulation, the approach of Bennett[5] to direct quantization systems is considered, as are the problems encountered in general with delta modulation analysis. This section, and also the computer simulation and experimental investigation, primarily consider the simple, single integration case. This gives a simpler base for analysis and simulation which will provide a guide to general delta modulation performance analysis and give insight into the nature of the noise.

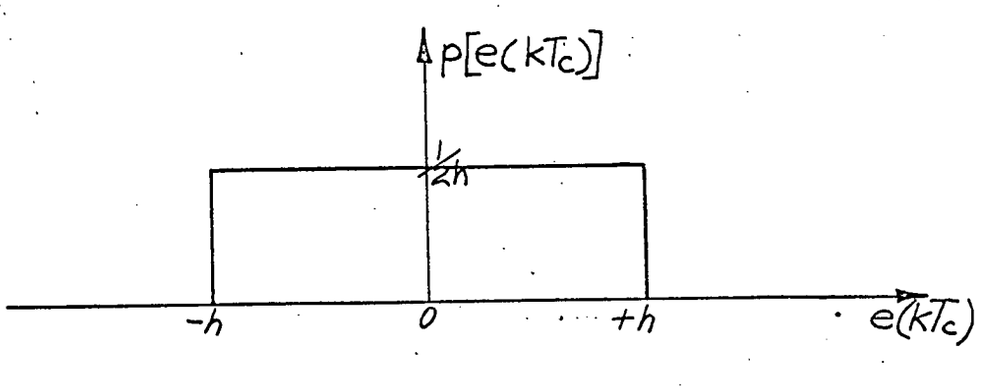
2.3.1 Initial Consideration

If we consider a simple delta modulator with an input signal such that the system is operating well below overload the reconstructed signal will be a shown in Fig. 2.1(b). The same line of analysis will now be applied that led to the expression for the noise power with PCM, as outlined in Section 1.3.3.

The value of f(t) at any clocking instant (i.e. the input signal sample value $f(kT_c)$, is given by the sum of the reconstructed signal value just after a clocking instant $r(kT_c)$, plus the discrete quantizing error value, $e(kT_c)$

That is: $f(kT_c) = r(kT_c) + e(kT_c)$ where $e(kT_c) = e(kT_c+)$ and $r(kT_c) = r(kT_c+)$

It can be seen from Fig. 2.1(b) that $-h < e(kT_c+) < +h$, providing the system is operating below overload. In fact it is convenient for the single integration case to define the overloading condition as occuring when the magnitude of $e(kT_c)$ exceeds h, the step height. This corresponds exactly to the definition that no overloading will occur provided the maximum signal slope is less than or equal to hf_c, the maximum slope capability of r(t). For a signal with an equal probability of taking any value over a large range of step heights (i.e. for $A_{max} \gg h$) it would seem reasonable to assume that $e(kT_c)$ has an equal probability of taking on any value between $\pm h$ as shown below.



For a continuous function denoted by x(t) with a uniform probability density function as above, the mean square value is given by:

$$\overline{x^2} = \int_{-h}^{h} \frac{x^2}{2h}\, dx = \frac{h^2}{3}$$

This can be written as $\int_{-\infty}^{\infty} P_x(\omega)d\omega = h^2/3$, where $P_x(\omega)$ is the power spectrum of x(t). For the discrete quantizing error (given by the continuous function, x(t), sampled at a rate of f_c) the above power will be the power in the region $0 - f_c/2$. (See Bennett[5]).
That is we can write:

$$\int_{-\omega_c/2}^{\omega_c/2} P_{ek}^{*}(\omega)d\omega = \int_{-\infty}^{\infty} P_x(\omega)d\omega = \frac{h^2}{3} \qquad \text{.....2.1}$$

where $P_{ek}^{*}(\omega)$ is the power spectrum of $e(kT_c)$. If the received quantized signal, $r(kT_c)$, were being low-pass filtered (ideal) with a cut-off at $f_c/2$, this expression would provide an estimate of the noise power due to quantizing. However as the low pass filtering for DM invariably has a cut off frequency much lower than $f_c/2$, a knowledge of the spectrum of $e(kT_c)$ would be required in order to estimate its power after typical low pass filtering. Hence such analysis does not lead to the same satisfactory conclusion as is achieved for PCM where the Nyquist sampling rate is used. Further, it can be seen from observation of Fig. 2.1 and from a consideration of the restrictions placed on the discrete values that r(t) can assume, that $e(kT_c)$ does not, in general, take on any value between $\pm h$ in a truly random manner, but each value will depend on the previous values and the signal characteristics to a significant degree.

For an initial consideration of the quantizing noise, it will be assumed that $e(kT_c)$ is a random variable bound by $\pm h$ and with some non-zero probability density function between $\pm h$. In this case the autocorrelation function of $e(kT_c)$ will have the value $\overline{e^2(kT_c)}$ at $\tau = 0$ and zero elsewhere, and thus the power spectrum, $P_{ek}^{*}(\omega)$, of $e(kT_c)$, will be flat over all frequencies, with a value denoted by A. Assuming secondly that the probability density function of $e(kT_c)$ is uniform between $\pm h$, then by applying the previous result that the power in the band $0 - f_c/2$ equals $h^2/3$ we get:

$$\int_{-\omega_c/2}^{\omega_c/2} P_{ek}^{*}(\omega)d\omega = A\omega_c = \frac{h^2}{3}$$

$$\therefore \quad A = \frac{h^2}{3\omega_c} = \frac{h^2}{3x2\pi} T_c \qquad \text{.....2.2}$$

If $e(kT_c)$ is passed through an ideal low pass filter with a cut-off

FIG 2.2 Finite Width Pulse Function and its Amplitude Spectrum

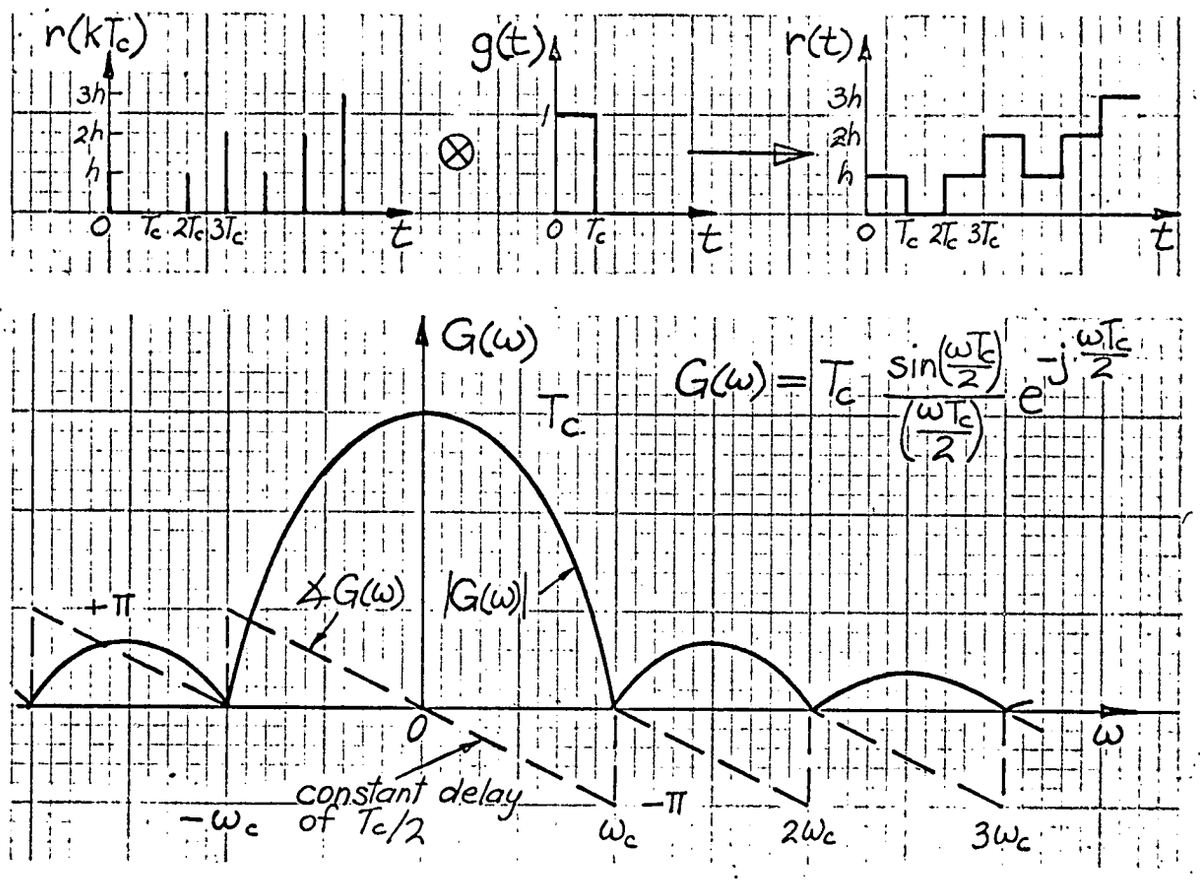
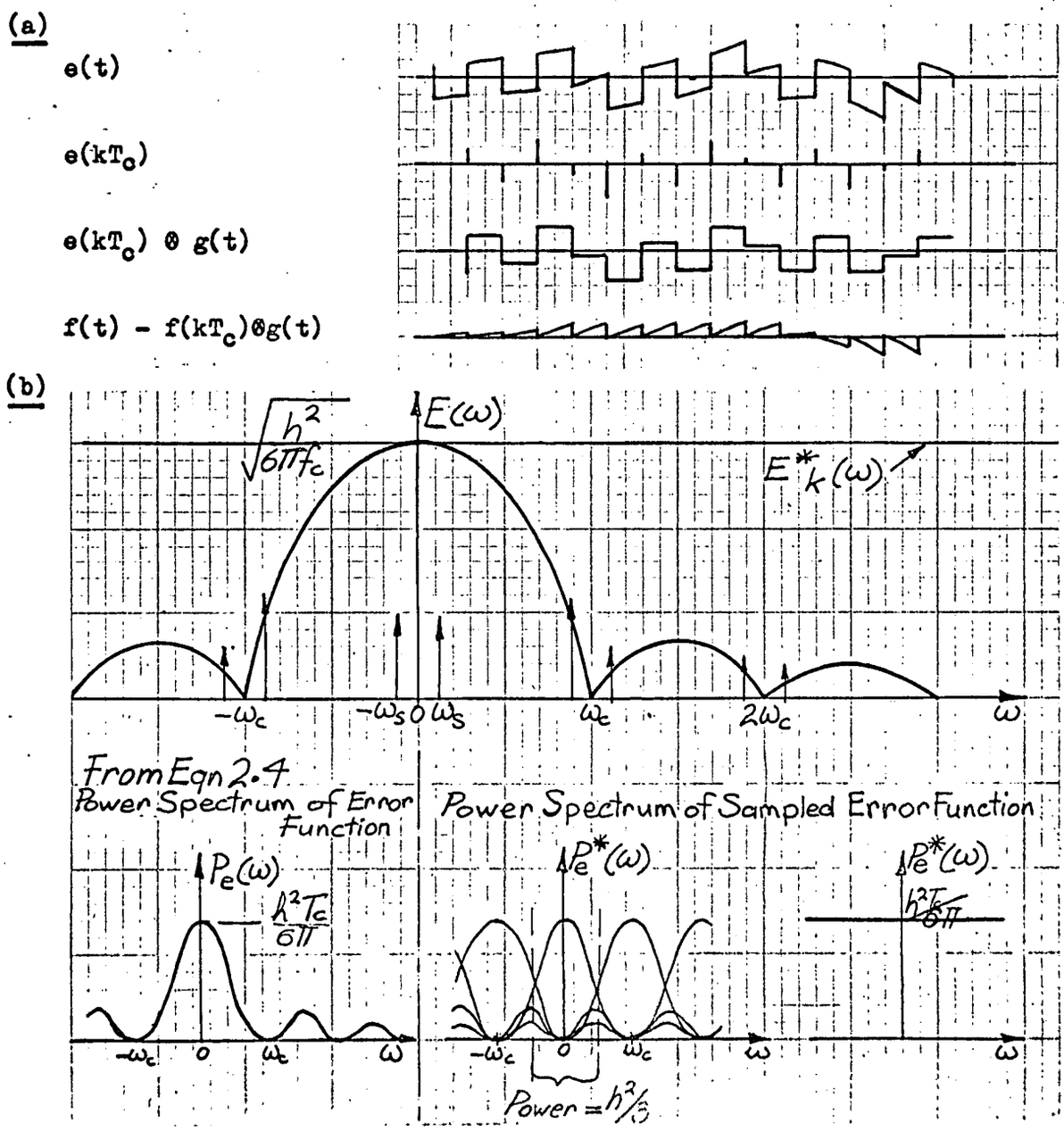


FIG 2.3 The Effect of Finite Width Pulses on the Sampled Error Function , $e(kT_c)$



frequency, f_m, this would give an estimate of the output noise power (correlated) as:

$$\text{Noise power in band } 0-f_m = \int_{-\omega_m}^{\omega_m} P_{ek}^{*}(\omega)\,d\omega = \frac{h^2}{3\omega_c^2\omega_m} = \frac{2h^2 f_m}{3f_c^2} \quad \ldots\ldots 2.3$$

This corresponds to a delta modulation system operating with r(t) being demodulated from p(t) as a discrete signal. It is therefore equivalent to a system where r(t) is ideally sampled at a rate f_c before low pass filtering.

In order to bring this analysis into line with the reality of the continuous nature of r(t) and e(t), the effect of finite width pulses will be considered. As shown in Fig. 2.2, r(t) equals $r(kT_c)$ convolved with the finite width pulse function, g(t), and thus the amplitude spectrum of r(t) is given by the product of the spectrum of g(t) (denoted by $G(\omega)$). The convolution of g(t) and $e(kT_c)$ however, will not give e(t) exactly as shown in Fig. 2.3 due to some signal component of e(t). The difference appears to be small and so, approximating e(t) as $e(kT_s) \circledast g(t)$ we obtain the estimate of the power spectrum of e(t), (using Eqn. 2.2), as:

$$P_e(\omega) = P_{ek}^{*}(\omega)\,\left|G(\omega)\right|^2 = \frac{h^2}{3x2\Pi}\,T_c\,\frac{\sin^2(\omega T_c/2)}{(\omega T_c/2)^2} \quad \ldots\ldots 2.4$$

Thus the amplitude spectrum, $E(\omega)$ of e(t) will be of the form

$$\frac{\sin(\omega T_c/2)}{(\omega T_c/2)}\; e^{-j(\omega T_c/2)} \quad \text{as shown in Fig. 2.3}$$

The exact expression for e(t) from $e(kT_c)$, as shown graphically in Fig. 2.3(a), is

$$e(t) = e(kT_c) \circledast g(t) + f(t) - f(kT_c) \circledast g(t) \quad \ldots\ldots\ldots\ldots 2.5$$

Hence the amplitude spectrum of e(t) is given by

$$E(\omega) = E_k^{*}(\omega)G(\omega) + F(\omega) - F^{*}(\omega)G(\omega)$$

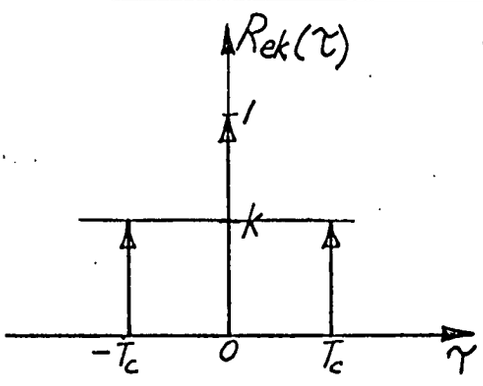
The additional term $F(\omega) - F^{*}(\omega)G(\omega)$ will give small signal components, distorted in amplitude and phase, as cross-products with the sampling frequency and its harmonics (as shown in Fig. 2.3(b)) for a sinusoidal signal. For a given amplitude spectrum such terms could be calculated.

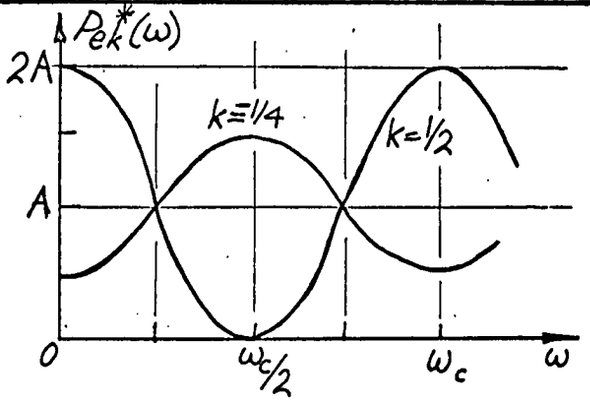
On the basis of this analysis we would expect the qunatizing error to have

(i) an amplitude spectrum of the form $\dfrac{\sin\omega T_c/2}{\omega T_c/2}\, e^{-j\omega T_c/2}$

and a power spectrum of the form $\dfrac{\sin^2(\omega T_c/2)}{(\omega T_c/2)^2}$

FIG 2.4 Some Models for the Autocorrelation Function of the Sampled Error Function, $e(kT_c)$ and the Resulting Power Density Spectrum



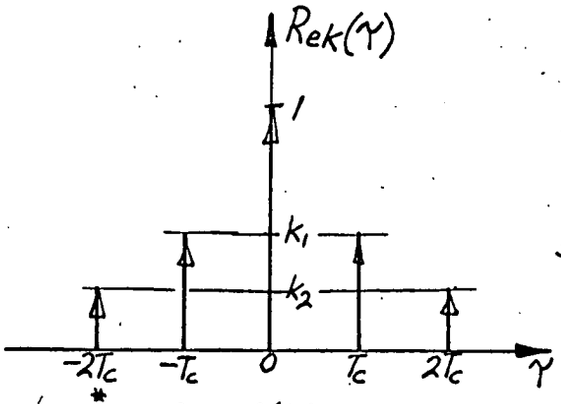


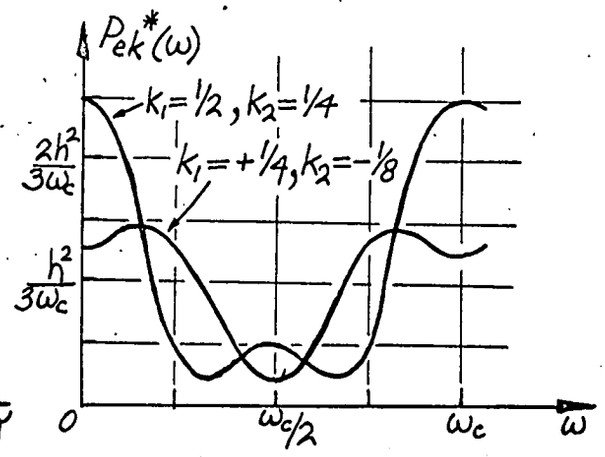
$$P_{ek}^*(\omega) = \int_{-\infty}^{\infty} R_{ek}(\tau) e^{-j\omega\tau}\, d\tau = A(1+ke^{+j\omega T_c}+ke^{-j\omega T_c})$$
$$= A(1 + 2k\cos\omega T_c)$$

From Equation 2.1: $\int_{-\infty}^{\infty} P_x(\omega)\, d\omega = \int_{-\omega_c}^{\omega_c/2} P_{ek}^*(\omega)\, d\omega = h^2/3$

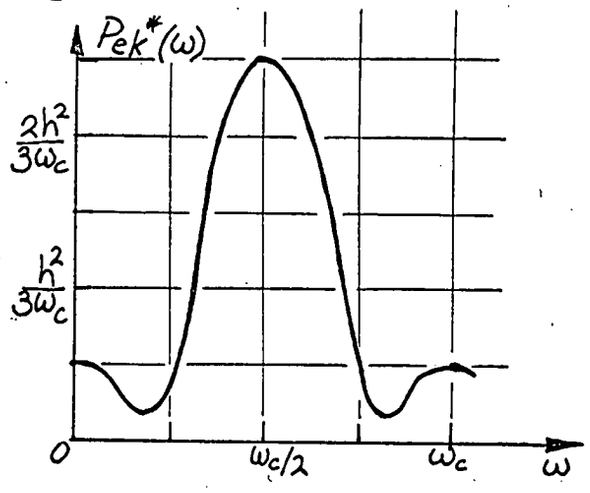
Since $\int_{-\frac{\omega_c}{2}}^{\omega_c/2} P_{ek}^*(\omega)\, d\omega = A\left(\omega_c + 2k\int^{2} \cos\omega T_c\, d\omega\right) = A\omega_c$

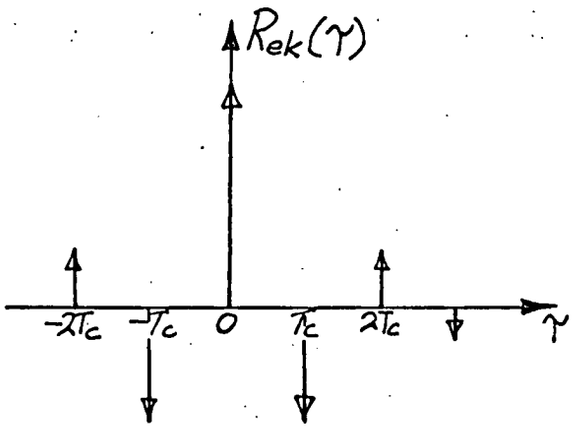
Therefore $A = \frac{h^2}{3\omega_c}$

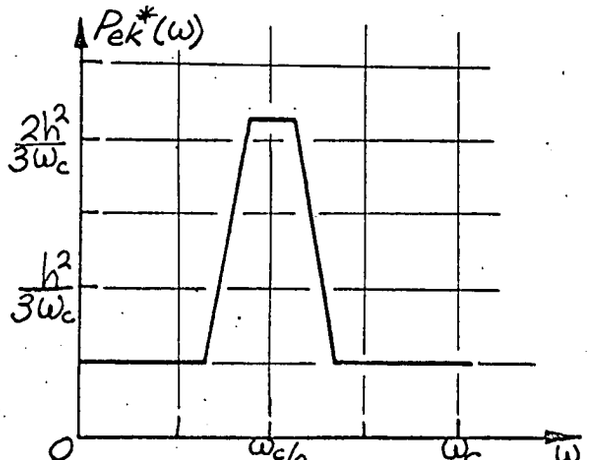




$$P_{ek}^*(\omega)\, d = A(1 + 2k_1\cos\omega T_c + 2k_2\cos 2\omega T_c)$$







A model of the situation for negative correlation between adjacent values of $e(kT_c)$.

(ii) with ideal low pass filtering with cut-off at f_m, a quantizing noise power given by:

$$Nq = \frac{h^2}{6\pi f_c} \int_{-\omega_m}^{\omega_m} \frac{\sin^2(\omega T_c/2)}{(\omega T_c/2)^2}\, d\omega \qquad \ldots\ldots\ldots\ldots 2.6$$

$$\simeq \frac{h^2}{6\pi f_c}\, 2\omega_m \text{ , provided } f_c \gg f_m$$

$$= \frac{2h^2 f_m}{3f_c} \qquad \text{as in Eqn. 2.3}$$

Hence this analysis predicts a quantizing noise power proportional to h^2 and $1/f_c$ and also proportional to f_m for $f_m \ll f_c$.

2.3.2 The Effect of Error Signal Correlation

Re-examining the assumptions made in arriving at the quantizing noise power and the quantizing error spectrum, it would appear that the assumption that $e(kT_c)$ is a random variable has the least justification. As mentioned earlier in this section each value of $e(kT_c)$ shows significant dependence on the previous values for the typical waveforms in Fig. 2.1. This is particularly so for the deterministic type of input signal considered, but for a random signal the assumption would be well justified. To obtain some idea of the effect of a non-random $e(kT_c)$ on the conlusions made above, autocorrelation functions for $e(kT_c)$ as shown in Fig. 2.4 will be considered, representing models of the type of autocorrelation function which might be expected. It is shown that any significant autocorrelation over the first few clock periods has a very significant effect on the power spectrum. In particular if the correlation between adjacent values of $e(kT_c)$ is negative the power spectrum in the region of frequencies less than $f_c/4$, (which would include the signal frequency region for normal DM operation) is reduced.

As the tendency to a reversal in sign of $e(kT_c)$ between successive samples increases, the noise power in the low frequency region decreases. Such a situation can be seen to be approached in Fig. 2.1 as the maximum signal slope is decreased or alternatively, for all other parameters fixed, as the clocking frequency, f_c is increased. Although it is difficult to estimate the rate of reduction in noise power (for increasing f_c) in a given band $0 - f_m$ due to shape changes in the power spectrum of the error function; it is possible to make two significant predictions for the behaviour of the error function.

While the dominant tendancy of r(t) remains to change direction at each clocking instant; adjacent values of the autocorrelation function of $e(kT_c)$ will alternate in sign and this will result in:

(i) a power spectrum for $e(kT_c)$ which remains basically flat over the lower frequency range, say for $f < f_c/4$; and

(ii) a power spectral density in this region which is considerably less than $h^2/(3\omega_c)$ for significant negative correlation between successive values of $e(kT_c)$.

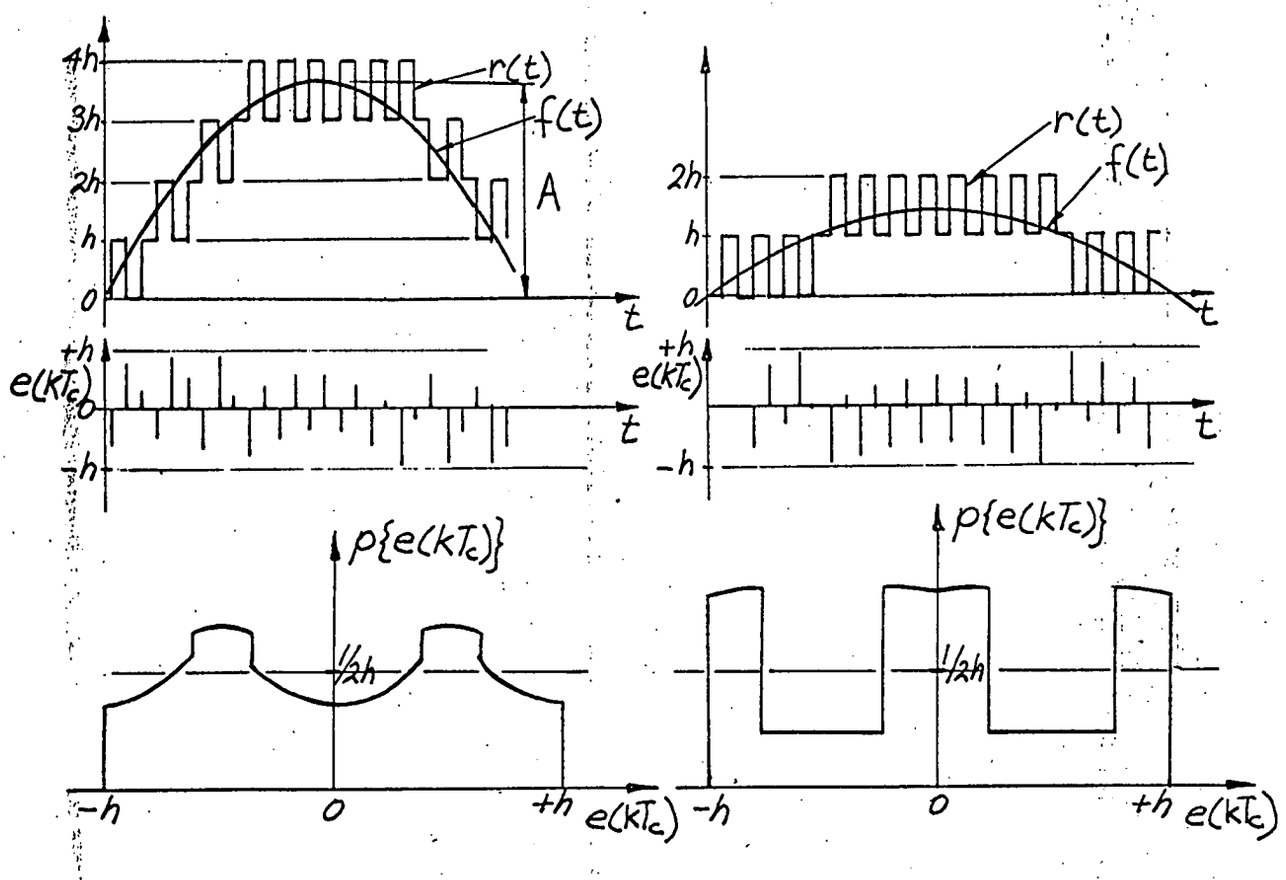
Over the range of conditions where r(t) is sufficiently far from slope overload for adjacent values of $e(kT_c)$ to have a negative correlation, the relationship $Nq = 2h^2 f_m/(kf_c)$ (provided $f_c \gg f_m$) of Equation 2.6 would still be expected to hold, except that the constant k would be somewhat greater than three.

On the other hand, as the slope overload condition is approached, the onset of positive correlation between adjacent values of $e(kT_c)$ would be expected to give a power spectral density, over the lower frequencies which is non-uniform and of greater than $h^2/3\omega c$. Under these conditions the same relationship for Nq would not necessarily be expected to hold. This ties in with observation of r(t) (Fig. 2.1c) as slope overload is approached. Although $f'(t)_{max} < hf_c$ and hence $|e(kT_c)| < h$ and no slope overload occurs, a state of "partial" or "instantaneous" slope overload can be seen to exist at times when r(t) fails to produce a change of sign in e(t) after two or more successive steps in the same direction. Furthermore it is under these conditions that the terms $f(t) - f(kT_c) \circledast g(t)$ of Eqn. 2.5, neglected in the analysis of e(t) by the consideration of $e(kT_c)$ alone, becomes increasingly significant.

2.3.3 The Effect of the Error Probability Density Function

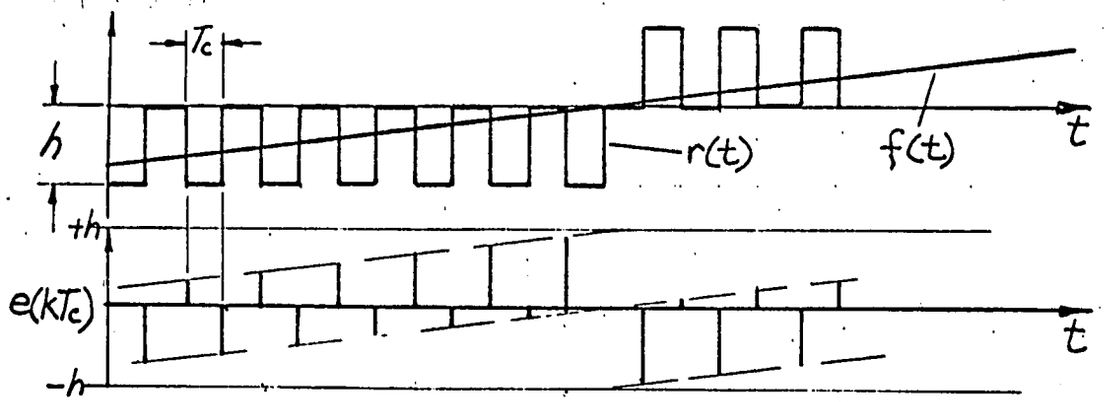
The assumption of a uniform probability density function for $e(kT_c)$ between $\pm h$ appears well justified under most, non-overload, operating conditions. The main situation where significant deviation from a uniform distributuion appears most likely to occur, is with deterministic types of input signals such as a sinusoidal signal. Considering the diagrams below, of fixed amplitude sinusoidal signals, it can be seen that non-uniform probability density functions will result due to the error function $e(kT_c)$, taking on values in a certain range for a disproportionate number of times.

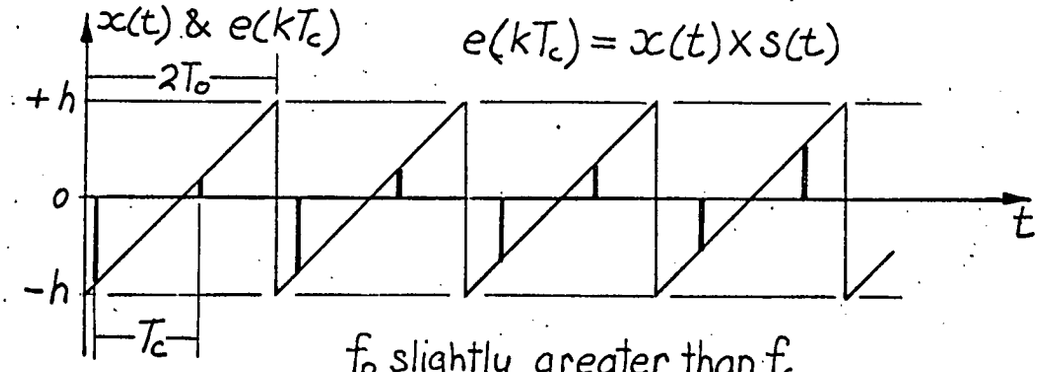
FIG 2.5 Illustrating the Expected Nature of the Probability Density Function of the Discrete Error Function for a Sine Wave Input



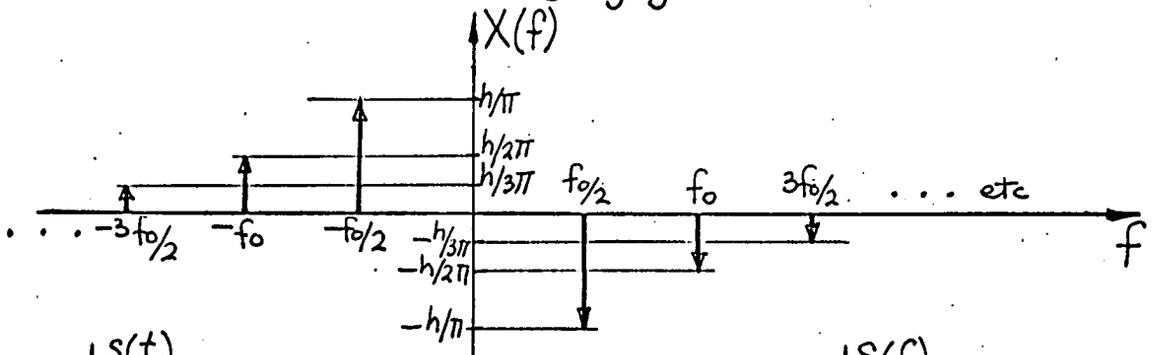
The dominant range of error values will be determined by the signal value relative to the discrete amplitude levels when the signal derivative is near or at zero. For the sinusoidal signal case the most probable values of $e(kT_c)$ will be A - kh and A - (k+1)h, where kh is the discrete amplitude level closest to, but less than A in magnitude. Discontinuities in $p(e(kT_c))$ would be expected at these values of $e(kT_c)$, and $p(e(kT_c))$ will be symmetrical about zero for signals with zero dc. content as shown in Fig. 2.5. The deviation from a uniform probability density function will be more pronounced for signals with low amplitudes relative to the step height due to the greater proportion of each period spent by r(t) between the levels of greatest magnitude.

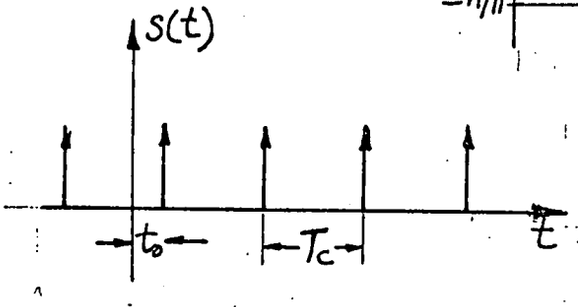
FIG 2.6 A Model for the High Clocking Frequency Situation

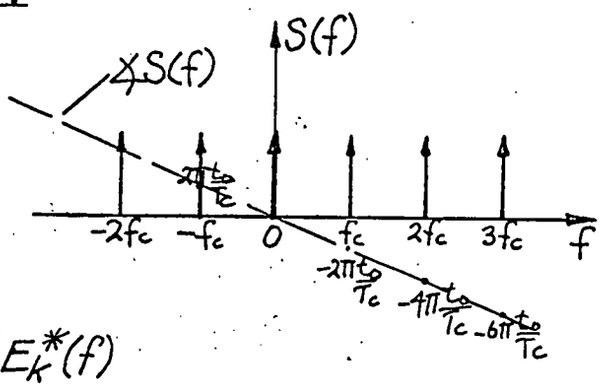




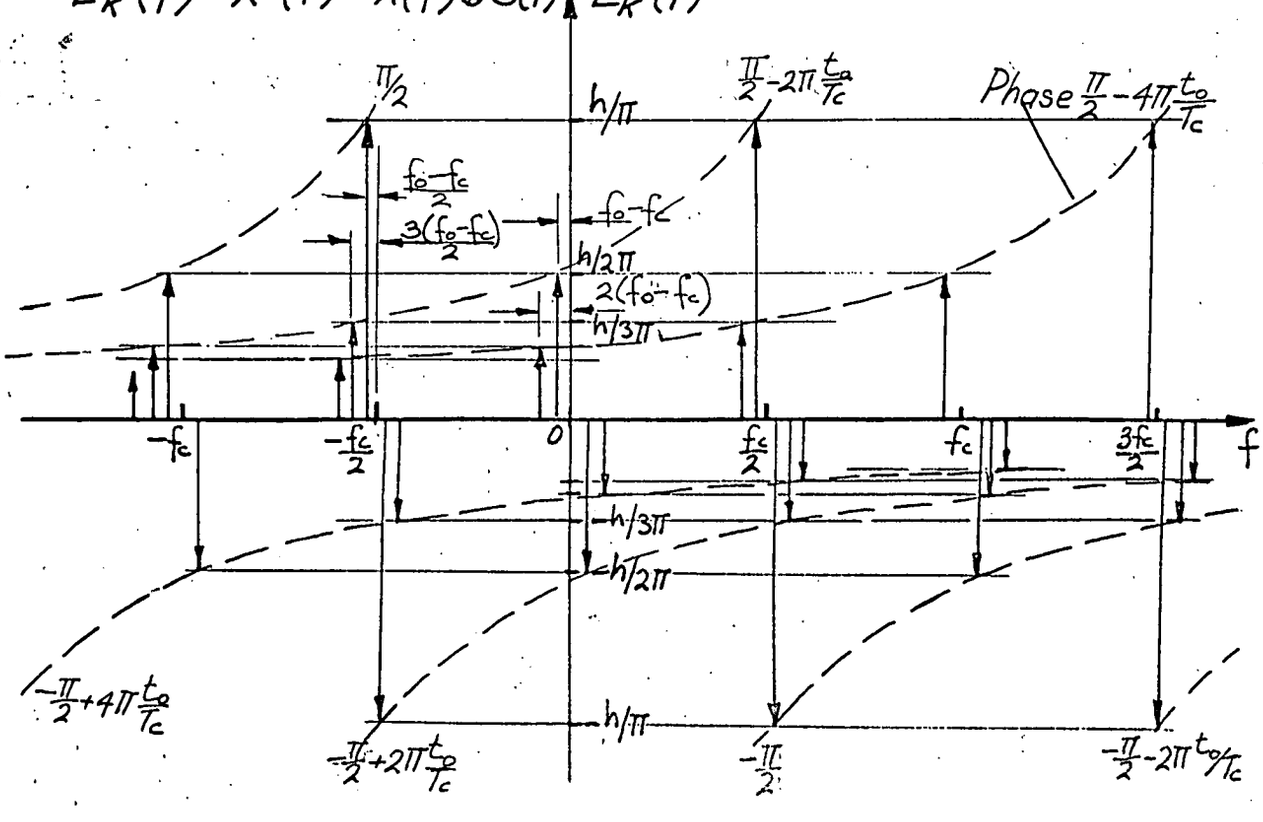
f_o slightly greater than f_c







$E_k^*(f) = X^*(f) = X(f) \circledast S(f)$ $E_k^*(f)$



The only effect of the variation of the probability density function in itself, is to give a mean square value for $e(kT_c)$ other than the value of $h^2/3$ which results from a uniform probability density function. By taking what would be expected to be a rather extreme distribution for $e(kT_c)$, it can be shown that the maximum variation of the mean square value from $h^2/3$ would only be of the order of 20% for the type of signals usually considered. This is in line with the conclusions of Zetterberg[28] and Sakrison[1] who indicate that the expression for the mean square value of the quantizing error for PCM based on a uniform distribution for the error, holds good for other distributions, provided the signal value is much greater than the quantizing level spacing.

2.3.4 The Effect of a High Clocking Frequency

A model for the situation when the clock frequency becomes very large relative to the maximum signal frequency can be established as shown in Fig. 2.6. The discrete error function, $e(kT_c)$, can be represented by the clock frequency sampling of a sawtooth function, $x(t)$, of amplitude h and period denoted by $2T_o$. This gives an accurate description of $e(kT_c)$ for all clock frequencies provided the input signal is a constant slope signal, with the slope given by $h(T_o - T_c)/T_oT_c$. $= h(f_c - f_o)$ Such a model provides for analysis of $e(kT_c)$ under the conditions of a very small input signal change between successive clocking instants (i.e. $T_o \simeq T_c$).

The amplitude spectra for $x(t)$, $s(t)$, (the sampling function) and the resulting spectrum for $e(kT_c)$, (given by $E_k^*(f) = X^*(f)$) are shown in Fig. 2.6. The components centred around zero frequency are given by:

$$e(kT_c)\Big]_o = \frac{h}{2\pi} e^{+j(2\pi \frac{t_o}{T_c} - \frac{\pi}{2}) + j(\omega_o - \omega_c)t} + \frac{h}{2\pi} e^{-j(\frac{2\pi t_o}{T_c} - \frac{\pi}{2}) - j(\omega_o - \omega_c)t}$$

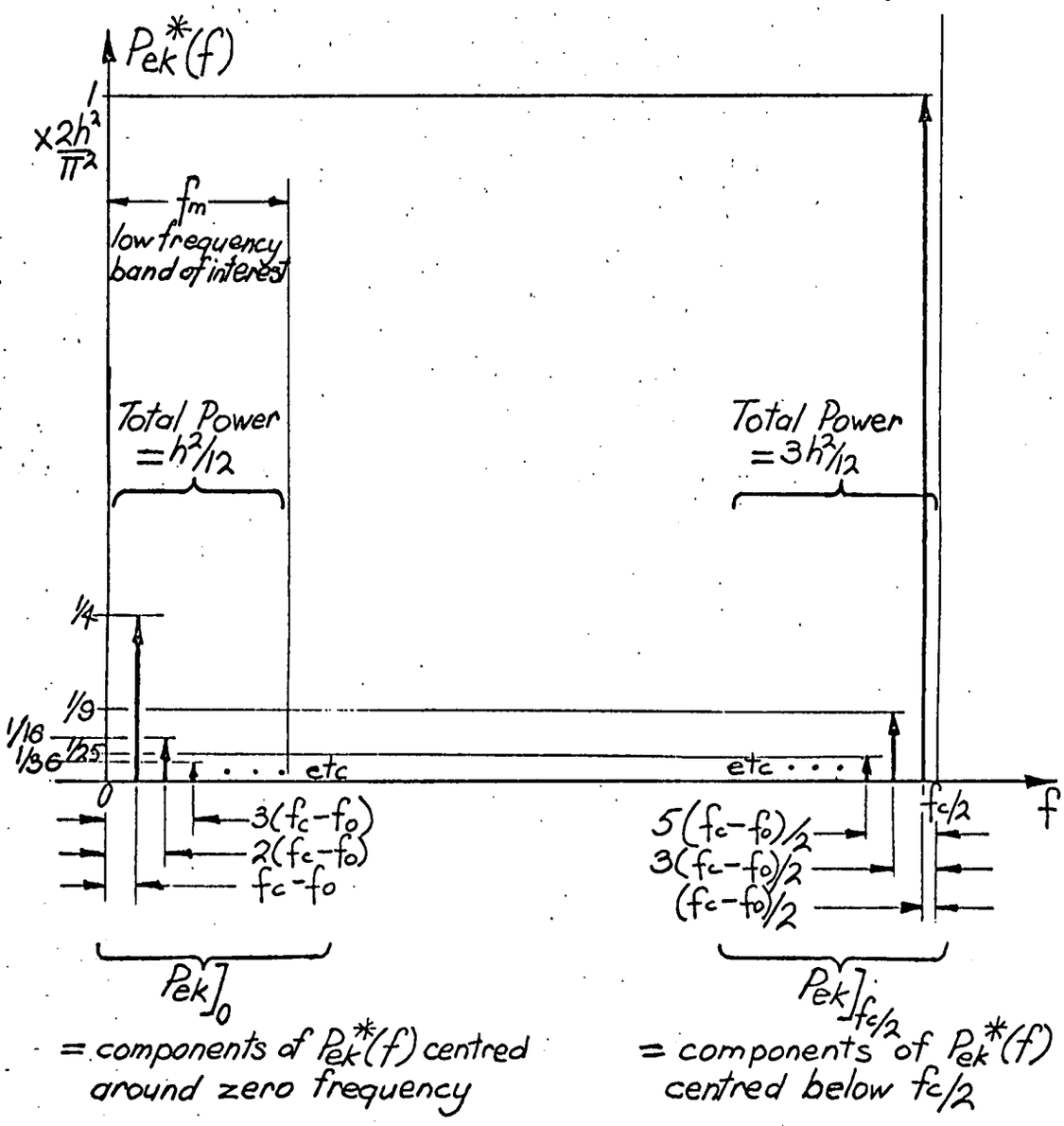
$$+ \frac{h}{4\pi} e^{j(4\pi \frac{t_o}{T_c} - \frac{\pi}{2}) j2(\omega_o - \omega_c)t}$$

$$+ \frac{h}{4\pi} e^{-j(4\pi \frac{t_o}{T_c} - \frac{\pi}{2}) - j2(\omega_o - \omega_c)t} + \ldots .$$

$$+ \frac{h}{2n\pi} e^{j(2n\pi \frac{t_o}{T_c} - \frac{\pi}{2}) jn(\omega_o + \omega_c)t} + \frac{h}{2n\pi} e^{-j(2n\pi \frac{t_o}{T_c} - \frac{\pi}{2})} \times$$

$$e^{jn(\omega_o - \omega_c)t} + \ldots .$$

FIG 2.7 One Sided Power Spectrum of the Discrete Error Function, $e(kT_c)$ in the Frequency Band 0 to $f_c/2$, for a Constant Slope Signal with a slope = $h(f_c - f_o)$ and $f_c \gg f_c - f_o$ ie. a very low input signal slope relative to hf_c.



$$\therefore\ e(kT_c)\Big]_o = \frac{h}{\pi}\left[\sin\left((\omega_o-\omega_c)t + 2\pi\frac{t_o}{T_c}\right) + \frac{1}{2}\sin\left(2(\omega_o-\omega_c)t + 4\pi\frac{t_o}{T_c}\right)\right.$$

$$\left. + \ldots\ldots + \frac{1}{n}\sin\left(n(\omega_o-\omega_c)t + 2n\pi\frac{t_o}{T_c}\right) + \ldots\ldots\right]$$

Therefore the power of the components centered around zero frequency is given by:

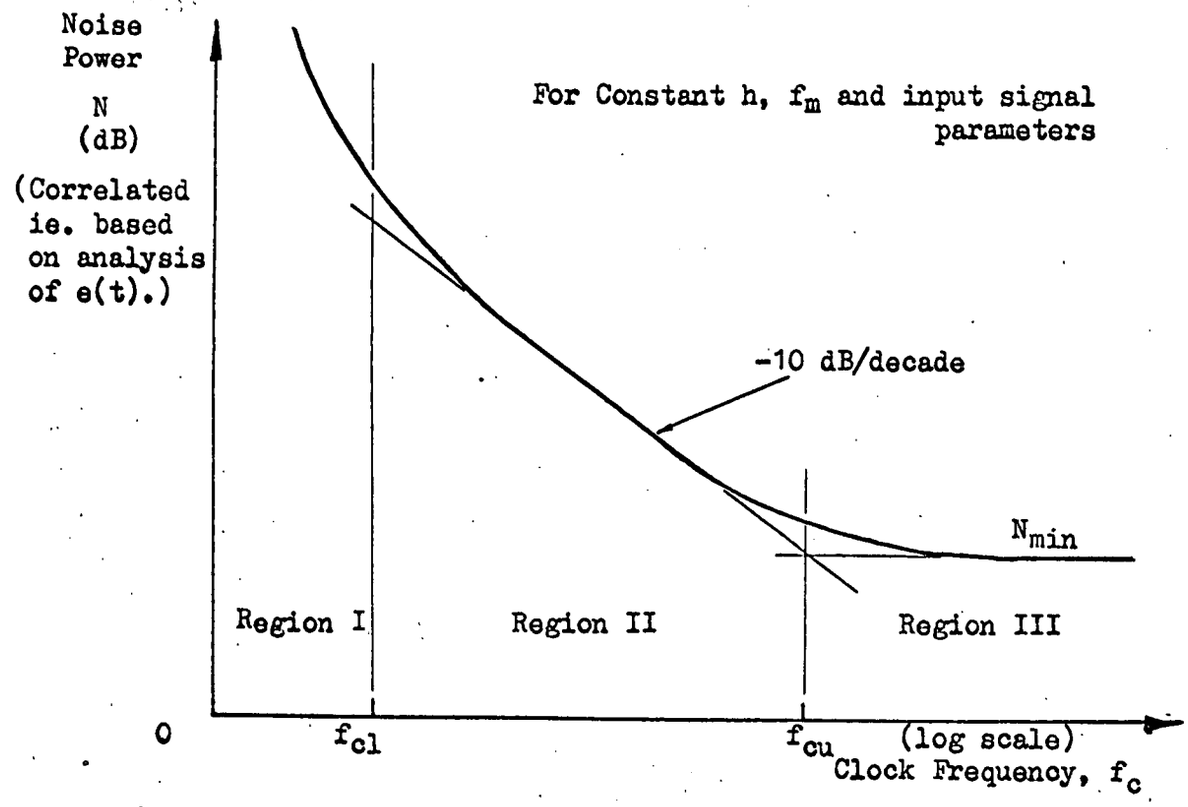
$$P = \frac{h^2}{2\pi^2}\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2} + \ldots\right)$$

$$= \frac{h^2}{12} \qquad \text{since} \quad \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$$

As $T_o \longrightarrow T_c$, the signal slope tends to zero, and the power in some given band, 0 to f_m, (where $f_m < f_c/4$,) tends to $h^2/12$ and likewise the power in the region $(f_c/2 - f_m)$ to $f_c/2$ tends to $h^2/4$. This indicates that for a zero slope (ie. dc) signal, the noise power in a given frequency band is independent of the clock frequency. This result is as would have been expected intuitively from observation of the error waveform for a dc signal.

The value of $h^2/12$ for the power $e(kT_c)$ does not represent a minimum power in the bandwidth $0 - f_m$ for all conditions. Fig. 2.7 shows the one sided power spectrum of $e(kT_c)$ in the region 0 to $f_c/2$ for a signal with a slope of constant magnitude $= h(f_c - f_o)$ where $h(f_c - f_o) \ll hf_c$. Therefore Fig. 2.7 shows the estimated power spectrum of $e(kT_c)$ for a constant triangular wave input. For a sinusoidal input or any varying slope input signal with a maximum slope, $f'(t)_{max.}$, of $h(f_c - f_o)$, the discrete spectral lines of Fig. 2.7 will represent the maximum deviation of each component from its respective axis (either $f = 0$ or $f_c/2$). A similarity between the general shape of Fig. 2.7 and Fig. 2.4 with alternating signs between adjacent autocorrelation values can be noted. It can be seen that if a number of high autocorrelation values of alternating sign were considered in Fig. 2.4 (which would correspond to the autocorrelation function of $e(kT_c)$ under the conditions being considered), the power spectrum in Fig. 2.4 would tend to the extreme case shown in Fig. 2.7. The immediate conclusion which can be made from this analysis is that for a given signal with a maximum slope limitation (i.e. a band limited signal) the increasing of f_c will only produce a reduction of the noise power in a given band, 0 to f_m, until the noise power components centered around $f_c/2$ (namely $P_{ek}\Big]_{f_c/2}$) fail to make any significant contribution to the power in the region 0 to f_m. Thus the noise power reduction of simple DM of 10dB/decade for increasing f_c, as estimated in Eqn. 2.6 from the earlier analysis, will have a cut-off point at some clock frequency, above which no noise

FIG 2.8 Estimated Nature of Noise Performance for simple (single integration) Delta Modulation.



where

f_{cl} = minimum clock frequency for no slope overload; defined by $f'(t)_{max} = hf_{cl}$.

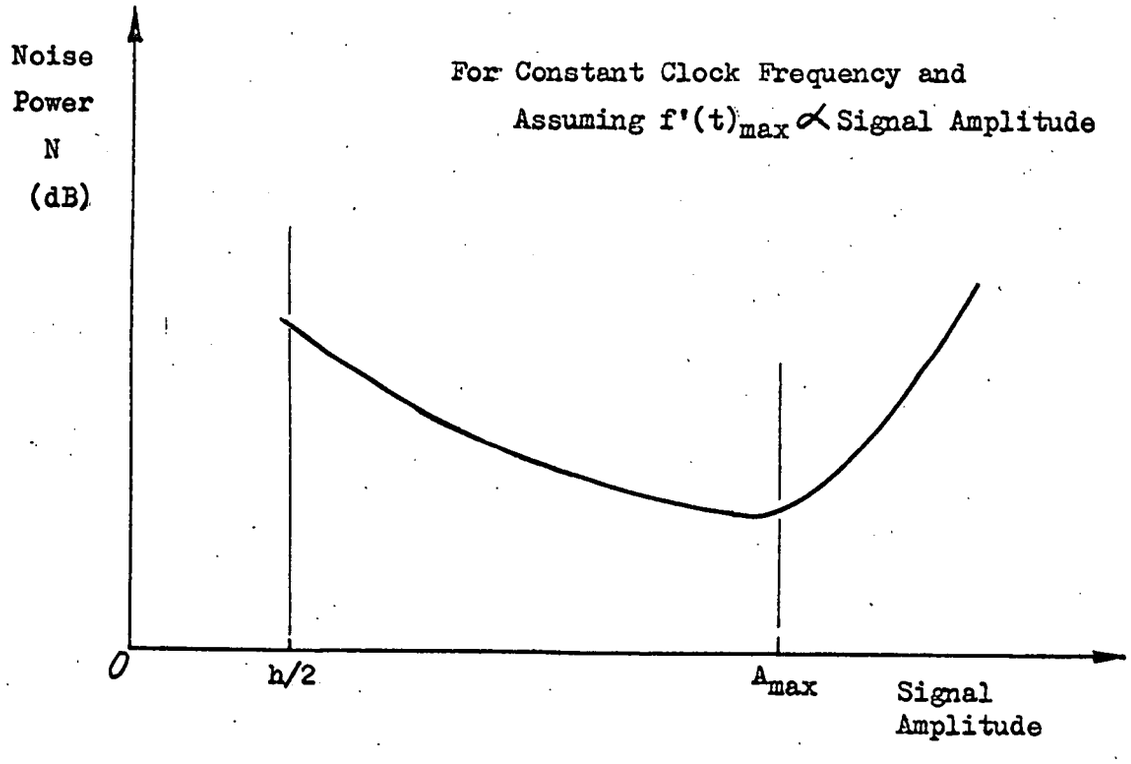
f_{cu} = function of f_m and $f'(t)_{max}$

N_{min} = function of $1/f'(t)_{max}$, h^2 and f_m and $N_{min} < h^2/12$ (for $f_m < f_c/4$). In general $N_{min} < h^2/2\pi^2$

In Region II, the asymptote is given by

$$N_q = \frac{2h^2 f_m}{k f_c}$$ for ideal LP filtering, with cut off frequency, $f_m \ll f_c$.

where $k > 3$ (and appears to be of the order of 6)



power reduction will occur. A general expression for the value of this upper"cut-off" clock frequency, f_{cu}, is not readily available from the preceeding analysis but it can be seen that f_{cu} depends mainly on f_m, and the maximum signal slope, $h(f_c - f_o)$ and also, to some degree on the probability density function of the signal derivative, f'(t).

A second conclusion from Fig. 2.7 is that for a given f_c and f_m the minimum noise power in the region 0 to f_m, denoted by N_{min}, will reduce for increasing $f'(t)_{max}$, due to an increasing proportion of $P_{ek}\big]_0$ falling outside the frequency band 0 to f_m. This reduction is again limited and will only occur until f_{cu} (dependant on $f'(t)_{max}$) increases to f_c. (This corresponds to the point at which the increase in $f'(t)_{max}$ gives rise to components of $P_{ek}^*(f)$ which are centered around $f_c/2$ (namely $P_{ek}\big]_{f_c/2}$), having a significant effect in the region 0 to f_m.)

2.3.5 Summary

Fig. 2.8 summarizes the expected nature of the noise power, with curves showing the predicted nature of its relationship to f_c and the input signal amplitude. To this stage the clocking frequency's effect on the noise power alone has been considered. In discussing overall performance it is to be noted that there will be a second effect of f_c on the available output SNR. For virtually all types of input signal the maximum signal slope will vary directly with the signal amplitude, A. Since the maximum slope capability of $r(t) = hf_c$, then for a constant state of loading, $f'(t)_{max}$ and hence A will vary directly with f_c. Therefore for increasing f_c there will be an additional 20dB/dec. increase in the output SNR available, while maintaining the same level of loading.

2.4 Review of Established Noise Analysis

The first performance analysis for delta modulation was proposed by de Jager[9] to describe the quantizing noise. De Jager's analysis is based on the following assumptions:

(i) The noise is considered to result from the error function e(t) after low-pass filtering.

(ii) Very little correlation exists with e(t) for $\tau \gg T_c$ and hence the power spectrum is flat for $f \ll f_c$. Therefore the quantizing noise power, Nq, is proportional to f_m, where f_m is the low-pass filter cut-off frequency.

(iii) The power density varies inversely with f_c since the total power of e(t) must remain constant for all f_c. Therefore $N_q \propto 1/f_c$.

(iv) The quantization noise power is totally independent of the input signal parameters.

Based on these assumptions and using numerical methods to calculate the constant of proportionality, de Jager arrived at the expression for the quantizing noise power, for single integration, as:

$$N_q = 0.316\frac{f_m h^2}{f_c}$$

Considering a sinusoidal signal (of frequency f_s) and taking the maximum signal amplitude before overloading (A_{max}) as given by: $f'(t)_{max} = 2\pi f_s A_{max} = f_c h$,

The maximum signal power, $S_{max} = f_c^2 h^2 / 2(2\pi f)^2$

Therefore $$\frac{S_{max}}{N_q} = 0.04\frac{f_c^3}{f_s^2 f_m}$$

Similarly for double * integration de Jager predicts

$$\frac{S_{max}}{N_q} = 0.000676\ \frac{f_c^5}{f^2 f_m^3}$$

Similar analysis of a single integration exponential delta modulator is made by Johnson [29] with the further assumptions

---

* This "double" integration actually being mixed integration as referred to in section 1.4.3.

that:

(i) If the noise function is obtained by subtracting the input, delayed by one clock period, from the output; the noise will be of a random nature and will be representative of the unwanted output signal.

(ii) The power spectral density of the noise will be of the form $\frac{\sin^2(\pi f/f_c)}{(\pi f/f_c)^2}$ which will be equal to 1, for $f_m \ll f_c$.

The constant of proportionality was determined from a computer simulation designed to give the total (unfiltered) noise power. This led to the expression for N_q of:

$$N_q = \frac{2Kf_m d^2}{f_c}$$

where d is the maximum step height and the value of K was estimated to be approximately 1/6 for most input amplitudes but increasing as slope overload is approached to reach the value of approximately 1/3 at the point of slope overload. Johnson's predicted output SNR expression with K = 1/6 is:

$$S_{max}/N_q = (3/8\pi^2)\ (f_c^3/f_s^2 f_m)$$

which is very close to that of de Jager.

A more thorough analysis of the quantizing noise is presented by van de Weg[4] for single integration delta modulation. Van de Weg derives an exact mathematical expression for the quantizing noise amplitude spectrum when a generalized input signal, f(t) is considered. To evaluate the noise power from the unresolved expression, van de Weg considers three limitations.

(i) ideal low pass filtering with cut-off frequency ω_m.

(ii) the rms value of the signal derivative (denoted by $D^{\frac{1}{2}}$ where $D = \overline{[f'(t)]^2}$) equals 1/4 of the overload slope.
ie. $D^{\frac{1}{2}} = hf_c/4$

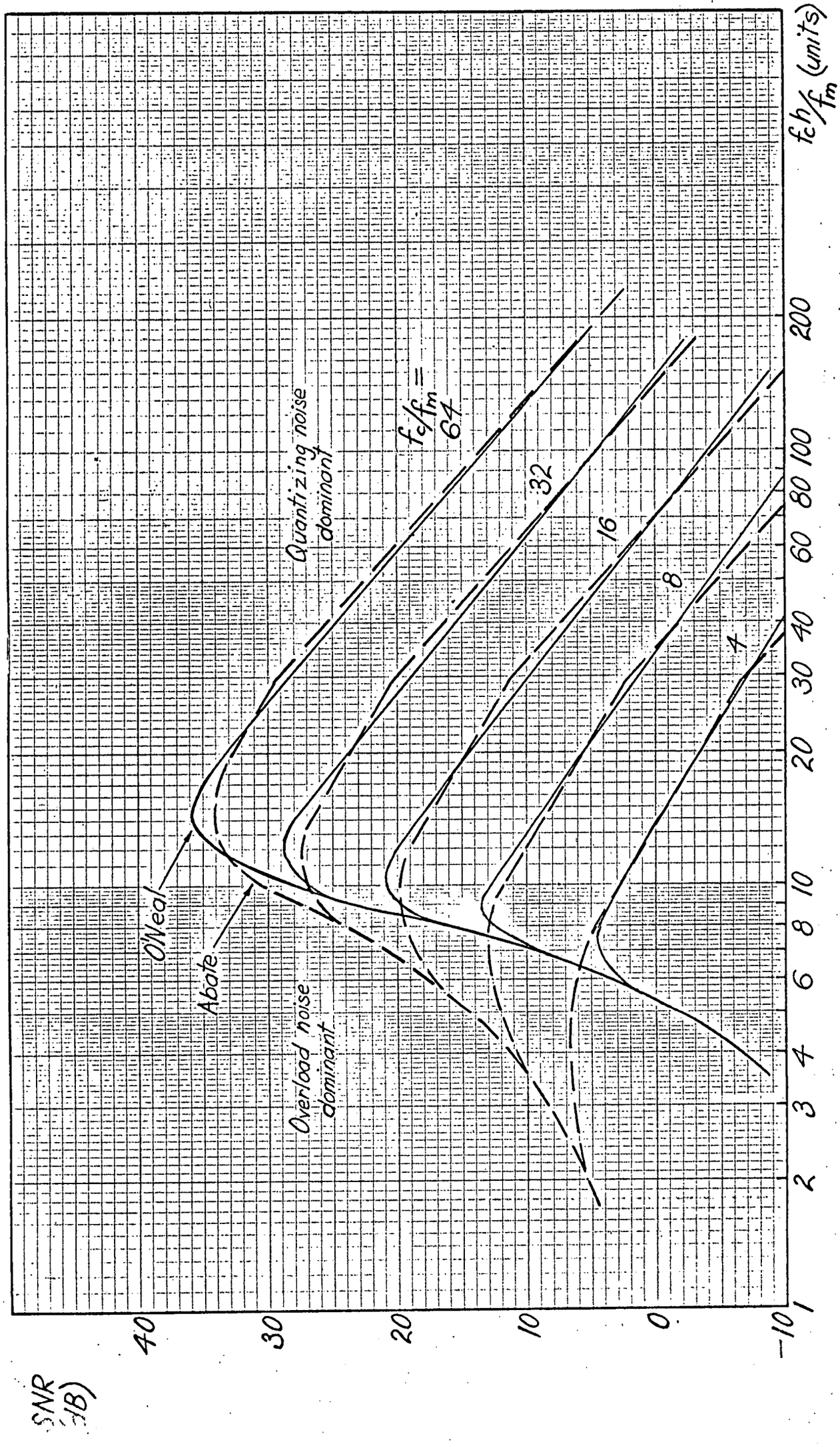
(iii) the input signal, f(t), having the characteristics of random noise with a power spectral density of constant value, ψ_o, for frequencies up to ω_m and zero above ω_m.

With suitable approximation, van de Weg arrives at the expression for the noise power of *:

---

*This equation is a slightly reduced form of van de Weg's equation[34] as presented by O'Neil[35].

Fig. 2.9 Theoretical Performance Curves, from O'Neal[35] and Abate[17] for a random bandlimited Gaussian input signal with a uniform spectrum and: mean square power, σ^2 = 1 unit.
step height, h = 1 unit.



$$N_q = \frac{2h^2 f_m}{3f_c} + \frac{8h^2}{\pi^2}\frac{f_m}{f_c}\sum_{n=1}^{\infty}\sum_{l=1}^{\infty}\frac{(-1)^{nl}}{l^2}\frac{\sin(2\pi n f_m/f_c)}{2\pi n f_m/f_c}\, e^{-\frac{\pi^2 l^2}{h^2}\psi_o(1-a_n)} \quad \ldots\ldots\ldots 2.7$$

where a_n is the value of the input signals autocorrelation function $R_f(\tau)$ for $\tau = nT_c$ (i.e. $a_n = R_f(nT_c)$) and the other symbols have their usual meaning. With the above limitations applied, numerical reduction gives the SNR approximation

$$S/N_q(dB) = 30\log(f_c/f_m) - 18.25, \quad \text{provided } f_c > 4f_m \quad \ldots\ldots\ldots\ldots 2.8$$

This equation is independent of any signal parameters or the step height since it applies for the particular input signal magnitude which gives an rms value of the signal derivative = $hf_c/4$.
Assuming the quantizing noise is independent of the type of signal, van de Weg applies this analysis for a full-load sine wave signal to give:

$$\frac{S_{max}}{N_q} = 10\log\left(0.040\,\frac{f_c^3}{f^2 f_m}\right) \quad dB$$

which is identical with the expression arrived at by de Jager. More specifically, for a sine wave signal, applying van de Weg's condition of the signal being 1/4 of the overload value, we get:

$$N_q = 0.317\, h^2\, f_m/f_c, \quad \text{for} \quad A = \sqrt{2} f_c h/4x2\pi f_s \quad \text{and} \quad f_c > 4f_m$$

Considering again the general equation for N_q of van de Weg (Equation 2.7) it can be seen that the first term is the same as the expression derived in Section 2.3.1 for the quantizing noise power. Van de Weg however, indicates that the summation term of Equation 2.7 has a very significant effect on the estimate of N_q; reducing the expression for N_q for the 1/4 load sine wave signal from $0.667\, h^2 f_m/f_c$ to $0.317\, h^2 f_m/f_c$.

Van de Weg's expression for the qunatizing noise power of Equation 2.7 is generally accepted as the most precise and useful prediction of quantizing noise. O'Neal[35] uses van de Weg's expression to determine the quantizing noise performance with varying system parameters by completing the analysis without restricting the signal to a specific level of loading. This is done for band limited Gaussian input signals with both flat and "integrated" (RC filtered) spectrums. O'Neil's estimation of the overloading noise power is based on the analysis of Rice (ibid) and the resulting curves for a bandlimited Gaussian signal with a uniform spectrum are shown in Fig. 2.9. These curves are given in terms of a "normalized step size", hf_c/f_m for an input signal with a constant rms

value of unity. Thus for a signal with an rms value of σ the step height, h, of Fig. 2.9 can be regarded as being the step height relative to the signal (i.e. h/σ).
The validity of the analysis is checked by comparison with computer simulation results. Good agreement (within about ½dB) was found in the qunatizing noise region but in the overload region O'Neil's curves underestimated the SNR very significantly.

Further analysis of slope overloading noise is presented by Protonotarios[36] whose expressions are generally accepted in preference to those of Rice/O'Neil as being more accurate.
Protonotarios uses van de Weg's description of the quantizing noise power. Further confirmation of van de Weg's description is provided by Goodman[37]. Working from a consideration of the delta modulation signals as sampled data random processes, Goodman provides an independent analysis of the quantizing noise. An expression is determined by deriving the correlation statistics of the input signal and this gives an expression which is virtually the same as van de Weg's.

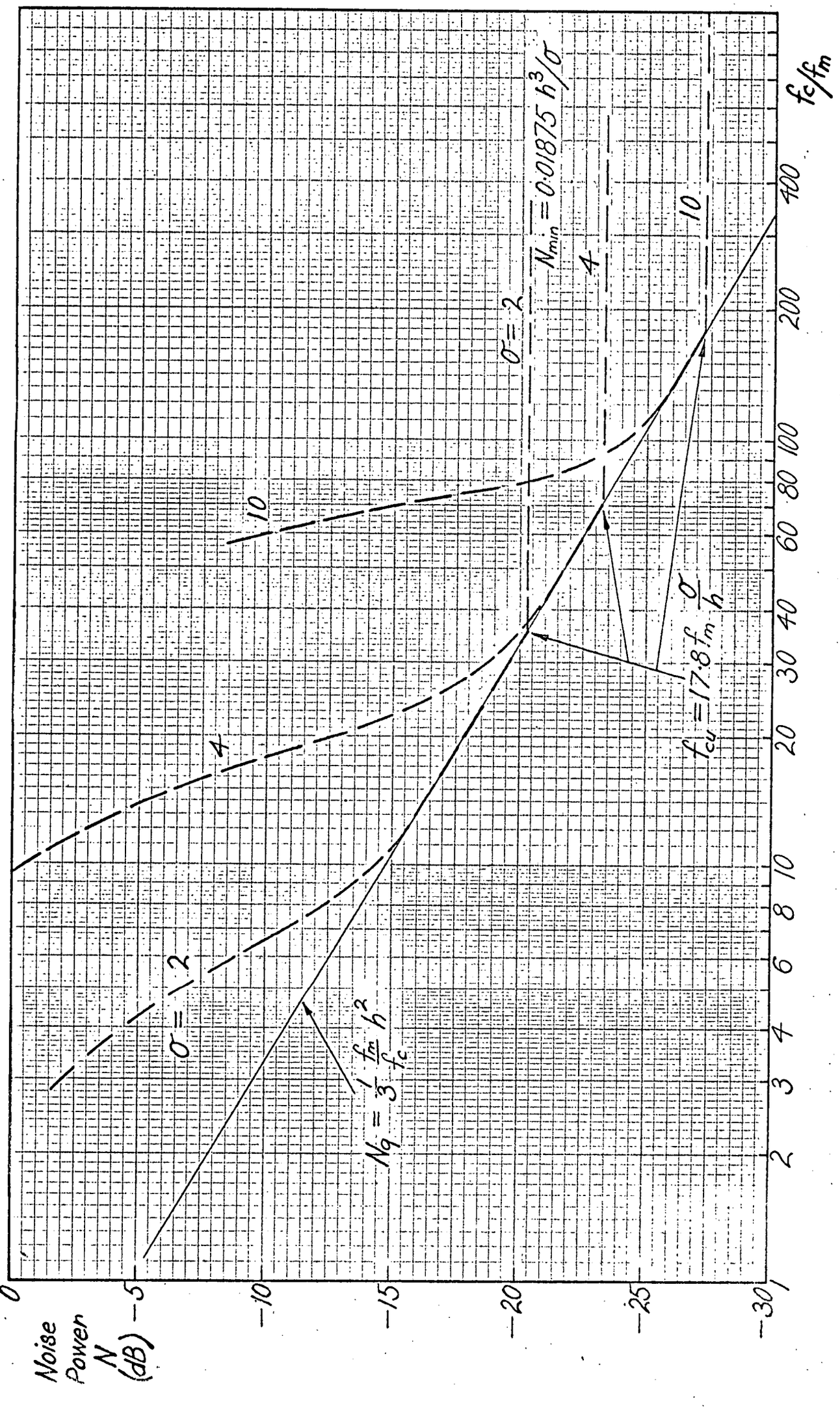
The difficulty with the use of van de Weg's expression (Equation 2.7) for the prediction of the quantizing noise performance is that further analysis is required to reduce the equation, so that a satisfactory description of the quantizing noise power for a particular type of input signal is provided. Abate[17] provides simple equations to give approximate descriptions of the quantizing and overload noise. These are based on empirical observation from the computer simulation results of O'Neil and not on analytical considerations. Abate describes the performance in terms of a "slope loading factor" s, which is directly related to O'Neil's "normalized step size". In terms of the mean power of the signal derivative, D, Abate defines s as:

$$s = hf_c/D^{\frac{1}{2}} \quad \text{where} \quad D = \overline{\left[f'(t)\right]^2} = \int_0^{\omega_m} \omega^2 P_f(\omega)\, d\omega$$

and $P_f(\omega)$ is the one sided power spectrum of the input signal. As with O'Neil's description of the noise power, Abate considers an input signal with an rms value of unity. For a signal with an rms value of σ, the value of s should be divided by σ (to give $s = hf_c/\sigma D^{\frac{1}{2}}$) and the value of D multiplied by σ^2.

Abate's empirical description of the noise power is given by the equations to three asymptotes, as below:

Fig. 2.10 Theoretical Noise Performance with Varying f_c/f_m, from Abate[17] for a random bandlimited input signal with an integrated spectrum and h = 1 unit.



Quantizing noise power, $N_q = \frac{\pi^2}{6}\left(\frac{D}{\omega_m^2}\right)\frac{f_m^3}{f_c^3}\,s^3$, for $s>8$ 2.9

$$N_q = \frac{4\pi^2}{3}\left(\frac{D}{\omega_m^2}\right)\frac{f_m^3}{f_c^3}\,s^2, \quad \text{for } s<8 \quad ....2.10$$

Overload noise power, $N_o = \frac{8\pi^2}{27}\left(\frac{D}{\omega_m^2}\right)(3s+1)\,e^{-3s}$ 2.11

For a bandlimited signal with a uniform spectrum (as considered by O'Neil for the curves of Fig. 2.9) the value of D/ω_m^2 is 1/3 and the equations become:

$$N_q = \frac{\pi^2}{18}\,\frac{f_m^3}{f_c^3}\,s^3, \qquad \text{for } s>8$$

$$N_q = \frac{4\pi^2}{9}\,\frac{f_m^3}{f_c^3}\,s^2, \qquad \text{for } s<8$$

and $$N_o = \frac{8\pi^2}{81}(3s+1)e^{-3s}, \quad \text{where } s = \frac{\sqrt{3}}{2\pi}\,\frac{f_c h}{f_m}$$

The SNR performance is predicted to be asymtotic to the curves given by $S/N(dB) = -10\log(N_q+N_o)$, (since s=1) which are shown by the dashed curves of Fig. 2.9. When the quantizing noise is dominant, the performance predicted by Abate agrees closely (within about 1dB) with that of O'Neil over the region of interest. Thus Abate's simple, closed form equations for N_q provide a good approximation to the analytical quantizing noise estimate of van de Weg. When the overload noise is dominant, Abate's predicted performance is greatly different from O'Neil's and it shows much greater agreement with the results from O'Neil's computer simulation;the discrepency being in general considerably less than 5dB.

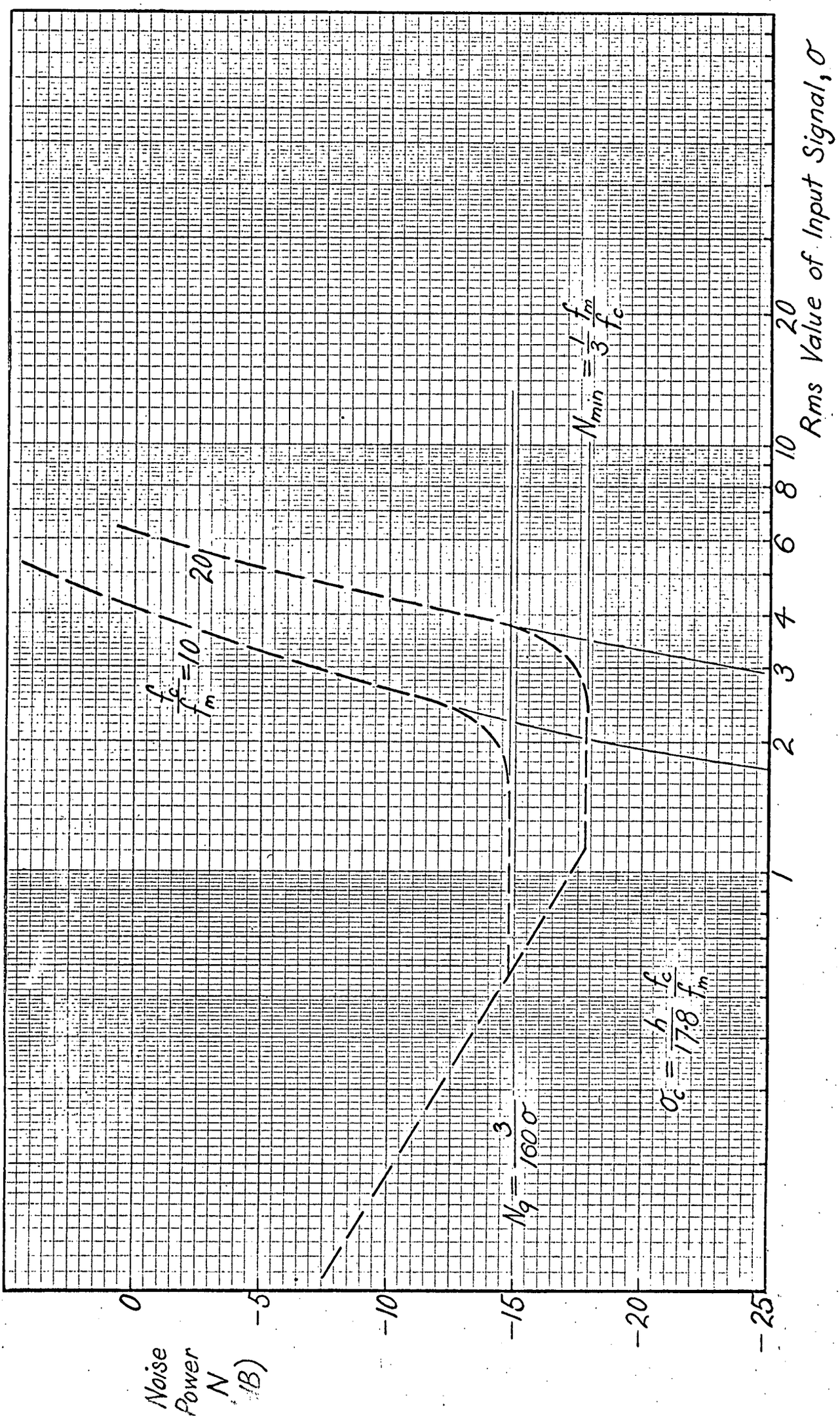
The particular case considered by van de Weg for the simplification of his general expression was for a uniform signal spectrum with $D^{\frac{1}{2}} = f_c h/4$ (i.e. signal loading = 1/4 of the overload level). This corresponds to s = 4 and gives by Abate's formulae

$$N_q = \frac{4\pi^2}{3}\,\frac{1}{3}\,\frac{f_m^3}{f_c^3}\,s^2 = 70.2\,\frac{f_m^3}{f_c^3}$$

Hence $S/N(dB) = 30\log(f_c/f_m) - 18.46$, for negligible N_o.

Fig. 2.11 Theoretical Noise Performance with Varying Input Signal, from Abate[17]

for random bandlimited input signal with an integrated spectrum and h = 1 unit.



This agrees closely with the relationship estimated by van de Weg (Equation 2.8).

The consideration of an input signal with an integrated* spectrum having a corner frequency, $f_\ell = f_m/4$ and with an exponential density function, is considered to provide a satisfactory approximation to single channel speech signals (i.e. non-multiplexed speech signals) (see Abate[17], p. 299). Abate found that his empirical equations were suitable for both Gaussian and exponential distributions. Thus adapting his equations for an integrated spectrum with $f_\ell = f_m/4$** we have:

since $$D/\omega_m^2 = \frac{f_\ell/f_m}{\tan^{-1}(f_\ell/f_m)} - \left(\frac{f_\ell}{f_m}\right)^2 = 0.1250$$

$$N_q = 0.125 \frac{\pi^2}{6} \frac{f_m^3}{f_c^3} s^3, \quad \text{for } s > 8$$

$$N_q = 0.125 \frac{4\pi^2}{3} \frac{f_m^3}{f_c^3} s^2, \quad \text{for } s < 8$$

$$N_o = 0.125 \frac{8\pi^2}{27} (3s + 1) e^{-3s}$$

where $$s = \frac{1}{2\pi\sqrt{0.125}} \frac{f_c h}{f_m} = 0.450 \frac{f_c h}{f_m}$$

The curves specified by the above three asymtote equations could be expected to provide one of the best and simplest predictions of the noise performance of simple delta modulation.

If an input signal with an rms value of σ is considered the above equations become (with the expansion of s to $0.450\, f_c h/\sigma f_m$ and D/ω_m^2 to $0.125\sigma^2$)

$$N_q = \frac{3}{160} \frac{h^3}{\sigma}, \quad \text{for } f_c > \frac{8}{0.45} f_m \sigma/h \quad \ldots\ldots 2.12$$

$$N_q = \frac{1}{3} \frac{f_m}{f_c} h^2, \quad \text{for } f_c < \frac{8}{0.45} f_m \sigma/h \quad \ldots\ldots 2.13$$

$$N_o = \frac{\sigma^2}{2.74} \left(1.35 \frac{h}{\sigma} \frac{f_c}{f_m} + 1\right) e^{-1.35 \frac{h}{\sigma} f_c/f_m} \quad \ldots\ldots 2.14$$

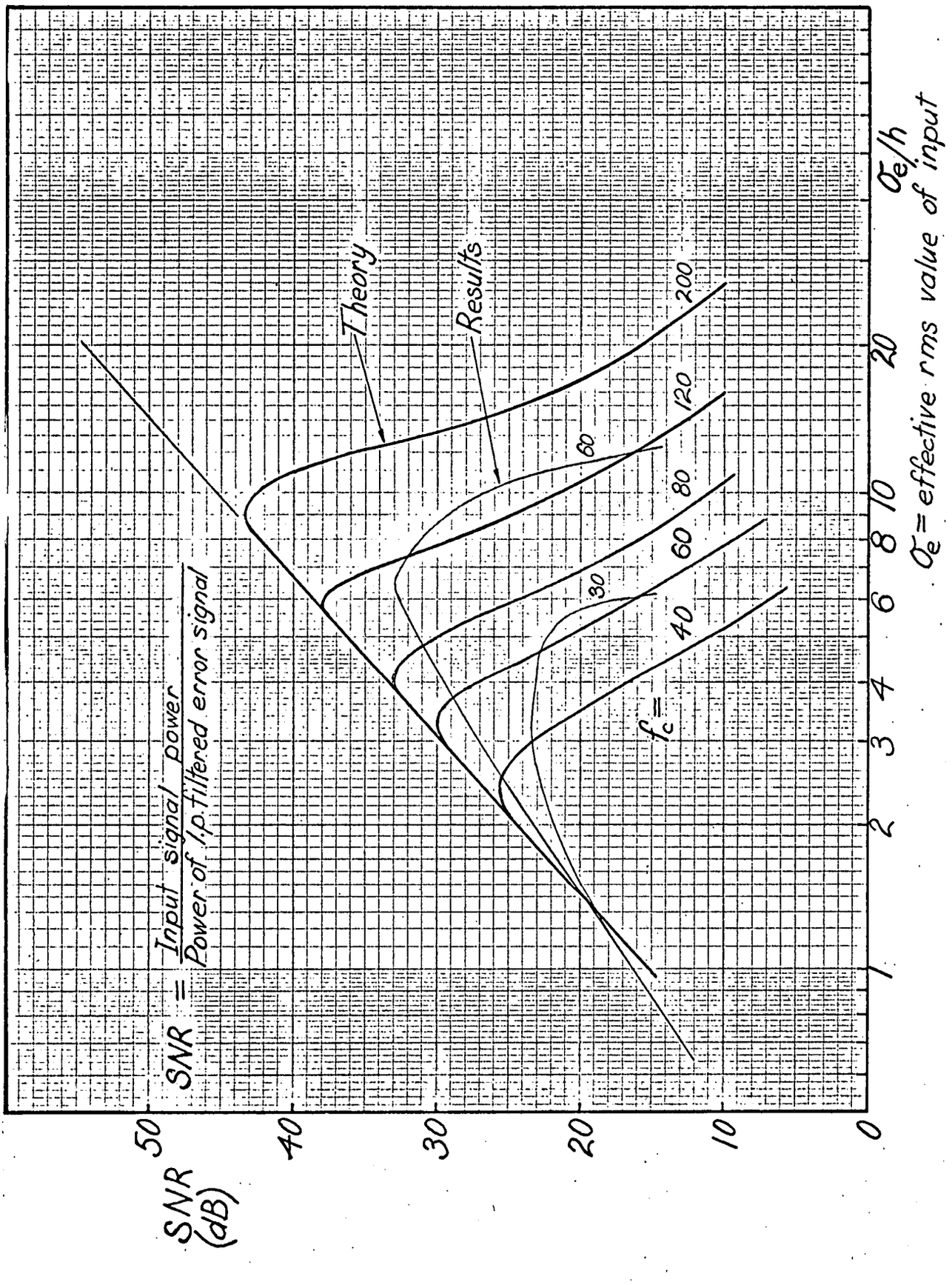
---

* The integrated spectrum is flat up to a frequency of f_ℓ and falls at 20dB/decade from f_ℓ to f_m, being bandlimited at f_m.

** $f_\ell = f_m/4$ provides a suitable speech spectrum approximation for f_m in the region usually considered for telephony.

Fig. 2.12 Theoretical Performance Curves from Zetterberg[28], for Varying Input Signal.

for a random noise input signal with a uniform spectrum, bandlimited at 200Hz and 4kHz.



Figs 2.10 and 2.11 show the variation in noise power estimated by these equations for varying f_c/f_m and σ respectively with the step height h = 1 unit. These curves indicate the two quantizing noise states specifically predicted by Abate. This feature of the quantizing noise is not discussed in any of the other delta modulation analyses although it is inherent in the general expression of van de Weg and the subsequent curves of O'Neil. This expected quantizing noise performance has two consequences which are significant in determining the conditions for optimum performance.

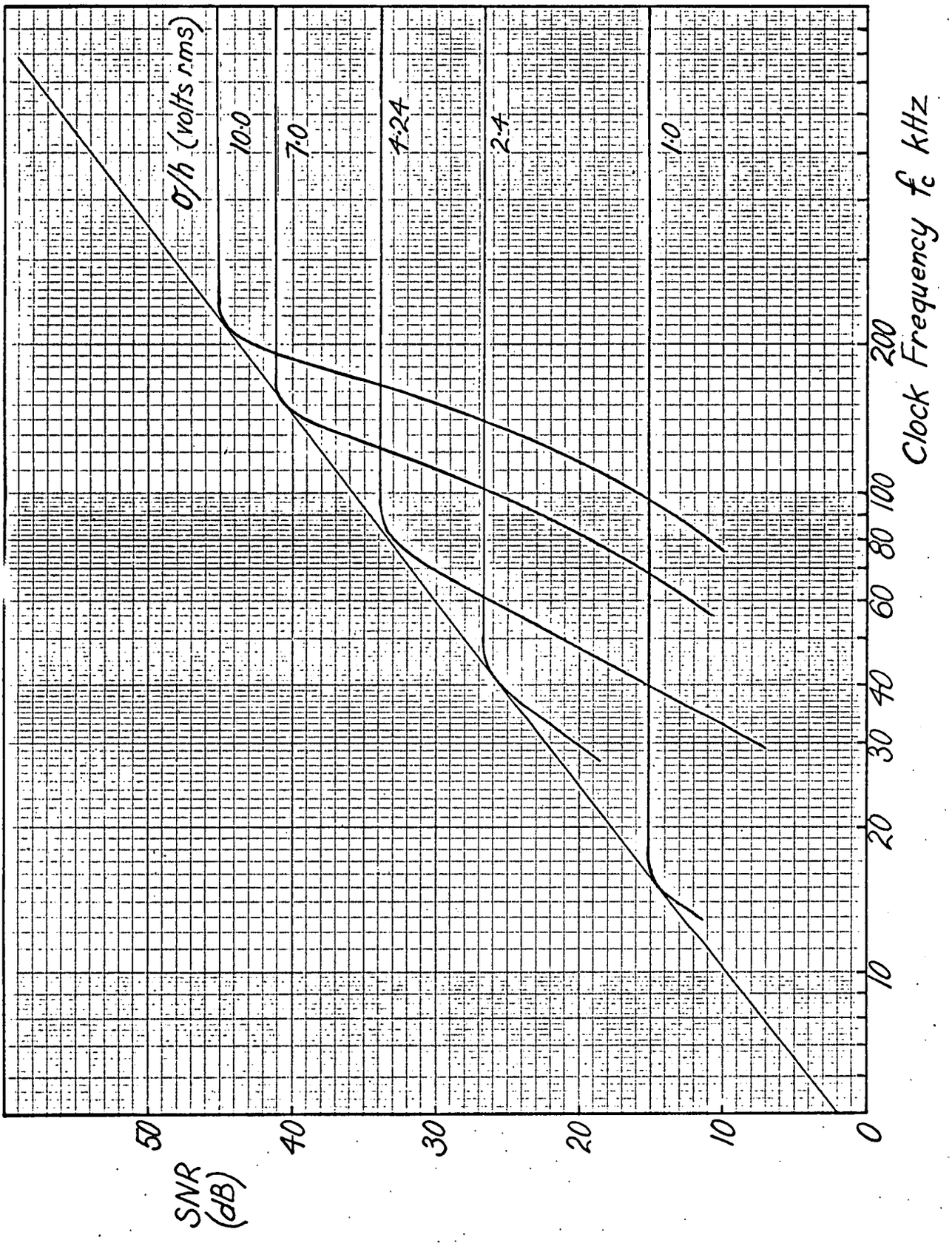
(i) Under the conditions of a constant magnitude input signal (Fig. 2.10) there is an upper clock frequency, f_{cu}, above which little reduction in noise power is expected. Most other analysis, such as that of de Jager and Johnson, does not predict any limit to the 10 dB/decade noise reduction with increasing f_c. The general shape of the quantizing noise against clock frequency characteristic shown in Fig. 2.10 was predicted in Section 2.3 (see Fig.2.8).

(ii) For a constant f_c and f_m (as in Fig. 2.11) Abate does not predict a constant quantizing noise for all signal values within the overload level but predicts a 10 dB/decade rise in N_q for signal values reducing below some critical level, σ_c. Thus in order to maintain near optimum output SNR the maximum signal should be near the overloading level, not only to maximize the signal power , but also to keep the noise power at the minimum level for as large a range of input signal values as possible.

A complete analysis of both the quantizing and overload noise of delta modulation was made in the relatively early stages of the investigation of delta modulation by Zetterberg[28]. Zetterberg's analysis has since received little attention as his performances estimates show little agreement over a large range of operating conditions with either his own or subsequenct experimental results. (This is indicated in Fig. 2.12). The basis of Zetterberg's overload noise analysis is to consider the product hf_c being held constant while h, and hence the quantizing noise, diminishes towards zero. The expression for N_o thus derived is deemed to hold good for all values of h and f_c. Similarly for the quantizing noise, hf_c is considered fixed while f_c tends to become infinite and hence the overload noise tends to zero. The failure of Zetterberg's analysis to provide an adequate description of the noise, would be due to the failure of the description of the noise in the extreme conditions considered, to hold over the full range of values of h and f_c. However one interesting result of Zetterberg's analysis is that for a constant input magnitude relative to the step height (i.e. σ/h constant), it predicts a constant N_q with varying f_c as was predicted in Section 2.3 and also by Abate's equation for

Fig. 2.13 Theoretical Performance Curves from Zetterberg[28], for Varying Clock Frequency

for a random noise input signal with a uniform spectrum, bandlimited at 200Hz and 4kHz.



$f_c > f_{cu}$. This is indicated in Fig. 2.13 showing Zetterberg's estimated SNR against f_c relationship for various σ/h values. Zetterberg's quantizing noise expression(for a bandlimited random noise signal) of:

$$N_q = 2\sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^3} \quad \frac{\sqrt{2/\pi}}{(2\pi)^3} \quad \frac{f_m}{f_3}\left(\frac{h}{\sigma}\right)^3$$

can be simplified by substituting for f_3, for a uniform bandlimited spectrum, where:

$$\omega_3^2 = \int_0^{\omega_m} \omega^2 P_f(\omega)\, d\omega \Big/ \int_0^{\omega_m} P_f(\omega) d\omega \quad = \sigma^2 D/\sigma^2 = \omega_m^2/3$$

$$N_q = \frac{2.404}{(2\pi)^3} \sqrt{\frac{2}{\pi}} \sqrt{3}\, h^3/\sigma \quad = \frac{\sqrt{3}}{(2\pi)^3} \; 1.918 \; h^3/\sigma$$

From Abate's expression for $s > 8$ of:

$$N_q = \sigma^2 \frac{\pi^2}{6} \left(\frac{D}{\omega_m^2}\right) \frac{f_m^3}{f_c^3} \; s^3$$ since $D/\omega_m^2 = 1/3$ for a uniform bandlimited spectrum.

and $$s = \frac{hf_c}{\sigma\sqrt{D}} = \frac{\sqrt{3}}{2\pi} \frac{hf_c}{f_m}$$

then $$N_q = \frac{\sqrt{3}}{(2\pi)^3} \; 1.643 \; h^3/\sigma$$

Therefore for large f_c Abate's N_q expression is the same as Zetterberg's but with a value of 0.67 dB lower. From Fig. 2.10 it can be seen that Abate's noise description predicts that the central quantizing noise region (Region II) becomes relatively less significant as σ increases. In doing so, Abate's description tends towards the Zetterberg predicted situation of a direct transition from the overload region (Region I) to the high f_c, quantizing noise region (Region III). However the lack of validity of Zetterberg's analysis for the large range of conditions for normal (near-optimum) operation, severely limits its usefulness.

Further discussion of delta modulation noise and SNR performance expected by established analysis in the light of experimental and computer simulation results is included in Chapter 5.

CHAPTER 3

COMPUTER SIMULATION OF A SIMPLE DELTA MODULATION SYSTEM

3.1 Introduction

The establishment of a computer simulation of a simple delta modulator using the Elliott 503 digital computer at the University of Tasmania is discussed. The purpose of the simulation was:

(i) to determine the feasability of establishing a computer simulation for satisfactory analysis and performance estimation of delta modulation.

(ii) to determine the types of problems and the limitations involved with the computer simulation.

(iii) to provide data likely to give greater insight into the nature and performance of noise in a simple delta modulation system and suitable for comparison with predicted performance, experimental results and established theory.

A satisfactory computer simulation, in addition to providing and processing results rapidly once the programme is established, has the advantage of being able to readily provide data on such things as the error amplitude spectrum. Simulation is also useful in investigating idealized systems which may have been considered in theoretical analysis but can only be approximated to, in an experimental realization.

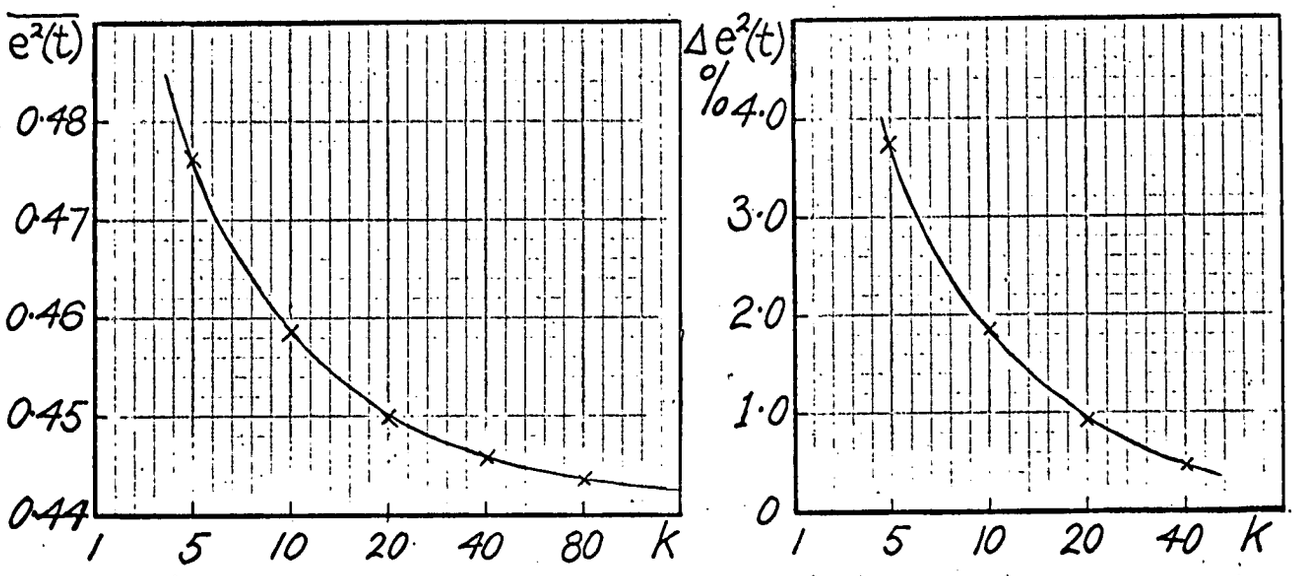
A simple, single ("ideal") integration delta modulator was considered for the computer simulation because the predicted performance and most frequently, the existing theory, is based on the single ideal integration case. In addition it provides simpler programming and the conclusions can be extended to apply to the "leaky" integration case or other cases of delta modulation.
An 800 Hz sinusoid was used as the input test signal, the actual frequency being only significant relative to the other time based parameters involved. A sinusoid signal was considered because of ease of generation and ease of separation of the output into signal and noise components. The frequency of 800 Hz was used for the test signal because this is the sine wave frequency usually accepted as being the most representative of speech signals as considered for telephony*. The step height, h, is taken to be one unit throughout.

---

*The normalized standard deviation frequency for the long term spectrum of bandlimited speech (say 0 to 3.5kHz) is given by:

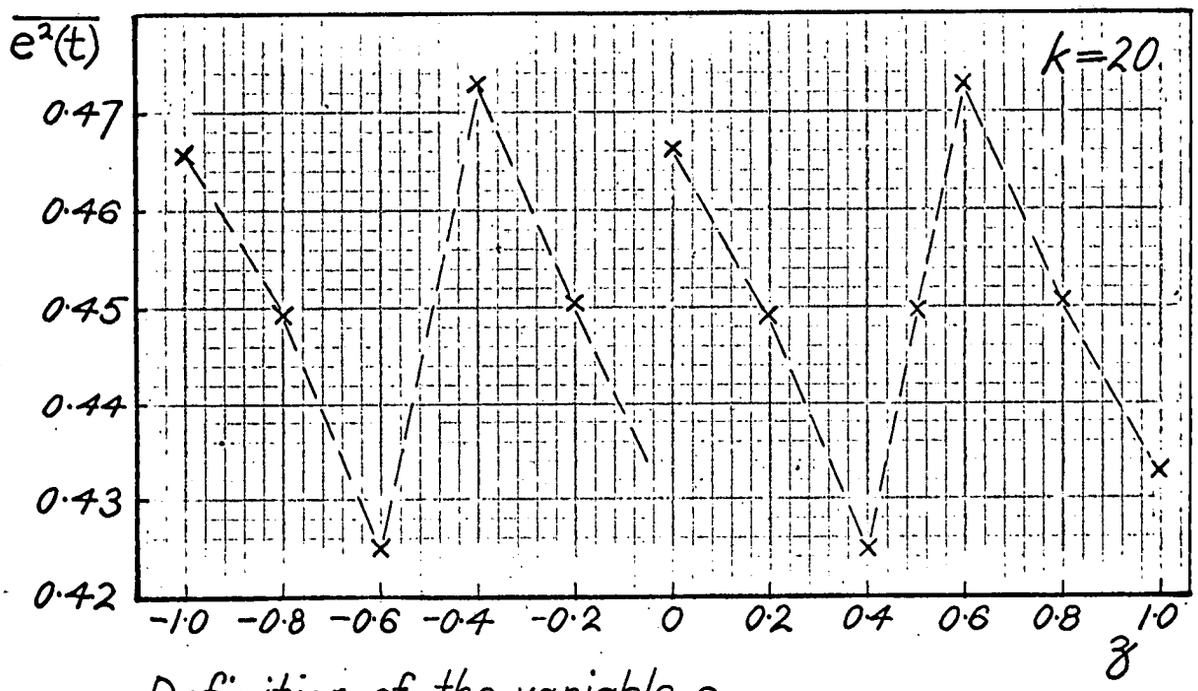
$$\bar{f} = \left(\frac{\int_{-\infty}^{\infty} S(f) f^2 \, df}{\int_{-\infty}^{\infty} S(f) \, df} \right)^{\frac{1}{2}} \simeq 800 \text{ Hz}$$

FIG 3.1 Variation of Mean Square Value of Error Function with Number of Computations per Clock Period

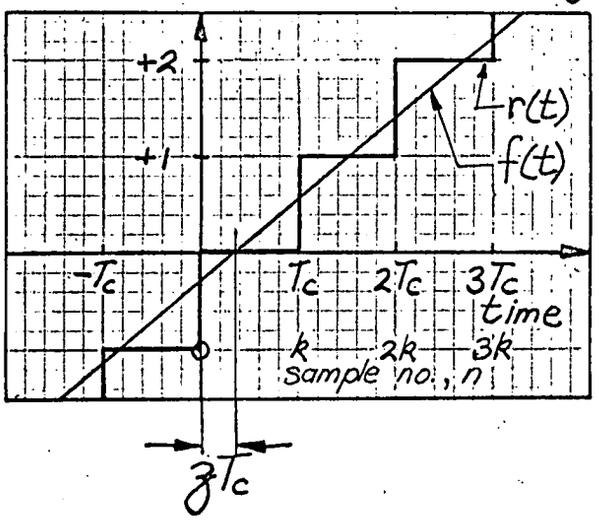


k = number of computations per clock period.
$f_c = 40$ kHz., $h = 1$ unit, $\delta = 1/2$,
Input signal $= \frac{20}{\pi} \sin(2\pi 800t)$ units

FIG 3.2 Variation of Mean Square Value of the Error Function with Initial Conditions



Definition of the variable, δ



$f_c = 400$kHz, $h = 1$unit
Input signal
$= \frac{20}{\pi} \sin(2\pi 800t)$ units

3.2 Establishment of Programme - Problems and Limitations

The first programme established was designed to give the total value of the error function, $e(t)=f(t)-r(t)$, which corresponds to the (correlated) noise with no output filtering. The important parameters relating to the simulation, whose effects were investigated, are:

(i) the number of input signal cycles considered

(ii) the computer sampling rate, f_s

(iii) the initial conditions of the system variables, f(t) and r(t).

The analysis of either r(t) or e(t) by taking sample values at a rate f_s, depends on the power of these functions falling off a high frequencies. If this is the case, the error due to the neglected power (namely the power at frequencies greater than $f_s/2$) can be made small by the use of a sufficiently large f_s. With the system's clock frequency, f_c, equal to an integral multiple of the signal frequency, the error function will repeat itself every one, or at the most two cycles of the input signal. Under such conditions only one or two input cycles are needed to obtain a description of e(t) which would be the same as that from a large number of cycles of input.

Initially, $e^2(t)$, the mean square value of e(t), was determined for various computer sampling rates and for a range of initial conditions. The system parameters used were f_c = 40kHz and the input signal amplitude, $A=20/\pi$ =80% A_{max} (where A_{max} = theoretical maximum sinusoid amplitude before overload = $f_c h/2\pi f$). Fig. 3.1 shows the percentage change in $\overline{e^2(t)}$, for various k, which would result if k were doubled. (k = number of computations per clock period. ie. $f_s=kf_c$). The low percentage changes in $\overline{e^2(t)}$ indicates that the noise power is low at these high frequencies and that computation rates of 10 or 20 times f_c should give a good description of the error function (with $\overline{e^2(t)}$ within about 4% and 2% respectively, of the actual value). If only the power up to same frequency, less than $f_s/2$, is being considered then these computation rates would be expected to give an even closer estimate of the actual power.

The effect of varying the delta system's function values at the start of computation is shown in Fig. 3.2 where z is as defined in the diagram. The range of z shown is the range that would be expected in an actual system. Over the full range of z the expected value of $\overline{e^2(t)}$ would be about 0.45 with a possible variation of up to about 5% for any particular initial conditions. Under actual operating conditions each value of z would not necessarily be expected with equal probability. In addition the power of e(t) in some limited frequency range may be more susceptible to variation with different

initial conditions than $\overline{e^2(t)}$, the total power of e(t). However the values of $\overline{e^2(t)}$ in Fig. 3.2 give an indication of how good the approximation to the noise power will be if a particular set of initial conditions are used in the simulation.

The second programme investigated was designed to give the amplitude spectrum of e(t) over range of frequencies of several times the clock frequency and to give the frequency component values of the reconstructed signal, r(t) over the lower range of frequencies so that. the noise power is a given frequency range, 0 to f_m, can be calculated. The frequency analysis of r(t), in preference to e(t), was performed because this facilitates simpler programming and removal of the signal component with delay and amplitude variation taken into account as desired.

The frequency analysis was achieved by incorporating into the simulation programme a standard procedure for the computation of the Fourier coefficients of a function specified at n sample points. The procedure initially used was the HUCC Library Procedure MH01, at the University of Tasmania. This procedure was found to give rise to significant error in the computed Fourier components which was typically from 0 to 10% but for some components was well over 100%. The error resulted from the use of iterative methods in the procedure for the calculation of the sine and cosine values required. The cumulative multiplication from one initial sine and cosine value, many thousands of times, allowed the computer round off error to take on very significant proportions. For this reason a modified Fourier analysis programme was used which computed one sine and one cosine value for each Fourier coefficient, thus reducing the length of the interative processes and therefore limiting the loss of accuracy, at the expense of slightly increased computation time.

The first frequency analysis (Run 3) used an f_c of 40 kHz. Since this is an integral multiple of the signal frequency (f_s=800 Hz) it means that one input cycle gives a total description of r(t). Twenty computations were performed per clock period (that is, $f_s = 20f_c$ = 800kHz) and thus the analysis gave frequency components up to 400kHz. The results from Run 3 indicated the following problems involved with the simulation.

(i) The values of the frequency components fluctuate widely (by up to 80% from the mean) for adjacent frequencies. This problem results from the use of a single frequency test signal which gives rise to a multitude of harmonics and harmonic cross-products of the signal and clocking frequencies. This effect has been noted previously by Bennett[5], de Jager[9] and others.

(ii) Because r(t) was periodic, at 800Hz in this case, the frequency components were spaced at 800Hz intervals thus giving only a small number of components in a typical audio band, say 0 to 4kHz.

This small number of components combined with the large fluctuations in value would result in low reliability for the power indicated in a given band.

(iii) The use of only one cycle of input, or of a clock frequency equal to an integral multiple of the signal frequency, gives only one set of parameter values for the start of an input cycle. Clearly this does not give results which are representative of the total situation with the given system parameters.

These problems could be overcome by using a large number of cycles of the input signal as is done by Johnson[29] in his computer simulation to find the total (unfiltered) noise power. Also, by adding small magnitude, low frequency components to the sinusoid signal, the periodicity of r(t) can be removed by giving much more random conditions at the start of each cycle. This would also smooth the large fluctuations due to signal and clock frequency cross-products and for these reasons these methods have frequently been used for experimental measurements (de Jager[9], Hauser and Zarda[24], and others). The use of many cycles of input however, faces two major computing difficulties when frequency analysis is required.

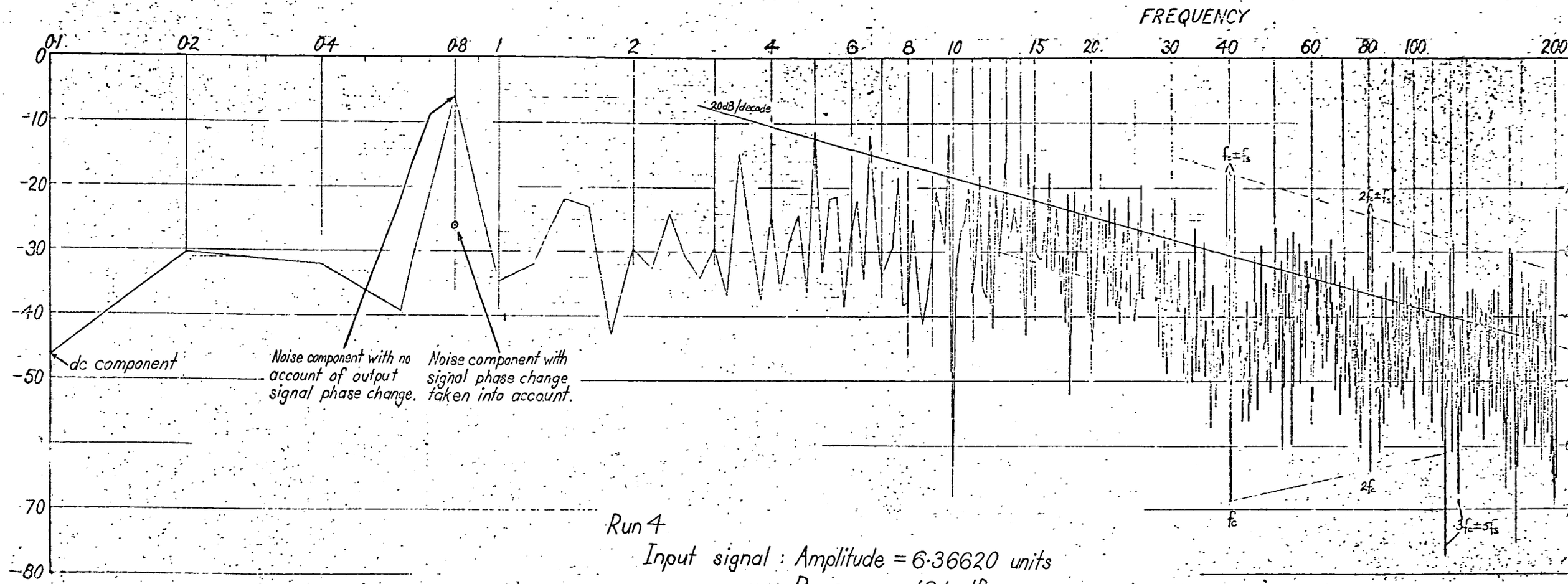
(i) The computing time for the Fourier analysis programme would become excessive. (A computing time of about 20 minutes with the Elliott 503 was required for Run 3 in which 1 cycle of the input and 1000 function sample values were considered.)

(ii) The required storage (for the function sample values) would be such that the internal store of the computer would be insufficient and external storage (such as punched tape) would be required. This would greatly increase the handling difficulties in running the simulation / frequency analysis programme.

If only the noise power after low pass filtering is required then a digital filter could be included in the simulation. This would allow the computation of the noise power by the continuous process (not requiring storage), of summing the squares of the output sample values. To consider the ease of incorporating a digital filter with this intent, a design was considered, for a typical audio filter, and is given in Appendix B. However, a digital filter was not implemented in the simulation as it is beyond the immediate aim of establishing a satisfactory simulation frequency analysis programme.

A simulation is required, which gives with a small number of input cycles, an output, r(t), approaching in nature the output which would be achieved using a large number of input cycles, each with randomized starting values. This would allow a frequency analysis to be performed and the noise power with idealized low-pass filtering

FIG 3·5 COMPUTER PRINTOUT OF ERROR FUNCTION AMPLITUDE SPECTRUM



NOISE AMPLITUDE dB (zero to peak amplitude relative to one step height)

Run 4

Input signal : Amplitude = 6·36620 units
Power = 16·1 dB
Frequency, f_s = 800 Hz

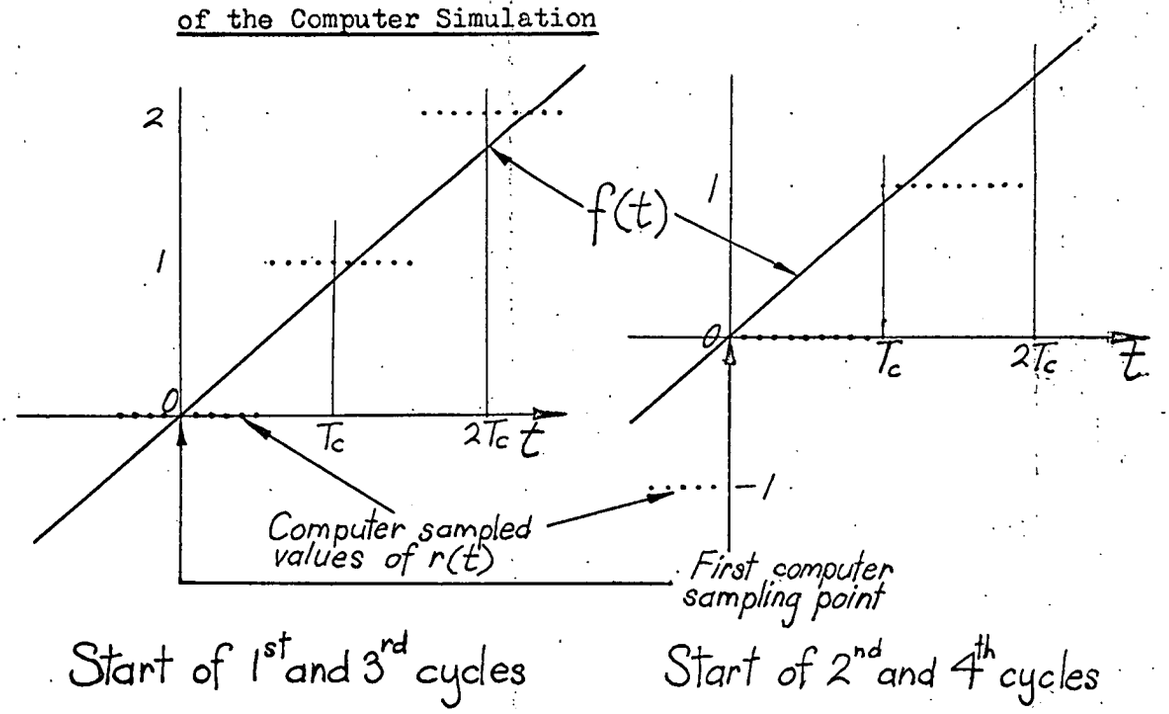
Clock frequency = 40·4 kHz

Step height = 1 unit

to be determined. In order to investigate the feasability of such a simulation giving satisfactory results, various programmes were run with varying conditions for the simulation implementation. All the parameters of the delta modulation system being simulated were kept virtually constant, with f_c close to 40kHz and an input signal amplitude of $20/\pi$ units, where h=1 unit.

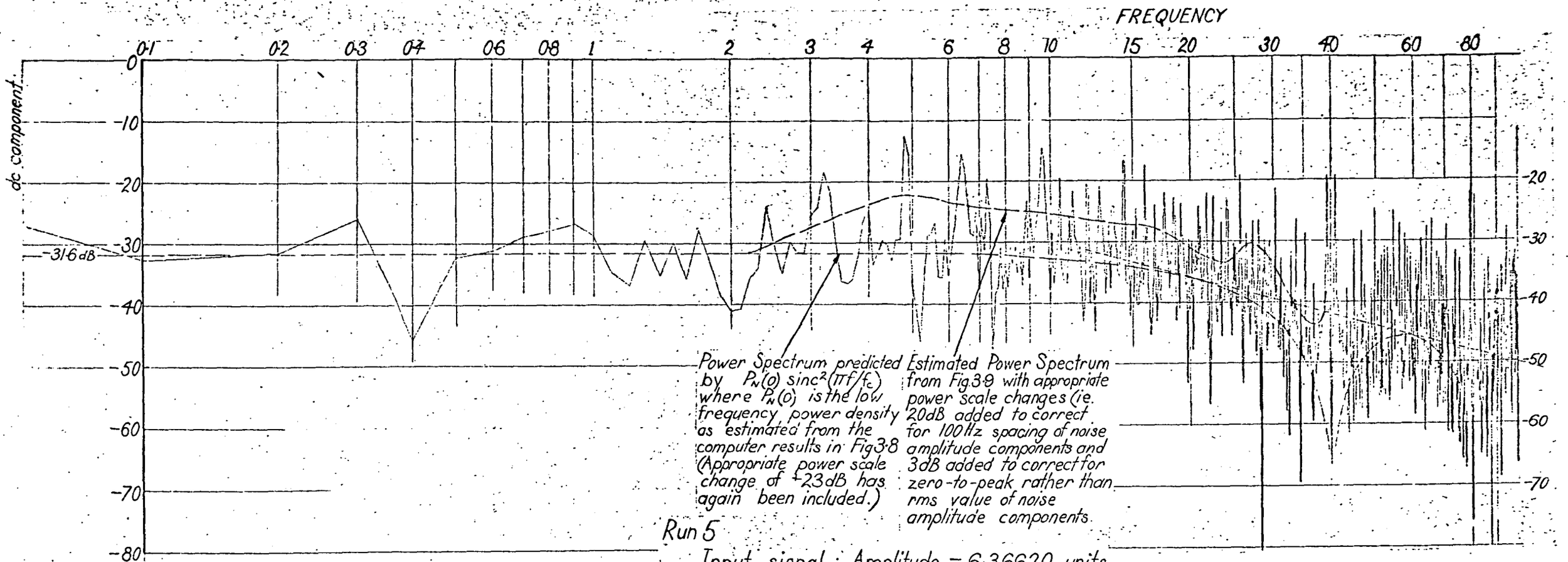
The second simulation/frequency analysis (Run 4) used 4 cycles of input (thus giving Fourier components with 200 Hz spacing) and the discrete amplitude spectrum of e(t) was plotted by the computer (see Fig. 3.5). The computer store limited the number of computations per clock period to 10 giving a computation sampling frequency of 40.4kHz and thus a frequency analysis up to 202kHz. The system's clock frequency, f_c, of 40.4kHz was selected to give Fourier components at f_c and its harmonics and also so that cross products of f_c and its harmonics, with the input signal are readily distinguishable (ie. components at $if_c \pm 800k$; i and k integral.) Also the use of an f_c which is not an integral multiple of 800 Hz gives varying system parameter values at the start of each cycle; although for Run 4 only two unique starting conditions existed, as shown below in Fig. 3.3.

FIG 3.3 Conditions at the Start of Each Cycle for Run 4 of the Computer Simulation



Run 4 was repeated with f(t) advanced by 1/20th of a clocking period (T_c) relative to the clocking instant. This changed the value of f(t) at the first computer sampling instant from 0 (as in Fig. 3.3) to $\frac{20}{\pi} \sin 2\pi \left(-\frac{1}{20} \quad \frac{800}{40,400}\right)$. This small variation in the initial conditions produced a significant change in the amplitude of the harmonic components. The fluctuations in the amplitude spectrum remained approximately the same in nature although the amplitude of any particular component varied by typically 0 - 20% and by at the

FIG 3·6 COMPUTER PRINTOUT OF ERROR FUNCTION AMPLITUDE SPECTRUM



Run 5

Input signal : Amplitude = 6·36620 units

Power = 16·1 dB

Frequency, f_s = 800 Hz

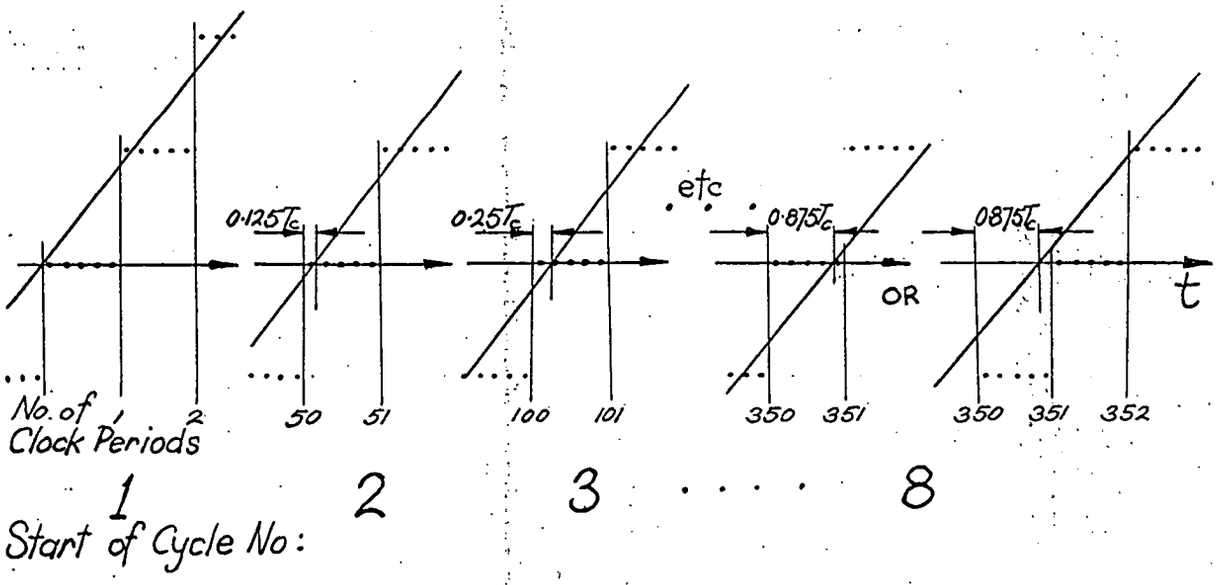
Clock frequency = 40·1 kHz

Step height = 1 unit

most over 100%. This indicates that the significant effect of the initial conditions on the amplitude spectrum of r(t) [and e(t)] must be considered in the simulation results.

The simulation / frequency analysis was repeated (Run 5) with 8 cycles of input (giving Fourier components with 100 Hz spacing). The amplitude spectrum as plotted by the computer is shown in Fig. 3.6. In order to keep within the computer's storage capacity only 5 sample values for computation could be taken per clock period. A clock frequency of 40.1kHz was used giving 50.125 clock periods per input cycle. This gave a broader range of conditions at the start of each cycle, as shown in Fig. 3.4 below

FIG 3.4 Conditions at the Start of Each Cycle for Run5 of the Computer Simulation



3.3 Simulation Effectiveness.

Despite the fluctuations of the amplitude spectra (shown in Figs. 3.5 and 3.6) the basic form of the spectrum can be observed and this is discussed in the next section. For the calculation of the power in a given band it would be required that the fluctuation of the amplitude of the frequency components be limited, so that the power does not depend to an undue degree on any particular components. Smoothing of the amplitude spectra could be achieved by assigning to each component a magnitude determined by some weighted average of the component and its adjacent component values. Such a smoothing procedure should not be applied where discrete amplitude components are clearly present, as for example at the frequencies of $f_c \pm 800$Hz. Fig. 3.6 also indicates that the fluctuation of the amplitude spectrum shows considerable 800Hz periodicity, particularly in the region 2kHz to 8kHz which is of

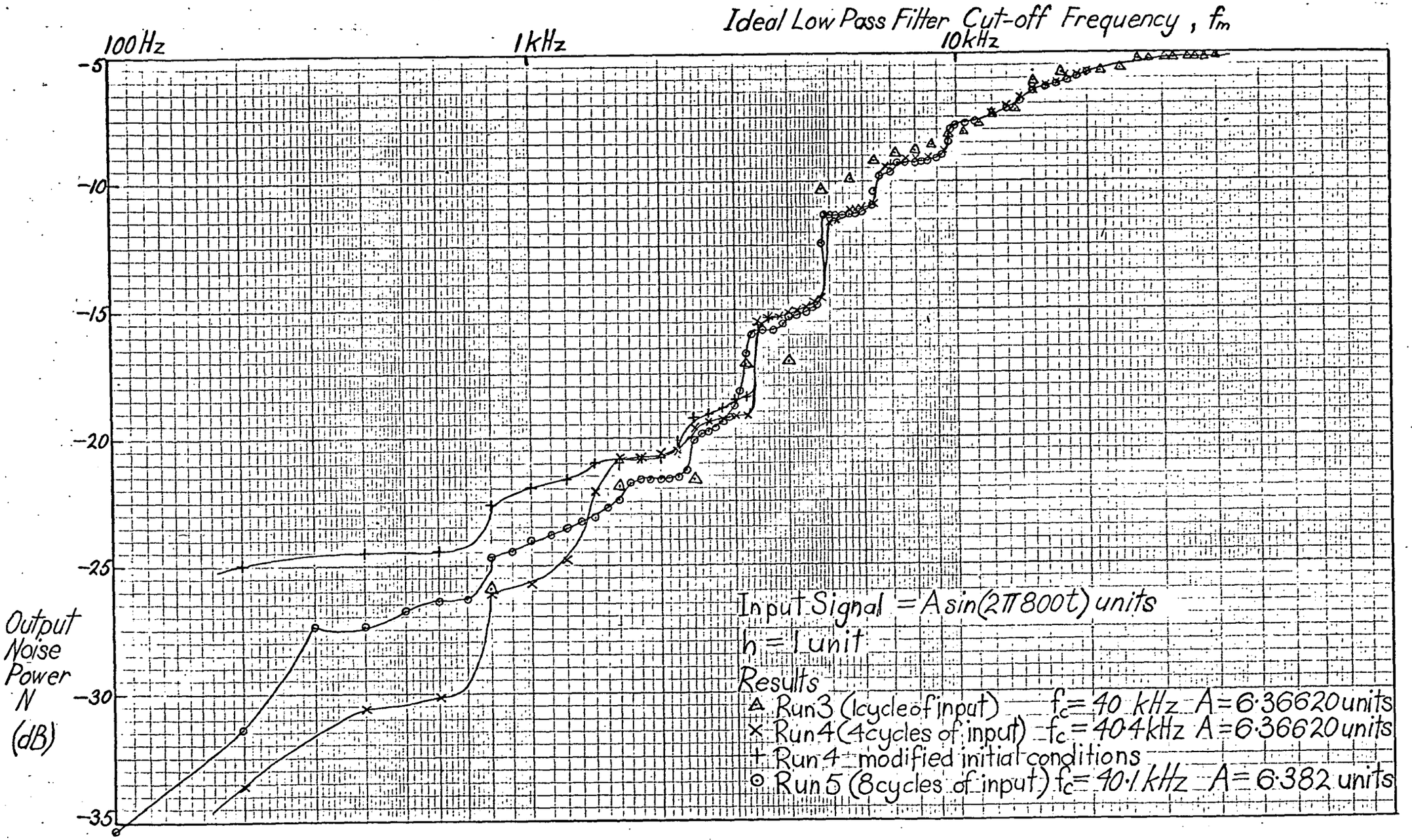


FIG 3.7 Computer Simulation Results for Noise Power Variation with Cut-off Frequency for Various Trial Runs

particular importance in the estimation of the power in the bandwidth 0 to f_m.

As a smoothing procedure, and also to investigate the effect of the different simulation conditions on the total power in a given low frequency band, curves of the total power of the noise function $n_1(t)$ in a frequency range 0 to f_m are shown in Fig. 3.7 for the various simulations. Therefore Fig. 3.7 shows the estimated noise power, N, in a frequency band, 0-f_m, for a range of f_m where N is given by:

$$N = \sum_{fk=0}^{f_m} \frac{[AMPL(fk)]^2}{2}, \quad AMPL(fk) = \text{Fourier component of frequency fk.}$$

The noise function $n_1(t)$ is not merely $e(t)=f(t)-r(t)$, after low pass filtering, since e(t) will contain a large signal component as discussed in Section 2.2 and can be seen to be present in Fig. 3.5. The uncorrelated noise, n(t) defined by:

$$n(t)=y(t)-kf(t-t_o),$$

where k and t_o are such that n(t) has minimum power is, in general, the most meaningful noise function. However in the case of a sinusoidal input, n(t) will have no component at the input frequency and thus does not give a good representation of the situation if either a continuous spectrum or a more generalized input signal were being considered. The noise function, $n_1(t)$, defined by only the phase of f(t) being varied to give a noise power minimum, gives a noise component at 800Hz of the same order as the adjacent components (see Figs. 3.5 and 3.6). Thus $n_1(t)$, defined by;

$$n_1(t)=y(t)-f(t-t_o), \quad \text{where } t_o \text{ is such that } n_1(t) \text{ has minimum power.}$$

is considered to be more representative of the uncorrelated noise for the simulation and for this reason will in general be the noise function considered.

The curves of N against cut off frequency, f_m (Fig. 3.7) provide a suitable format for smoothing the fluctuations of the discrete components of the Fourier analysis. These curves indicate that the simulations using more than one input cycle (Runs 4 to 5) give noise power estimates that are equal within 0.5 dB for f_m greater than about 3.5 kHz. The slight variation in initial conditions for Run 4 produced a large difference in N for low cut off frequencies, but for $f_m > 1.6$kHz their noise power curves are virtually identical. It can be seen that one input cycle (Run 3) does not provide a good indication of the N, f_m relationship due to insufficient Fourier components and that the use of only one set of initial conditions gives up to 2dB difference from the N estimated by the other

simulations. The large variation in the estimates of N for low f_m would be expected, as in this region the estimate of N is under the effect of the fluctuations in the Fourier components and the smoothing effect of the summation of the component powers has not taken effect.

Good agreement for the estimation of N exists between the simulations when four or more input cycles are used, providing the cut-off frequency is not considerably below that normally considered for audio signals. (The discrepency being 1dB or less for $f_m > 1.7$kHz). It can therefore be concluded that an effective simulation of a delta modulation system with a sinusoidal input, for the purpose of noise power estimation, can be achieved using only a small number of cycles of input. The use of four or more cycles would seem adequate with the system's function values at the start of each cycle spread evenly over the possible range of values. This will alleviate the problem of a highly periodic noise function with a strong dependance on the initial conditions. A suitable spread of conditions at the start of each cycle is obtained by selecting a clock frequency, f_c, which will give $n+1/c$ clock periods per input cycle; where n is any integer and c is the number of input cycles considered. Thus f_c should be given by $f_c/f_s = n+1/c$, where f_s = signal frequency.

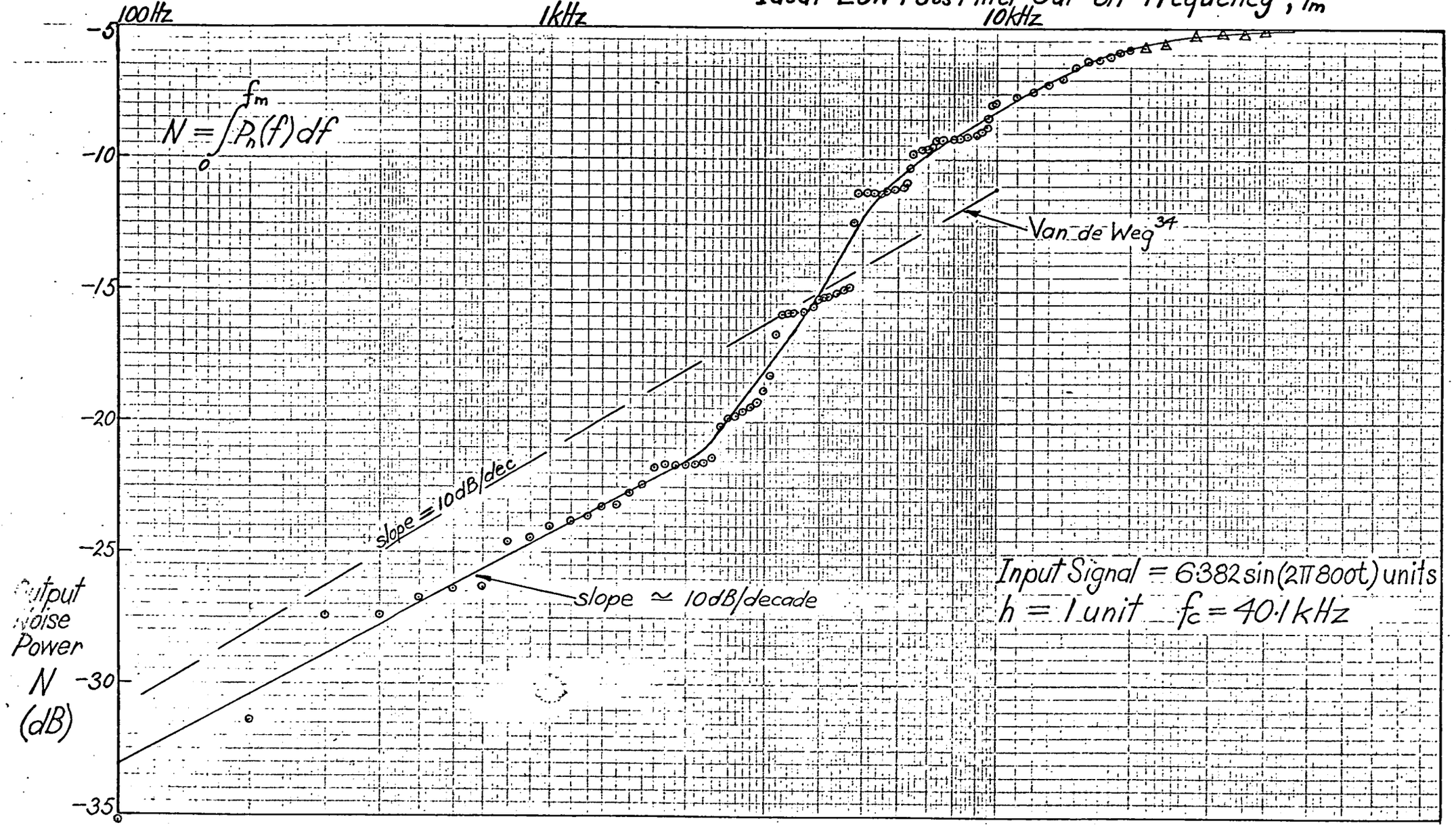


FIG 3.8 Curve of Noise Power against Cut-off Frequency from Computer Simulation Results

3.4 Noise Power and Spectral Considerations.

The curve of the noise power (N) against the cut-off frequency (fm) from a satisfactory simulation (eg. Fig. 3.7 for Run 5) is a good format for consideration of the spectrum of the noise function. The major contribution to the noise of harmonics of the signal frequency is evident from Fig. 3.7, with the 4th and 6th harmonies being dominant for this particular set of delta system parameters.

The relationship between N and f_m which would be expected for a more general input consisting of a range of frequencies, but which provides the same level of system loading, can be estimated by smoothing the effect of the discrete noise components as shown in Fig. 3.8. From this N against f_m curve the continuous spectrum of N can be estimated and compared with the discrete spectrum of Fig. 3.6. Considering again the expression for the computation of N: $N = \sum_{fk=0}^{f_m} \frac{[AMPL(fk)]^2}{2}$, an estimate of the power density spectrum (one sided) of the noise function, $P_n(\omega)$. can be made from

$$\int_0^{\omega_n} P_n(\omega)d\omega = N$$

The spectrum, $P_n(\omega)$, was estimated graphically from a curve of N against f_m using linear scales by taking the instantaneous derivative at various points. The resulting spectrum is shown in Fig. 3.9.

It should be noted that this process of "smoothing" the N against f_m curve before graphical differentiation will give only a rough estimate of the noise spectrum which would result from an input of a more general nature than the sinusoid, with a range of component frequencies. However, it should provide significant information about the nature of the noise spectrum. This estimated continuous spectrum is superimposed on Fig. 3.6 (with appropriate power scale changes) to indicate the correlation with the computer plotted discrete amplitude spectrum*. (The curve of Fig. 3.9 does not follow the mean of the discrete amplitude values of Fig. 3.6 since it represents a "smoothed" curve of the mean square value, but the correlation between the spectra shapes is exhibited.)

* It should be noted that the computer plot of the discrete amplitude spectrum is presented by the straight line joining of the discrete values, whereas the spectrum should actually be represented by impulse functions at the discrete Fourier component frequencies.

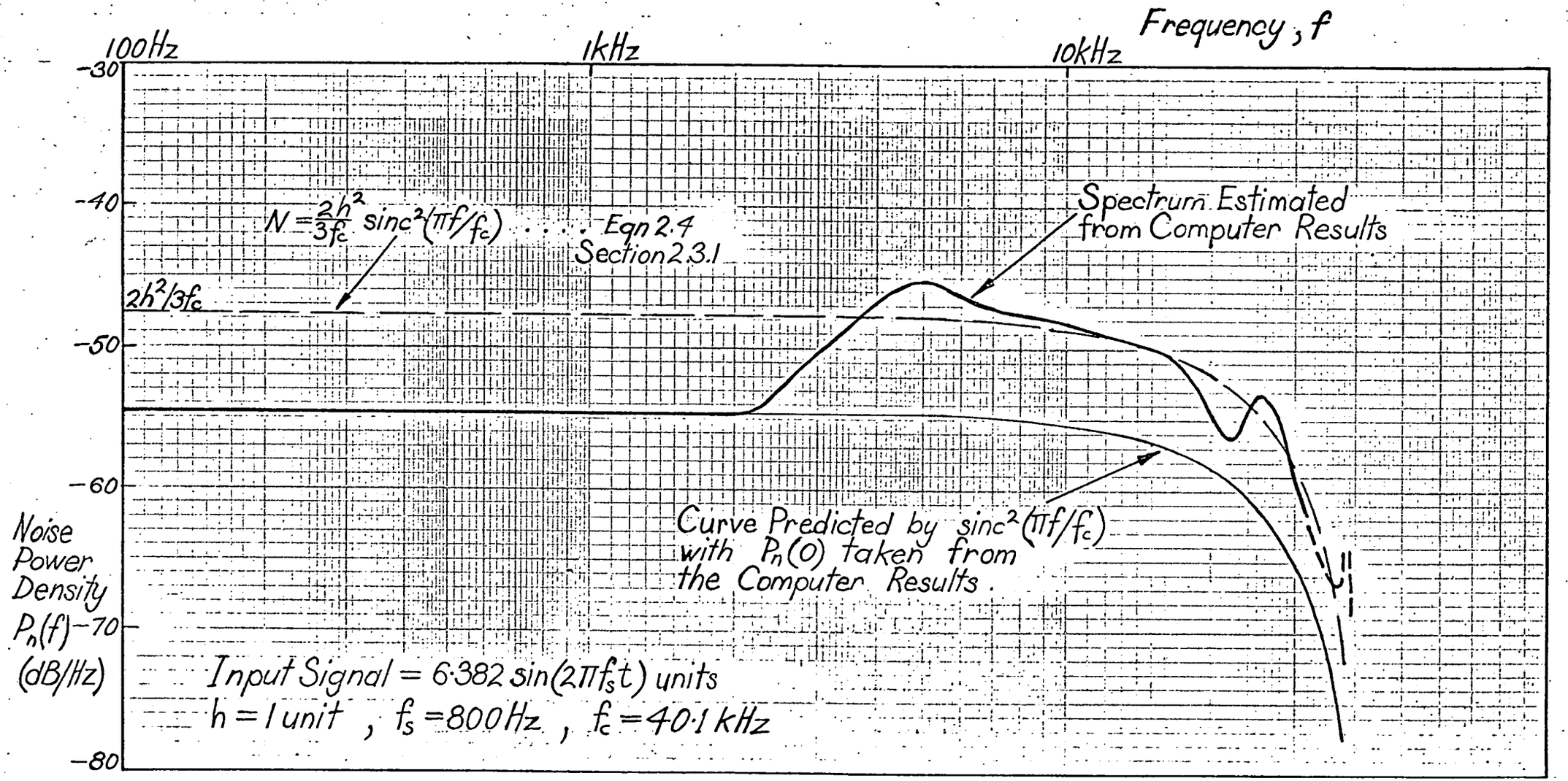


FIG 3.9 Power Spectrum of Noise Function from Computer Simulation Results

The most significant characteristic of the spectrum of the noise function, indicated by Figs. 3.5 and 3.6, is that the power spectrum (or more precisely the power of the discrete Fourier components) exhibits a distinct $\frac{\sin^2(\pi f/f_c)}{(\pi f/f_c)^2}$ shape. This characteristic is as would be expected from a consideration of the autocorrelation of r(t) which would have the triangular form, $1 - |\tau|/Tc$, if the value of r(t) were random between adjacent clock intervals. Alternatively (as discussed in Section 2.3.1) consider the error function sampled at a rate f_c, to give e*(t). If e*(t) is random in nature, a power spectrum of the form, $(\sin x/x)^2$, will result when a full pulse extension of e*(t) is considered. The quantizing noise power spectrum shape of $(\sin x/x)^2$ was assumed intuitively by Johnson[29] and enters into the mathematical expressions of van de Weg[34]. The strong noise components as sidebands of the clocking frequency and its harmonics, which are evident from Figs. 3.5 and 3.6 were also predicted by the initial considerations of Section 2.3.1.

The deviation of the noise spectrum from the $(\sin x/x)^2$ curve is indicated by the curves of Fig. 3.9 (and is also indicated by the dashed curves of Fig. 3.6). The zero frequency value of the $(\sin x/x)^2$ curve is taken from the low frequency power density as estimated from the computer results in Fig. 3.8. Also shown in Fig. 3.9 is the power density spectrum anticipated by the initial considerations of the delta modulation quantizing noise in Section 2.3.1. This spectrum is indicated by the dashed line for which suitable correction for the one sided nature of spectra of Fig. 3.9 has been made. As indicated by Fig. 3.9 the computed estimate of the noise power spectrum, $P_n(f)$, for the particular system conditions considered, is flat up to about 2kHz from where it shows considerable variation from the $(\sin x/x)^2$ form; the computed curve being up to about 10dB greater. Therefore, under these conditions, the quantizing noise in the frequency band $0 - f_m$ will only be proportional to the cut-off frequency, f_m up to about 2kHz. This contrasts with the performance analysis, for a sinusoidal input; of van de Weg and de Jager who predicted that N_q would be proportional to f_m up to about $f_c/4$ (10kHz in this case). The N_q, f_m relationship of van de Weg (de Jager's being virtually identical) is indicated by the dashed line in Fig. 3.8. For f_m in the range 3.2 to 4.7 kHz (which includes the range for normal audio cut-off frequencies) van de Weg's predicted N_q is very close to that of the computer simulation. Elsewhere however, (in the range of f_m, 0 to 10kHz) the discrepency is significant, being in general about 3 dB. Although de Jager's analysis was for the quantizing noise under any loading conditions between threshold and slope overload; van de Weg's move precise analysis was based on a specific and relatively low level of loading.

Van de Weg considered a system loading given by the rms value of the signal derivative being equal to 1/4 of the maximum slope capability of the delta system. i.e. $\left\{\overline{[f'(t)]^2}\right\}^{\frac{1}{2}} = hf_c/4$.
As such, van de Weg's mathematical analysis could not be expected to hold under the loading conditions considered in the simulation, which are tending towards slope overload (the maximum signal slope = $0.8\ hf_c$)

The failure of the "N_q proportional to f_m" relationship in the 0-10kHz region, is due to the non-uniform nature of the power density spectrum of the quantizing noise over the low frequency range of interest. This non-uniform spectrum was anticipated in Section 2.3.2. to occur, as the result of significant correlation between adjacent error values at the clocking instant. The power density below the value of $2h^2/3f_c$ was also anticipated for the low frequency region where a uniform spectrum exists. However the frequency of the departure of the spectrum from a uniform spectrum was not expected to be at such a low frequency where it in fact affects the power in a typical audio band. This point is discussed further in the following section.

3.5 Noise Power Results.

In general, theoretical analysis has been aimed at investigating the effect on N_q or S/N_q of varying f_c while maintaining a constant input signal loading of the system. To maintain a constant load with varying f_c the input signal amplitude must be varied with f_c. This does not indicate the nature of the quantizing noise change with changing loading. Nor does it indicate the relationship of N_q to f_c for a constant signal amplitude; which would be of value in determining the optimum f_c for a given signal. Noise power results were obtained from the computer simulation for varying f_c with a constant input signal.

Simulation and noise frequency analysis programmes were written and processed for the delta modulation system employing a range of 11 clock frequencies. All other parameters were kept constant as for previous simulations with the amplitude of the sinusoidal input being set of 6.0 units. Based on the conclusions of Section 3.3 for the most satisfactory simulation with a limited sample length of input, the clock frequencies were chosen according to the formula $f_c=(n+1/c)f_s$ where n is any integer, f_s is the sinusoidal frequency (800 Hz) and c is the number of cycles of input considered. Eight cycles of input were considered for the simulations at all clock frequencies; this being the maximum number readily handled by the computer storage. The use of eight cycles of input gave a 100 Hz spacing to the components of the noise as computed by the Fourier analysis section of the programme.

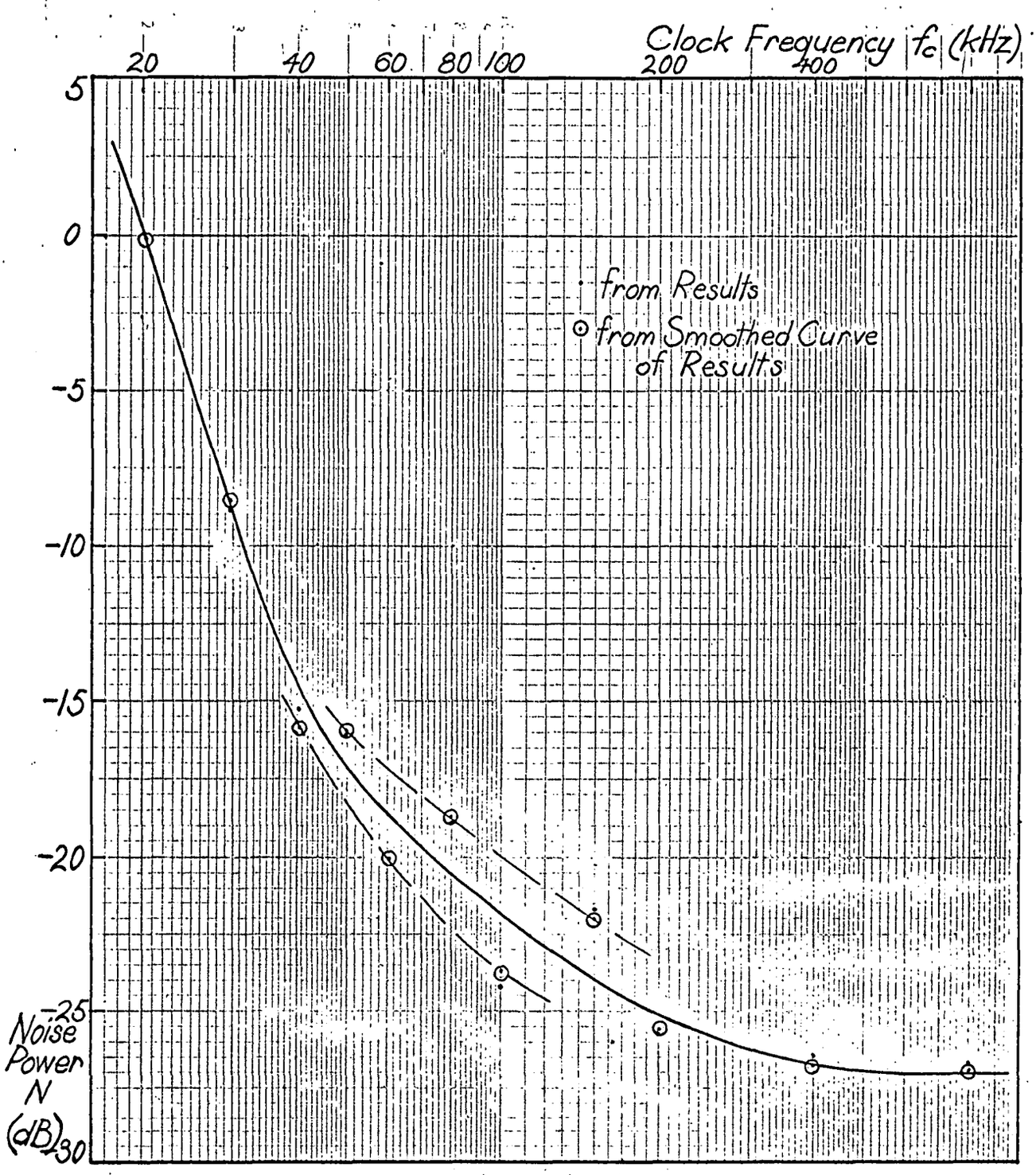
The simulation analysis programme was designed to give five computer samplings of the noise function per clock period for clock frequencies below 70kHz. This was reduced to four for clock frequencies between 70 and 110 kHz and to two for clock frequencies above 110kHz. The reduction in the number of computations per clock period was necessary to limit the storage and computation time requirement of the programme to a reasonable level. The Backing Store of the Elliott 503 computer was required for the programmes which considered clocking frequencies of 200kHz or above. Although a low number of computer samplings of the noise function gives inaccuracies in the noise power analysis, the error will only be significant for the power at frequencies of the same order as the clock frequency. This is due to the rapid roll-off of the power density at higher frequencies as indicated in Section 3.3 and also in Chapter 2. Since the limitation in the number of computations per clock period is applied only as the clock frequency increases there should be no significant loss of accuracy in the calculation of the power in a given low frequency band.

The computer simulation programmes were designed to give printouts of the amplitude of the Fourier components and the corresponding frequency. (with 100Hz spacing between components) The printout also included the amplitudesquared, the sum of the squares of the amplitudes and the noise

FIG 3.11 Noise Power Variation with Clock Frequency from Computer Simulation Results

with: Ideal Low Pass Filter with Cut-off Frequency, f_m = 4kHz

Input Signal = 800Hz sinusoid with 6.0 volts amplitude (ie 4.24 volts rms)



power in dB from zero up to each frequency. The noise power in dB corresponding to a frequency f_m is given by:

$$N = 10 \log \sum_{fk=0}^{f_m} \frac{[AMPL(f_k)]^2}{2}$$

where $AMPL(f_k)$ is the amplitude of the Fourier component at the frequency f_k.

The curves of N against f_m derived from the computer results for each clock frequency were plotted and these are shown in Appendix C. The points show the computer analysis output and these indicate the major contribution to the noise power of harmonics of the signal frequency. As discussed in Section 3.4, the curves are drawn to give a smoothing of the computer analysis noise power results and these are shown in Fig. 3.10. Each curve gives, for a constant clock frequency, the variation of N with varying cut-off frequency f_m, for an idealized low pass filter. From Fig. 3.10 the relationship between the noise power, N and the clock frequency, f_c as indicated by the simulation for a given idealized low pass filter can be determined. The curve of N against f_c for a typical audio low pass filter cut-off frequency of f_m = 4kHz is shown in Fig. 3.11.

As discussed previously and anticipated in Section 2.3 the major noise power is contributed by harmonics of the signal frequency. The computer simulation results (Appendix C) indicate that the noise power spectrum changes with changing clock frequency, mainly as the result of changes in the contributions of harmonics of the signal frequency. The N against f_m curves of Fig. 3.10 indicate that for f_c = 30kHz and above, the noise power tends to a maximum of about -5 dB as f_m increases to approach f_c. ie. the noise power in the frequency band 0 to f_c appears to be constant for all f_c, provided f_c is above a certain value.

Referring again to Section 2.3.1 where, based on the assumption that the discrete error function e(kTc) is a random variable with no correlation betweem adjacent values of e(kTc) (ie. the power spectrum of e(kTc) is uniform) it was estimated that the quantizing error power spectrum would be of the form $\text{sinc}^2(\omega Tc/2)$, (Eqn. 2.4) and hence the quantizing noise power in the frequency band 0 to f_m would be given by:

$$N_q = \frac{h^2}{6\pi f_c} \int_{-\omega_m}^{\omega_m} \text{sinc}^2(\omega Tc/2)\, d\omega \quad \ldots\ldots 2.6$$

Putting $\omega_m = \omega_c$ we get,

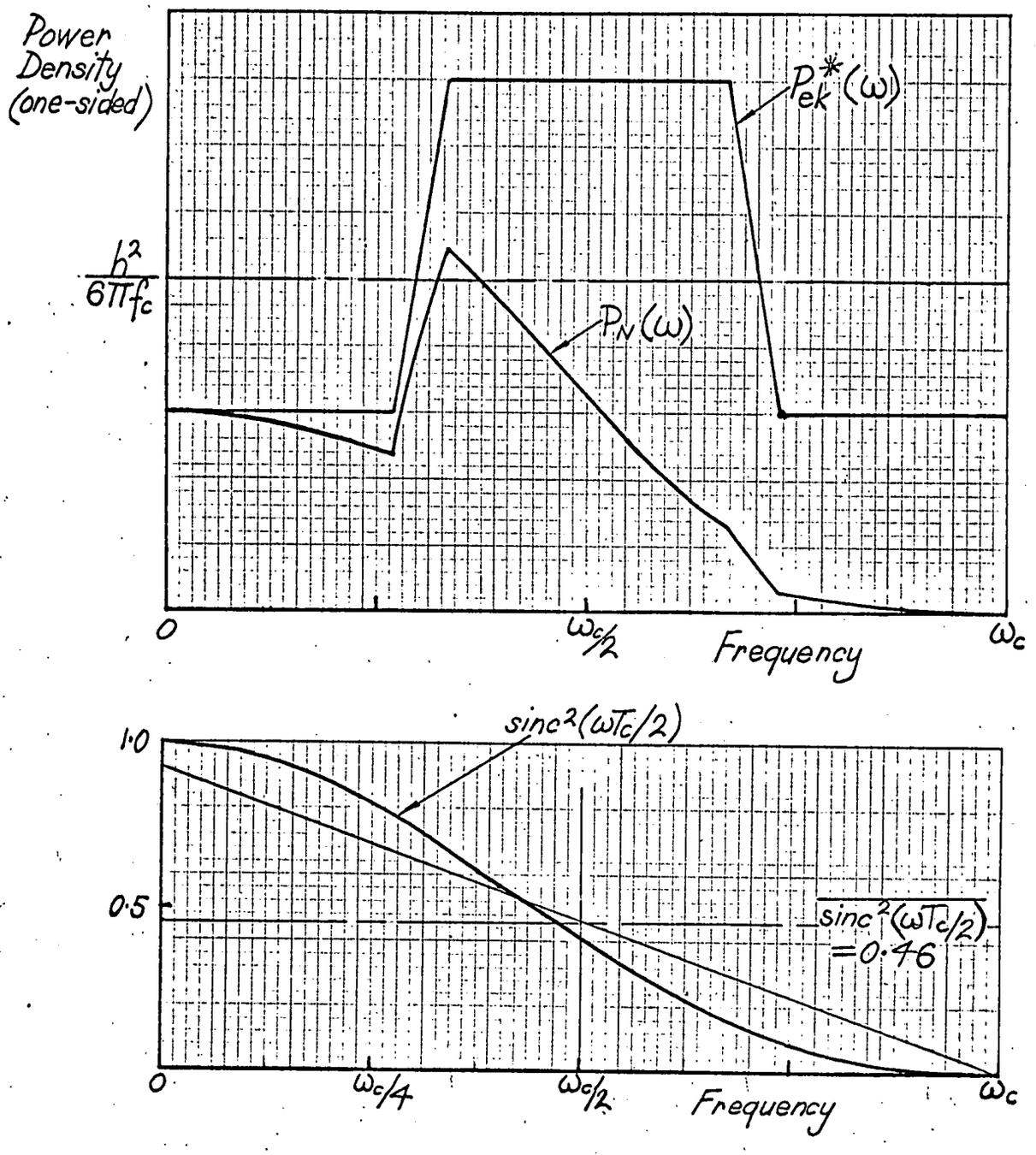
Noise Power in the frequency band 0 to f_c $\simeq \frac{h^2}{6\pi f_c} 0.92\ \omega_c = 0.92 \frac{h^2}{3} = 0.307$

$= -5.13$ dB

Since $\int_{-\omega_c}^{\omega_c} \text{sinc}^2(\omega Tc/2)\, d\omega \simeq 0.92\omega_c$

(Refer: Carlson[2], p.33)

FIG 3.12 The Effect of a Non-Uniform Power Spectrum for the Discrete Quantizing Error on the Noise Power in the Frequency Band 0-fc



Thus the initial quantizing noise spectral considerations of Section 2.3.1 predict that the quantizing noise power in the band 0 to f_c is constant, regardless of the clock frequency. This agrees with the behaviour observed from the computer simulation results. Furthermore, the predicted constant noise power of -5.1dB shows very good agreement with the observed value from Fig. 3.10. This agreement is not entirely as expected because the validity of Eqn 2.6 was questioned in Section 2.3.2 where the effect of error signal correlation was considered. Section 2.3.2 showed that the constant, k in the equation derived from Eqn. 2.6. for $f_c \gg f_m$ ie. $N_q = \frac{2h^2 f_m}{k f_c}$ should be somewhat greater than 3 and has been estimated by others (refer Section 2.4) to be close to 6. This was described as being due to a reduction in the noise power spectral density in the region of $f \ll f_c$ resulting from the error signal correlation (See Fig. 2.4).

A further consideration of the power spectrum of the quantizing noise shows why Eqn. 2.6 is reasonable for estimating the noise in the frequency band 0 to f_c while being inaccurate for the band 0 to f_m where $f_m \ll f_c$. From Section 2.3.2 it was expected that the discrete quantizing error power spectrum Pek* (ω) could be described in terms of a cos ωTc power series. Fig. 3.12 shows a Pek* (ω) of the nature expected and the quantizing noise power spectrum, $P_N(\omega)$ which would result, where $P_N(\omega)$ is given by:

$$P_N(\omega) = Pek^*(\omega)\ \mathrm{sinc}^2(\omega Tc/2) \ldots . \text{Eqn. 2.4}$$

Putting $Pek^*(\omega) = \frac{h^2}{6\pi f_c} + P(\omega)$, then $P(\omega)$ in the region 0 to ω_c, is an even symmetrical function about $\omega_c/2$.
Consider $Pek^*(\omega)\ F(\omega)$, where $F(\omega) = \overline{F(\omega)} + F_1(\omega)$, and $F_1(\omega)$ is any function which is odd symmetrical about $\omega_c/2$ in the region 0 to ω_c. Then

$$\int_0^{\omega_c} Pek^*(\omega)F(\omega)d\omega = \frac{h^2}{6\pi f_c}\int_0^{\omega_c} F(\omega)d\omega + \int_0^{\omega_c} P(\omega)F(\omega)d\omega$$

$$\text{Since } \int_0^{\omega_c} P(\omega)F(\omega)d\omega = \overline{F(\omega)}\int_0^{\omega_c} P(\omega)d\omega + \int_0^{\omega_c} P(\omega)F_1(\omega)d\omega$$

$$= 0$$

$$\text{then } \int_0^{\omega_c} Pek^*(\omega)F(\omega)d\omega = \frac{h^2}{6\pi f_c}\int_0^{\omega_c} F(\omega)d\omega$$

Since $\mathrm{sinc}^2(\omega Tc/2)$ can be approximated as having the general properties of $F(\omega)$ then a reasonable approximation from Eqn 2.4 is:

Noise power in the frequency band 0 to f_c

$$= \int_{-\omega_c}^{\omega_t} P_N(\omega)d\omega = \int_{-\omega_c}^{\omega_c} Pek^*(\omega)\ \mathrm{sinc}^2(\omega Tc/2)d\omega$$

$$\simeq \frac{h^2}{6\pi f_c}\int_{-\omega_c}^{\omega_c} \mathrm{sinc}^2(\omega Tc/2)d\omega \quad \text{(as in Eqn 2.6)}$$

$$\simeq \frac{h^2}{6\pi f_c}\ 0.92\,\omega_c = 0.92\ h^2/3 = -5.1\text{dB} \quad \text{(as before)}$$

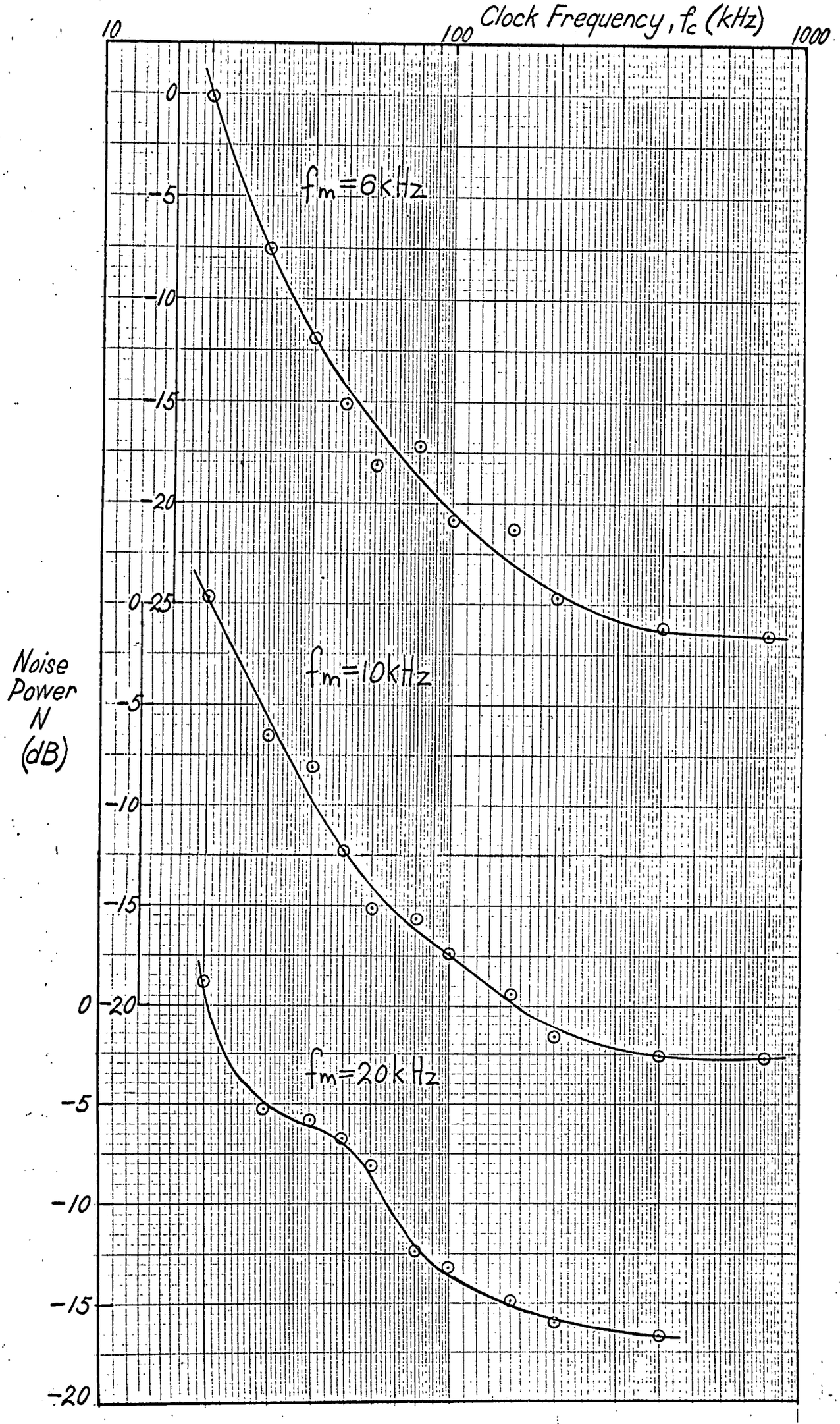
Fig. 3.10 indicates an N against f_m curve for f_c = 20.1kHz strongly different from the curves for f_c = 20.1kHz and above. This is due to slope overload for f_c = 20.1kHz. Under slope overload conditions the discrete error function e(kTc) takes on values outside the range $\pm h$ and thus the value of $h^2/3$ for the mean square value of e(kTc) (derived in Section 2.3.1) becomes an underestimation. For the simulation, the clock frequency at which the maximum signal slope ($2\pi fA$) equals the maximum reconstruction signal slope($f_c h$) is f_c = 30.2kHz. This is the point of theoretical slope overload and since the total noise power in the frequency band 0 to f_c has a constant value of $0.92h^2/3$ for clock frequencies down to 30.2kHz it appears that the uniform probability density function of e(kTc) between $\pm h$ (assumed in Section 2.3.1) holds for clock frequencies down to the onset of theoretical slope overload. It would therefore be expected that the contribution of overload noise to the total noise as theoretical slope overload is approached,(as predicted by Abate and O'Neil; refer Section 2.4) is due to changes in the power spectrum of the discrete error function, e(kTc) and is not due to changes in the uniform probability density function of e(kTc).

A third and most important observation from Fig. 3.10 is that for a constant noise power cut-off frequency, f_m there is not a sequential decrease in noise power with increasing clock frequency. This result is reflected in Fig. 3.11 of the N, f_c relationship for f_m = 4kHz where the N against f_c curve is drawn as what appears to be the mean of two other curves (shown dotted). The dotted curves show a high and unsatisfactory degree of variation of N which results from the simulation technique. This fault in the simulation technique was discussed in Section 3.3 where it was pointed out that for low f_m there would be large variations in the estimates of N_q due to the low number of Fourier components from the analysis. It was concluded in Section 3.3 that this factor, combined with a low number of cycles of the input and with specific initial conditions would limit the effectiveness of the noise power estimates for an f_m of less than about 4kHz. Fig. 3.10 indicates this increasingly irregular estimation of N with decreasing f_m. However Fig. 3.10 would indicate that there is still considerable variation in the N estimate due to the simulation technique above 4kHz. It can be seen from Figs. 3.10 and 3.11 that the fluctuations in N resulting from the method of simulation occur predominantly in the middle clock frequency regions where the system is operating between theoretical slope overload and clock frequency saturation*. This is because the middle clock frequency region is where the error function pattern and hence the noise power spectrum is most dependant on the initial conditions of the simulation. Thus it is the region of greatest fluctuation in the noise power spectrum due to the limited input sample length and specific initial conditions.

---

* Clock frequency saturation being the region where N fails to decrease further for increasing f_c.

FIG 3.13 Noise Power Variation with Clock Frequency from Computer Simulation Results for Various Low Pass Filter Cut-off Frequencies



Other curves of N against f_c derived from Fig. 3.10 for various cut-off frequencies (f_m) are shown in Fig. 3.13. Although the simulation results are not free of the effects of using a short sample of input, it provides useful results particularly when the higher values of f_m are being considered.

If an indication of the noise spectrum is required, the problems with the simulation are not readily overcome without extremely large computer storage or the lengthy process of intermediate data output. If only the N, f_c characteristic were required, a beter simulation process incorporating digital filtering in the program could be used as discussed in Section 3.3.

A discussion of delta modulation performance as indicated by the computer simulation, relative to theoretical analysis and experimental results is included in Chapter 5.

CHAPTER 4

EXPERIMENTAL DELTA MODULATOR

4.1 Introduction

The design and construction of an experimental delta modulator was undertaken for the following reasons:-

(i) To have available an experimental system on which the computer simulation is based so that evaluation of the simulation method has a more positive reference. (To this stage, evaluation of the worth of the simulation has been based on a consideration of the simulation results themselves and on how they compare with established theory). In addition the construction of an experimental modulator will give insight into aspects of the system which may have been overlooked in the establishment of the simulation model.

(ii) To enable output noise and other relevant measurements to be taken, which will provide further insight into the performance of delta modulation. This will also provide data on which assessments of existing theory can be made.

(iii) To indicate the type of problem likely to be encountered in the realization of a practical delta modulation system.

The design of the experimental modulator aims for the realization of a simple delta modulator with a single stage of ideal integration feedback from an ideal impulse output. This is done in order to keep the experimental work in line with previous discussion and with the basis of the computer simulation.

4.2 Circuit Design Considerations

The experimental system was limited to the modulator alone. Since the local demodulator gives a reconstructed signal, r(t) which will be the same as that at a receiver (provided there is errorless transmission), then all performance measurements can be made without the construction of a receiver demodulator. This removes the need for any receiver pulse regeneration and clock synchronizing circuitry. Although not necessary for the performance measurements, circuitry for the generation of an output pulse train suitable for transmission was included.

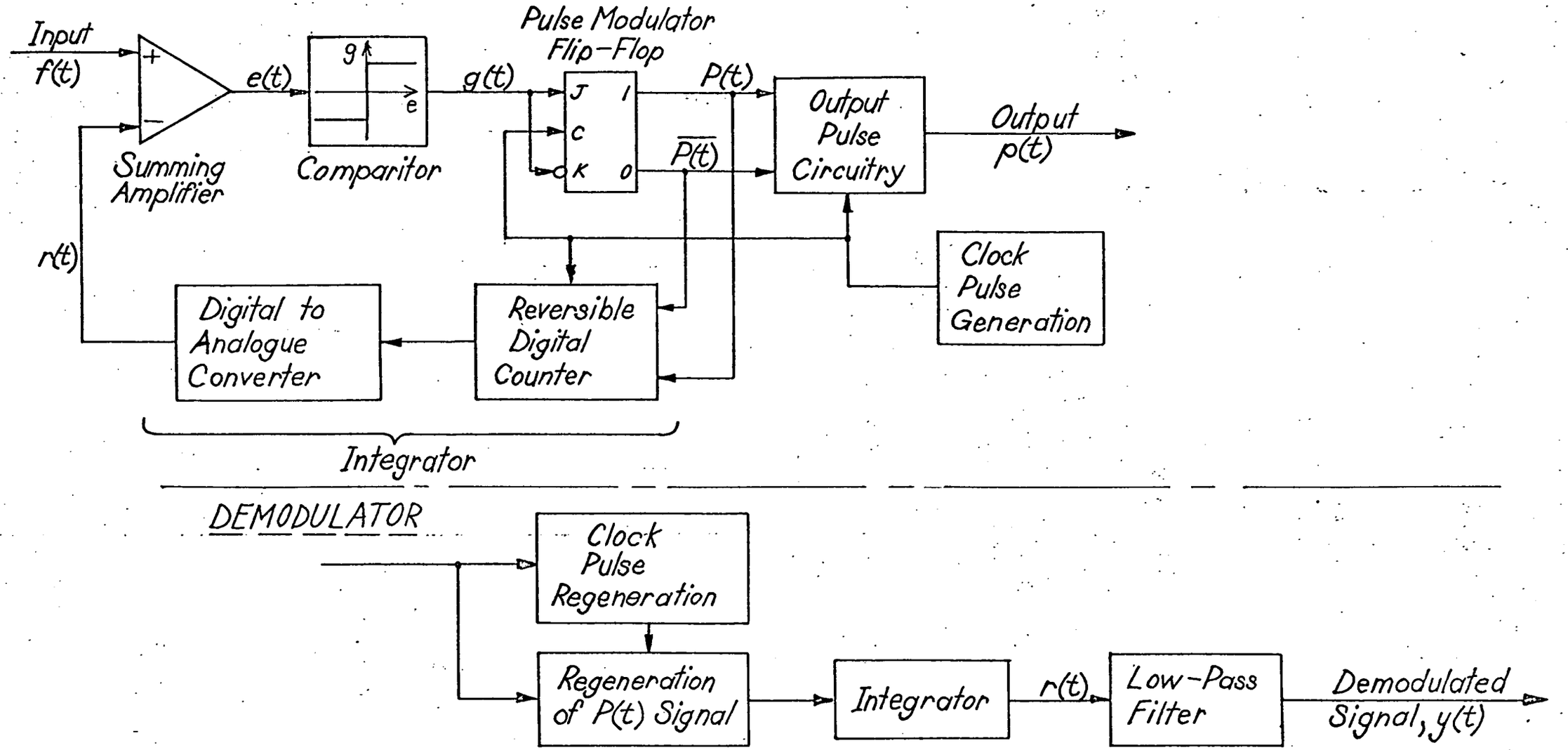


FIG 4.1 Block Diagram of Experimental Delta Modulator

In previous experimental systems the realization of the feedback integration network has been through the use of an RC network approximation to an ideal integrator. In the experimental delta modulator the effect of ideal impulse inputs into a perfect integrator was achieved with the use of an up-down (or reversible) counter (giving the function r(t) in digital form) with parallel output to a digital-to-analogue converter. The more complex digital circuitry for the feedback was employed for two reasons:-

(i) To give feedback integration which matches as closely as possible the perfect integration considered in analysis and the computer simulation. This makes the experimental results more valid for comparison.

(ii) To consider the feasability of implementing a fully integrated modulator suitable for L.S.I. by the elimination of large capacitance requirements in the feedback network.

A reversible counter capability of 16 output levels was used giving the modulator the capacity to handle signals with extreme values of up to 8 x step height. This capacity was used to give a signal handling ability comparable with that of previous discussion and would not necessarily be sufficient for speech or some other particular input signals.

A fixed frequency pulse generator, as would be required for a practical system, was not constructed. Instead, an external signal generator was used in order to give a variable clocking frequency capability.

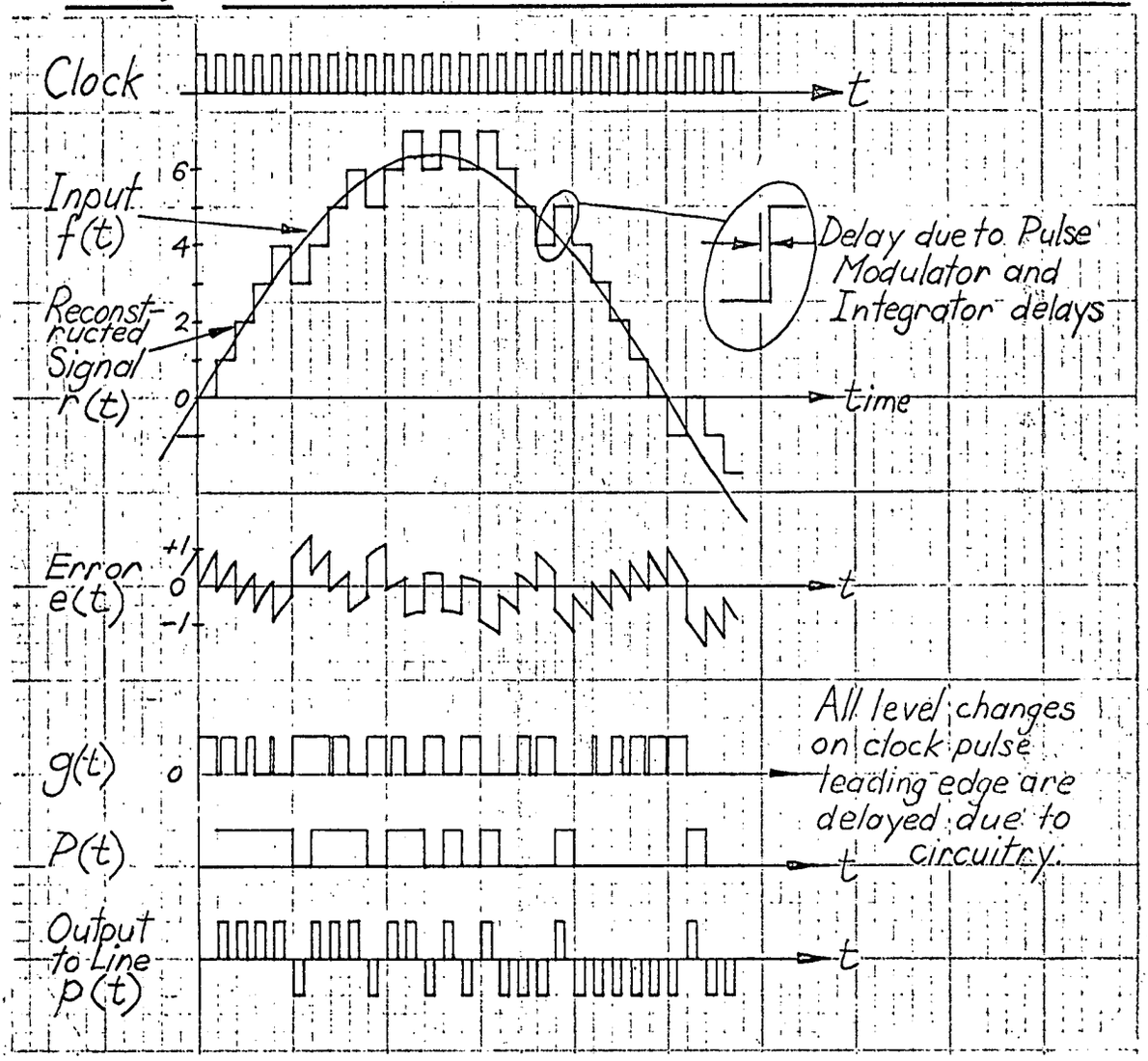
A block diagram of the experimental modulator is shown in Fig. 4.1. Also included in Fig. 4.1 is the block diagram of the receiving demodulator which would be required in an actual delta system.

4.3 Circuit Realization

Positive logic was used for the realization of digital circuitry. Fairchild RTμL was used for logic elements and μA709C (or its equivalent) for operational amplifiers wherever suitable. The circuit diagram of the experimental delta modulator is shown in Fig. 4.2 The waveforms explaining the operation of the modulator are shown in Fig. 4.3.

Fairchild RTμL and μA709C components were used as they were cheap and the most readily available. However, for a practical system design, more advanced micrologic elements having the improved

FIG 4.3 Waveforms Illustrating Operation of Experimental Modulator



features of higher switching speed and lower power consumption would be used. In particular, the low leakage currents of MOSFET elements would be of advantage in the digital-to-analogue converter. At the time of design (several years prior to the submission of this thesis) the μA709C operational amplifier and RTμL elements were the better choice of available elements for the purpose required. The introduction of improved IC's in the meantime would result in a different circuit realization for a contemporary design.

4.3.1 Reversible Counter

For the reversible counter with 16 output levels, 4 JK flip-flops connected in the "trigger" mode were used. The gated flip-flop outputs, P(t) and $\overline{P(t)}$, were used for the count up or down control, with the logic conclusions after a change in P(t) being carried from each flip-flop to the next one. This "feed-on" of the control signal, minimizes the hardware and is possible due to the delay in level change introduced by the flip-flop and gates.

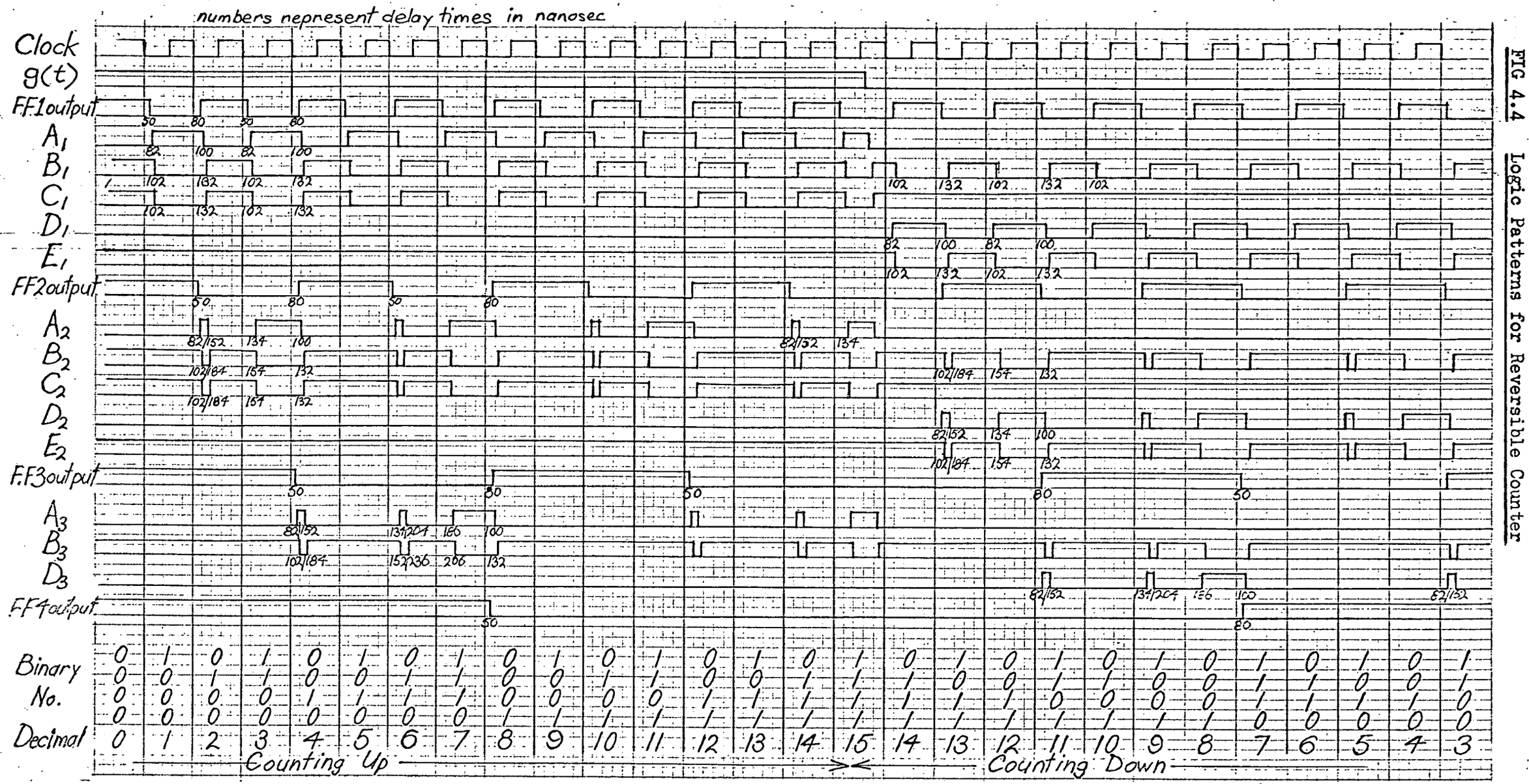


FIG 4.4 Logic Patterns for Reversible Counter

During the development of the modulator, difficulty was encountered with "race" conditions arising from the delay involved in the logic elements. After investigating various methods of eliminating this problem the use of separate clocking lines for the reversible counter and the pulse modulator flip-flop was decided on. A delay of about 700 nsec was introduced into the reversible counter clocking line by the incorporation of the RC circuit and RTμL 9900 buffer. After a consideration of the delay times involved, the 700 nsec delay was found to be sufficient to allow time for the P(t) change to take effect at all the reversible counter flip-flop inputs before the onset of the clock pulse. This method was adopted as being the best implementation, from the aspects of minimum hardware and maximum switching reliability.

The waveforms explaining the operation of the reversible counter are shown in Fig. 4.4, with the delay times shown being taken from the manufacturer's specifications of maximum delays.

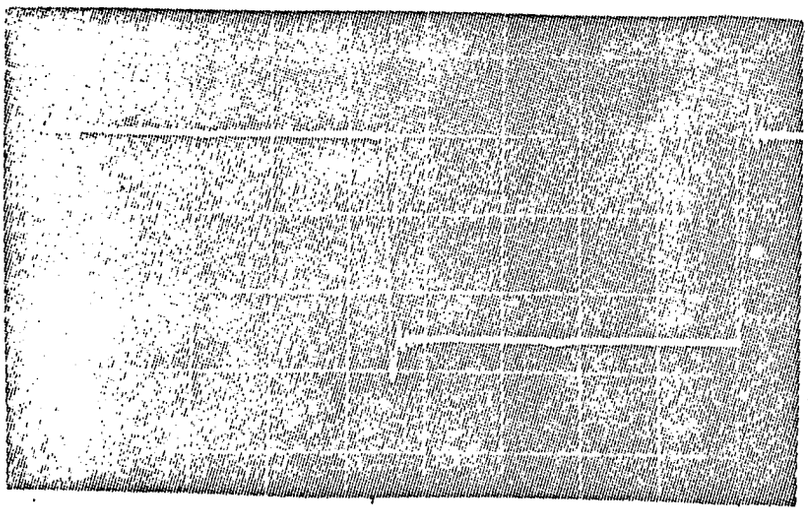
4.3.2 Digital-to-Analogue Converter

Fairly stringent requirements are needed for the D/A converter as it is here that inaccuracies in the reconstucted signal, r(t), are most likely to arise. A step height of 1 volt was chosen, giving r(t) a possible range of 15 volts peak to peak, which is within the range of a μA709C. The μA709C was chosen as the operational amplifier because of its low bias and leakage currents relative to other operational amplifiers available at the time. Also, silicon diodes were used to reduce the current leakage problem to a minimum.

The use of a cascade D/A converter was considered, the main advantage being the constant resistive load it presents to the power supply. However, the accuracy of the cascade D/A converter depends on the accuracy of the switching levels controlling it and thus a simple diode switch could not be readily used. The diode controlled D/A converter shown in Fig. 4.2 was selected, with the accuracy being controlled by ensuring a well regulated power supply and the stability of the resistor values. All the D/A converter resistors were chosen to have power ratings of the order of 10 times the required minimum,in order to limit the resistor value variation due to resistive heating. Good tolerance on the resistor values is not required as it is the ratio of the resistance values of the various branches which is important. These ratios were adjusted with the variable resistors to give the required operating characteristics.

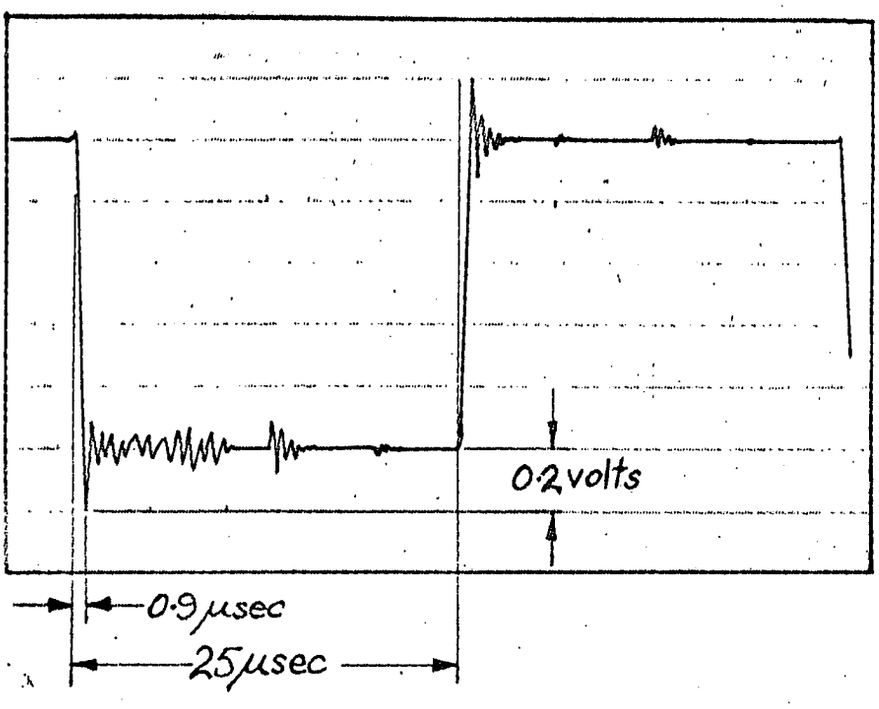
The branch resistor values were chosen so that the leakage curent accounts for, at the most, less than 3% of the output voltage. Thus any variation in the leakage current, which is one of the main

FIG 4.5 Transient Response of Digital to Analogue Converter



Scales; (approx)

Vert. 0.4 volts/div.
Horiz. 5 sec/div.



Step Response of Reconstructed Signal, r(t) with f_o=40kHz.

potential sources of unwanted output variation, can have only a very small effect on the output.

The RTμL927 inverters were used as an interface between the reversible counter and D/A converter to give sufficient drive for the diode switches. The voltage bias at the D/A converter operational amplifier input was set at + 1.5 volts to give an optimum voltage level for the diode switch operation. This gave a 6.5 volt bias to the output.

Since the D/A converter gave an inverted output, the comparitor and summing junction could not be incorporated in a single differential amplifier. The use of a separate summation operational amplifier however, readily provided for the addition of bias to counter the D/A converter output bias. The input signal and reconstructed signal (ie. B in Fig. 4.2) summation junction input resistances of 2.7kΩ were matched in value to within 0.2%. This was done in order to provide as close to equal weighting as possible for the summation of the two signals. The removal of one operational amplifier could be achieved by inverting the logic of the reversible counter thus making the incorporation of the summation junction and the comparitor possible.

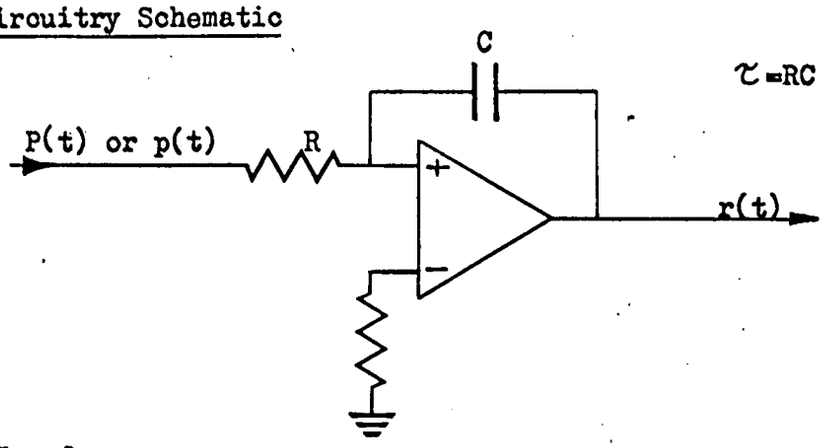
A μA710C was selected for the comparitor as it provides an output which is comparable with the ensuing logic circuitry. The 3.9 volt zener diodes were inserted between the inputs as a safeguard to limit the differential input voltage. Because the low input impedance of the μA710C would load the summing junction μA709C to a sufficient degree to significantly affect its output voltage, a buffer stage was added between them.

Considerable difficulty was encountered during the development of the circuitry in arriving at the best frequency compensation for the μA709C operational amplifiers. In particular the D/A converter μA709C was most susceptible to high frequency oscillations which had to be eliminated, while maintaining a good frequency response for the output. The frequency compensation shown in the circuit diagram (Fig 4.2) was found to be satisfactory, giving a step change in the output, r(t), as indicated in Fig. 4.5.

Two regulated modular power supplies were used for all voltage supply rails except the + 12 volt and the + 3.6 volt supply required for the RTμL elements. These supplies are indicated in the circuit diagram (Fig. 4.2).

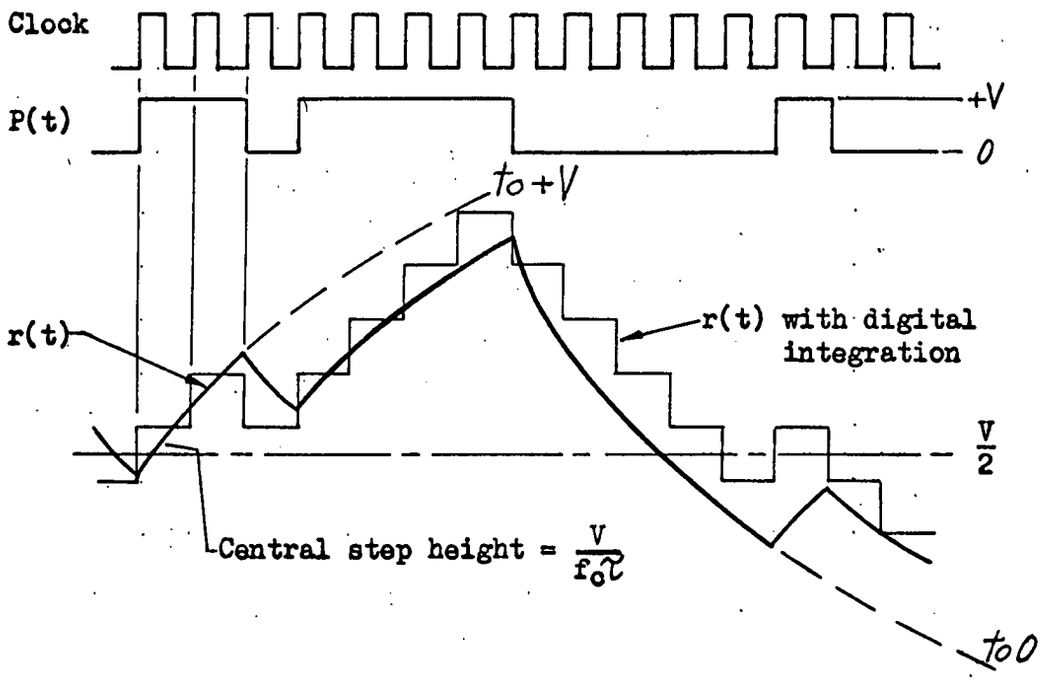
FIG 4.6 Alternative Integration using RC Network

Circuitry Schematic

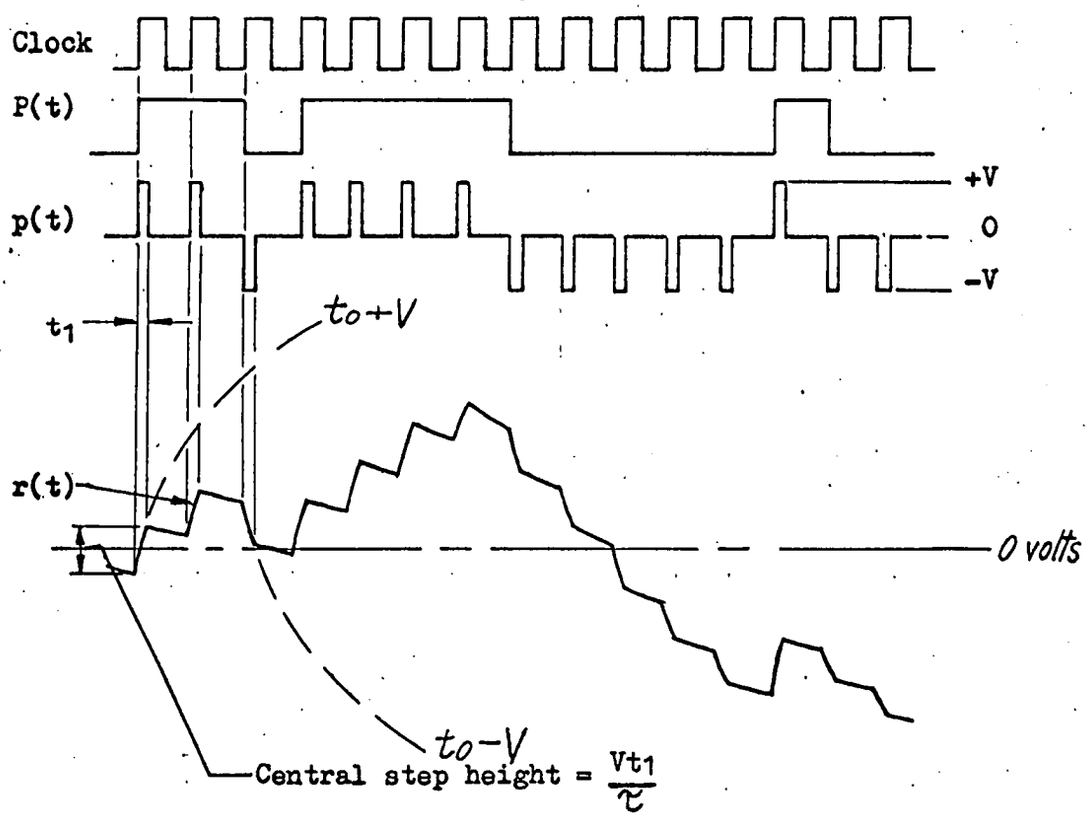


Waveforms

with Non Return to Zero Input to Integrator



Or : with Return to Zero Input to Integrator



The circuit adjustments required for correct operation were:

(i) The D/A converter bias resistance adjusted to give 1.5 volts of bias with all the inputs "off" (ie. no input current to the μA709C).

(ii) The D/A converter feedback resistor adjusted to give a one volt step height.

(iii) The D/A converter branch resistor values adjusted to give the correct output voltage levels.

(iv) The summation μA709C input bias current resistor adjusted to give an output (e(t)) with zero dc content for zero input signal.

4.4 Practical System Considerations

The experimental modulator developed could provide the basic design for a practical system, with the number of flip-flops in the reversible counter determined by the required number of output levels. The type of feedback network employed would allow complete I.C. implementation of the modulator making it suitable to realization by L.S.I. A practical demodulator however would employ a linear low-pass filter unless a satisfactory and economically competitive digital low-pass filter was developed. There are three features of the experimental delta modulator which are of significance for practical implementation.

(i) The digital style of the feedback integrator circuitry makes it highly suited to the introduction of totally digital companding. This could be achieved by the introduction of additional clock pulses to the reversible counter. The decision on the occurrence and number of additional clocks would be determined by logic circuitry observation of the output pulse train (as discussed in Section 1.5).

(ii) The digital nature of the feedback integrator eliminates the problem of idling noise. Idling noise results when the 1010...output pulse pattern varies to a more complex pattern which introduces lower frequency idling noise components. These idling noise components can become noticeable in the band of the signal frequencies. The change in idling pattern results from a drift in the ratio of the current feed to current drain in a linear feedback network (see Wang[38]) and is thus eliminated when a digital feedback network is employed.

(iii) The good resistor tolerances for the D/A converter and the summing junction are not required on the absolute values but are required for the values relative to other resistors. For I.C. production only poor accuracy (to within about 5%) can be achieved for the matching of resistor values. This means that either adjustment potentiometers or accurate external resistors would be required in an I.C. D/A converter implementation.

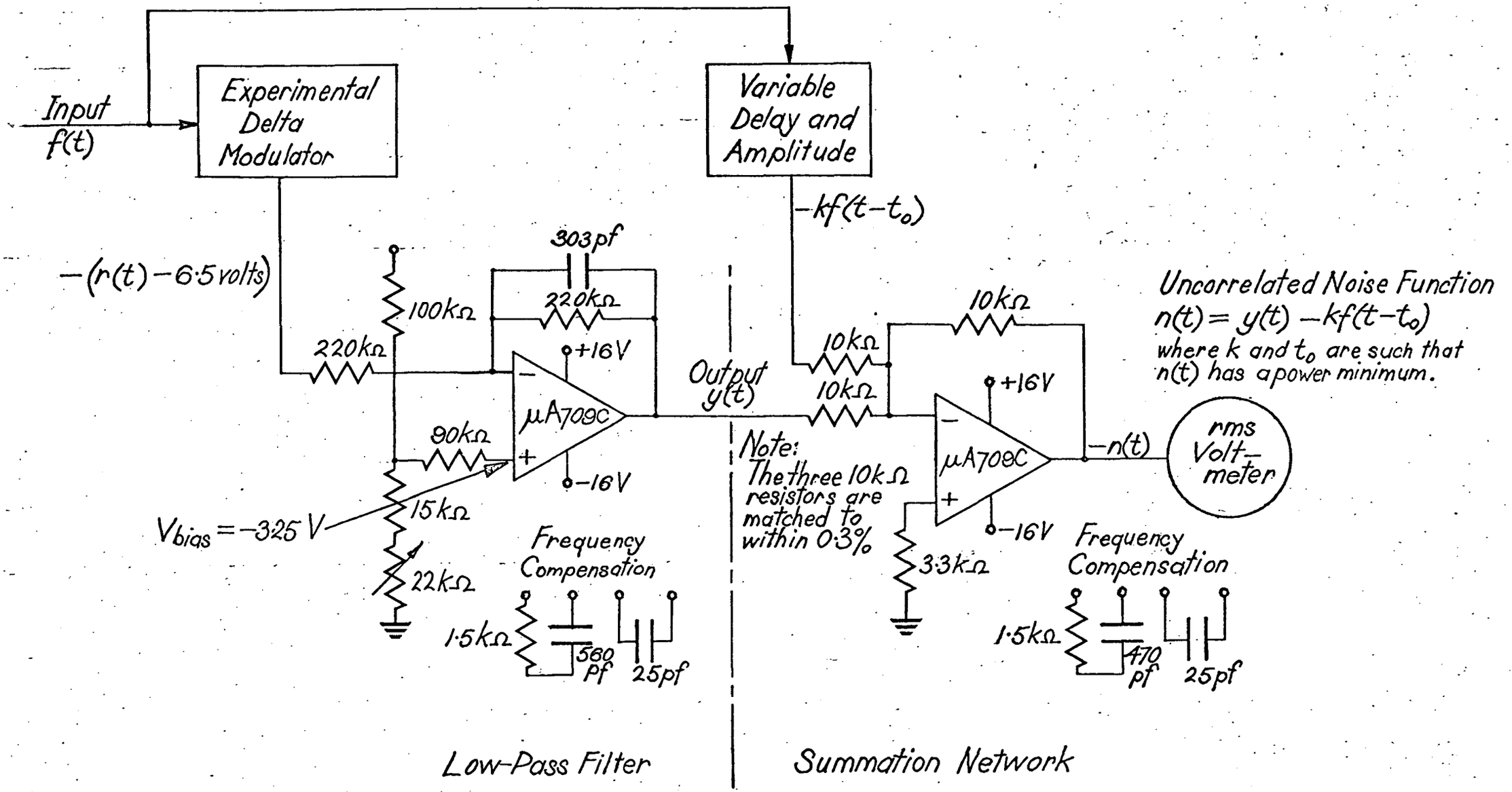


FIG 4.7 Noise Measurement Circuit

The considerably greater complexity of the feedback circuit of the experimental modulator compared with an RC "integrator" is a severe limitation with the design of the experimental modulator. Furthermore, as was discussed in Section 1.4.3, the optimum feedback transfer characteristic is one that matches the spectrum of the input signal. Therefore for speech signals, the flat transfer characteristic over the lower frequencies, of the low-pass filter integrator is in fact desirable. Fig. 4.6 shows two types of reconstructed signal which can result from the use of an active RC integrator. It would only be with the development of economic L.S.I. implementation of the experimental modulator circuitry that the type of feedback circuit used in the experimental modulator would be chosen in preference to a linear filter network.

4.5 Noise Measurement

4.5.1 Method of Noise Measurement

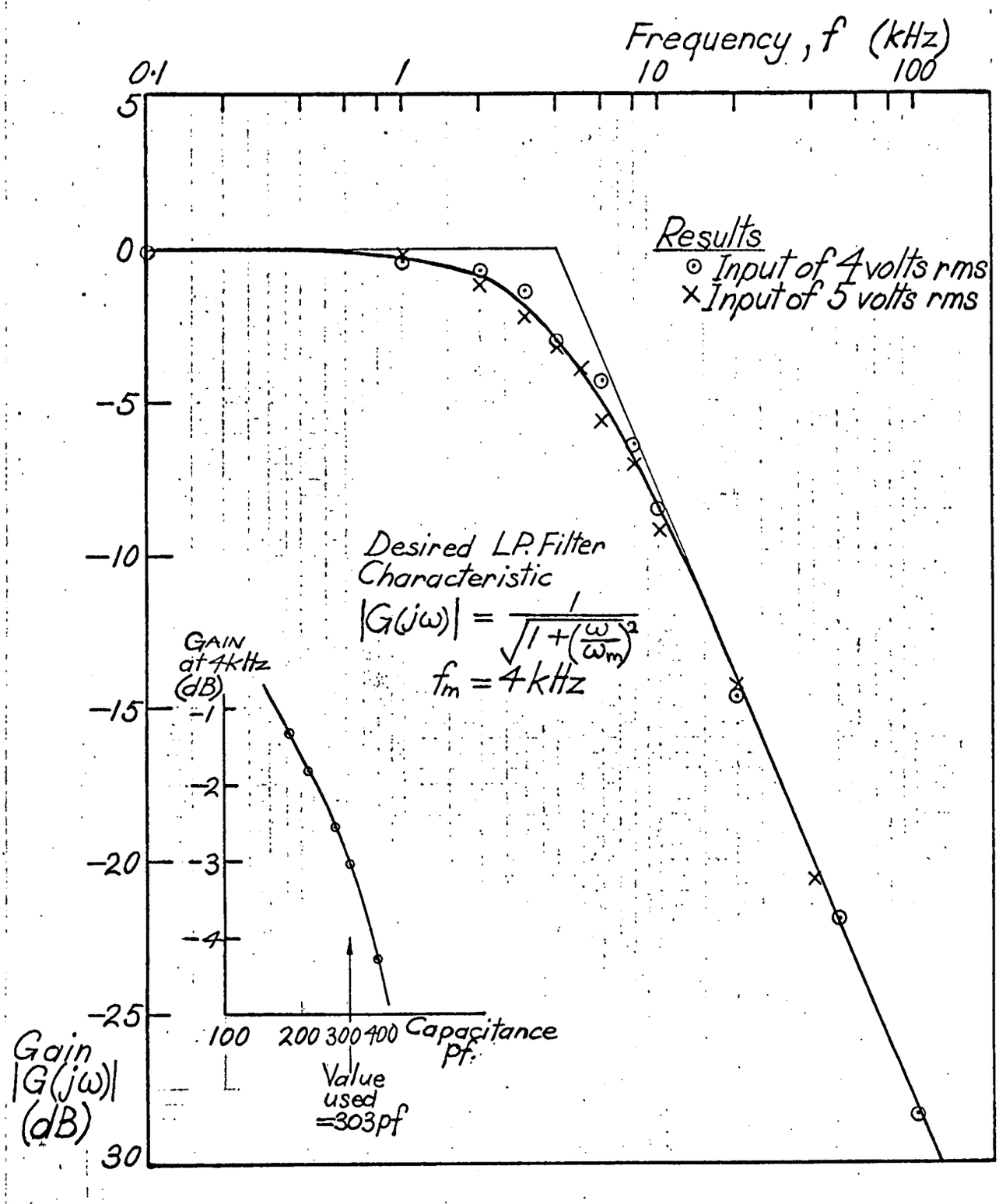
For the measurement of the noise perforamance of the experimental delta modulator the noise measurement curcuit of Fig. 4.7 was developed. A signal generator with two outputs, each of variable magnitude, with a variable delay between each, was used to provide a sinusoidal test input signal with a frequency of 800 Hz. From the output signal y(t), one of the signal generator outputs was subtracted. By varying the amplitude and phase of this test signal until the power of the difference was a minimum, the uncorrelated noise function, n(t), as defined in Section 2.2 was obtained. The signal, $kf(t-t_o)$, which on subtraction from the output, y(t) gives the noise power minimum, is the output signal. The specification of the output signal and output noise gives the output SNR defined by:-

$$\text{Output } \frac{S}{N} = \frac{\text{Output power at the signal frequency (800 Hz)}}{\text{Output power at all other frequencies.}}$$

The noise minimizing process was conducted prior to taking each noise power reading.

The output, y(t), (which corresponds to the output which would be obtained at a distant receiver provided errorless transmission occurred) was obtained by passing r(t) through the RC, single stage low pass filter. The filter frequency characteristic, shown in Fig. 4.8, was designed to give a 3 dB cut-off frequency of 4kHz. The 20dB/decade attenuation roll-off rate provided a simple filter which was considered adequate for measurement of the noise for the purpose of determining the system performance. In a practical system the noise power could be reduced further by the use of a sharp cut-off low-pass filter.

FIG 4.8 Low Pass Filter Characteristic



The offset of r(t) of 6.5 volts as it comes from the experimental modulator was removed by applying an input bias voltage to the filter operational amplifier of -3.25 volts. The three 10kΩ resistances in the summation network were selected to have values matched to within 0.3% in order to provide noise power readings which were as accurate as reasonably possible.

All noise and signal readings were made with a Dawe true rms valve voltmeter. Long term variation in the output voltage of the laboratory power supplies was observed, the variation being up to 2% in one day. For this reason the power supplies and circuit biasing were checked and adjusted as necessary before each set of measurements. All voltages were adjusted to within 0.5% of their correct value.

4.5.2 Results

All measurements were made using an 800 Hz sinusoid as the test input signal. Initially the experimental modulator performance with a constant clocking frequency was determined.

(i) Constant Clock Frequency

For a clock frequency kept constant at 40 kHz, readings of the rms values of the input signal f(t), total output signal y(t), and the output noise and signal components, n(t) and $kf(t-t_o)$ respectively, were taken for a range of input signal amplitudes, from below threshold to the maximum capacity of the system. In addition the delay, t_o, and the number of levels of r(t) employed, were determined for each value of the input signal. From the results, curves of output signal power, noise power and SNR;against input signal amplitude were drawn. These are shown (for f_c = 40 kHz) in Fig. 4.11.

Photographs of various waveforms generated during the operation of the modulator were taken. These are shown in Fig. 4.9 for constant parameters of f_c = 40 kHz and input signal $\simeq$ 4.6 volts peak to peak. In addition, photographs illustrating the different waveforms for various input signal amplitudes, with f_c = 40 kHz, are shown in Fig. 4.10.

The noise measurement readings were repeated at a clock frequency of 80 kHz. From these results another set of curves of output signal power, noise power and SNR against input signal amplitude were obtained for f_c = 80 kHz. These are given in Fig. 4.12.

(ii) Constant Input Signal

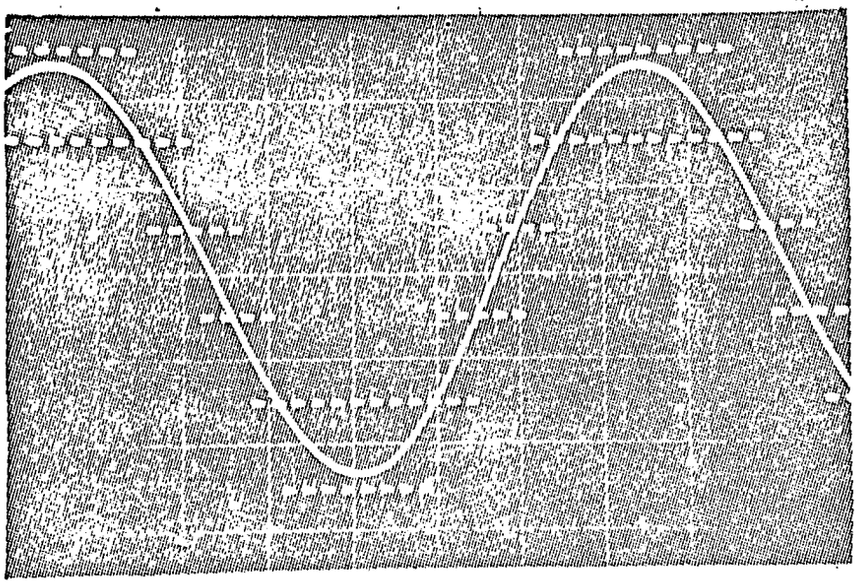
Further results were taken to indicate the experimental modulator performance for a constant input signal with varying clock frequency. As in Section 3.5 for the computer simulation noise power results, the relationship of N to f_c for a constant input signal rather

FIG 4.9 Waveform Photographs for the Experimental Delta Modulator

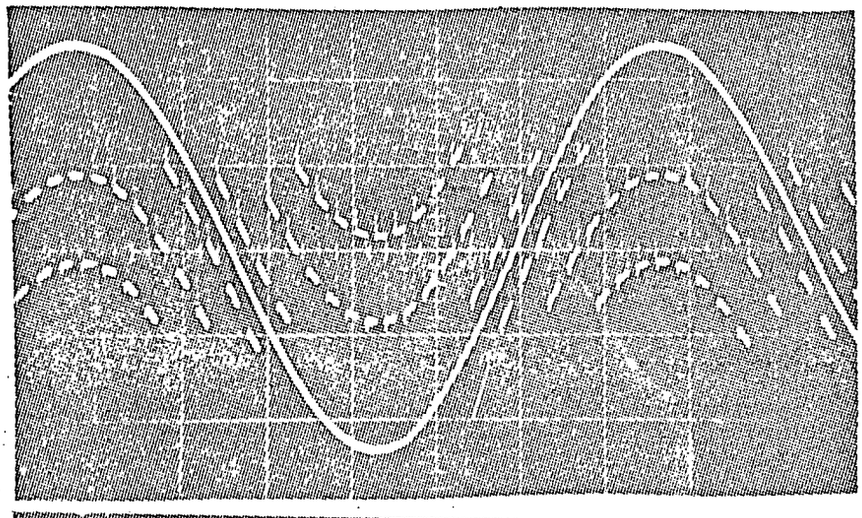
Input = 4.6 volt peak to peak, 800Hz sinusoid.

f_o = 40 kHz.

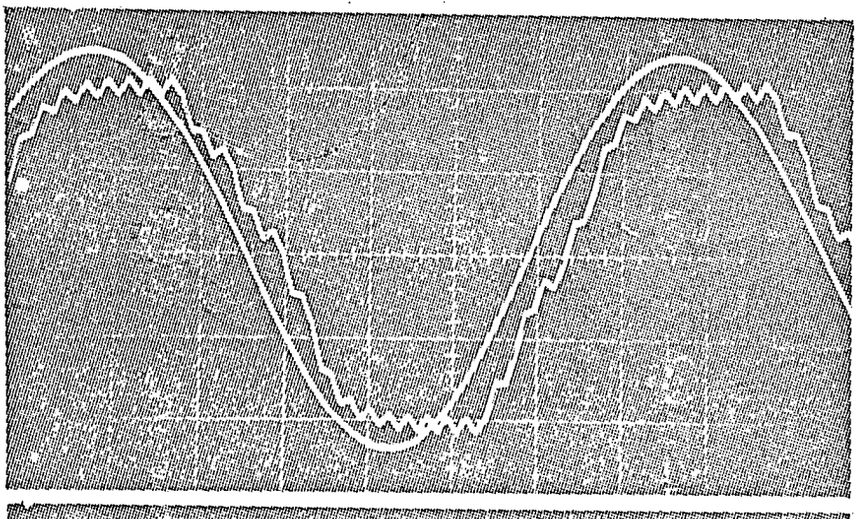
Scales (approx.) : Time: 0.2 msec/div.
Amplitude: 1 volt/div.



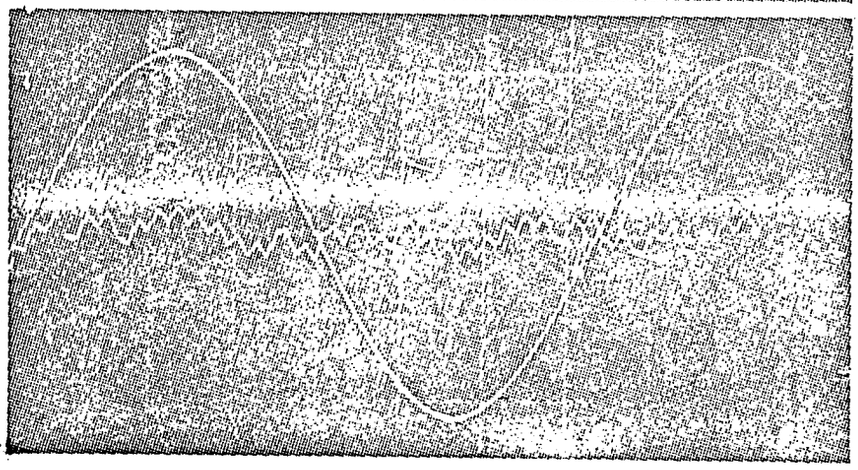
f(t) and r(t)



f(t) and e(t)



f(t) and y(t)



$-kf(t-t_o)$ and n(t)

than a constant signal loading was investigated. Constant signal loading has in general been most commonly considered and it provides easier theoretical analysis. However the N, f_c relationship for a constant signal amplitude is of more practical value in indicating the optimum f_c for a given set of signal parameters.

The maximum input signal before the experimental modulator reaches amplitude overload is 15 volts peak to peak. This is an arbitary overload point determined by the number of levels of output available in the experimental modulator. Measurements were taken to determine the modulator performance relative to the clock frequency with the input signal set at constant values of 80% (12.0 volts peak to peak) and 25% (3.75 volts peak to peak) of the maximum input before amplitude overload. Values of the uncorrelated (minimum) noise and output signal were measured for clock frequencies from 4 kHz to 500 kHz. This clock frequency range covers all levels of modulator loading. In addition the delay, t_o, and the number of levels of r(t) employed were determined for each value of the clock frequency. From the results, a curve of the output SNR against clock frequency for each of the input signal amplitudes was obtained. These are shown in Fig. 4.16. Curves of the output noise power, N against f_c for each of the constant input signal amplitudes were also determined and are shown in Fig. 4.17.

Photographs of the various waveforms generated during the operation of the modulator at different clock frequencies are shown in Fig. 4.15 for a constant input signal amplitude of 12.0 volts peak to peak.

4.6 Observations and Discussion of Results

4.6.1 Constant Clock Frequency

The curves of output SNR against input signal in Fig. 4.11 and 4.12 indicate a regular variation in SNR with each 2 volt change in the peak to peak input signal voltage. This results from the significant effect on the output readings, of the introduction of two new levels of the reconstructed signal, r(t). The new levels occur as the peak to peak input signal voltage increases to reach each odd multiple of the step height. The effect is, principally on the noise power and to a lesser degree, the output signal power.

The introduction of new levels of r(t) and the resulting sudden change in SNR can be seen to occur at slightly lower peak to peak input signal values than the odd integral values expected. This would be due to slight error in the D/A converter setting of the various voltage levels of r(t) and also due to error in the overall setting of the step height (ie. the D/A converter gain). However, the introduction of the new

FIG 4.10 Photographs of Waveforms for Various Input Amplitudes

with f_c = 40 kHz , f_s = 800 Hz

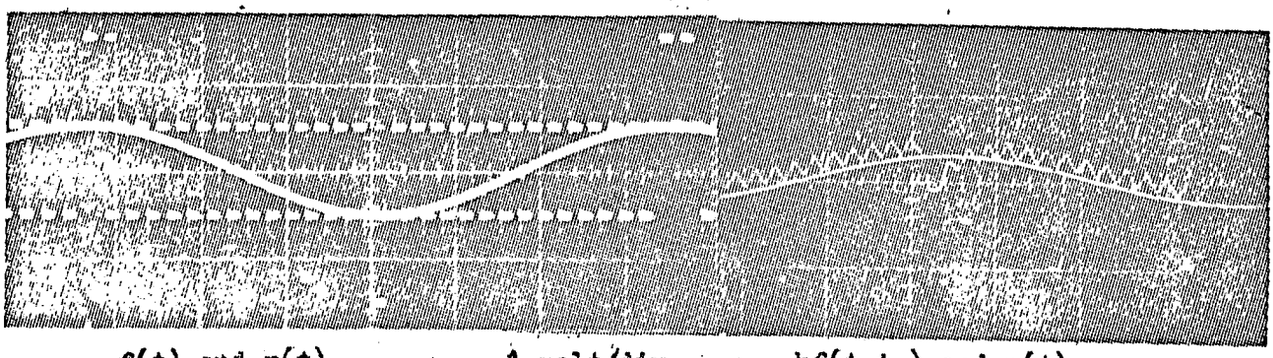
Scale: (approx) Time: 0.2 msec/div.

Amplitude: 2 volts/div.

(unless otherwise specified)

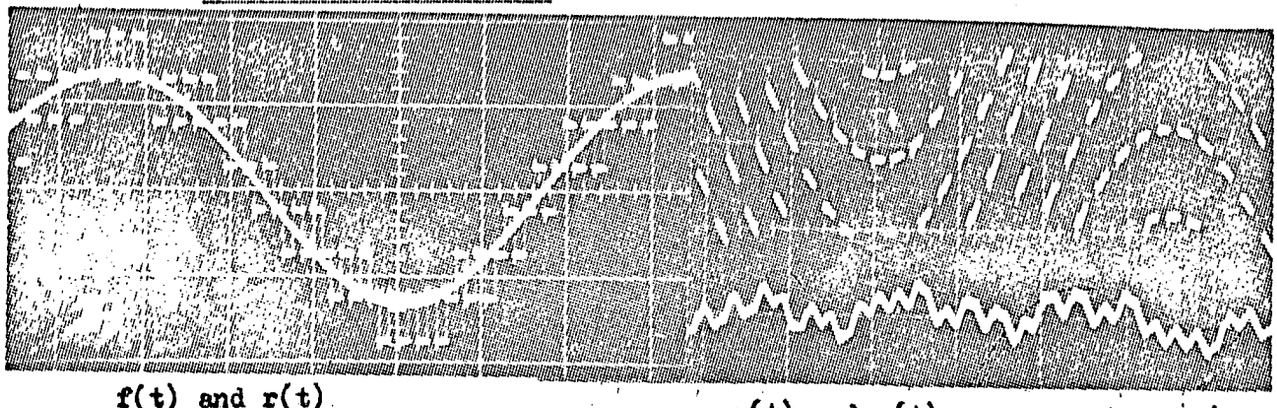
Input

1.0 volts peak to peak



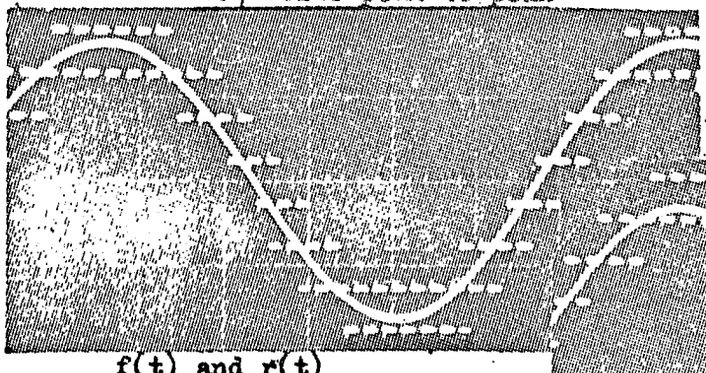
f(t) and r(t) 1 volt/div -kf(t-t_o) and n(t)

5.3 volts peak to peak



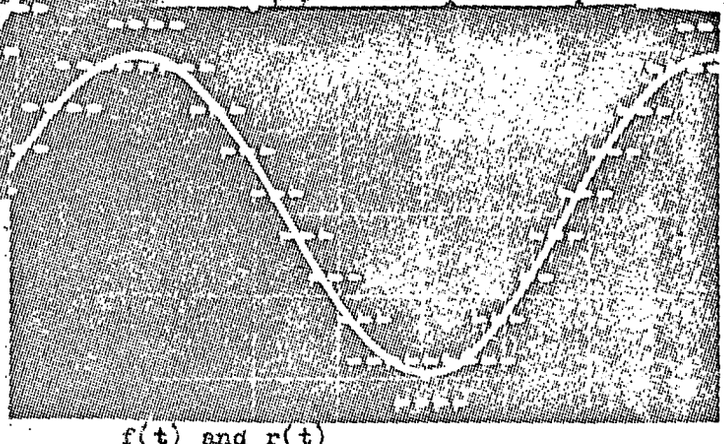
f(t) and r(t) e(t) and n(t) 1 volt/div

6.4 volts peak to peak



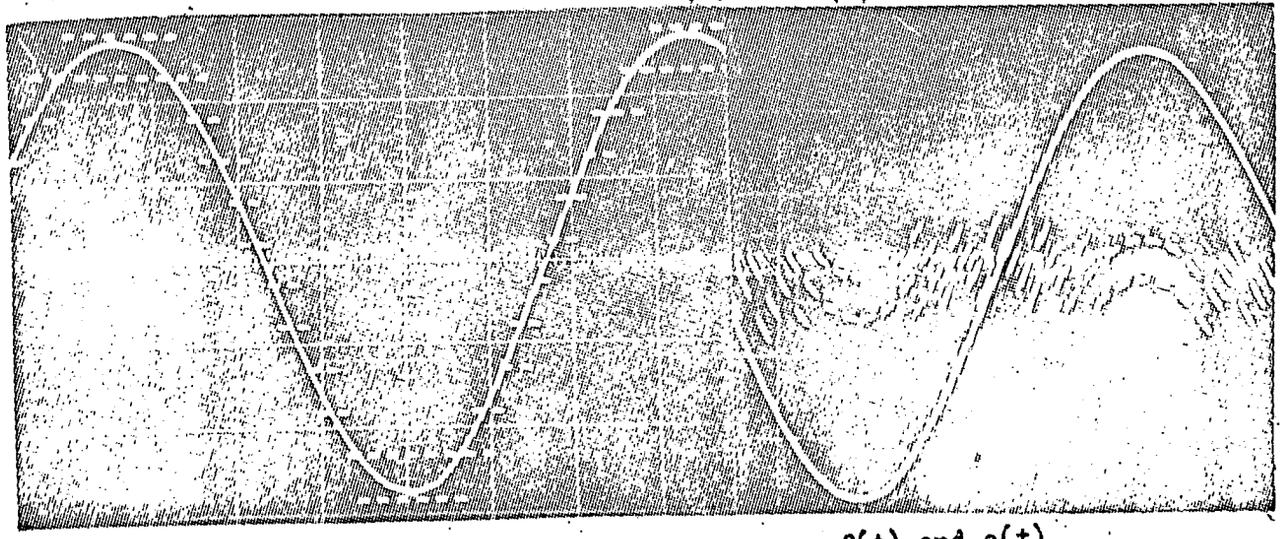
f(t) and r(t)

7.56 volts peak to peak



f(t) and r(t)

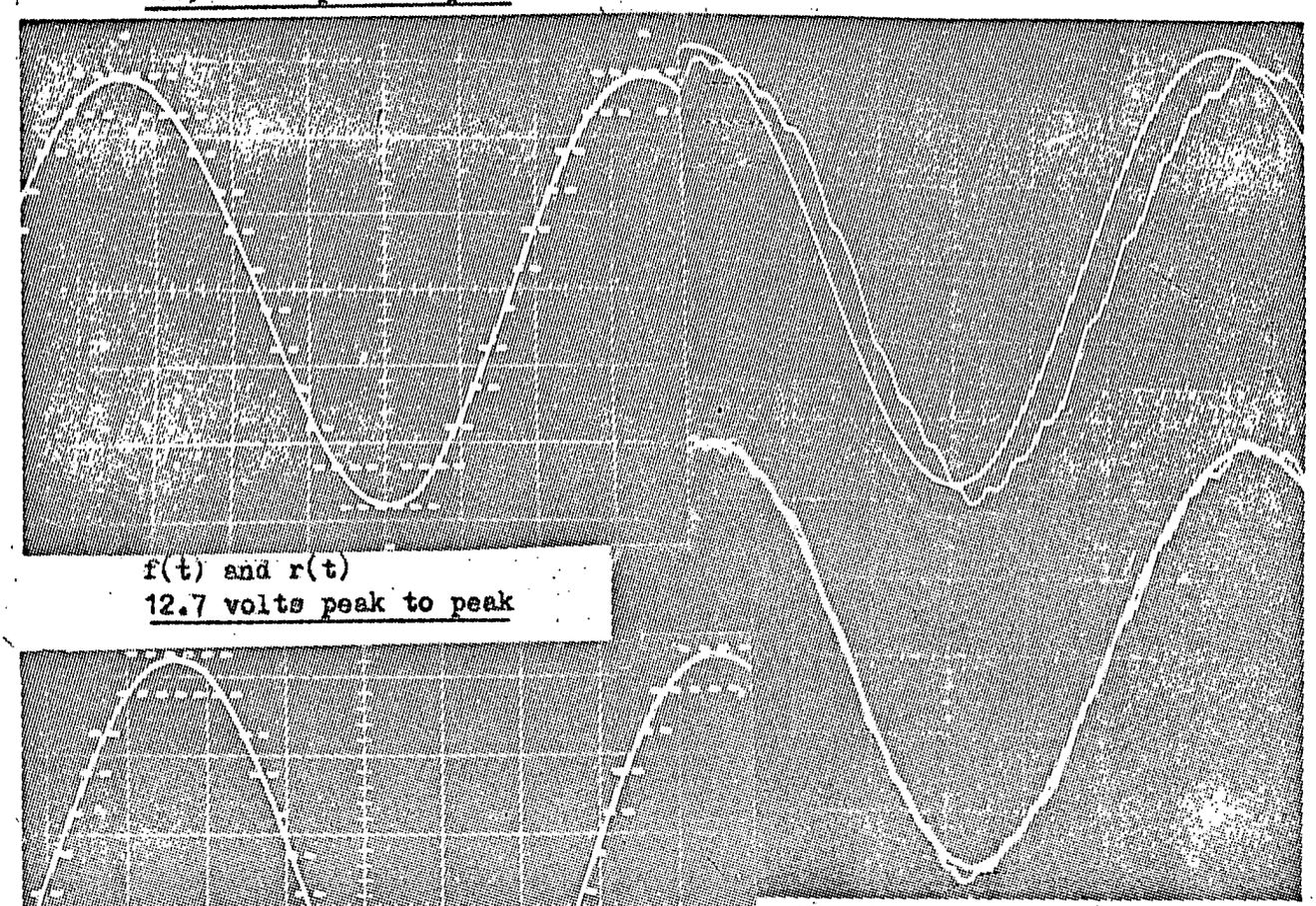
10.7 volts peak to peak



f(t) and r(t) f(t) and e(t)

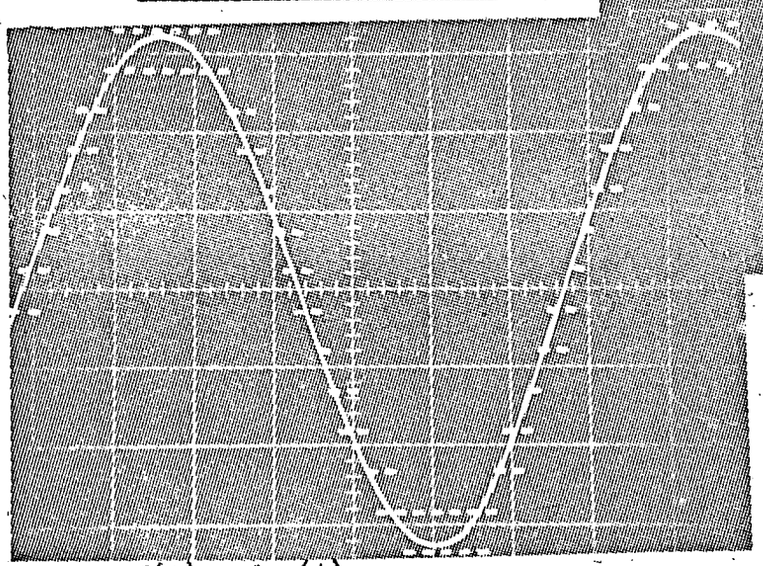
FIG 4.10 (Cont.)

10.9 volts peak to peak



f(t) and r(t)

12.7 volts peak to peak

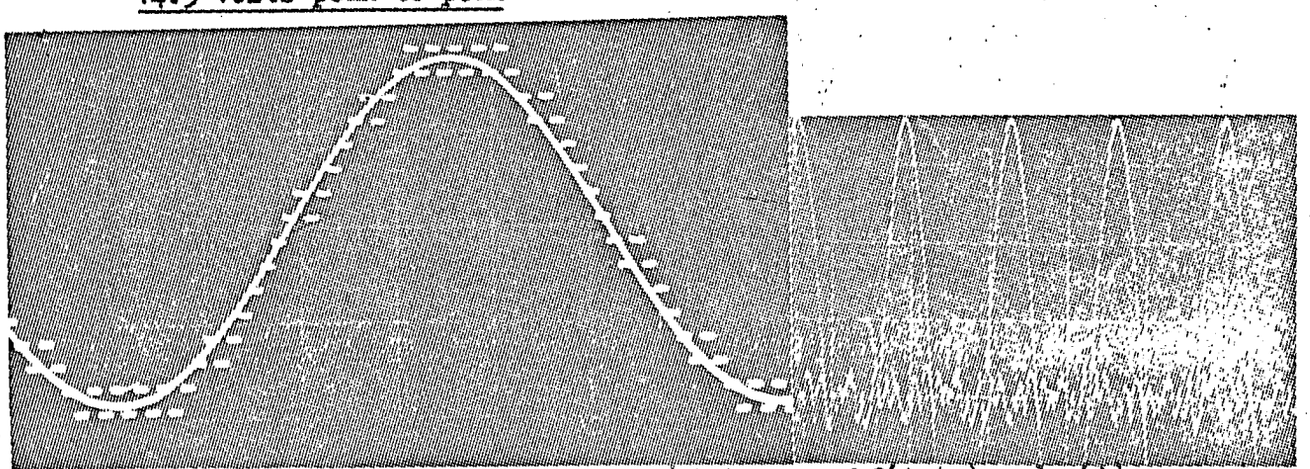


f(t) and r(t)

f(t) and y(t) . . Top

$kf(t-t_o)$ and y(t) . Bottom

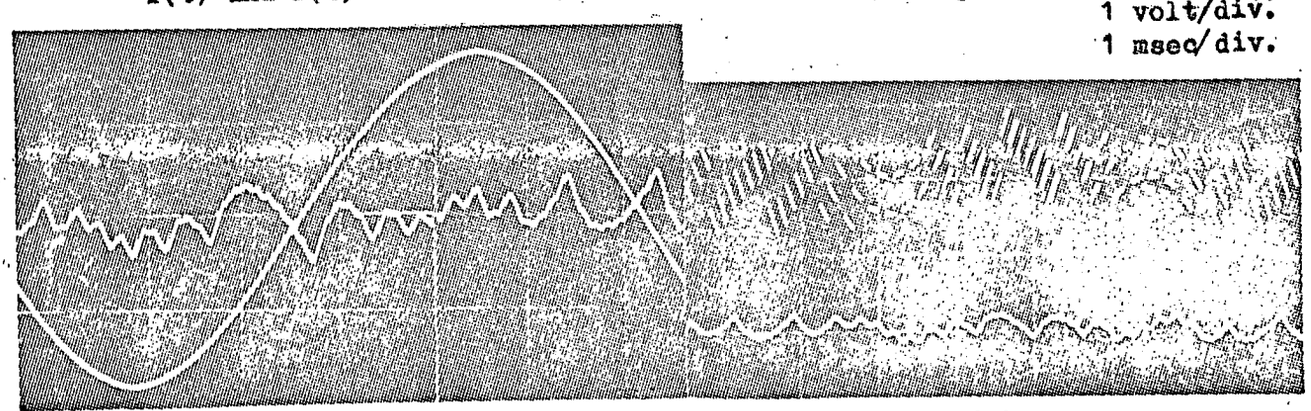
14.3 volts peak to peak



f(t) and r(t) 4 volts/div. $-kf(t-t_o)$ and n(t)

1 volt/div.

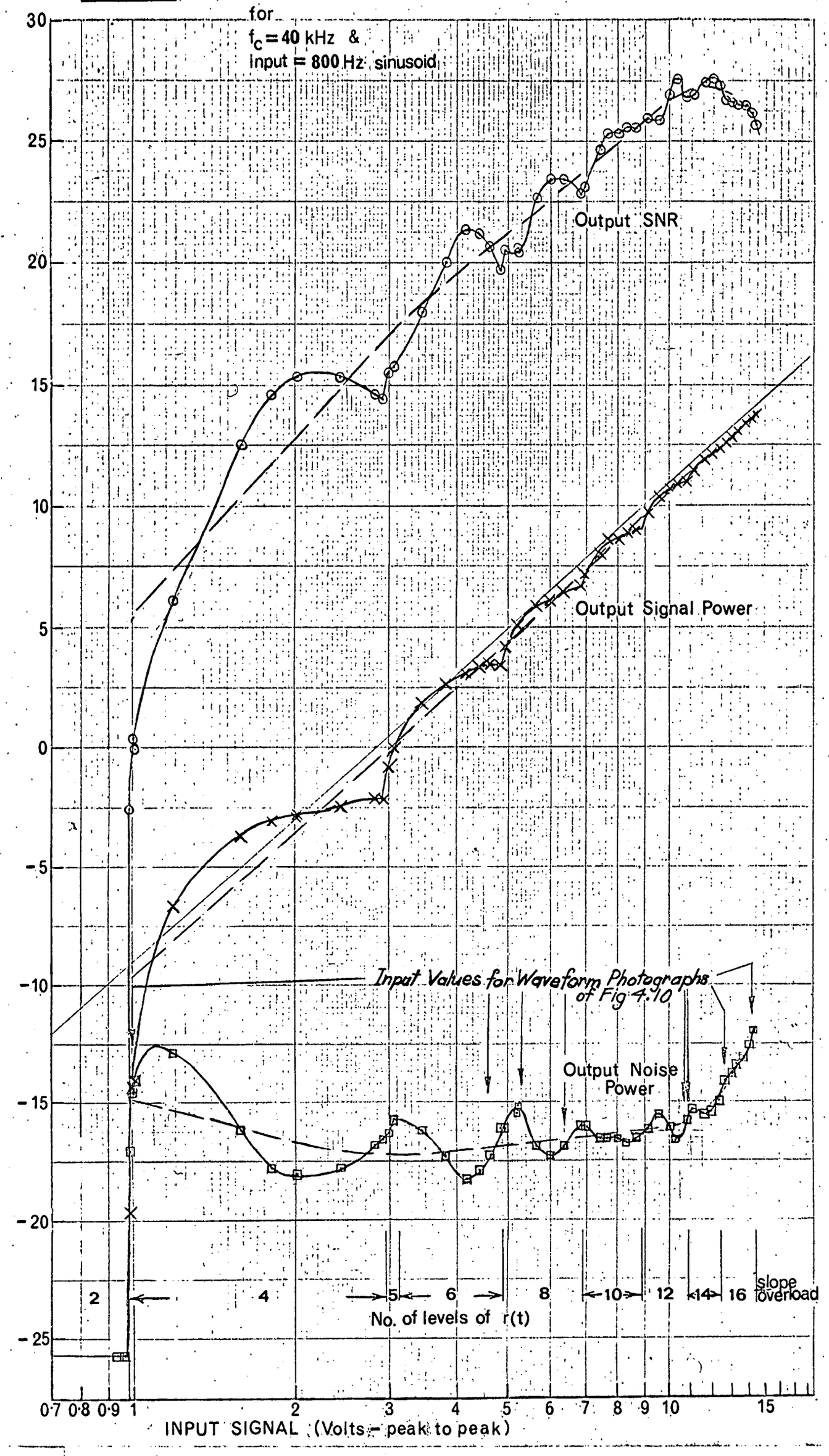
1 msec/div.



$kf(t-t_o)$ and n(t) 1 volt/div. e(t) and n(t)

4 volt/div. for f(t)

dB. **FIG 4-11** Measured Experimental Modulator Performance for $f_c = 40$ kHz & input = 800 Hz sinusoid



levels of r(t) at input signal values which are within about 3% of the expected values, indicates that the modulator is operating within the tolerances that would be expected.

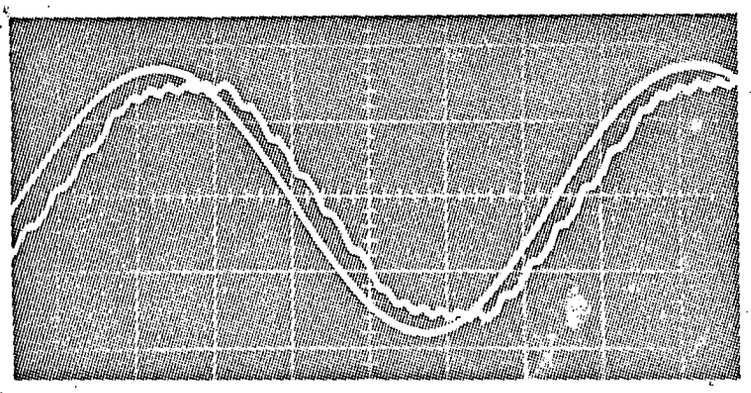
The regular variation in output readings can be observed (particularly in Fig. 4.12 for f_c = 80 kHz) to take the form of an immediate drop in noise power and rise in output signal power which accompanies the introduction of new levels to the output. The noise power then rises, falls and then rises again according to the input signal value relative to the value of the reconstructed signal, r(t). This variation of the noise power is considerable, varying from about 6dB for low levels of input to about 1 db for the higher inputs. Similarly the output signal power variation due to the effect of the introduction of new levels of r(t) is considerable, varying from about 4db at low levels of input to no variation for the higher inputs. Fig. 4.13 illustrates the considerable effect on the output signal of that particular small change of input which causes new levels of r(t) to be introduced.

The variation in the output SNR with each 2 volt change in the peak to peak value of the input signal, results mainly from the output noise variation, with reinforcement from the output signal variation. These variations in the observed results arise due to the use of a sinusoidal test signal. This phenomena would not occur for a non-deterministic type of input signal.

FIG 4.13 Photographs of Input and Output Signals

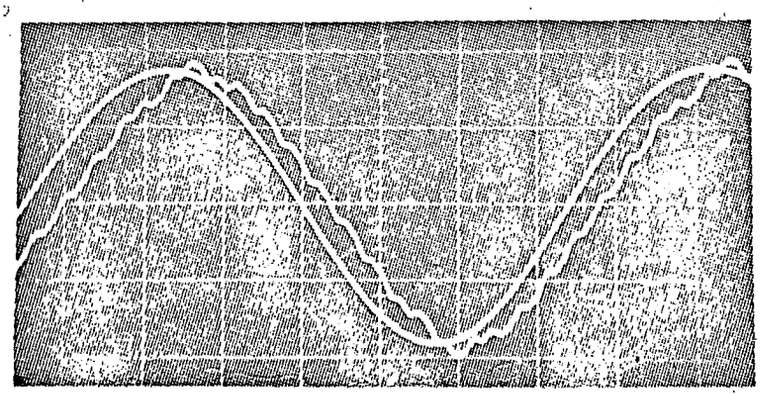
Input = 800 Hz sinusoid f_o = 40 kHz.

Scales : (approx) Time : 0.2 msec/div.
Amplitude : 2 volts/div.



Input : 6.8 volts peak topeak

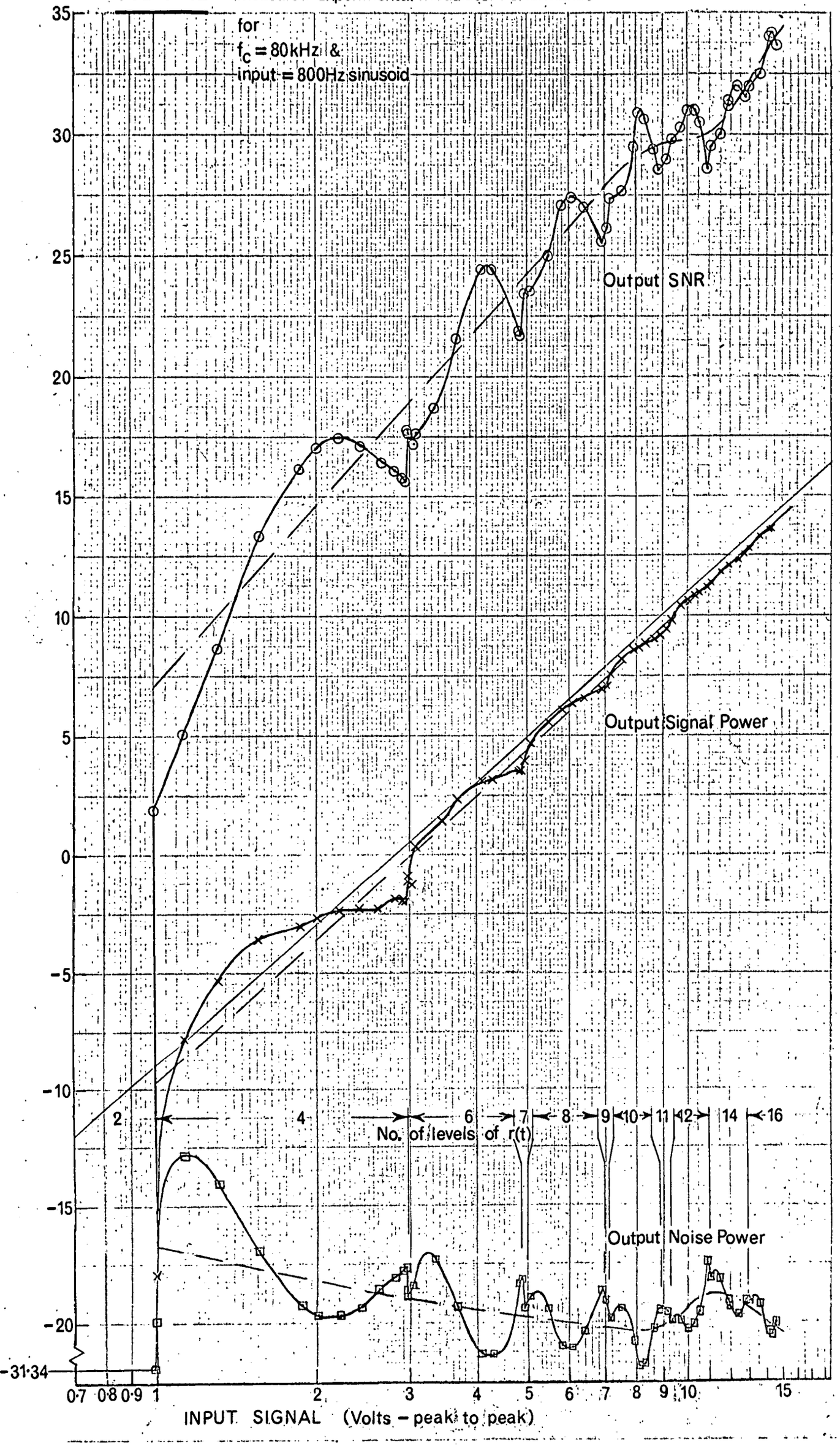
System near the limit of 8 levels of r(t).



Input : 6.9 volts peak to peak

System just using 10 levels of r(t).

FIG 4-12 Measured Experimental Modulator Performance



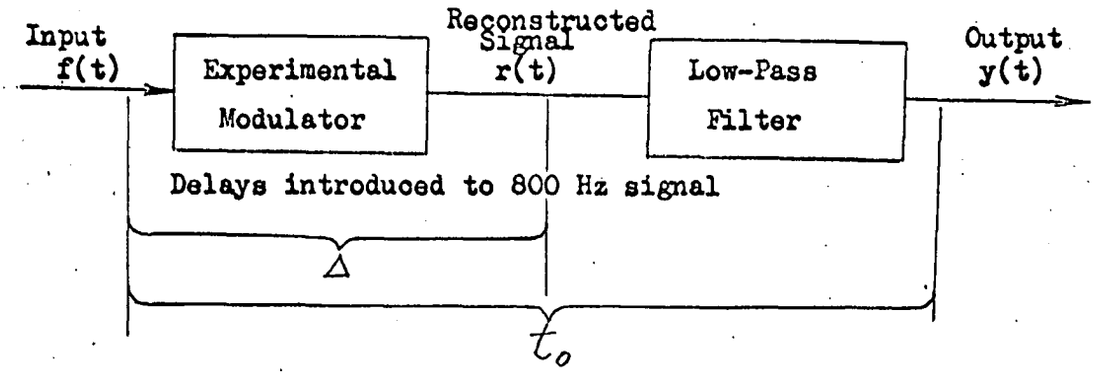
In applying previous analysis and conclusions to the constant clock frequency results, some method of eliminating the variations due to the deterministic nature of the input is necessary. In Chapter 2 a random input signal with an equal probability of taking any instantaneous value over a limited range was considered for the purposes of analysis. The effect of an input which was random but otherwise had all the characteristics of a sine wave, would be to give a "smoothed" or mean line for the performance curves as indicated by the dashed line in Figs. 4.11 and 4.12. It is these mean curves which will be given the greatest attention in discussion and analysis of the experimental modulator performance.

As expected from prior consideration of the delta modulator, the major contribution to the SNR versus input characteristic for a constant clock frequency is the output signal power, for the expected operating range of input signal amplitudes which were considered. It is when the input signal increases to the value that slope overload of the system starts to occur, that the contribution of the output noise to the characteristic of the SNR performance curve starts to become more significant. From Fig. 4.11 for f_c = 40 kHz it can be seen that the mean noise starts to increase significantly for inputs above about 11 volts peak to peak. It would therefore appear that this is the value at which slope overload commences. However, the maximum slope of the input, $f'(t)_{max}$, is less than the slope capability of the modulator, (hf_c) even for an input of 15 volts peak to peak. Thus slope overload in the strict sense does not occur. This increase in noise for inputs above about 11 volts peak to peak would be the result of the "partial" or "instantaneous" slope overload described in the last part of Section 2.3.2. In this Section, instantaneous slope overload was deemed to occur when r(t) fails to produce a change of sign in e(t) after two or more successive steps in the same direction. From Fig. 4.10 it can be seen that this situation occurs in the photographs where the input is greater than 11 volts peak to peak and does not occur in the photographs for inputs less than about 11 volts. For an input of 10.9 volts the photograph indicates that the waveforms are on the verge of the instantaneous slope overload situation and this value can be seen to correspond approximately to the value at which the significant rise in the noise power starts.

Delay Considerations

Delay of the output signal, y(t) behind the input, f(t) is introduced by both the experimental modulator and the low pass filter, as indicated in Fig. 4.14. This figure also indicates the theoretical phase delay of the low pass filter. For the signal frequency of 800Hz

FIG 4.14 Delay Introduced by Delta Modulator



Approximate measured values for f_c = 40 kHz

Δ = 20μsec t_o = 80μsec

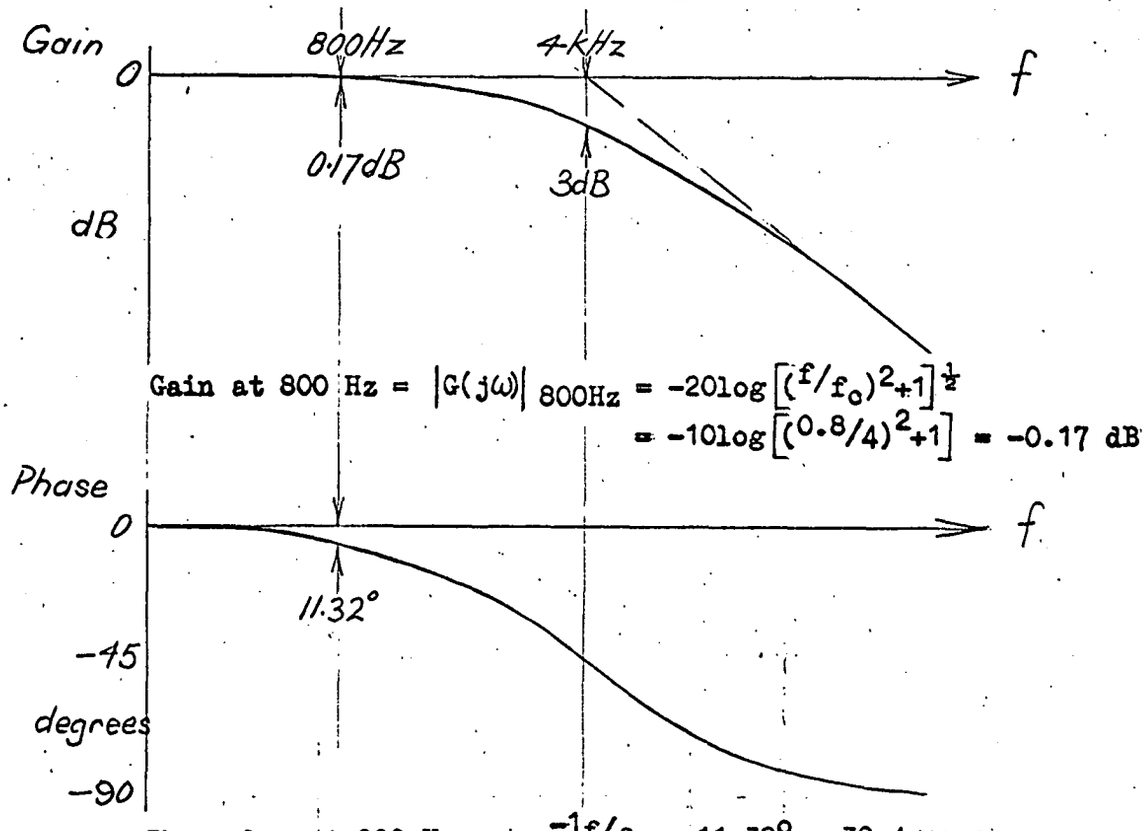
Therefore delay due to low-pass filter = 60μsec (approx.)

Period of input signal = $\frac{1}{800\ \text{Hz}}$ = 1250μsec

Clock period, for f_c of 40 kHz = 25μsec

80 kHz = 12.5μsec

Theoretical Low Pass Filter Characteristic



Gain at 800 Hz = $|G(j\omega)|_{800Hz} = -20\log\left[(f/f_c)^2+1\right]^{\frac{1}{2}}$

$= -10\log\left[(0.8/4)^2+1\right] = -0.17$ dB

Phase lag at 800 Hz = $\tan^{-1} f/f_c = 11.32^\circ = 39.4\mu$sec.

For f_c=40 kHz

Measured delay of output, $\left(y(t) = L\left[r(t)\right]\right)$ behind input, $\left(f(t)\right)$

$= 22^\circ = \frac{22}{360} \times 1250\mu\text{sec} = 76.4\mu\text{sec}$.

Therefore delay due to experimental modulator

$\simeq (76.4 - 60) = 16.4\mu\text{sec} = \frac{16.4}{25} = 0.66$ Clock periods.

For f_c=80 kHz

Measured delay of output behind input

$= 20^\circ = \frac{20}{360} \times 1250\mu\text{sec} = 69.5\mu\text{sec}$

Therefore delay due to experimental modulator

$\simeq (69.5 - 60) = 9.5\mu\text{sec} = \frac{9.5}{12.5} = 0.76$ Clock periods.

the theoretical delay is 11.3° or 39.4 μsec. However, as indicated in Fig. 4.14, approximate measurements using the C.R.O. indicate a low pass filter delay of about 60 μsec. This difference would result from the effect on the filter characteristics of the loading due to the circuitry following the filter. (Because the filter was emperically designed to give an amplitude characteristic close to the theoretical, the phase characteristic could not be expected to also be close to the theoretical).

The delay of the output, y(t) behind the input, f(t) was measured along with the other values, for each set of parameters, using more accurate methods than direct readings from the C.R.O. The delay, t_o, which gave minimum noise was determined both from the dial setting of the $-kf(t-t_o)$ output (see Fig. 4.7) and from Lissajou figures on the C.R.O. This gave delays of 22° for f_c = 40kHz and 20° for f_c = 80kHz for all values of the input, provided the system was not in a state of slope overload. As indicated in Fig. 4.14 this gives a rough value of 0.7 clock periods for the delay introduced by the experimental modulator.

The "mean" performance curves of Figs. 4.11 and 4.12 show the output signal level to be from about 0.3 dB below the input signal level for high values of input to about 0.75dB below the input signal level for lower inputs. This loss of power from the input to output resulting from the modulation/demodulation process is partly attributable to the loss due to the low pass filter. The theoretical loss at 800 Hz of 0.17 dB would be close to (better than +0.2, -0.1dB) the actual loss; as the filter's measured amplitude characteristic corresponds closely to the theoretical (refer Fig. 4.8). The remaining loss (approximately 0.1 to 0.6dB) would be due to the effect of the delay of r(t) behind f(t) introduced by the experimental modulator. The photographs comparing the input, f(t), with the output, y(t), indicate how the delay would prevent the output from reaching the full value of the input.

4.6.2. Constant Input Signal

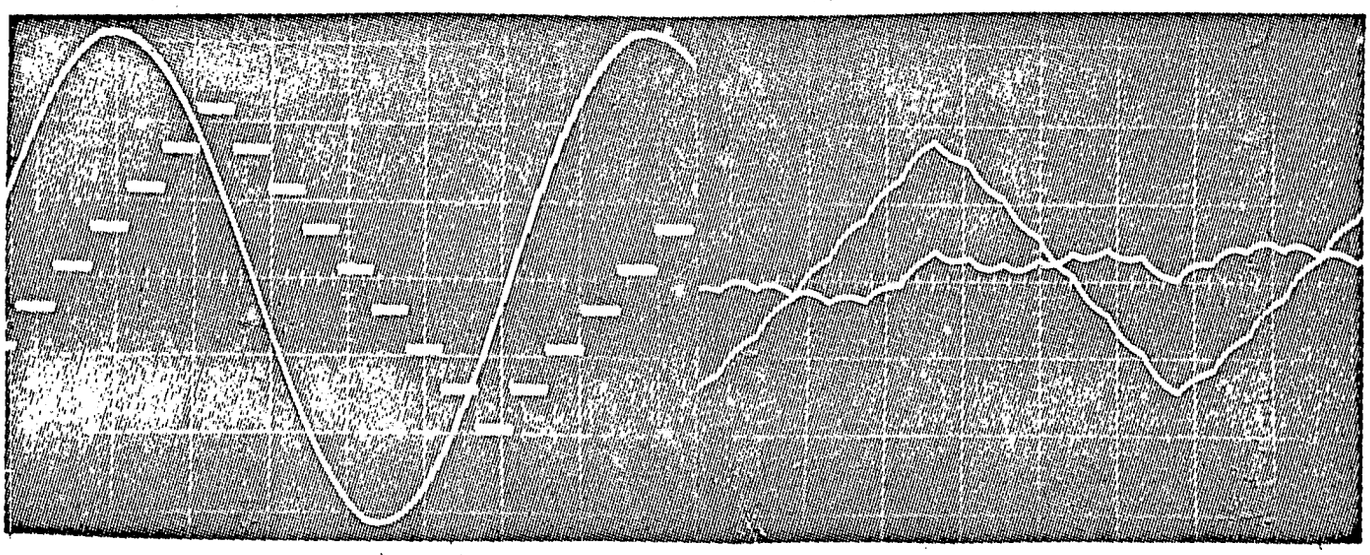
From the results taken for constant input signal amplitudes of 6 volts and 1.875 volts, the curves of the experimental modulator performance with varying clock frequency were determined. These results will be subject to the same error effect, due to the deterministic nature of the input, as the constant clock frequency results. With the constant input signal results, the error will be of the form of a bias of the results from the "mean", in one direction. This error can be estimated from Fig. 4.11 and 4.12 for two of the clock frequency values. The estimated values of the errors are shown below the appropriate points in Fig. 4.16 and 4.17. From these values and from Figs. 4.11 and 4.12 it was observed that the maximum variation from the mean due to the deterministic nature of the input is subject to the following

FIG 4.15 Photographs of Waveforms for Various Clock Frequency

with Input = 4.24 volts rms (ie 12 volts peak to peak)
800 Hz sinusoid

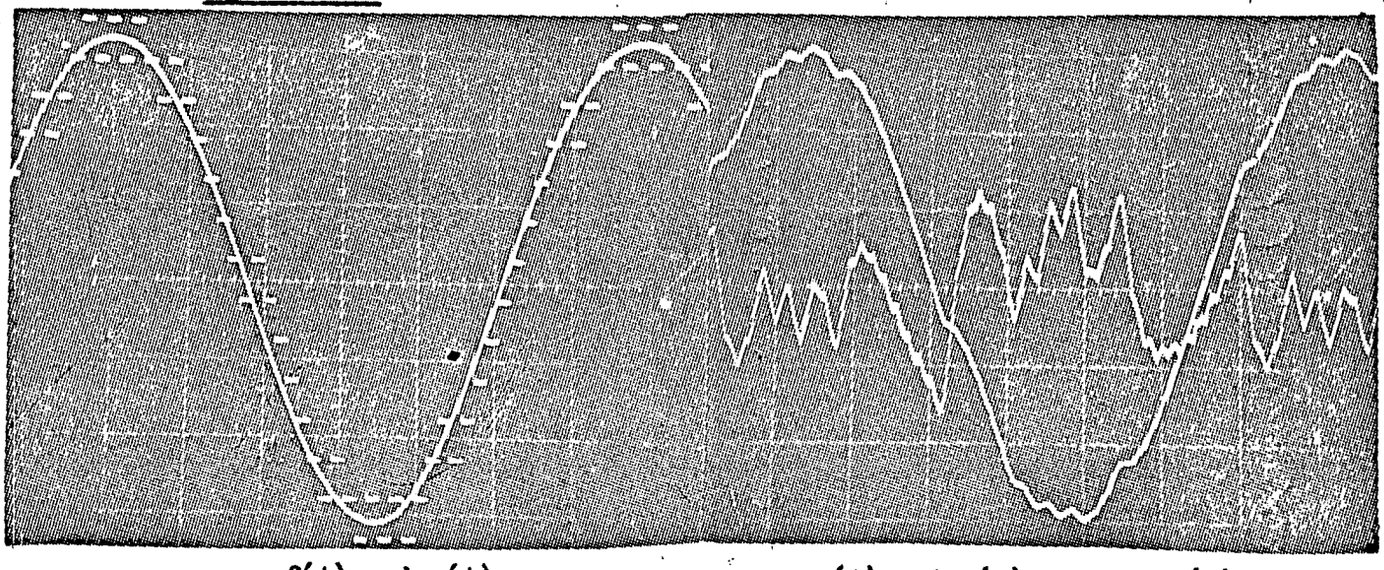
Scales: (approx) Time: 0.2 msec/div.
Amplitude: 2 volts/div.
(unless otherwise specified)

f_o = 10 kHz



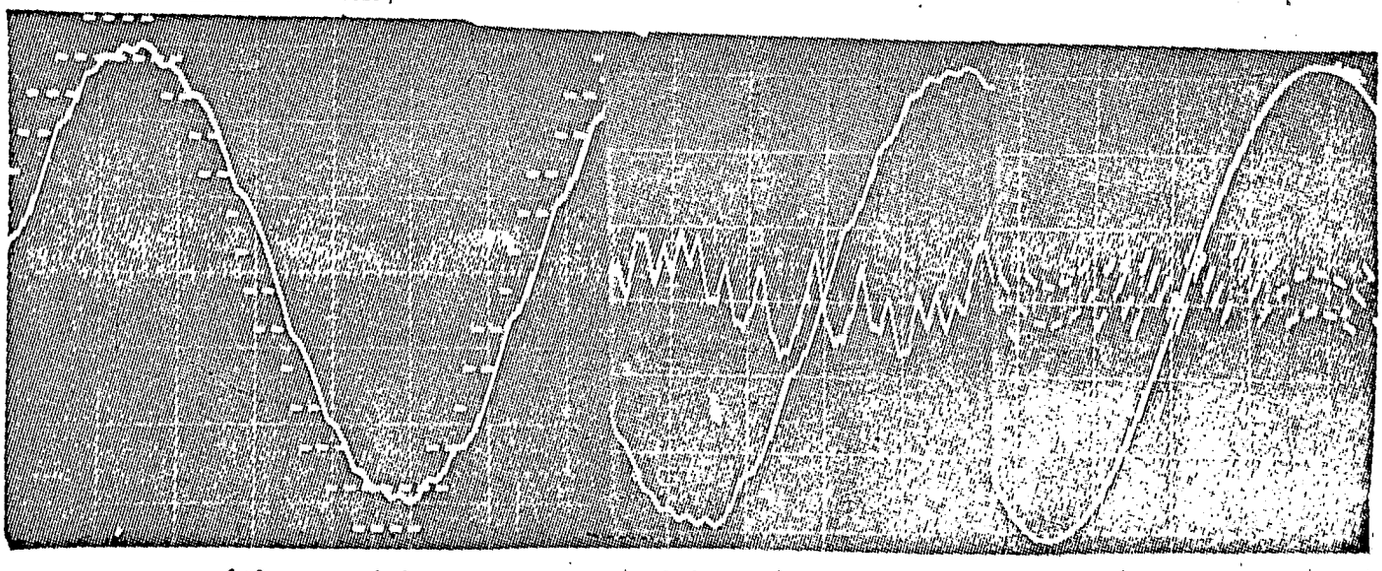
f(t) and r(t) y(t) and n(t)

f_o = 30 kHz



f(t) and r(t) y(t) and n(t) n(t) scale
0.2 volt/div.

f_o = 40 kHz



y(t) and r(t) y(t) and n(t) f(t) and e(t)
n(t) scale
0.4 volts/div.

FIG 4.15 (cont)

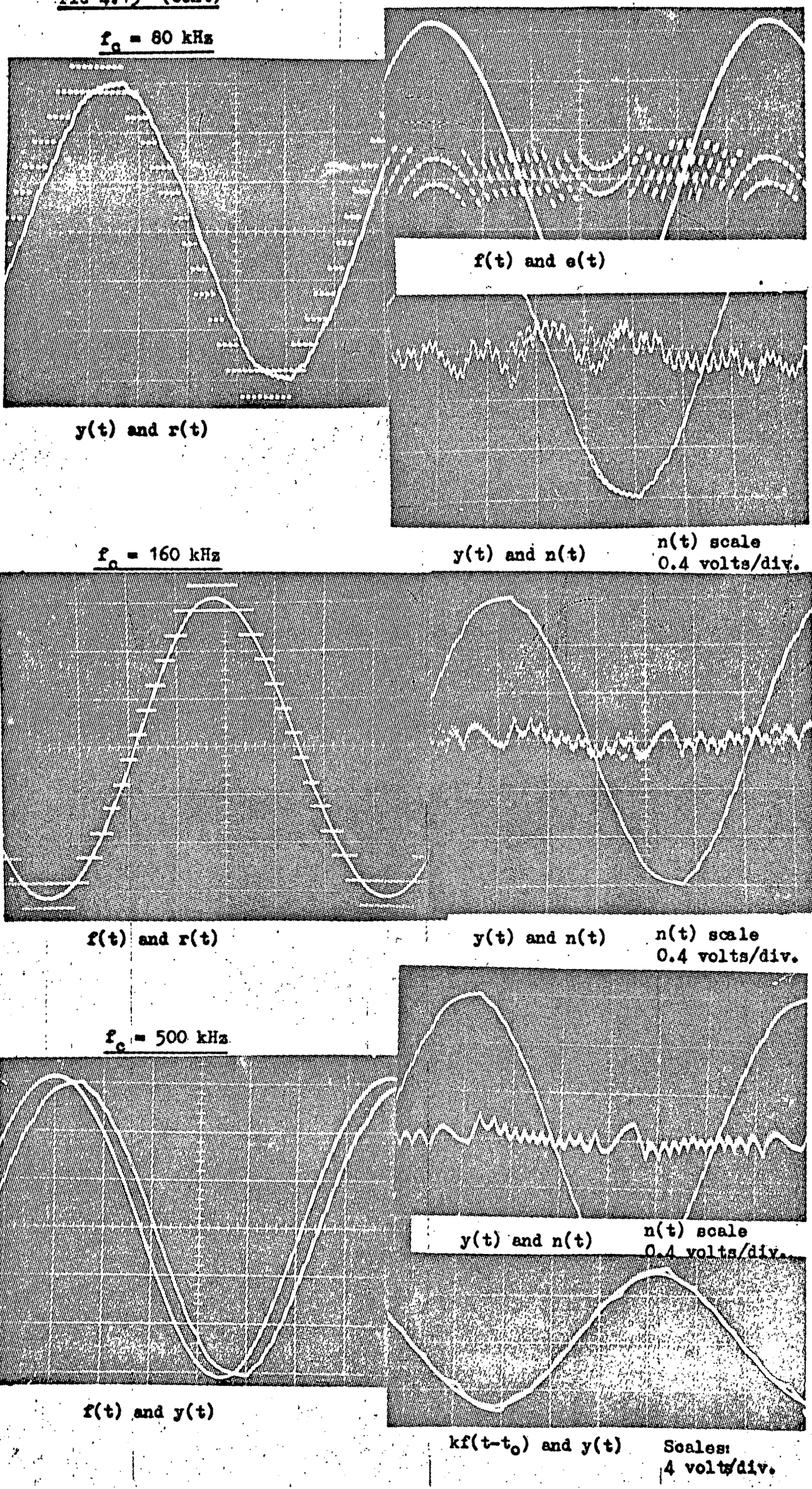
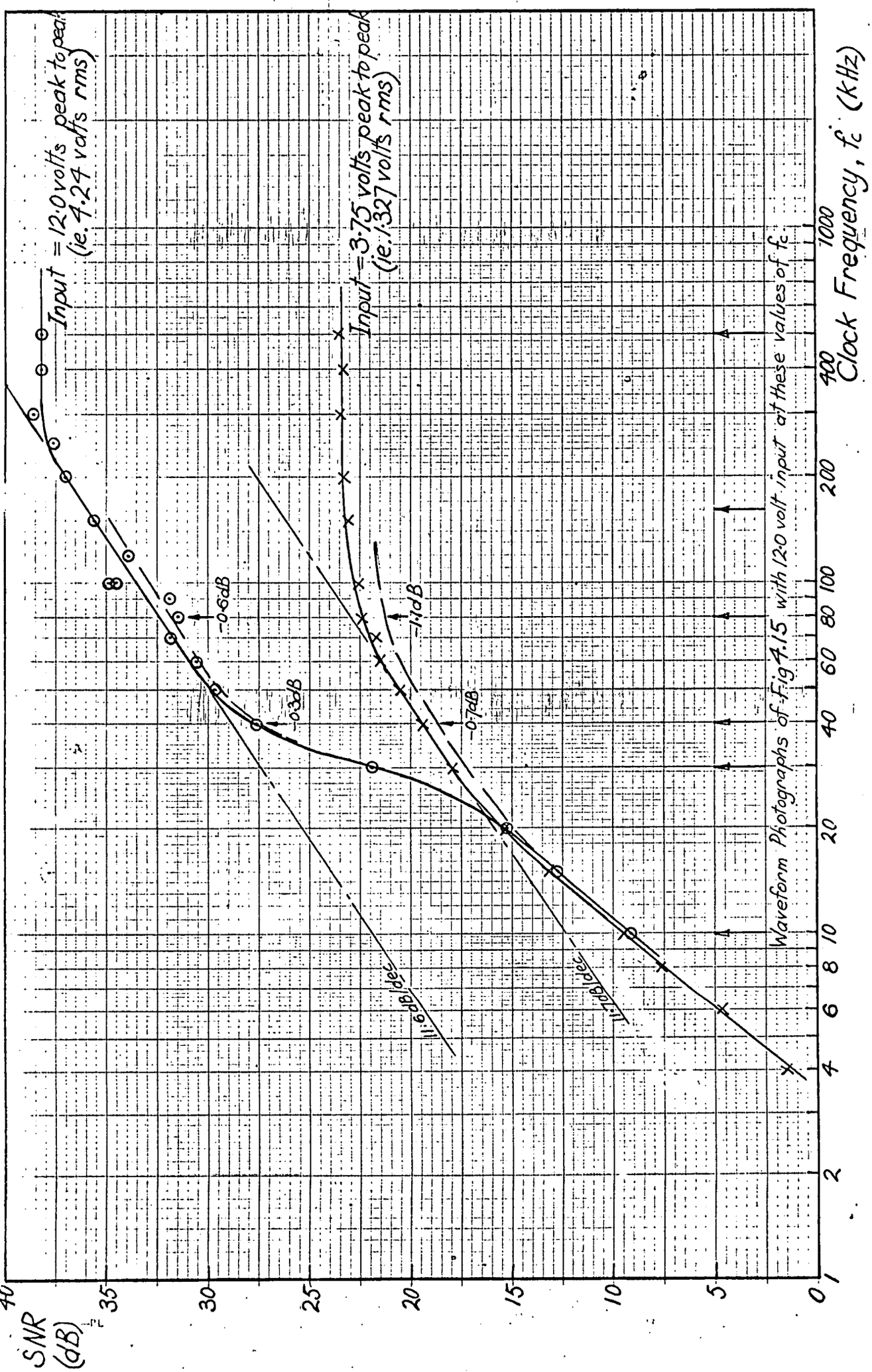

f_c = 80 kHz
f(t) and e(t)
y(t) and r(t)
y(t) and n(t)
n(t) scale
0.4 volts/div.
f_c = 160 kHz
f(t) and r(t)
y(t) and n(t)
n(t) scale
0.4 volts/div.
f_c = 500 kHz
y(t) and n(t)
n(t) scale
0.4 volts/div.
f(t) and y(t)
kf(t-t_o) and y(t)
Scales:
4 volts/div.

FIG 4.16 Output SNR Variation with Clock Frequency from Experimental Results

Input = 800 Hz sinusoid , Step height = 1 volt.



conditions:-

(i) for the output signal power, the maximum variation decreases with increasing input but does not appear to vary with clock frequency in the non-overload region.

(ii) for the output noise power, the maximum variation decreases with increasing input, tending to zero variation as slope overload is approached. The variation also decreases with decreasing clock frequency.

From the above observations it appears that the fluctuation in the results and hence the bias on the curves of Fig. 4.16 and 4.17 would taper off to zero as slope overload was approached. The dashed curves of Figs. 4.16 and 4.17 indicate the estimated performance with the bias taken into account. This same variation from the mean, resulting from the deterministic nature of the input, would account for the fluctuation in the results of Fig. 4.16, particularly noticeable for the larger input.

Fig. 4.17 showing the output signal and noise performance with varying clock frequency, indicates the relative contribution of the signal and noise to the SNR performance curves of Fig. 4.16. Three distinct performance regions, for various ranges of clock frequency, emerge from Figs. 4.16 and 4.17. Considering the curves for the 12 volt peak to peak input it is observed that for clock frequencies above 30 kHz, the output signal is constant and thus the output noise determines the characteristic of the SNR curve. Below 30 kHz the output signal is attenuated due to slope overload in the modulator and both the output signal and noise contribute towards the characteristic of SNR curve.

For an input of 12 volts peak to peak the three regions for the performance are as shown in Fig. 4.17, and they can be defined as follows:-

Region I - Slope overload and partial slope overload region, consisting of clock frequencies up to about 45 kHz. This region can be regarded as consisting of:

(i) Region IA with clock frequencies up to a value of about 25 kHz in which the system experiences slope overload. In this region the noise power appears to be constant at its maximum value of about -5db and the output signal power falls at approximately 20 db/decade with decreasing f_c.

(ii) Region IB, between about 25kHz and 45kHz in which the system experiences partial slope overload. In this region the output signal power tends to its maximum (of about 0.4 dB below the input) and the noise power falls rapidly with increasing f_c.

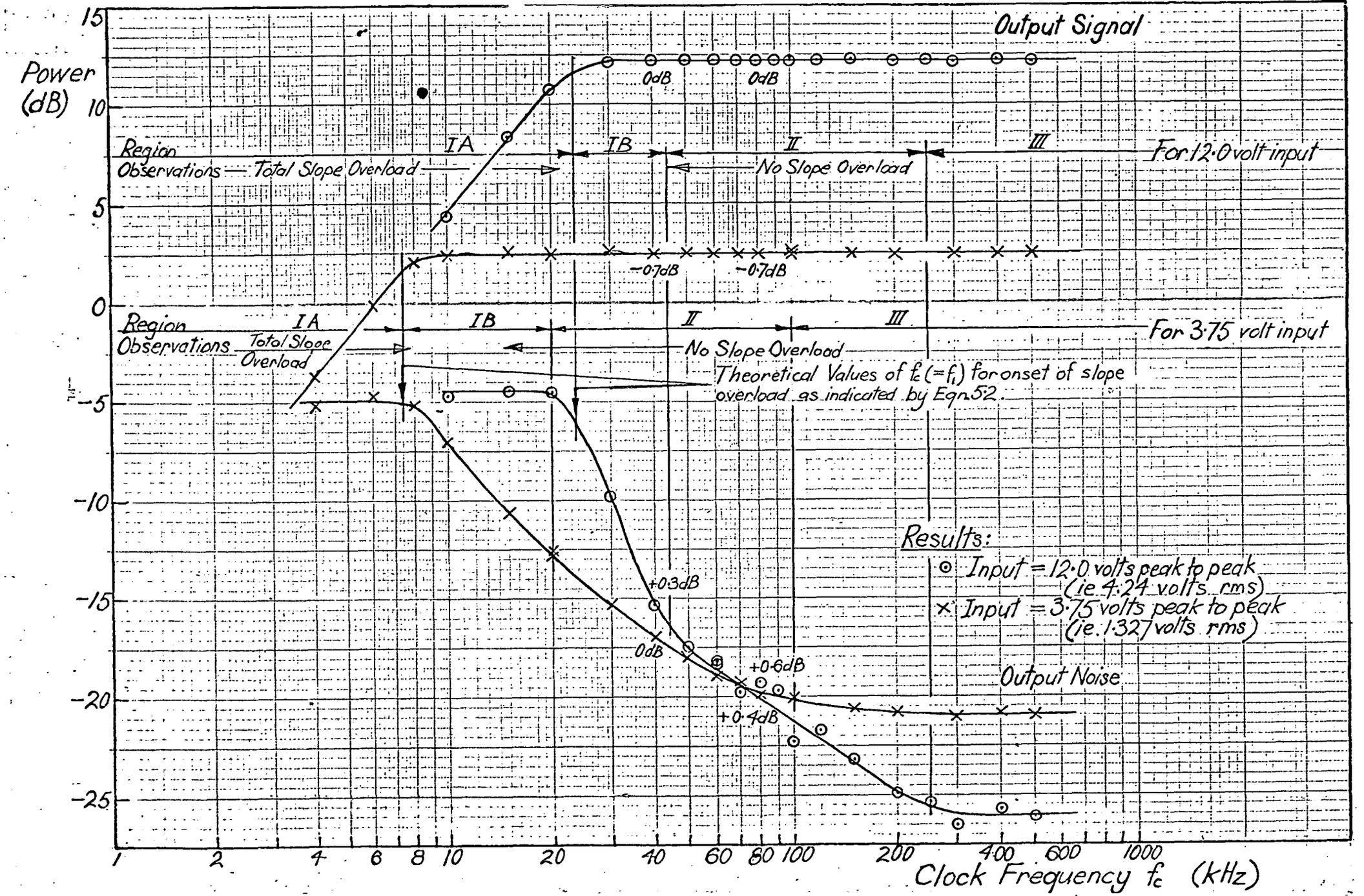


FIG 4.17 Output Noise and Signal Variation with Clock Frequency from Experimental Results

Input = 800 Hz sinusoid, Step height = 1 volt.

Region II - would be the optimum operating range and includes clock frequencies from approximately 45 to 250 kHz for the 12 volt peak to peak input. In this region the noise power falls at a rate of roughly 10 db/decade with increasing f_c and the output signal remains constant.

Region III - includes the range of clock frequencies from about 250 kHz and above for which the noise power is at a minimum and no further noise reduction is achieved through increasing f_c. This minimum noise region was anticipated, and is discussed in Section 2.3.4. The output signal also remains constant in this region and thus the SNR remains constant at its maximum value for the given input signal.

The performance regions as described are most suitably defined by the approximate level of loading of the delta modulator at the boundary of the regions. A suitable measure of loading for single integration delta modulation is the value of D as used by van de Weg[34] and later Abate[17], where $D^{\frac{1}{2}}$ equals the rms value of the signal derivative. ie. $D = \overline{[f'(t)]^2}$. (For sine wave signal $D^{\frac{1}{2}} = 2\pi f_s A/\sqrt{2}$). This gives (from Fig. 4.17) the following emperically determined definitions for the performance regions:

Region IA $D^{\frac{1}{2}} > 0.9hf_c$

Region IB $0.9hf_c > D^{\frac{1}{2}} > 0.4hf_c$

Region II $0.4hf_c > D^{\frac{1}{2}} > 0.09hf_c$

Region III $D^{\frac{1}{2}} < 0.09hf_c$

Arrows in Fig. 4.16 indicate the values at which the waveform photographs of Fig. 4.15 were taken. These photographs illustrate the waveforms observed and the nature of the noise signal, for the various clock frequencies and their associated levels of loading. From the photographs for f_c = 10kHz (and from observation of the waveforms for other clock frequencies in Region IA) it was observed that for the system in total slope overload:

(i) the output is basically triangular in shape and is reduced in amplitude and considerably delayed relative to the input.

(ii) the noise signal consists fundamentally of odd harmonics of the signal frequency plus components of the clock frequency and its harmonics.

Observations (as noted in Fig. 4.16) of the number of levels of r(t) and the degree of overload of the system, also helped to define the boundaries of the overload regions used to describe the performance characteristics.

It was observed that as the system moved further into the total slope overload condition, the output signal amplitude and the

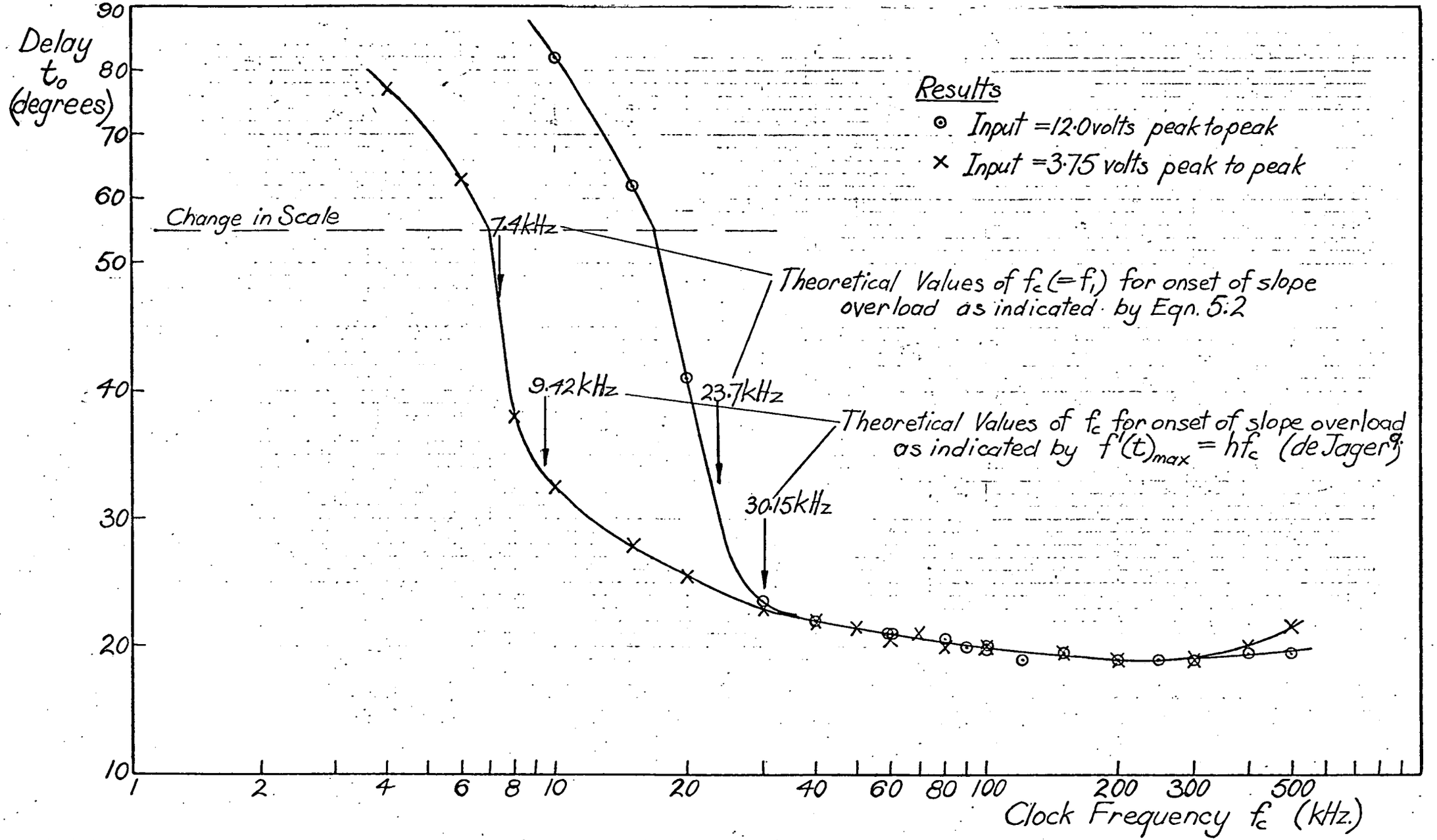


FIG 4.18 Variation in Delay of Output, y(t) behind Input, f(t) with Clock Frequency from Experimental Results

output noise were determined solely by the clock frequency and the step height. It is under these conditions that the output signal and noise performance become independent of the input signal value. This situation is reflected in Fig. 4.16 and 4.17, where, for an f_c of about 8kHz and less (ie. total slope overload for both inputs) the S_o, N and SNR performance curves for both input values are shown to coincide. A theoretical description of the performance in this region, derived from the observed waveforms, is provided in Section 5.2.

From the photographs for f_c = 30 kHz and 40 kHz, it can be seen that a state of "partial" slope overload exists (as defined earlier by r(t) failing to produce a change of sign in e(t) after two or more successive steps in the same direction). Associated with the partial slope overload state it was also observed that the magnitude of the error function, e(t), reached values considerably greater than 1 volt at times. This observation will have an effect on the theoretical considerations of Section 2.3; as discussed in Chapter 5.

Observations for the higher clock frequencies indicate that as the clock frequency increases, the noise signal becomes a regular wave shape determined by the ratio of the signal amplitude to the step height. Superimposed on this basic shape are components of the clock frequency and its harmonics. As f_c becomes much greater than f_m, (the filter cut-off frequency) the contribution of harmonics of f_c to the noise power falls. This explains why in Fig. 4.17 the noise power stops decreasing with increasing f_c beyond a particular value (denoted by Region III) (Refer Section 2.3.4).

Delay Considerations

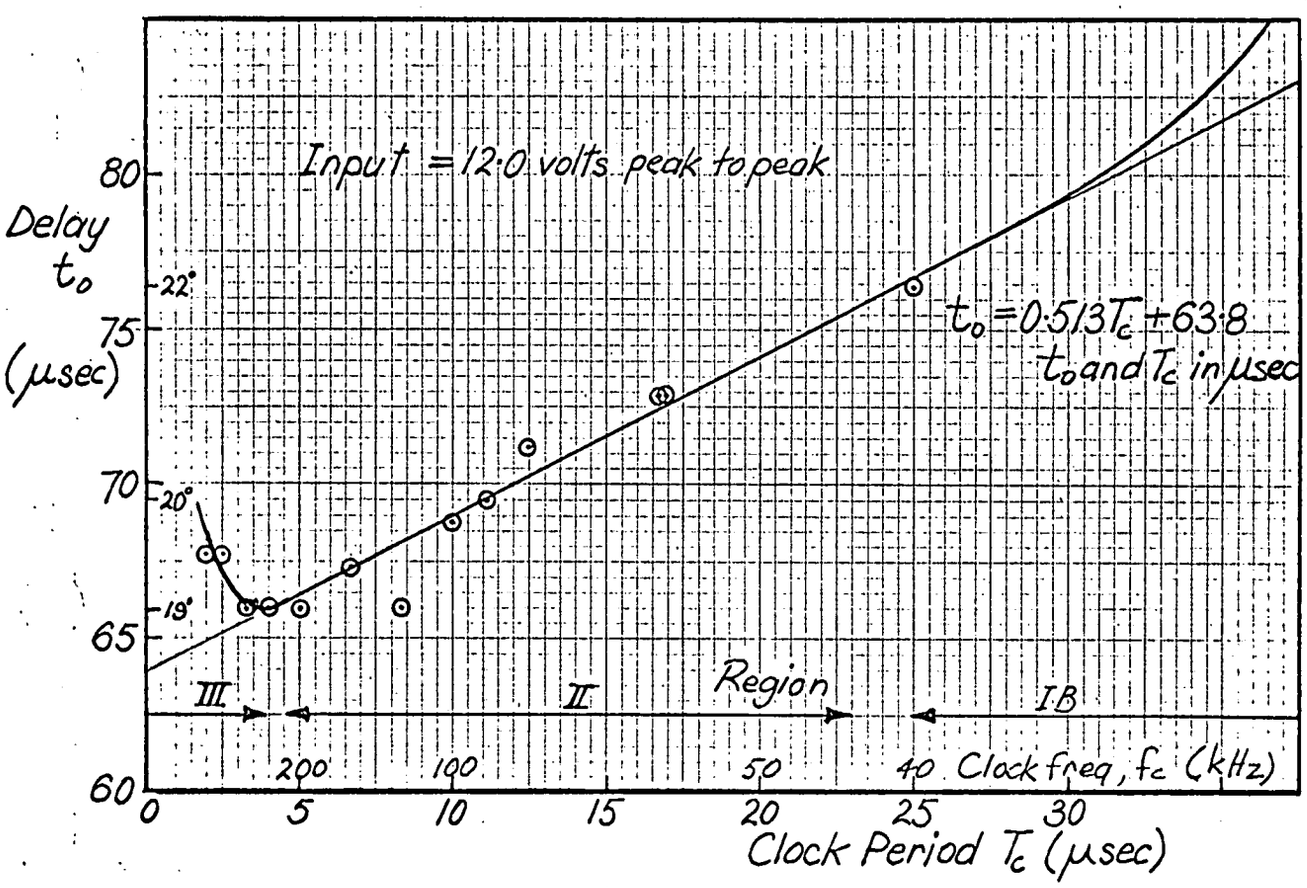
For the constant input signal case, values of the delay, t_o of the output, y(t) behind the input, f(t) were taken over the range of f_c's. These are shown in Fig. 4.18. From the curves of Fig. 4.18 the following observations can be made:-

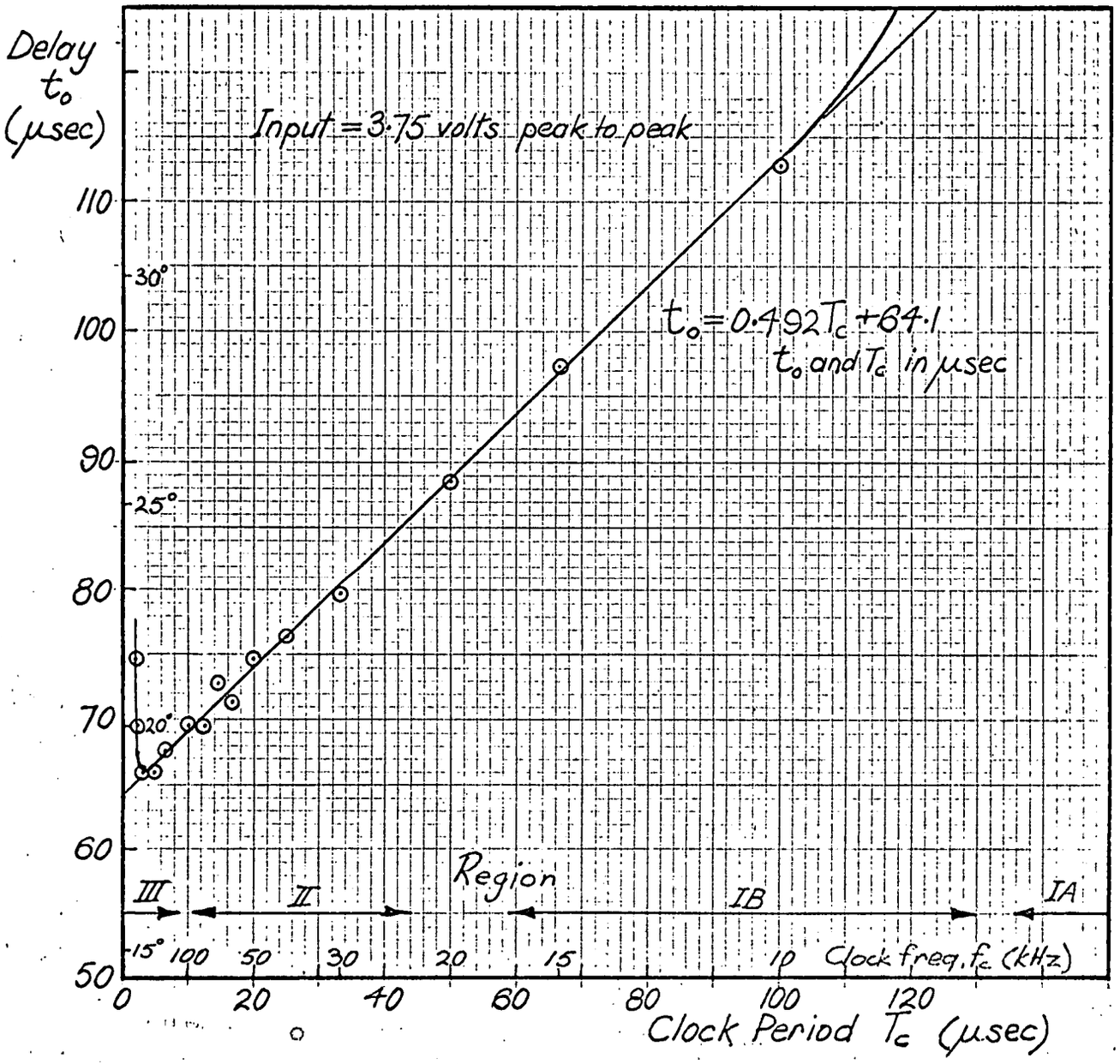
(i) Over the range of f_c from 30 to 300 kHz the delay curves are the same for both input signal amplitudes.

(ii) The curves exhibit sharp cut-off points at clock frequencies of about 9 kHz and 30 kHz for inputs of 3.75 volts and 12.0 volts peak to peak respectively.

(iii) For clock frequencies below the cut-off values the delay increases rapidly with decreasing f_c due to the additional and over-riding delay effect of total slope overload. For f_c's above the cut-off value there is a gradual decrease in delay with increasing f_c, until a value of about 300 kHz is reached.

FIG 4.19 Variation in Delay with Clock Period





The delay curves could not be expected to yield any information to help define the region of partial slope overload, as the state of partial slope overload effects the noise function but not the output signal. However, the cut-off value of the clock frequency for the delay curves does provide a good indication of the onset of total slope overload. Once the maximum signal slope exceeds the modulator slope capability, increased delay of the output behind the input would be expected. The theoretical values of the clock frequency for the on-set of slope overload (based on de Jager's description*) are 30.15 and 9.42 kHz for inputs of 12.0 and 3.75 volts peak to peak respectively. These values shown good agreement with the observed cut-off frequencies from Fig. 4.18 for the experimental delay curves.

It was anticipated in the delay consideration of Section 4.6.1 that for the modulator operating in the non-overload condition, the delay introduced by the modulator alone would be a constant fraction of the clock period, regardless of the clock frequency. To test this hypothesis the delay results were replotted in Fig. 4.19 as delay against clock period, T_c, for the non overload results.
The delay introduced by the low pass filter will be constant for the 800 Hz sinusoid signal regardless of the values of the input signal and the clock frequency. Fig. 4.19 indicates a straight line delay versus clock period characteristic for both values of input, over the modulator's operating range (Region II). This confirms the suggestion that the modulator delay, Δ, is a fixed proportion of the clock period.

An equation for the total delay between the output and input of:

$t_o = 0.5T_c + 64$, for t_o, T_c in μsec, gives a good description of the delay, clock period characteristic for both inputs. It can be concluded that the delay introduced by the experimental modulator is about 0.5 T_c, regardless of the input signal value or the clock frequency. The low pass filter delay, as indicated by Fig. 4.19, for an 800 Hz sinusoidal input, is about 64 μsec, (ie. 18.4°).

There is no apparent explanation for the increase in delay when f_c exceeds 300 kHz. This increase occurs in the region of minimum noise operation (i.e. Region III) for both of the input signal amplitudes but is apparently independent of the onset of Region III. The only difference between the delay increase for the two inputs is that it is greater for the smaller input. It is not considered likely that a continuing increase in delay would occur for increasing f_c. Further investigation would be necessary to determine the cause of this increase in delay for high clock frequencies.

---

* For a sinusoidal signal, the maximum signal slope:
$f'(t)_{max} = 2\pi f_s A$, equals the modulator slope capability hf_c, for the onset of slope overload. Therefore the onset of slope overload is defined by $2\pi f_s A = hf_c$.

CHAPTER 5.

CONCLUSION

5.1 Introduction

This chapter seeks to relate the results, observations and conclusions of the preceeding three chapters. Points of agreement between the analysis, the computer simulation results and the experimental results will be summarized to provide a description of the performance of delta modulation under the conditions considered. Explanations for points of variance between the analysis, computer simulation and experimental results will be given wherever possible. Conclusions on the nature of the noise in delta modulation will be made and the value of a sine wave as a test signal for more general inputs will be discussed.

5.2 Theoretical Considerations of Slope Overload

Based on the observations of the modulator waveforms for the slope overload situation, the description of the output signal and noise power performance as outlined in Fig. 5.1 was proposed. The equation for the output signal power thus determined, combined with the non slope overload S_o equation give the following description of the S_o performance:-

$$\left.\begin{aligned} S_o &= \frac{2}{\pi^4}\left[\frac{h^2 f_c^2}{f_s^2} - S_G(\omega)\right]_{f_s} \qquad \text{for } f_c < f_1 \\ S_o &= A^2/2 - \left[S_G(\omega)\right]_{f_s} \qquad \text{for } f_c > f_1 \end{aligned}\right\} \text{..Eqn. 5.1}$$

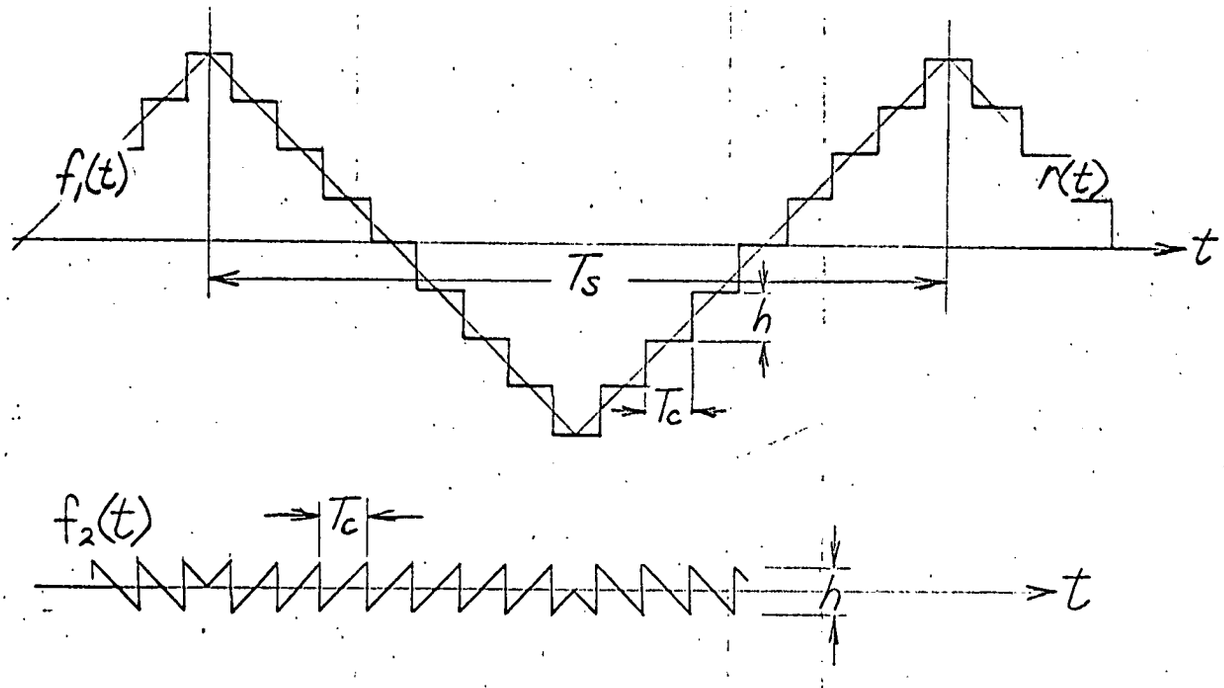
where $S_G(\omega)\big]_{f_s}$ is the power loss of the low-pass filter at the signal frequency, f_s; A is amplitude of the input signal f(t); and f_1 is the clock frequency at which the boundary of the slope overload region is deemed to exist, ie. f_1 defines the transition from RegionIA to RegionIB.

From the experimental results (refer Fig. 4.17) it can be seen that Equation 5.1 provides a very good description of the output signal power characteristic for a simple delta modulator with a sinusoidal input. In addition, from Equation 5.1 an expression for the onset of slope overload (ie. an expression for f_1) can be defined:-

$$\frac{2}{\pi^4}\frac{h^2 f_c^2}{f_s^2} = \frac{A^2}{2} \qquad \text{for } f_c = f_1$$

$$\text{ie. } hf_c = \frac{\pi^2}{2} A f_s \qquad \text{.....Eqn 5.2}$$

FIG 5.1 Theoretical Model for Slope Overload
- based on waveform observation



As an approximation, consider the reconstructed signal, r(t), as consisting of the sum of a triangular wave at the signal frequency plus a sawtooth wave at the clock frequency, as shown above.

i.e. $r(t) = f_1(t) + f_2(t)$

The amplitude of $f_1(t)$ is given by: $hT_s/4T_c = hf_c/4f_s$

Hence as an approximation, r(t) can be written as:

$$r(t) = \frac{2hf_c}{\pi^2 f_s}\left(\cos \omega_s t + \frac{\cos 3\omega_s t}{9} + \frac{\cos 5\omega_s t}{25} + \frac{\cos 7\omega_s t}{49} + \ldots\right)$$

$$+ \frac{h}{\pi}\left(\sin\omega_c t \pm \frac{\sin 2\omega_c t}{2} + \frac{\sin 3\omega_c t}{3} \pm \frac{\sin 4\omega_c t}{4} + \ldots\right)$$

This predicts an output signal before filtering of $\frac{2hf_c}{\pi^2 f_s}\cos\omega_s t$ with a power of $2h^2f_c^2/\pi 4f_s^2 = -16.88 + 20\log(hf_c/f_s)$ dB.

Applying the experimental modulator conditions of, h = 1 volt, f_s=800 Hz and a low pass filter loss of 0.17dB at 800 Hz. we get:

Output signal power , $S_o = -75.1 + 20\log f_c$ dB.

Similarly, after applying the low pass filter characteristic of $|G(j\omega)| = \frac{1}{\sqrt{1 + (\omega/\omega_m)^2}}$ to the expression for r(t) and subtracting the output signal , we get :

Mean square value of noise function , n(t)

$$= \frac{2h^2f_c^2}{\pi^4 f_s^2}\left\{\frac{1}{9^2\left[1 + (3f_s/f_m)^2\right]} + \frac{1}{25^2\left[1 + (5f_s/f_m)^2\right]} + \ldots\right\}$$

$$+ \frac{h^2}{2\pi^2}\left\{\frac{1}{1 + (f_c/f_m)^2} + \frac{1}{2^2\left[1 + (2f_c/f_m)^2\right]} + \ldots\right\}$$

Applying the experimental modulator conditions of h = 1 volt, f_s = 800 Hz and f_m = 4 kHz this expression reduces to :

$$\overline{n^2(t)} = 3.23\times10^{-10}f_c^2 + 0.0506\left(\frac{1}{1 + (f_c/f_m)^2} + \frac{1}{4\left\{1 + (2f_c/f_m)^2\right\}} + \ldots\right)$$

The noise power predicted from Fig. 5.1 is given by:-

$$N = 3.23 \times 10^{-10} f_c^2 + 0.0506 \left(\frac{1}{1+(f_c/f_m)^2} + \frac{1}{4(1+4(f_c/f_m)^2)} + \cdots \right)$$

..... Eqn. 5.3

being the sum of the signal frequency components and the clock frequency components respectively. The resulting curve for the overload noise performance with clock frequency is shown dotted in Fig. 5.2. From this figure it can be seen that in constrast to the signal performance, the predicted noise performance does not agree with the experimental results. Equation 5.3. predicts a noise minimum of -15.5 dB,with harmonics of the clock frequency being dominant for clock frequencies up to about 4kHz and harmonics of the signal frequency being dominant for clock frequencies above about 8kHz. It is apparent that some major source of overload noise is neglected by the theoretical model considered.

Upon reconsidering Fig. 5.1 it can be seen that the slope overload model was based upon the situation when the clock frequency is an even integral multiple of the signal frequency. Although interpolation for clock frequencies between these particular values is satisfactory for the description of the output signal, this is not so for the description of the output noise. Waveform observations on the C.R.O. indicate, that as well as the basic stepped triangular wave (as indicated by Fig. 5.1) there is an additional component with a frequency which is less than the frequency of the fundamental of the triangular wave (ie. $< f_s$). This additional component is illustrated in Fig. 5.3 and results from changes in the logic levels of the triangular wave by one unit. It can be regarded as a change by one step height in the d.c. level of the reconstructed wave and is due to f_c not being, in general, an exact even multiple of f_s. This prevents the establishment of a regular pattern of the nature considered for Fig. 5.1. The rate of the d.c. variation is lower as f_c tends to nf_s (n even) and it varies from zero when $f_c = nf_s$ to a rate of $f_s/2$ when $f_c = mf_s$, m odd. It was observed that as well as a two level rectangular wave, the superimposed low frequency noise may also take the form of a 3 level wave as indicated in Fig. 5.3. Under certain conditions 4 or 5 levels were observed in the rectangular wave, apparently resulting from two different low frequency rectangular waves being superimposed. In these cases one of the superimposed low frequency rectangular waves was observed to have a particularly low frequency (of the order of $f_s/100$).

The effect of the low frequency component of the noise during slope overload can be taken into account by considering a two level only rectangular wave of sufficeintly low frequency to allow it to

FIG 5.2 Overload Noise Performance

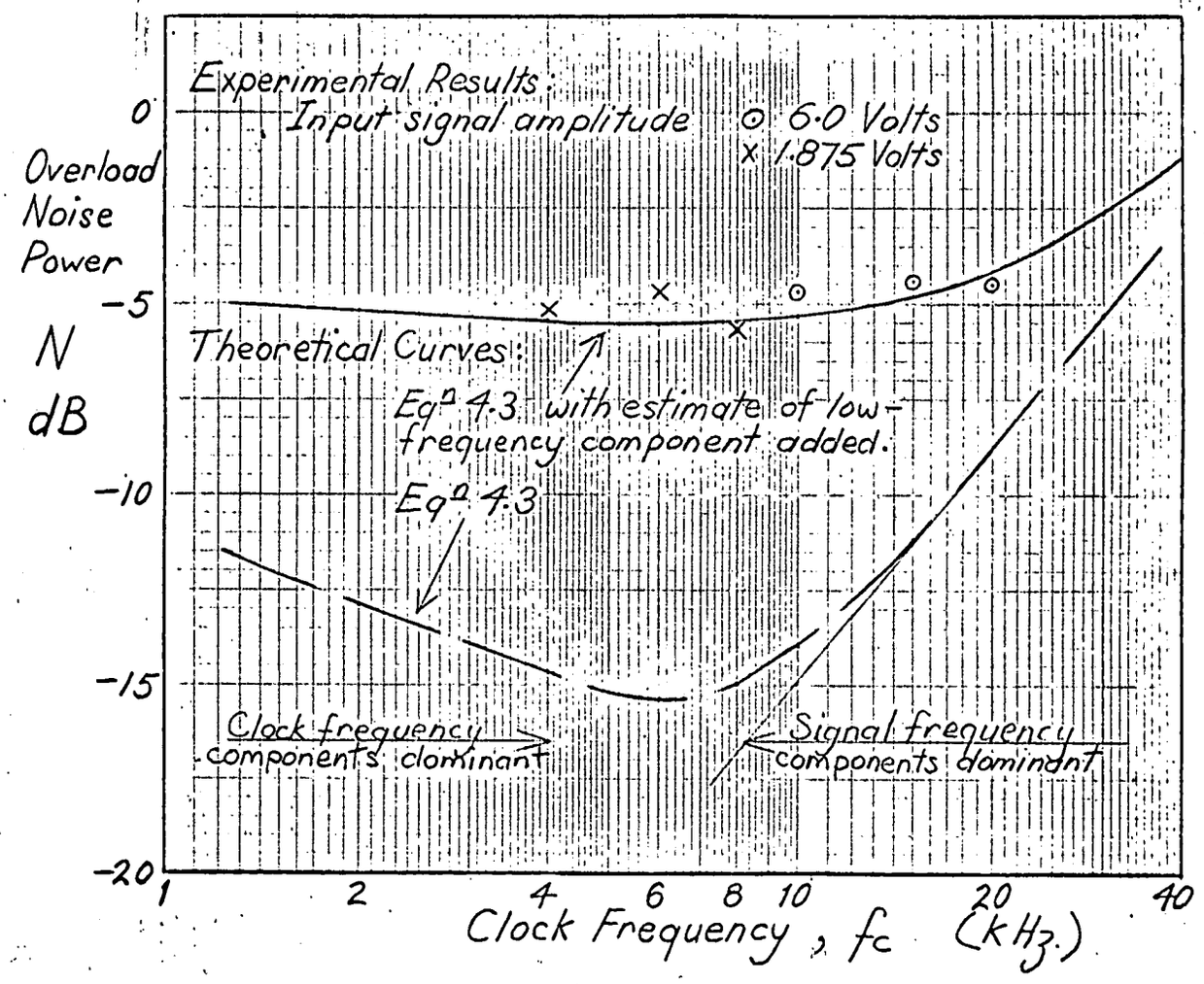
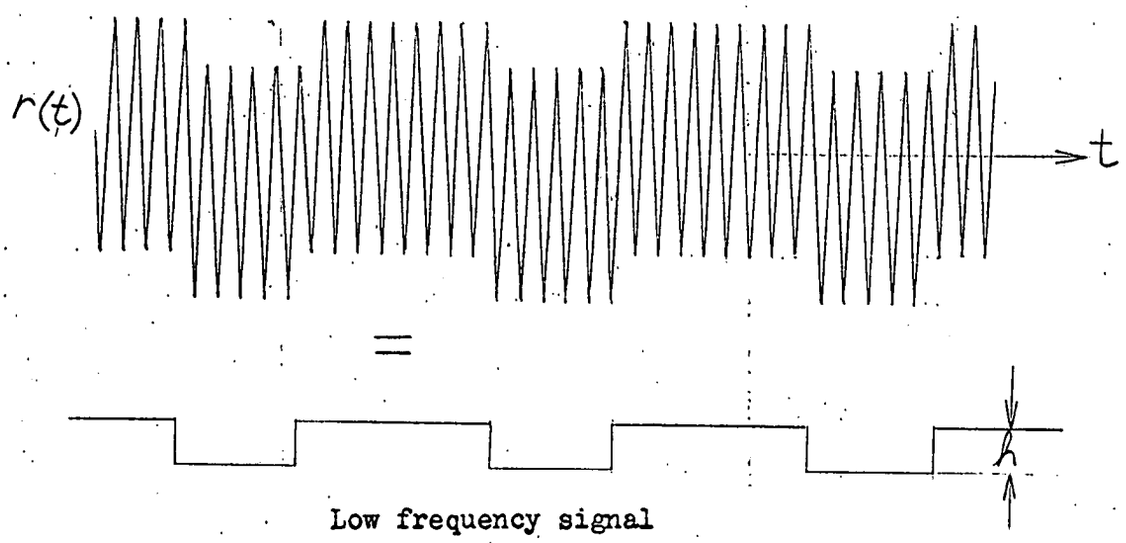


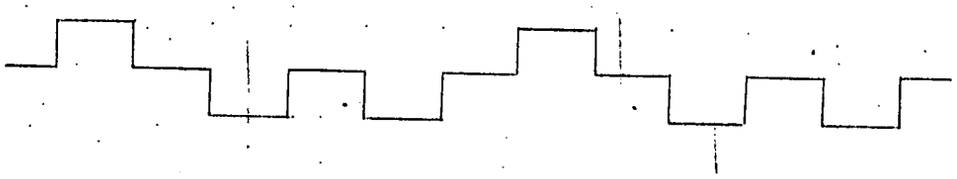
FIG 5.3 Observed Slope Overload Reconstructed Signal



Low frequency signal

+

Clock frequency signal , $f_1(t)$



Another form of superimposed low frequency signal observed.

pass through the lowpass filter without significant attenuation. This would result in an additional contribution to the overload noise of $h^2/4$. When added to Equation 5.3 this results in the overload noise power performance as indicated by the solid line in Fig. 5.2. The resulting predicted overload noise performance can be seen to show good agreement with the experimental results. Any additional noise due to the presence of 3 or more levels in the d.c. variation of the overload reconstructed signal does not appear to be present in the experimental results. This would be so for two reasons:-

(i) The infrequent appearance of more than two levels in the d.c. variation, and

(ii) The low frequency cut-off of the "true rms value" voltmeter used for the measurements, which would eliminate the lower frequency components resulting from a 3 or more level d.c. variation in the reconstructed signal.

The foregoing slope overload noise considerations lead to the following conclusions:

(i) The major contribution to slope overload noise is from low frequency components (from $f_s/2$ down) which can be described as a low frequency variation in the d.c. level of the reconstructed signal.

(ii) The mean square value of the low frequency component is given by $h^2/4$. This represents about 80% of the noise power maximum; which was empirically determined and theoretically confirmed at -5dB ($=h^2/3.16$).

(iii) Significant overload noise power reduction could be expected by the use of a band pass instead of a low pass filter for the elimination of unwanted noise from the output signal. This third conclusion was not experimentally confirmed and would be a good point for further investigation.

5.3 A Comparison of Experimental and Computer Simulation Results

Fig. 5.4 shows both the experimental modulator and computer simulation noise power results for various clock frequencies with an input signal amplitude of 6.0 volts. Also shown are the resulting estimated performance curves. Good agreement between the results is exhibited over all clock frequency regions except the slope overload region. It should be noted that the computer simulation results will be subject to the same bias effect as the experimental results, due to the fluctuations with varying input amplitudes caused by the deterministic nature of the input signal (as discussed in Section 4.6.2).

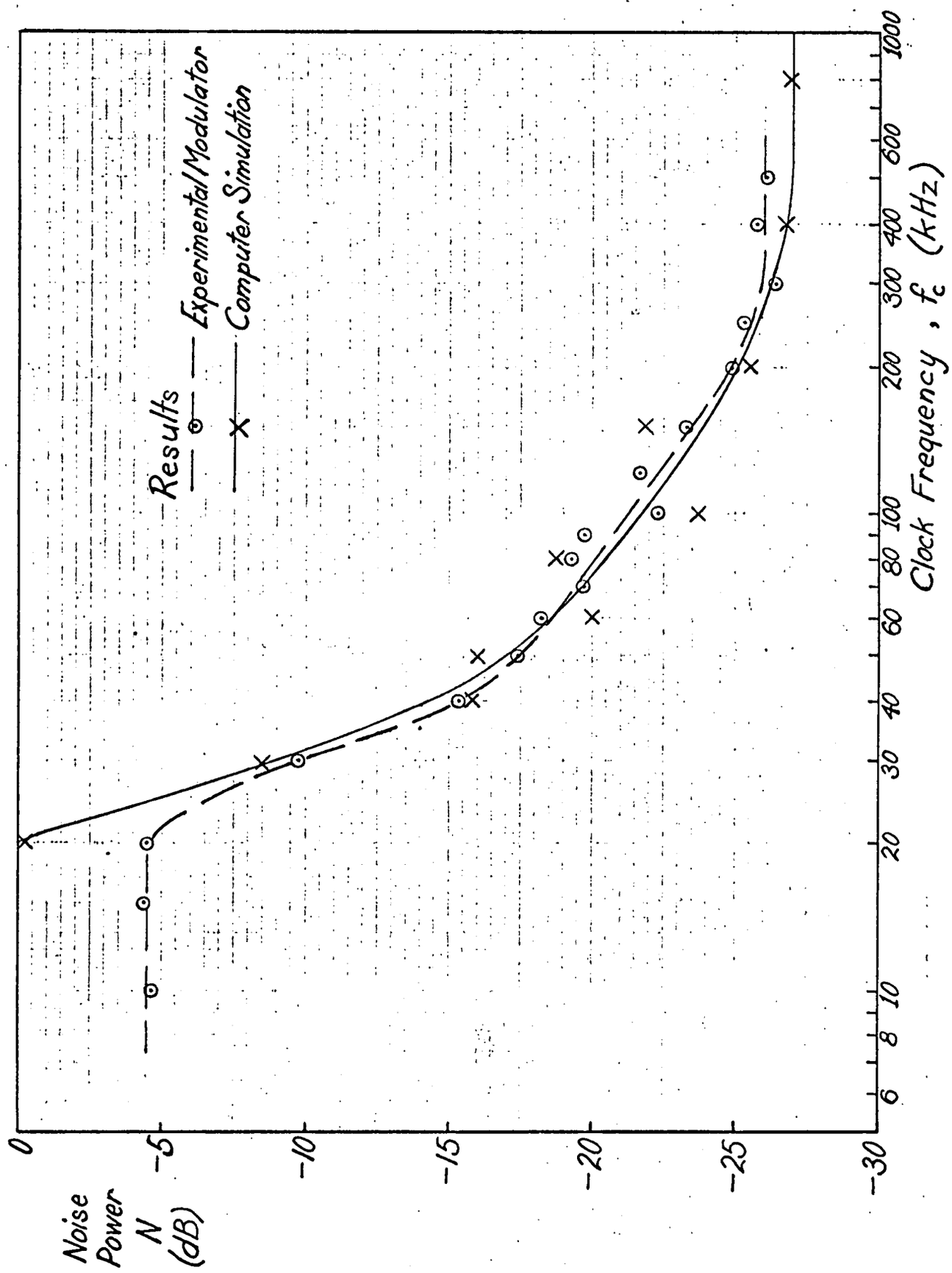
Two main differences exist for the noise measurements between the computer simulation and the experimental modulator. These are:-

(i) For the computer simulation the noise power component

Fig. 5.4. A Comparison of Experimental and Computer Simulation Results

Input signal: 800Hz sinusoid with an amplitude of 6.0 units.

Step height, h = 1 unit.



at the signal frequency was obtained by subtracting the input from the reconstructed signal with the phase of the input varied to give an output noise power minimum.

ie. Output noise function is given by (refer Section 3.3):-

$n_1(t) = y(t) - f(t-t_o)$; where t_o is such that $n_1(t)$ has minimum power.

For the experimental modulator results the noise power was obtained by subtracting the input from the output signal with both the phase and amplitude of the input varied to give an output noise power minimum.

ie. Output noise function is given by (refer Section 4.5.1) the uncorrelated noise function:

$n(t) = y(t) - kf(t-t_o)$; where k and t_o are such that n(t) has minimum power.

(ii) For the experimental modulator the output was obtained by passing the reconstructed signal through a simple RC low pass filter with a cut-off frequency of 4kHz. (Refer Fig. 4.7). For the computer simulation an ideal low-pass filter with a 4kHz cut off was considered.

The effect of (i), the different noise function definitions will only be significant where the contribution of the signal frequency component to the results is significant. Under such circumstances the computer simulation results would indicate a greater noise power than the experimental results. The simulation results indicated that the noise component at the signal frequency was significant, particularly for the system operating in slope overload and the high clock frequency region. The uncorrelated noise definition used for the experimental results will reduce the simulation noise power curve of Fig. 5.4 by about 3dB at slope overload (i.e. f_c = 20kHz); by less than ½ dB over the partial slope overload region and most of the operating region; and by about 2½ dB in the high clock frequency region.

The effect of the use of the simple RC low pass filter with the simulation was determined from the simulation results to be most significant for clock frequencies from about 100 kHz and above. A large increase would result for the higher frequencies (about 3½ dB at f_c = 400 kHz) due to the significant noise power components above the 4kHz cut-off frequency. If both of the different noise power measurement conditions of (i) and (ii) above were taken into account with the simulation results they would account for the difference from the experimental results in the high clock frequency region and for most (about 3.2 dB) of the difference between the overload (f_c = 20kHz) results.

The uncorrelated noise function, n(t) , used for the experimental modulator noise power readings,would provide the most useful definition of output noise desired in considering modulator performance with some typical input signal. However the use of this noise

Fig. 5.5. A Comparison of Performance Curves Estimated from Measurements with Theoretical Curves.

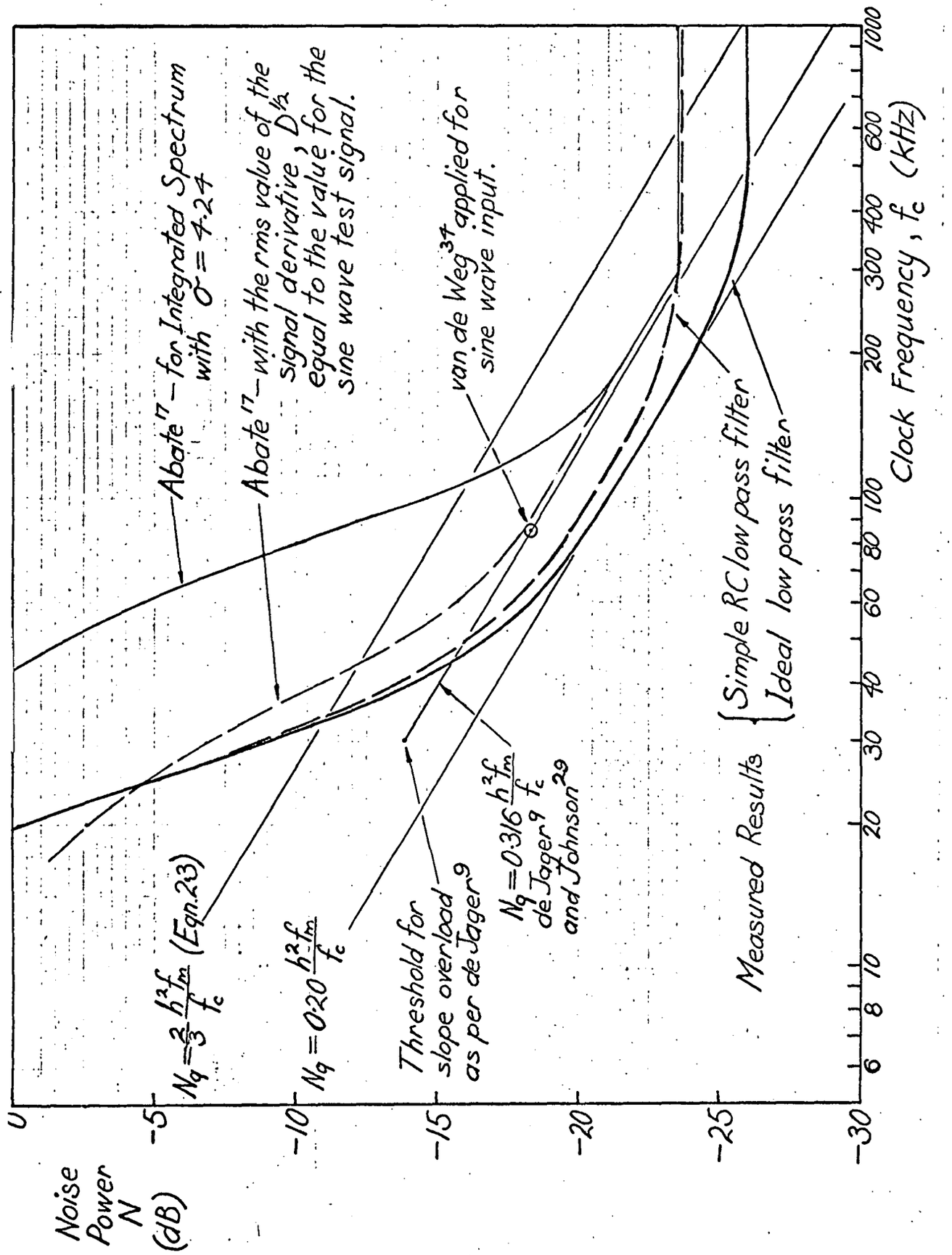
Input signal for Measurements:- 800Hz sinusoid with;

amplitude = 6.0 units

ie. σ = 4.24 units rms.

Step height, h = 1 unit.

Measured noise defined by, $n(t) = y(t) - f(t-t_o)$; where t_o is such that n(t) has minimum power.



definition for a sinusoidal signal with the intention of indicating the performance characteristic of a more general signal would be unsatisfactory as it would result in a zero noise component at the signal frequency. Therefore the flat noise power curve from the experimental results for the slope overload region would be indicative of the performance for a sinusoidal input only. The computer simulation curve in the slope overload region would more accurately describe the output noise power performance for a general input represented by the sinusoid.

The estimated noise power performance for an ideal low pass filter with a 6 volt, 800 Hz sinusoid representing a general input (eg. speech) would be as indicated in Fig. 5.5. This is based on the computer simulation results, with the $n_1(t) = y(t) - f(t-t_o)$ definition for the noise and with correction for the bias of the results due to the deterministic nature of the input (as discussed in Section 4.6.2). The similarly based estimated performance with a 4kHz cut off, RC low pass filter was estimated from the computer simulation results and is also indicated in Fig. 5.5.

As discussed, the sharp cut-off between the slope overload and partial slope overload noise performance regions (IA and IB respectively) for the experimental results is unrealistic for a general signal represented by the sinusoidal input. Similarly the sharp cut-off exhibited by the output signal power at the same point would not represent the situation for a more general input.

5.4 A Comparison of Measured Results with Established Theory

The earliest analytical description of delta modulation performance is that given by de Jager[9]. His description of the noise performance (Eqn. 2.7, Section 2.4) for a sinusoidal input gives a pessimistic prediction over the mid clock frequency region (Region II), being about 2 dB above the curve estimated from the measured results (see Fig. 5.5). In addition, de Jager's description of the noise does not predict the minimum noise region (Region III) of performance at high clock frequencies, the partial slope overload region (Region IB) or the noise power variation with varying input signal amplitude, which were observed in the measurements. De Jager's analysis was restricted to the quantizing noise situation and would not be relevant below his predicted slope overload point. His predicted clock frequency for the onset of slope overload, as given by $D^{\frac{1}{2}} = f_c h/\sqrt{2}$, is about 30% higher than the value indicated by the measurements. (where $D = \overline{[f'(t)]^2}$, the mean square value of the signal derivative) This is shown in Fig. 4.17 where the clock frequency for the onset of slope overload was found to be more accurately determined from the slope overload considerations of

Section 5.2 (Eqn. 5.2). De Jager's analysis does predict the 10dB/decade fall in noise power with clock frequency indicated in Fig. 5.5 for the mid clock frequency region.

Johnson[29] predicts a quantising noise power expression for a sine wave input which is very close to de Jager's, $N_q = 0.316 \frac{f_m h^2}{f_c}$ expression. One of the assumptions made by Johnston was that the noise is given by the output less the input delayed by one clock period. It was found (Section 4.6.2) from the experimental results that the delay of the output signal behind the input for a sinusoidal input is 0.5 clock periods.

When the analysis of van de Weg[34] is applied for a sine wave input and ideal low pass filtering, an expression which is again very close to de Jager's results for the quantizing noise. The additional limitation that the value of the signal derivative equals 1/4 of the overload slope (i.e. $D^{\frac{1}{2}} = hf_c/4$) is also imposed.

The difference of about 2dB between the above theoretical predictions and the noise power performance derived from the measurements (refer Fig. 5.5) for the middle range of clock frequencies (Region II) would be mainly due to a different output noise definition. Both de Jager and Johnson determined the value for the constant in their N_q expressions empirically. They did not use the uncorrelated noise function definition of $n(t) = y(t) - f(t-t_o)$ (refer Section 5.3) considered by the author to be the most relevant for a sine wave input. Therefore the output noise functions they considered would contain a significnat component which was coherent with the input signal. A similar situation for the noise signal definition would result from the application of van de Weg's analysis for a sine wave input because the evaluation of the van de Weg expression for N_q (Eqn. 2.7) assumes a random noise input with a uniform band-limited spectrum.

It has been shown (refer Section 4.6.2 and Fig. 4.17) that the assumption that the noise is independant of the value of the input signal and is given by $N_q = k \frac{f_m h^2}{f_c}$ is valid, provided the level of loading of the system falls in the approximate range $0.4hf_c > D^{\frac{1}{2}} > 0.09hf_c$ (i.e. Region II of the perforamnce curves). From the measurements made, a value for k in the above N_q expression of 0.20 has been determined for the ideal low pass filter, sine wave input situation. This value would be more suitable than the established value of 0.316 for the conditions considered, where the uncorrelated noise power as considered by the author is of interest.

As discussed in Section 2.4, van de Weg's[34] quantizing noise

power expression, as extended by O'Neal[35] to remove the restriction of a specific level of loading, is accepted as one of the most precise quantizing noise power performance descriptions. This description (Fig. 2.9) is for band-limited gaussian input signals. O'Neal's overload noise power description is not confirmed by subsequent measurements and for this reason the sinple, closed form equations of Abate[17] (Eqns. 2.9, 2.10 and 2.11) are considered for comparison with the measured performance. Abate's description of the noise power performance is empirically determined, can be applied for random, band limited input signals with various spectra, and agrees well (within 1dB) with the O'Neal/van de Weg description in the quantizing noise region. Fig. 5.5 shows Abate's estimated performance curve for a random input signal, bandlimited to 4kHz, with an integrated spectrum and an rms value of 4.24 volts (i.e. the same rms value as the sine wave used for the measurements). This curve is determined from the reduced form of the equation given is Section 2.4 by Eqns 2.12, 2.13 and 2.14.

A comparison of Abate's curve with the comparable (i.e. ideal low pass filter) measured curve,indicates that both demonstrate the two quantizing noise states. (i.e. Regions II and III). Although this feature of the quantizing noise performance is inherent in van de Weg's general expression and the subsequent curves of O'Neal, it is not specifically predicted by any other analysis. In particular the analyses for a sine wave input (e.g. de Jager[9] and Johnson[29]) predict no limit to the - 10 dB/decade, N_q reduction with increasing clock frequency. Qualitatively the description by Abate agrees with the curves obtained from measurements and with the general description of the noise performance regions from the noise analysis of Section 2.3. Each agrees in defining distinct regions of performance, with the exception that the partial slope overload region for sinusoidal inputs as proposed by the author is included in the one overload region by Abate.

Significant variation of the noise power values exists between Abate and the measurements, as indicated in Fig. 5.5. For the quantizing noise the difference (about 2½dB) would result mainly from the uncorrelated noise function definition used for the measurements as discussed earlier in this section. In addition, Abate's analysis, being for a random, wide band signal (up to f_m), necessarily predicts the onset of the high clock frequency region (Region III) at a different level of loading. (i.e. $D^{\frac{1}{2}} = hf_c/8$ for Abate cf. $D^{\frac{1}{2}} = hf_c/14$ for the curve from the measurements in Fig. 5.4). The fundamental difference between the type of signal considered by Abate and the sinusoidal signal used for the measurements is the reason for the much larger overload noise contribution and the higher clock frequency (or lower rms input signal value) for the onset of overload, as estimated by Abate. This can be seen by comparing the rms value of the signal derivative, $D^{\frac{1}{2}}$, for the two types of signal as below:

Sine wave used for measurements: $D^{\frac{1}{2}} = 2\pi f_s \sigma = 5{,}030\,\sigma$

Integrated spectrum considered by Abate: $D^{\frac{1}{2}} = \sqrt{0.125}\ 2\pi f_m \sigma = 8,900\sigma$ where σ = rms value of the input. Therefore, although the signal rms values are equal in Fig. 5.5, the signal derivative rms value for the integrated spectrum is about 80% greater than that used for the measurements.

Applying the value of $D^{\frac{1}{2}}$ used for the measured curve of Fig. 5.5 to Abate's expression for the overload noise for an integrated spectrum, an equivalent rms signal value of 2.4 volts (cf. 4.24 volts) is obtained and the resulting overload performance estimate is shown by the dashed curve. This curve provides a far better estimate of the measured curve (within 3dB in general) and predicts the onset of slope overload at about the same level of loading (i.e. at about the same value of f_c).

5.5 The Value of a Sinusoid as a Test Signal

It has been shown that the use of a sine wave input has value in indicating the qualitative features of the noise performance of delta modulation for more general types of input signal. However even for this purpose there is a limitation imposed by the deterministic nature of the input which causes an additional unwanted variation in the results, depending on the input amplitude relative to the closest integral multiple of the step height. The variation was discussed and taken into account in estimating performance in Section 4.6. This problem with the sine wave as a test signal has also been mentioned by Bennett[5], de Jager[9], Kikkert[39] and others. The alternative method of eliminating the fluctuations by the addition of a small low frequency component to the test signal (refer Section 3.2) has been found to be satisfactory[9,24]. The definition of the output noise function can also have a significant effect on the basic characteristics of the noise performance curves(as illustrated in Section 5.3)when a sine wave test signal is used.

The above problems can be overcome, as has been discussed in the appropriate sections, to allow a sine wave test signal to be used to indicate the nature, and the factors affecting the noise performance characteristics of delta modulation for a more general input. A much more significant limitation exists for the measurement of the quantitative performance of delta modulation when a sine wave input signal is used. A representative sine wave signal gives noise power results which are up to about 3dB lower than random bandlimited signals with the same rms value, over the quantizing noise regions. The difference depends on the definition of the noise signal. This represents reasonable agreement and suitability of a sine wave as a test signal for more general inputs,for quantizing noise The main failure of the sine wave as a representative signal is that it

gives very optimistic values for the onset of slope overloading, and gives sharper increases in noise power as slope overload increases, than occur with more general wide band signals. This was discussed in Section 5.4, and results from the radically different spectral density properties of the two types of signal which give rise to different rms values of the signal derivatives relative to the rms values of the signals. Therefore, since the overloading process for delta modulation is slope and not amplitude determined, a comparison of quantitative values for slope overload performance and the onset of slope overload is not possible for signals with substantialy different spectral density properties.

5.6 Conclusions on the Nature of the Noise

Measurements obtained from the computer simulation and the experimental modulator have indicated the following conclusions regarding the nature of the noise produced by a single integration delta modulator.

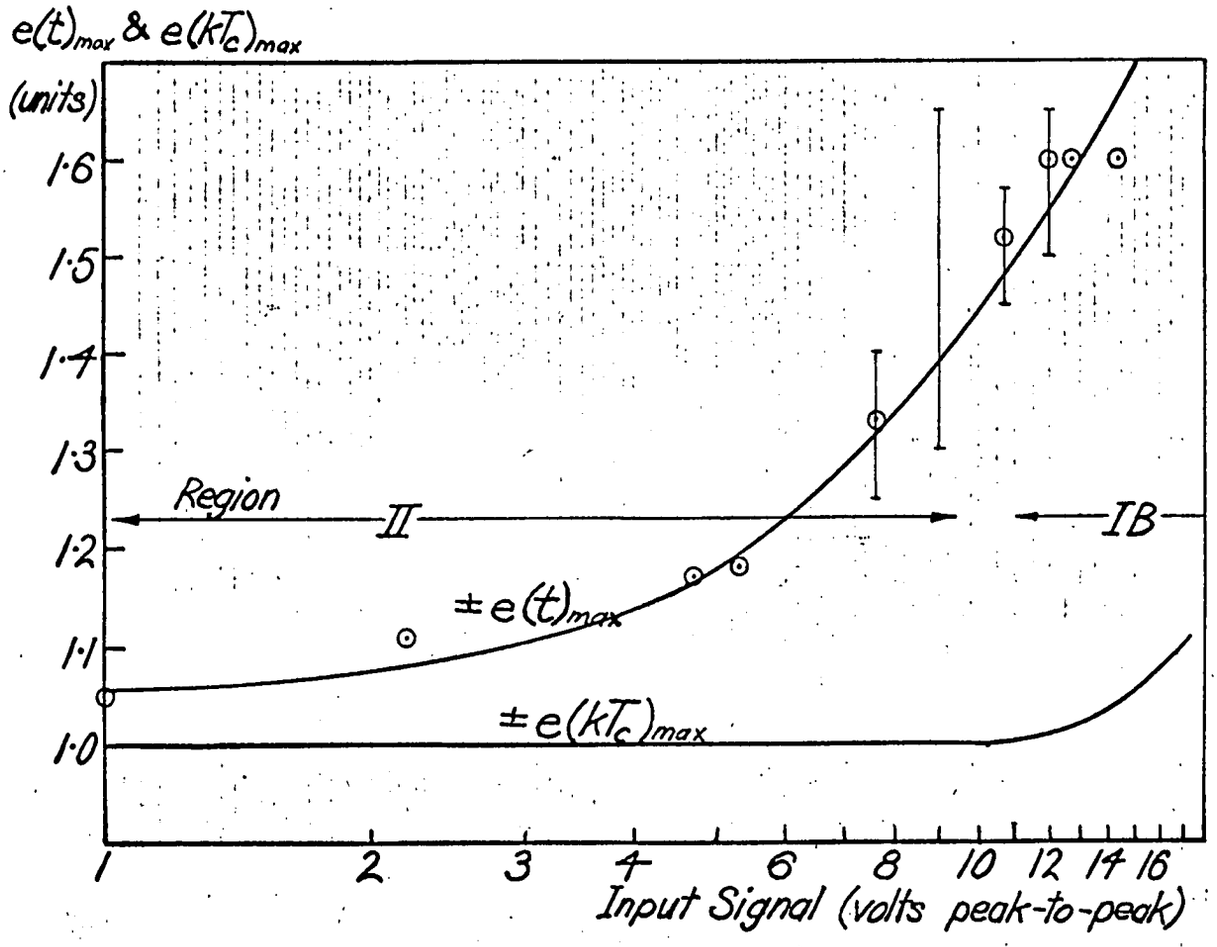
The power spectrum of the quantizing noise varies markedly from the $\text{sinc}^2(\omega T_c/2)$ description which was assumed by many previous investigators. It does however, exhibit a definite $\text{sinc}^2(\omega T_c/2)$ shape (or envelope characteristic) which results from the periodic, sample and hold, character of the output signal, r(t). (The $\text{sinc}^2(\omega T_c/2)$ quantizing noise power spectrum would be expected from a consideration of the autocorrelation function of the error function, e(t), which would have the triangular form $1-|\tau|/T_c$, if the value of e(t) at any instant was independent of the values in other clock intervals. ie. if the pulse signal to line, p(t), had a random polarity.

The power spectrum of the quantizing noise for the modulator operating in the middle region of loading (i.e. Region II) tends to be relatively uniform up to a frequency of roughly $f_c/10$ with a density of about $h^2/10\omega_c$ for a sine wave input. This density is about 2dB less than the value of about $h^2/6\omega_c$ which would be expected from a random band limited, simulated speech signal. It is approximately one third of the density which is indicated by the simplified model; that e(t) is uncorrelated between adjacent intervals.(which gave a density of $h^2/3\omega_c$ over the lower frequencies)

It has been shown that the reduction in the noise power density below the theoretical value of $h^2/3\omega_c$ over the lower frequencies relative to f_c, is compensated for by an increase above this value at other frequencies, provided the system is operating without slope overload,in the middle region of loading (ie. Region II). As a result, it was found that the quantizing noise power in the frequency range 0 to f_c is constant for all f_c (in Region II of operation) and is given by:

$$Nq = \frac{h^2}{3\omega_c}\int_{-\omega_c}^{\omega_c} \text{sinc}^2(\omega T_c/2)\,d\omega = 0.9\frac{h^2}{3} = 5.13\text{dB}$$

Fig. 5.6. Curves of Maximum Error Function and Discrete Error Function Values against Input Signal Value.
from observations of experimental modulator waveforms for f_c = 40kHz, input = 800Hz sine wave, and h = 1 unit.



The main change in the noise function associated with and contributing to the onset of slope overload (ie. in the partial slope overload region) has been shown to be a change in the shape of the power density spectrum of the noise function. In the partial slope overload region the uniform noise power density over the lower frequencies, increases towards the theoretical $h^2/3\omega_c$ value for a decreasing clock frequency and rapidly continues increasing to exceed this value. The change in the noise spectrum occurs without change in the total power. in the band 0 to f_c, but results from a shift in the noise power, from the high density region of the spectral distribution to the lower frequencies relative to f_c. It is not until the total slope overload region is approached that the total noise power in the band 0 to f_c increases. This change in the power density spectrum of the noise function associated with partial slope overload results from a decrease in the probability of a change in the polarity of the output pulse between successive pulses. It was indicated in Appendix A that as the probability of a change in the polarity decreases, a greater proportion of the power of the output pulse signal to line (and hence of the output signal, y(t)) falls within the lower frequency range.

The other factor contributing to partial slope overload was observed to be the variation in the probability density function of the error function, e(t). Fig. 5.6 shows the maximum range of error function values (and discrete error function values) for varying loading, as observed from the experimental modulator. The discrete error function, $e(kT_c)$ (ie. the function determined by the sampling of e(t) just after each clocking instant) has a maximum magnitude of h(= 1 unit) until the system has moved into the partial slope overload region of operation (ie. Region IB). With the onset of partial slope overload, $e(kT_c)_{max}$ increases beyond 1 unit in magnitude for an increasing input signal. Therefore the onset of partial slope overload modifies the probability density fucntion of $e(kT_c)$, which was assumed in the analysis of Section 2.3 be to uniform for $-h<e(kT_c)<+h$ and zero elsewhere. The analysis of Section 2.3 indicates that as the limit on the maximum value of $e(kT_c)$ extends beyond h, the noise power in a given low frequency band will increase as the square of the extended limit on $e(kT_c)_{max}$.

5.7 Optimum Operating Conditions

The optimum operating conditions for simple delta modulation can be investigated from two viewpoints.

1. For a given clock frequency and input signal spectrum, what is the input signal value (relative to the step height, h) which gives optimum performance?

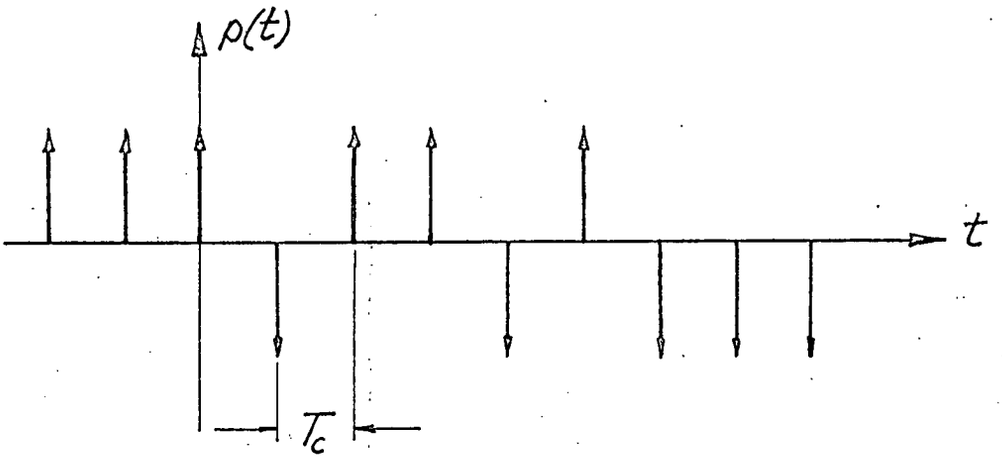
2. For a given range of input signal amplitudes and frequencies, what is the optimum clock frequency which provides for a required, minimum level of performance? ie. the minimum clock frequency (in order to have a minimum transmission bandwidth) which will provide a required SNR.

The answer to both these questions is determined by the regions of operation which give optimum conditions. For a given clock frequency, Figs. 2.9 and 4.11 indicate that the maximum SNR will be achieved for an input signal value which puts the system on the boundary of slope overload operation (ie. the boundary of regions IA and II). At this point the signal power is at its maximum before the noise power increases from its minimum (achieved in Region II) due to the onset of slope overload. For a broad dynamic range of inputs with maximum SNR the system should operate in the mid-frequency and partial slope overload regions (ie. Regions II and IB) Region III should be avoided as it is here that the noise power increases at 10dB/decade (as well as the signal power decreasing at 10dB/decade) with a falling input signal. Total slope overload (Region IA) should also be avoided as it is here that the rate of noise increase exceeds the rate of signal increase, resulting in rapid SNR degredation with increasing signal power.

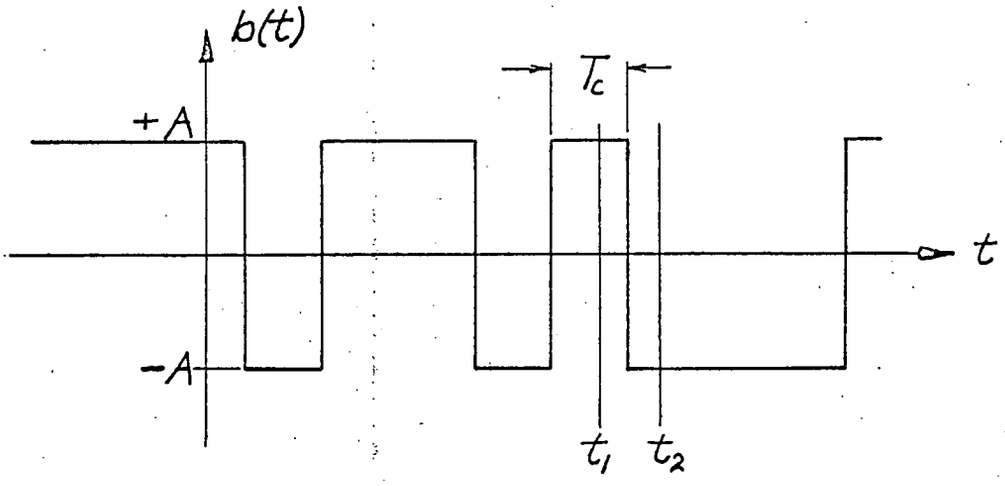
For a given range of input signals, Figs. 4.16 and 2.10 indicate that continual SNR improvement is available with increasing f_c, until the system is operating in Region III for the maximum input. From here the performance remains unimproved with increasing f_c. Therefore this point indicates the maximum required operating clock frequency. However, this maximum f_c could represent a waste of transmission bandwidth if the required maximum SNR is a lower value or if the 10dB/decade increase in SNR with f_c is an unwarrented use of bandwidth.

APPENDIX A

Consideration of the Power Spectrum, $P(\omega)$, of periodic train of ideal impulses, $p(t)$.



The periodic pulse signal shown above will result from the ideal sampling of the periodic binary signal, b(t), shown below.



Putting b_1 and b_2 as the possible values of b(t) at the times t_1 and t_2 respectively; where $t_2 > t_1$, $t_2 - t_1 = \tau$, and τ is the displacement variable of the correlation function.

Therefore the autocorrelation function of p(t) is given by;

$$R_b(t_1,t_2) = \overline{x_1(t)x_2(t)}$$

Assuming that the probability distribution of p(t), and hence of b(t), is invariant with time, then b(t) is a stationary process. Therefore the autocorrelation function for b(t) is given by:

$$R_b(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1 b_2 p(b_1,b_2)\,db_1\,db_2$$

For $|\tau| > T_c$ b_1 and b_2 are independent and therefore:

$$R_b(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1 b_2 p(b_1)p(b_2)\,db_1\,db_2 = \left[\overline{b(t)}\right]^2$$

For $|\tau| < T_c$,

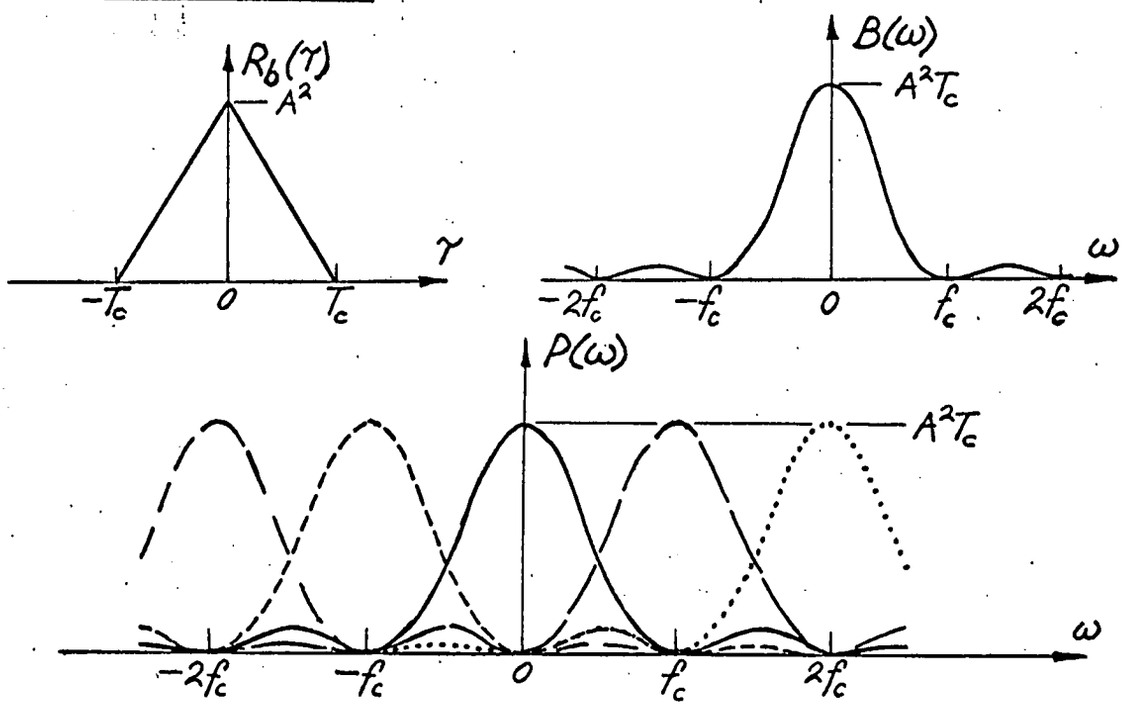
$$R_b(\tau) = A^2 P(b_1 = b_2) + (-A^2) P(b_1 \neq b_2)$$

$$= A^2 \left\{ P(t_1 \text{ and } t_2 \text{ are in same interval}) + P(t_1 \text{ and } t_2 \text{ are in different intervals}) \times P(\text{no value change between intervals}) \right\}$$

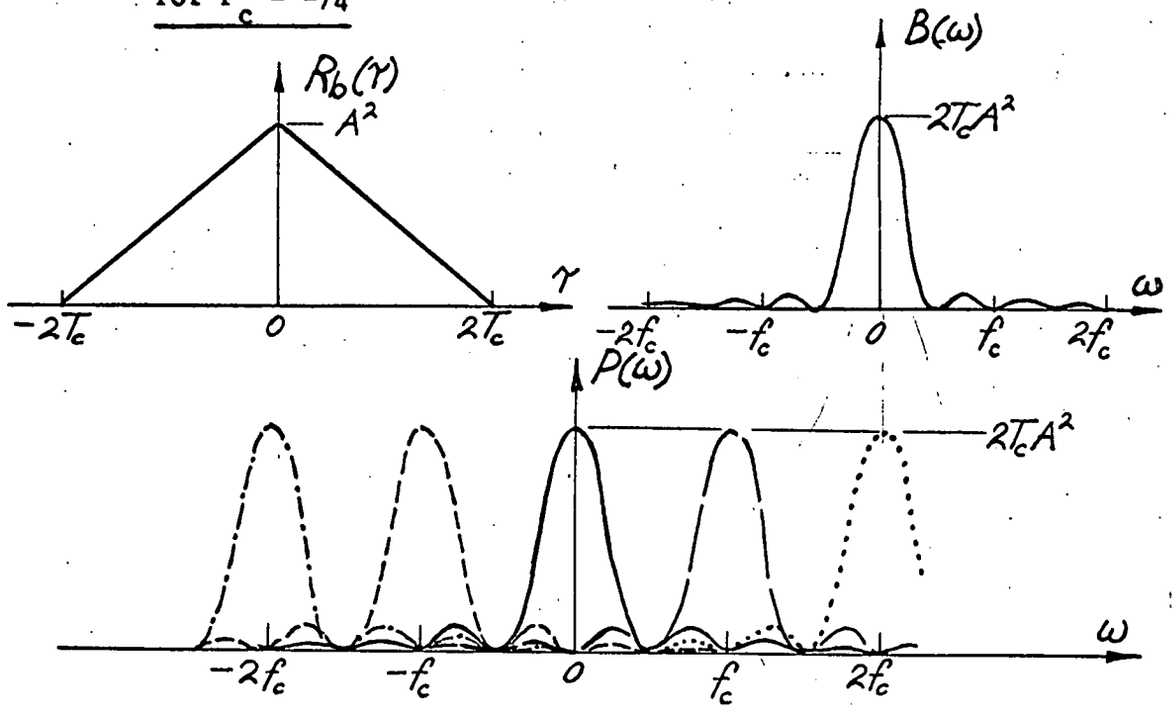
$$- A^2\, P(t_1 \text{ and } t_2 \text{ are in different intervals}) \times P(\text{value change between intervals})$$

Power Spectra of b(t) and p(t) for two values of P_c

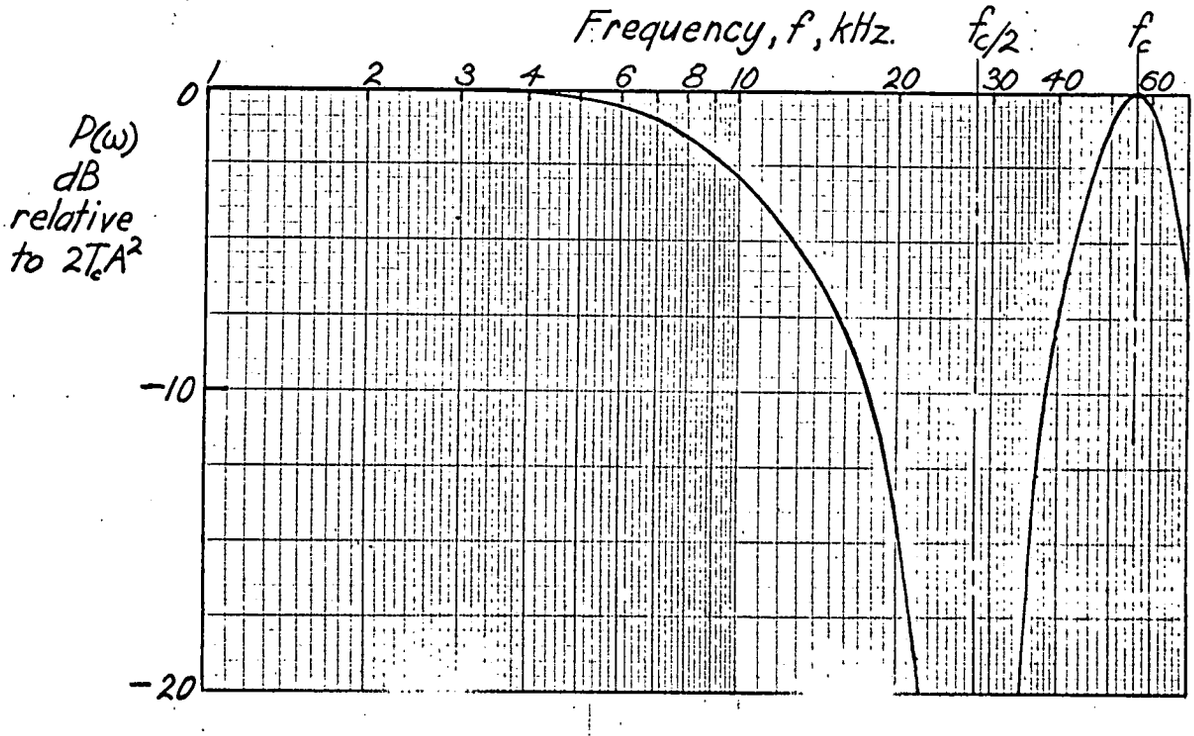
For $P_c = P_{nc} = \frac{1}{2}$



For $P_c = 1/4$



An enlargement of the power spectrum $P(\omega)$ in the range $0-f_c$



Putting Pnc = probabiltiy of no change in polarity between successive intervals.

and Pc = probability of a change in polarity between successive intervals.

Then $Pnc + Pc = 1$,

and $$R_b(\tau) = A^2\left\{(1 - \frac{|\tau|}{T_c}) + \frac{|\tau|}{T_c}\ Pnc - \frac{|\tau|}{T_c}\ Pc\right\}$$

$$= A^2\left\{1 - 2\,\frac{|\tau|}{T_c}\ Pc\right\}^c \qquad \ldots\ldots\ldots\ A1$$

If p(t) is a random binary pulse signal then $P_c = Pnc = \frac{1}{2}$, and $\overline{b(t)} = 0$ and therefore:

$$R_b(\tau) = \begin{cases} A^2(1 - \frac{|\tau|}{T_c}), & |\tau| < T_c \\ 0, & |\tau| > T_c \end{cases}$$

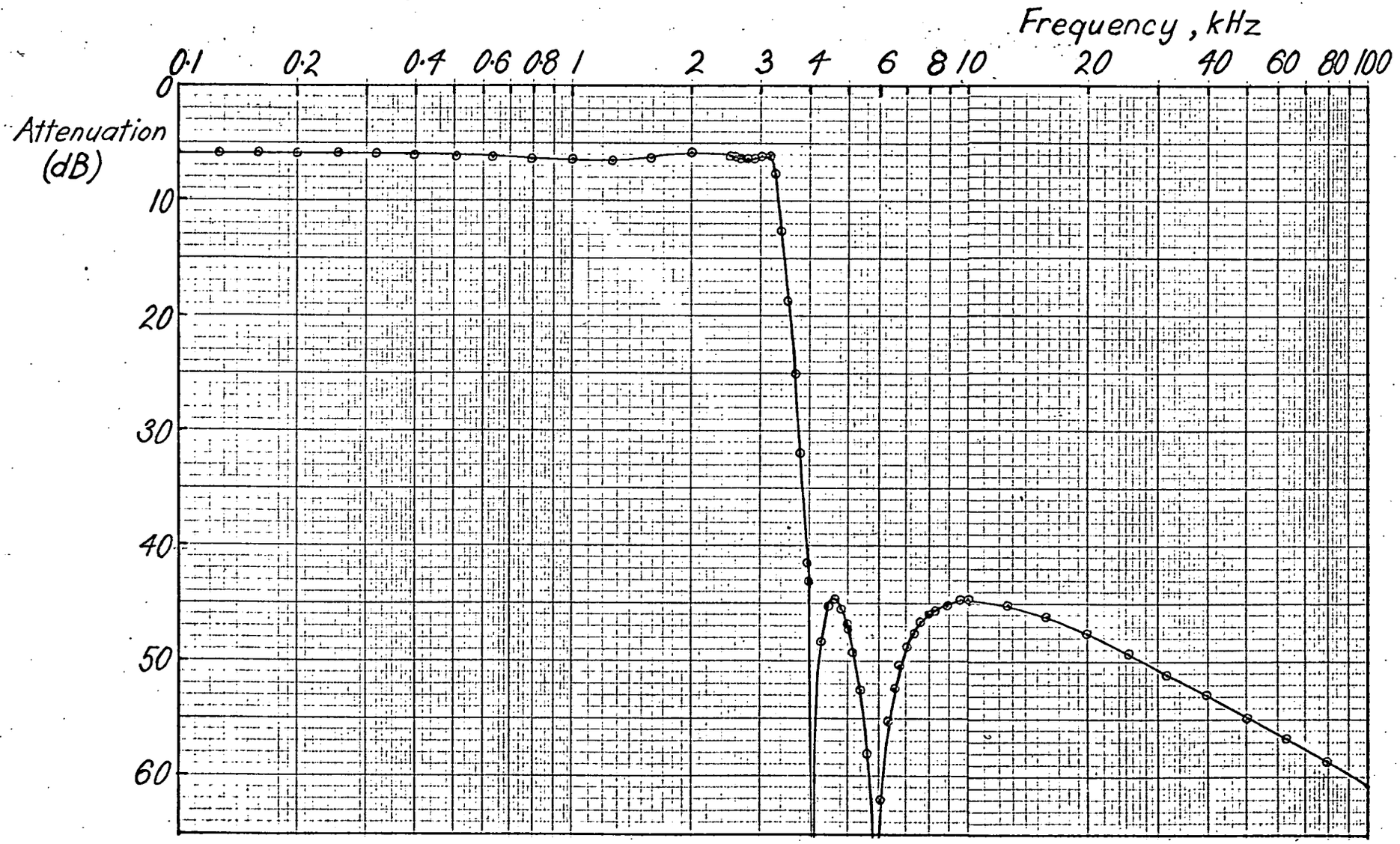
Therefore the power spectrum of b(t) is given by:

$$B(\omega) = A^2T_c\,\frac{\sin^2(\omega T_c/2)}{(\omega T_c/2)^2}$$

The power spectrum of p(t) is given by the convolution of $B(\omega)$ with the ideal sampling function spectum, $S(\omega)$, and is therefore flat over all frequencies.

ie. $P(\omega) = B(\omega) \circledast S(\omega) = A^2T_c$, as shown in the adjacent diagram.

Also shown in the diagram are the autocorrelation function and spectrum of b(t) and the resulting power spectrum of p(t) which would result if the probability of no change in the polarity of p(t) between successive pulses increased to 0.75 (ie. $P_c = 0.25$).

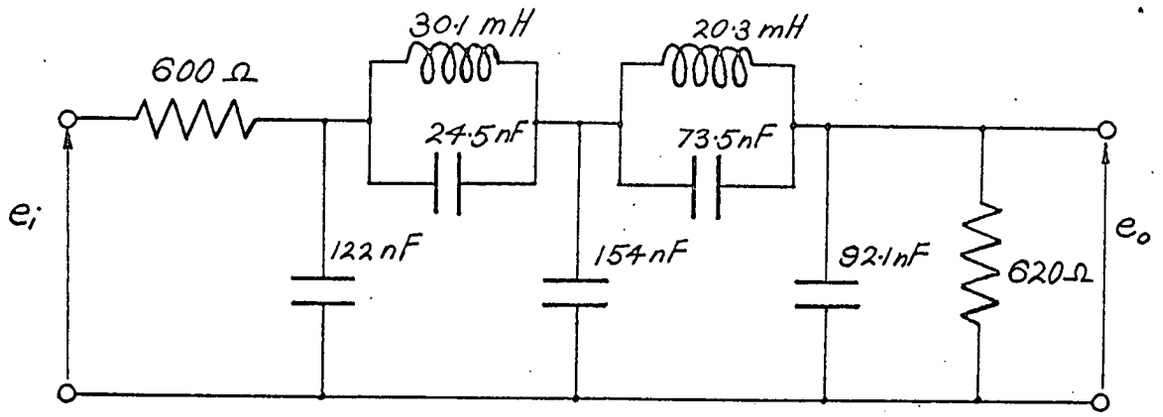


Audio Filter Frequency Response

APPENDIX B

The Design of a Digital Filter Simulation of an Output Low Pass Filter for Inclusion in the Computer Simulation of Delta Modulation.

The audio filter design of Fowler[40] (shown below) was considered to be suitable for output filtering where analysis of delta modulation performance for typical speech transmission is being considered.



The analysis of this audio filter was performed on an Elliott 503 computer using the network analysis programme of Brownell[41], developed at the Electrical Engineering Department of the University of Tasmania. The resulting frequency response is shown in the adjacent diagram and the Laplace transform transfer function is given by:

$$\frac{E_o(S)}{E_i(S)} = \frac{6.110\times10^{-4}S^4+1.238\times10^{6}S^2+5.553\times10^{14}}{1.063\times10^{-6}S^5+2.425\times10^{-2}S^4+8.844\times10^{2}S^3+1.257\times10^{7}S^2+1.678\times10^{11}S+1.093\times10^{15}}$$

The pole/zero evaluation programme of Brownell was used to give:

$$\frac{E_o(S)}{E_i(s)} = \frac{(S+2.589\times10^{4}j)(S-2.589\times10^{4}j)(S+3.682\times10^{4}j)(S-3.682\times10^{4}j)}{1.740\times10^{-3}(S+9.425\times10^{3})(S+5.419\times10^{3}-1.526\times10^{4}j)(S+5.419\times10^{3}+1.526\times10^{4}j)(S+1.273\times10^{3}-2.035\times10^{4}j)(S+1.273\times10^{3}+2.035\times10^{4}j)}$$

The "impulse invariant" method of digital filter design was adopted[42], this method being based on designing discrete responses to impulse which are the same as the sampled impulse response of the analogue filter. In order to obtain a frequency response similar to that of the analogue filter, the sampling frequency must be greater than twice the highest significant frequency of response of the analogue filter.

The Laplace transform function was simplified by partial fraction expansion using the residue method. The Z-transform[44] was then taken using a sampling frequency of 100kHz (ie. $T = 10^{-5}$ sec.)

This gives:

$$F(z) = \frac{E_o(z)}{E_i(z)} = F_1(z)+F_2(z)+F_3(z)+F_4(z)+F_5(z)$$

$$= \frac{a_1}{1-b_1z^{-1}} + \frac{a_2}{1-b_2z^{-1}} + \frac{a_3}{1-b_3z^{-1}} + \frac{a_4}{1-b_4z^{-1}} + \frac{a_5}{1-b_5z^{-1}}$$

where $a_1 = 5884.9$ $\quad b_1 = 0.91006$

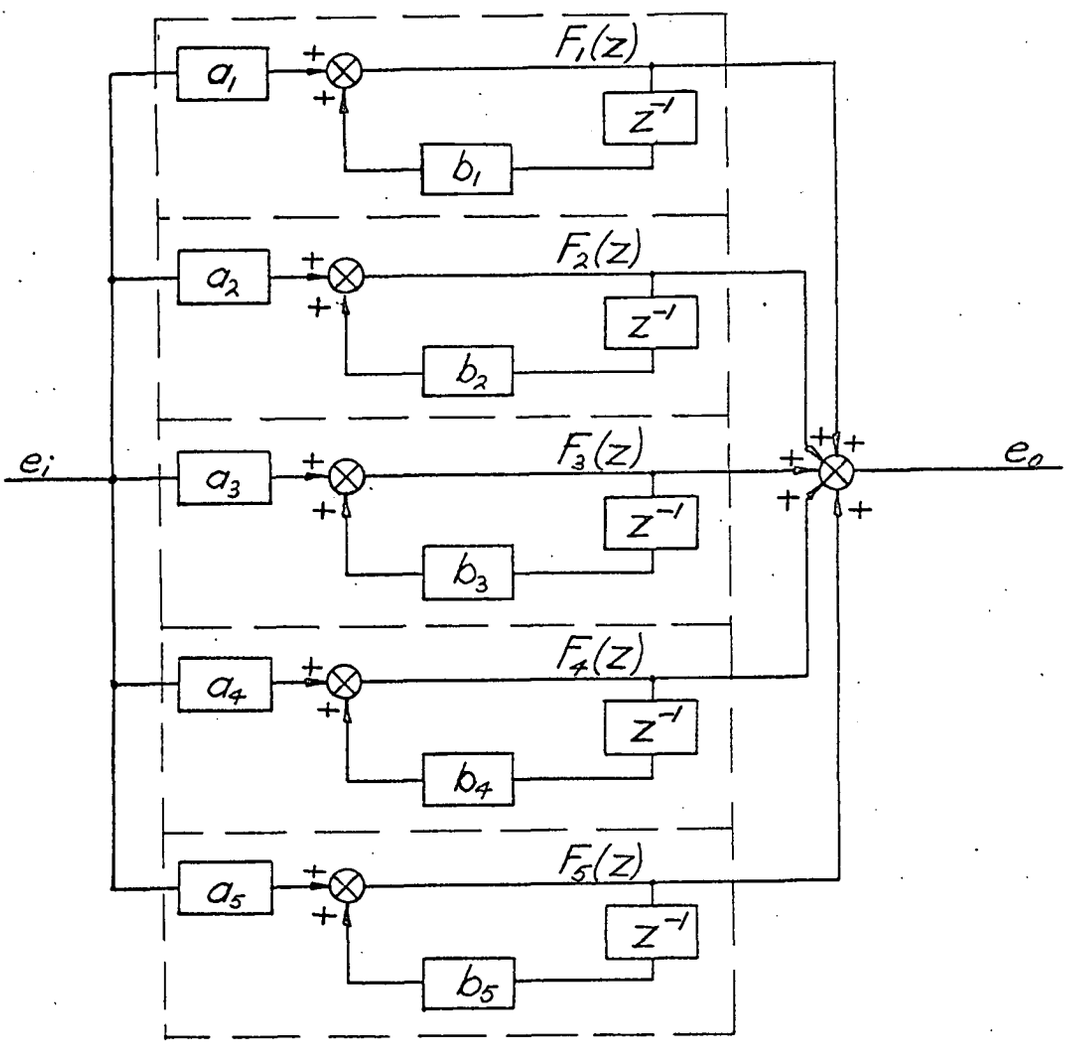
$a_2 = -2772.1 + 978.26j$ $\quad b_2 = 0.93624-0.14400j$

$a_3 = a_2^*$ $\quad b_3 = b_2^*$

$a_4 = 408.56-545.00j$ $\quad b_4 = 0.96698-0.19955j$

$a_5 = a_4$ $\quad b_5 = b_4^*$

This expression can be digitally implemented using the parallel form as shown in the following figure. The sub-filters, $F_i(z)$, are readily realized by the direct form,as the partial fraction expansion yielded terms which are all the ratio of zero to first order polynomials in z^{-1}.



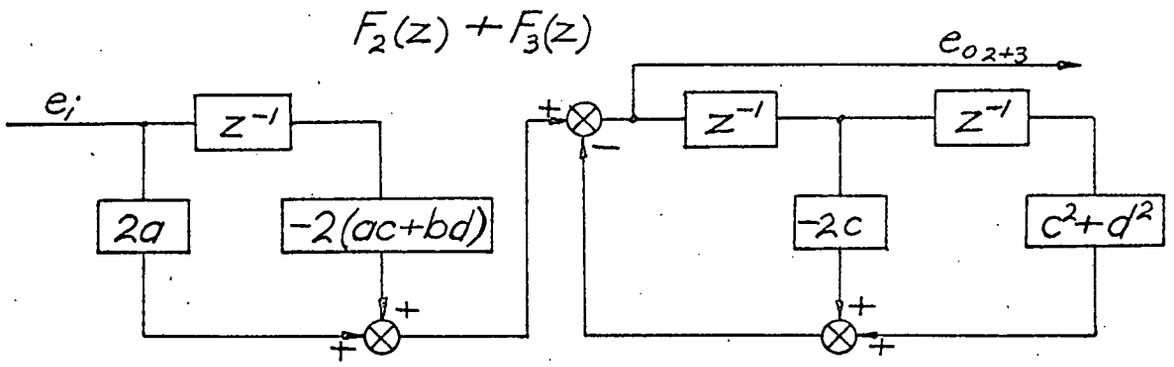
Radar[43] comments that the direct form of implementation for high order difference equations is undesirable for reasons of numerical accuracy and that the direct form is a lot more sensitive to quantization effects than the parallel form.

The first digital realization shown has complex coefficients, but as these occur in conjugate pairs the output will be real only. An alternative implementation without complex coefficients can be made as shown below.

Putting $F_2(z) + F_3(z) = \dfrac{a + jb}{1-(c+jd)z^{-1}} + \dfrac{a-jb}{1-(c-jd)z^{-1}}$

Then $F_2(z) + F_3(z) = \dfrac{E_{o2+3}(z)}{E_{i_{2+3}}(z)} = \dfrac{2a-2(ac+bd)z^{-1}}{1-2cz^{-1}+(c^2+d^2)z^{-2}}$

which can be realized as:



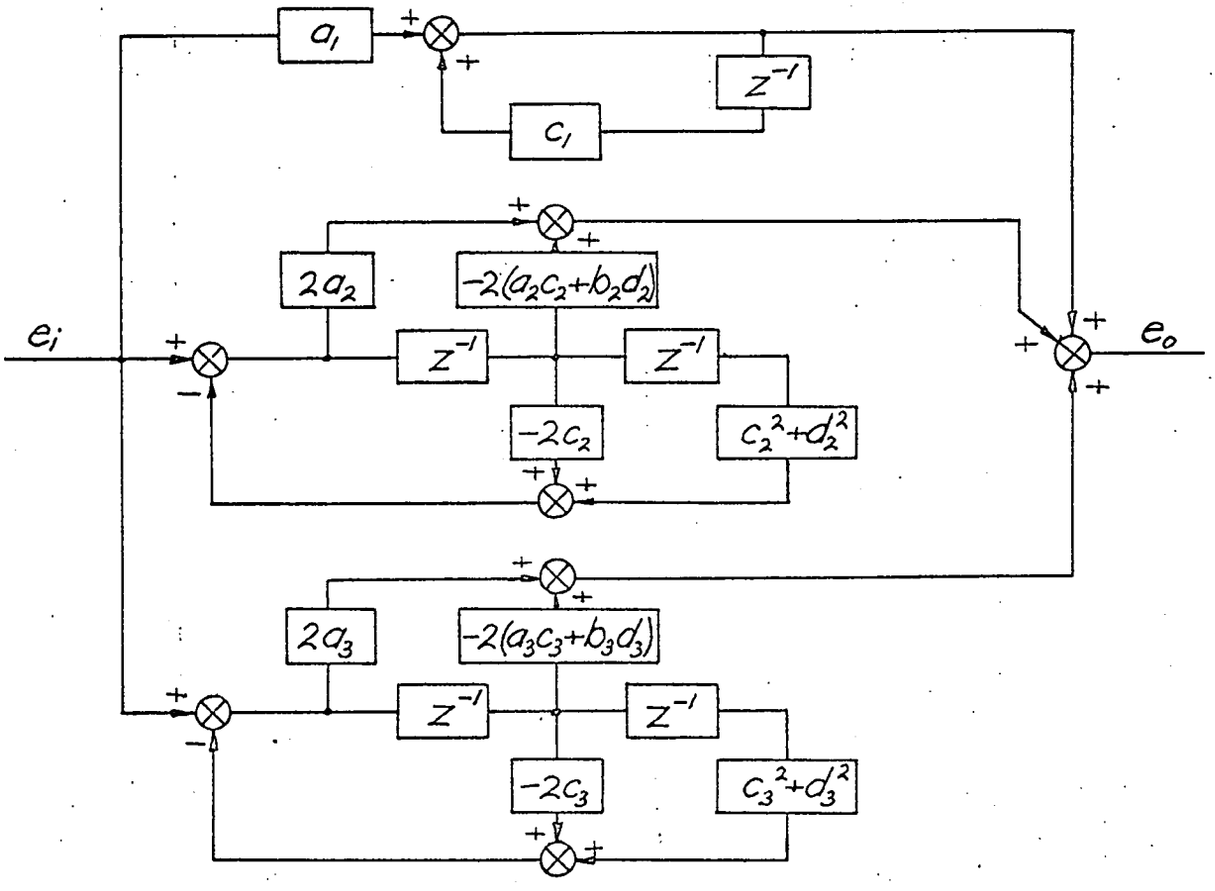
Such a representation has eliminated the need for programming the computer to handle complex numbers. However, it is in itself more complex. The above implementation can be simplified, giving an overall digital filter implementation as shown in the following figure, where:

$a_1 = 5884.9, \quad c_1 = 0.91006$

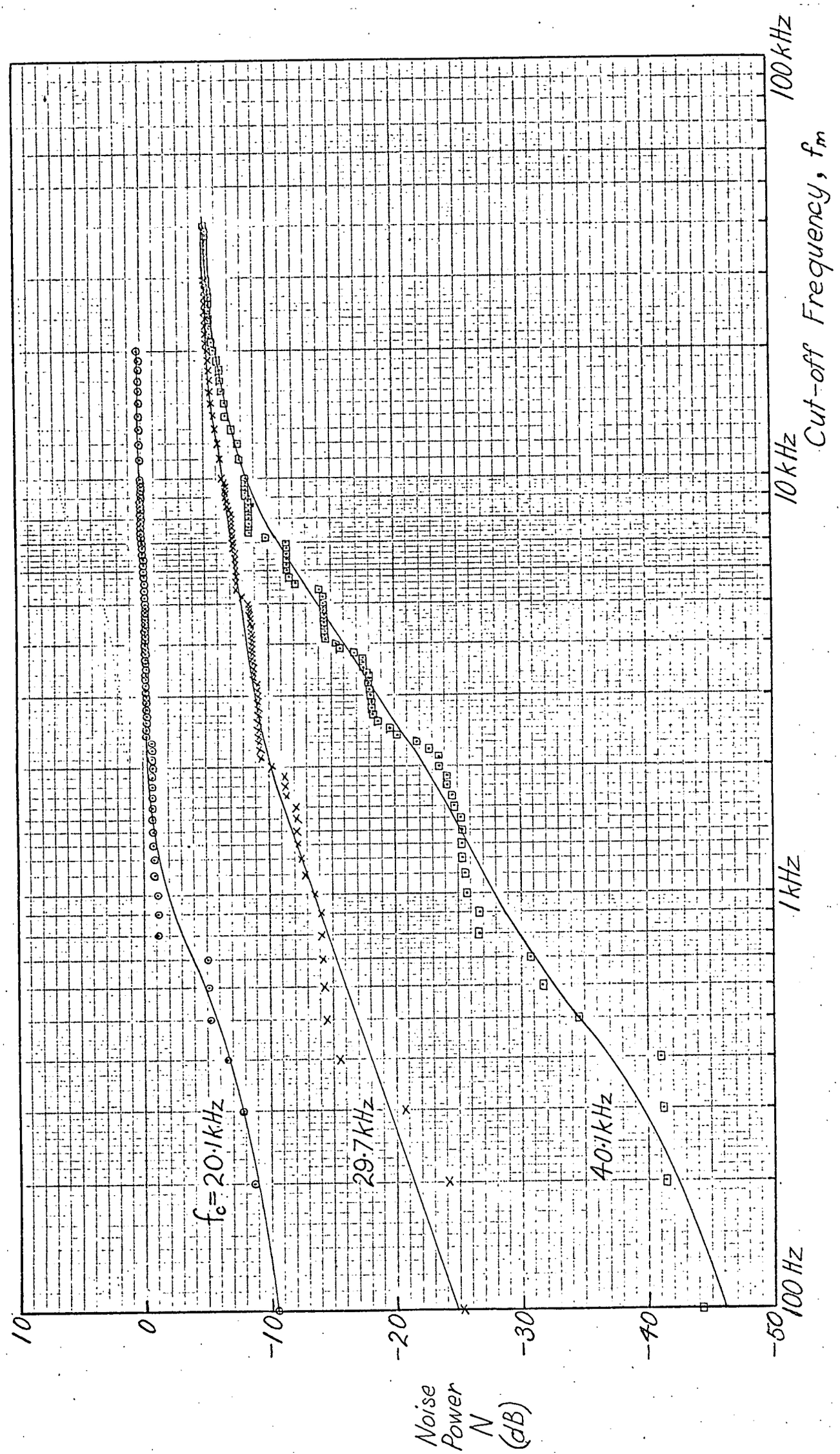
$a_2 = 2772.1, \; b_2 = 978.26, c_2 = 0.93624, \; d_2 = -0.14400$

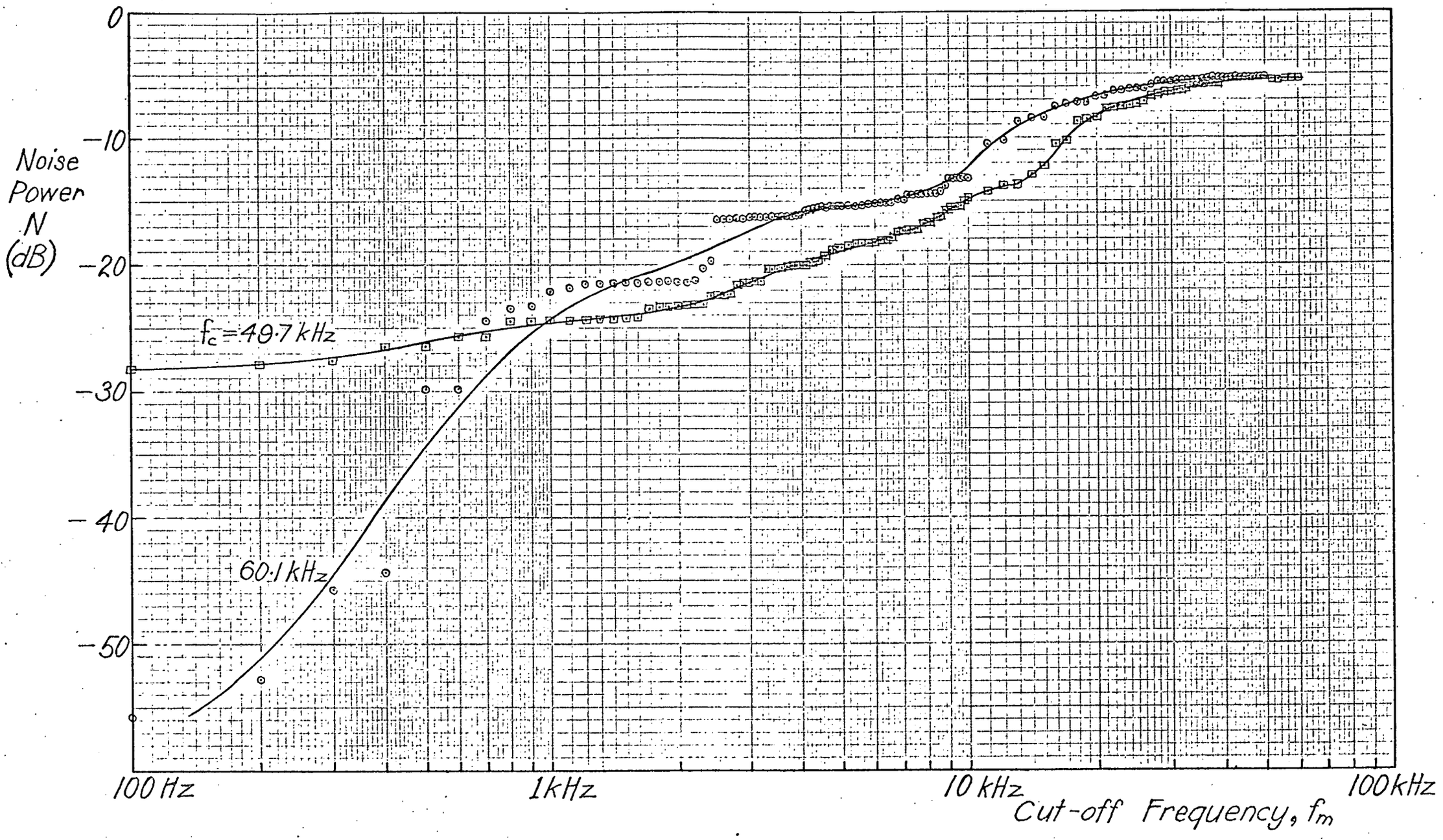
$a_3 = 408.56, \; b_3 = -545.00, c_3 = 0.96698, \; d_3 = 0.19955$

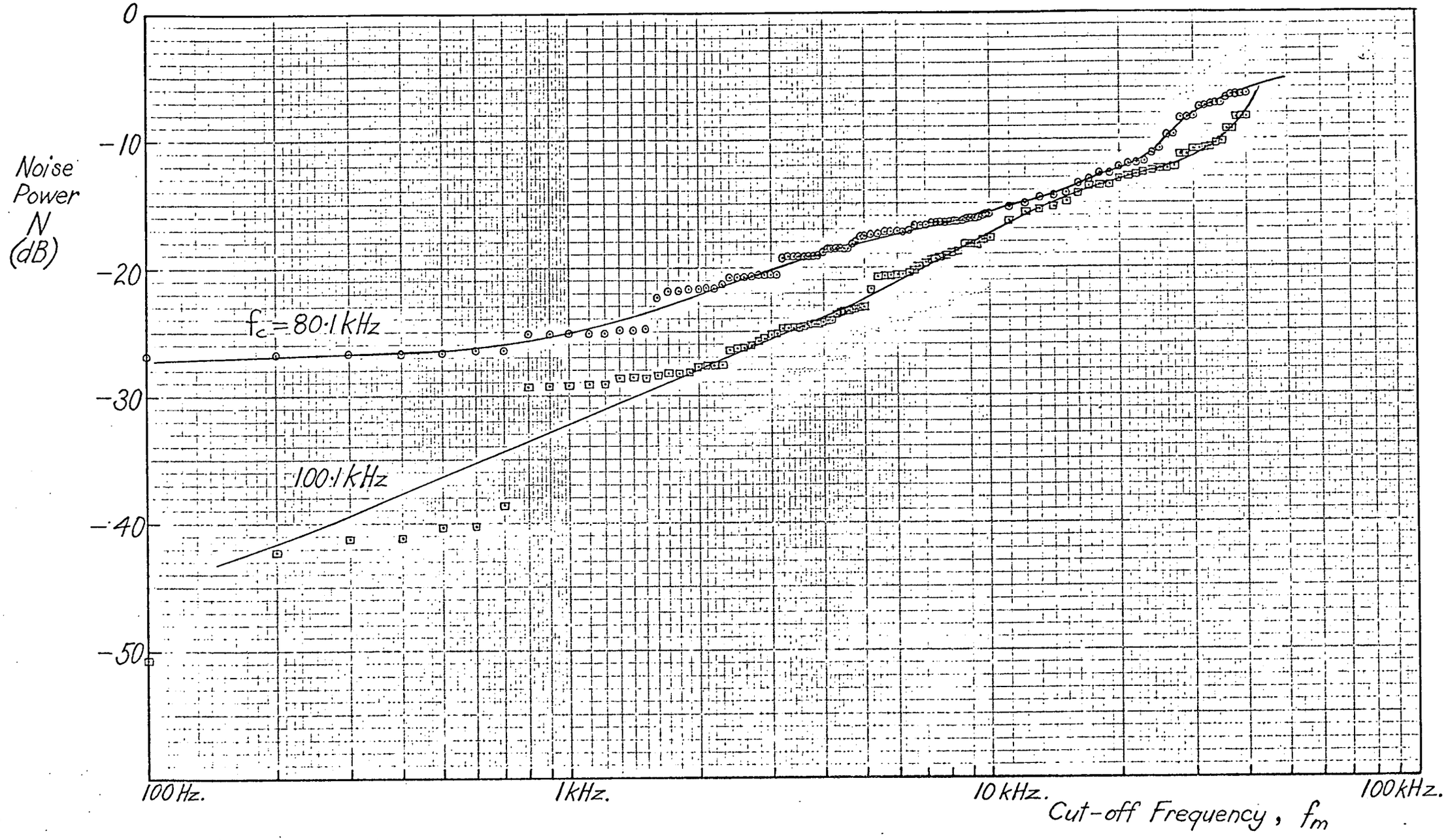
and z^{-1} = one sample delay = 10^{-5} sec.

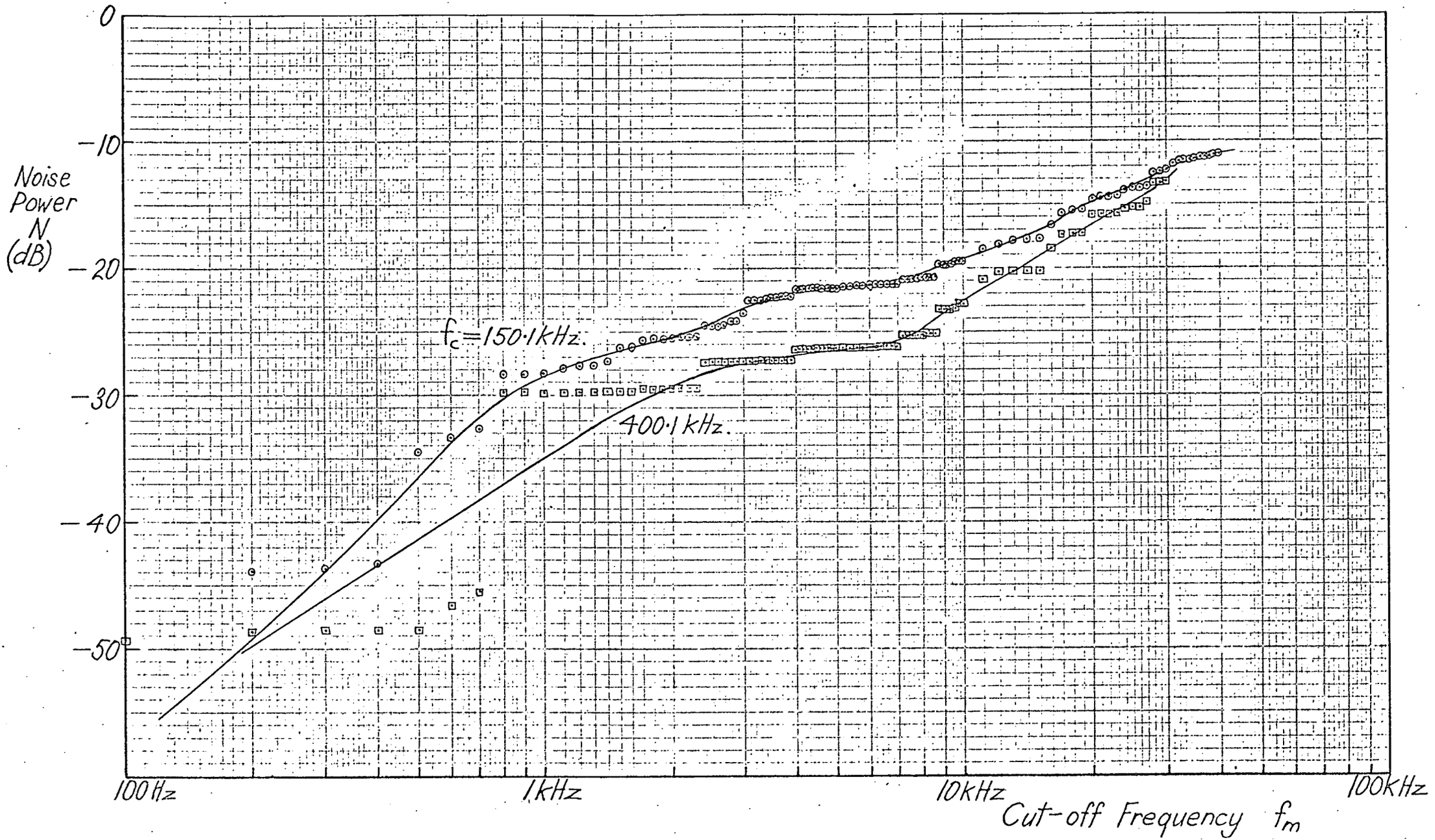


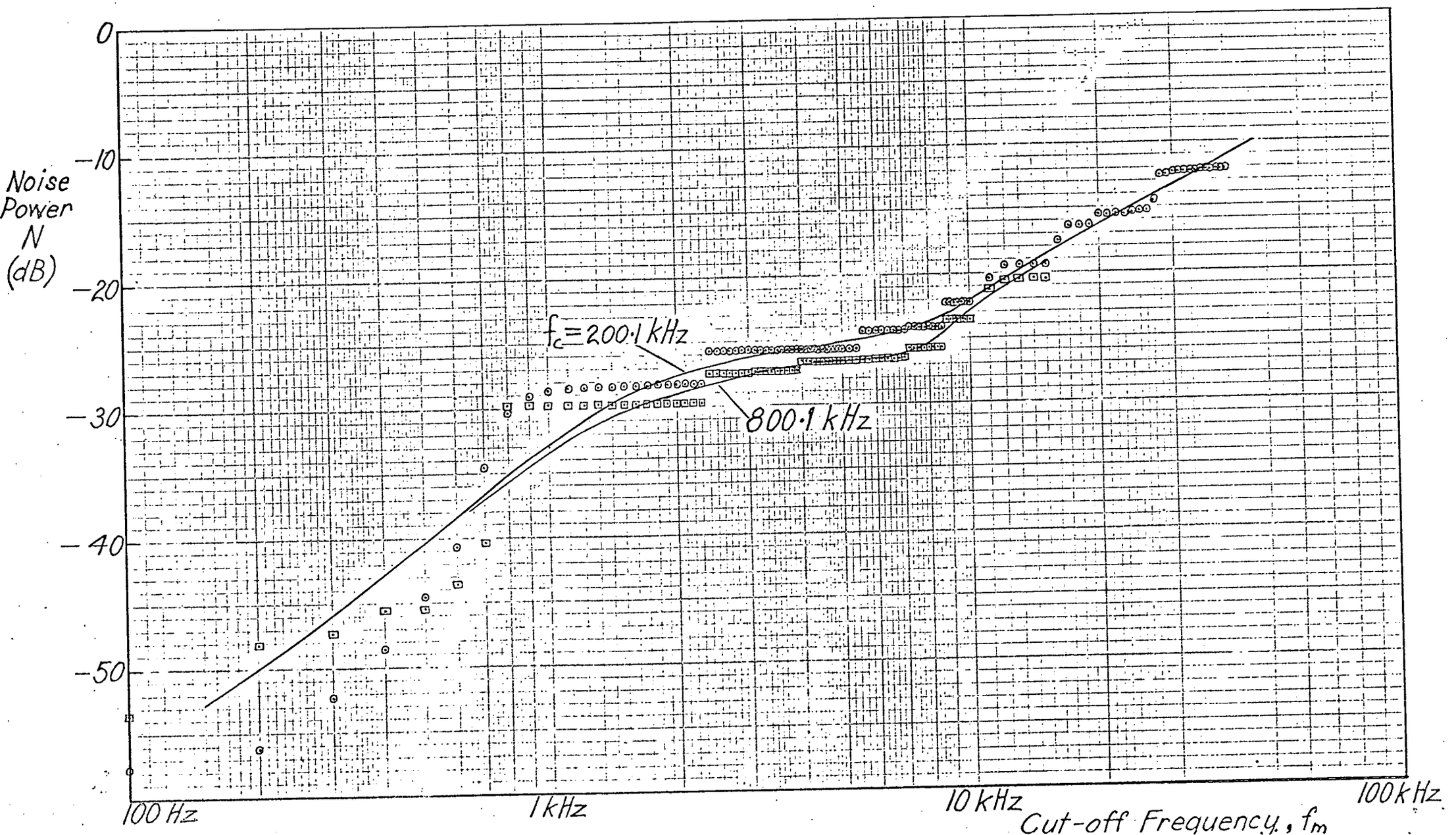
Computer Simulation Results - Curves of Noise Power against Ideal Low Pass Filter Cut-off Frequency for Various Clock Frequencies.




0
−10
−20
−30
−40
−50
Noise Power N (dB)
f_c = 49·7 kHz
60·1 kHz
100 Hz
1 kHz
10 kHz
100 kHz
Cut-off Frequency, f_m


0
-10
-20
-30
-40
-50
Noise Power N (dB)
f_c = 80·1 kHz
100·1 kHz
100 Hz.
1 kHz.
10 kHz.
100 kHz.
Cut-off Frequency, f_m


0
−10
−20
−30
−40
−50
Noise Power N (dB)
f_c = 150·1 kHz.
400·1 kHz.
100 Hz
1 kHz
10 kHz
100 kHz
Cut-off Frequency f_m


0
-10
-20
-30
-40
-50
Noise Power N (dB)
f_c = 200·1 kHz
800·1 kHz
100 Hz
1 kHz
10 kHz
100 kHz
Cut-off Frequency, f_m

BIBLIOGRAPHY

1. Sakrison, D.J., "Communication Theory: Transmission of Waveforms and Digital Information", Wiley, 1968.

2. Carlson, A.B., "Communication Systems, An Introduction to Signals and Noise in Electrical Communications" McGraw-Hill 1968

3. Hancock, J.C., "An Introduction to the Principles of Communication Theory", McGraw-Hill 1961

4. Lathi, B.P., "Signals Systems and Communication" Wiley 1965

5. Bennett, W.R., "Spectra of Quantized Signals" vol. 27, Bell System Technical Journal, 1948.

6. O'Neal, J.B., "Predictive Quantizing Systems (Differential PCM) for Transmission of Television Signals" Bell Syst. Tech. Journal, vol. 45, May-June 1966, pp. 689-721

7. McDonald, R.A., "Signal to Noise Performance and Idle Channel Performance of Differential PCM Systems with Particular Application to Voice Signals" Bell Syst. Tech. Journal, no. 7, p.1125, September 1966.

8. Schouten, J.F., de Jager, F., and Greefkes, J.A., "Delta Modulation, A New Modulation System for Telecommunication", Philips Technical Review, vol. 13 no. 9, March 1952.

9. de Jager, F., "Delta Modulation, A Method of PCM Transmission Using a One-unit Code", Philips Research Reports, vol. 7 November 1952, pp. 442-66

10. Bower, F.K., "What Use is Delta Modulation to the Transmission Engineer?", AIEE Trans., vol. 76, part 1, 1957

11. Greefkes, J.A. and de Jager, F., "Continuous Delta Modulation", Philips Research Reports, vol. 23, April 1968

12. Tomozawa, A. and Kaneko, H., "Companded Delta Modulation for Telephone Transmission", IEEE Trans. on Comm. Tech. vol. COM - 16, no. 1, February 1968.

13. Brolin, S.J., and Brown, J.M., "Companded Delta Modulation for Telephony", IEEE Trans. on Comm. Tech. vol COM - 16, no. 1, February 1968

14. Bennier, D.J., McKibbin , J., and Wollaston, T.C., "Multiplexing of Delta Modulated Signals" Final Report, Technical Report Project Mallard W.B.S. no. HA010405-04, WRE Salisbury, October 1968

15. Schindler, H.R., "Delta Modulation", IEEE Spectrum, October 1970, pp. 69-78

16. Winkler, M.R., "High Information Delta Modulation", IEEE International Convention Record, Part 8, 1963

17. Abate, J.E., "Linear and Adaptive Delta Modulation", Proc. IEEE vol. 55, no. 3, March 1967

18. Wing, P.A., "Adaptive Delta Modulation", Electronics Letters, vol. 5, 1969, pp. 191-192

19. Jayant, N.S., "Adaptive Delta Modulation with a One-Bit Memory," Bell Syst. Tech. Journal, vol. 49, no. 3, March 1970, pp. 321-342

20. Hosokawa , S., and Yamashita, K., "Companded Delta Modulation Coders of the ½ Power and 2/3 Power Types", Electron. Commun. Japan, vol. 51-A, 1968, pp. 18-26

21. Betts, J.A., and Ghani, N., "Adaptive Delta Modulator for Telephony", Electronics Letters, vol. 6, no. 11, May 1970

22. Bosworth, R.H., and Candy, J.C., "A Companded One-Bit Coder for Television Transmission" Bell Syst. Tech. Journal, vol. 48, no. 3, (May-June 1969), pp. 1459-1479

23. Greefkes, J.A., "Digitally Controlled Code Modulation", Philips Research Laboratories (Eindhoven) Report, m.s. 6045, 1969.

24. Hauser, K., and Zarda, S.J., "Design of Digitally Controlled Delta Modems ," Philips Telecommunications of Australia Ltd., (Hendon, S.A.), December 1968

25. Inose, H., Yasuda, Y., Murakami, J., "A Telemetering System by Code Modulation, Δ-Σ Modulation", IEEE Trans. on Space Electronics and Telemetry , vol. SET-8, no. 3, September 1962.

26. Inose, H., and Yasuda, Y., "A Unity Bit Coding Method by Negative Feedback", IEEE Proc. vol. 51, November 1963

27. Inose,H., Aoki , T., Watanabe,K., "Asynchronous Delta Modulation System" Electonics Letters, vol. 2, no. 3, p. 95, March 1966

28. Zetterberg, L., "A Comparison between Delta and Pulse Code Modulation," Ericsson Technics, vol. 2, no. 1, 1955, pp. 95-154

29. Johnson, F.B., "Calculating Delta Modulation Performance" IEEE Trans. Audio and Electroacoustics, vol. AU-16, no. 1, March 1968

30. Wolf, J.K., "Effects of Channel Errors on Delta Modulation", IEEE Trans. vol. COM-14, no. 1., February 1966

31. Donaldson & Chan, "Analysis and Subjective Evaluation of Differential P.C.M. Voice Communication Systems", IEEE Trans. COM-17, no. 1, p. 10, February 1969.

32. Donaldson, R.W., and Douville, R.J., "Analysis, Subjective Evaluation, Optimization and Comparison of the Performance Capabilities of PCM, DPCM, DM, AM and PM Voice Communication Systems", IEEE Trans. on Comm.Tech. vol. COM-17, no. 4, August 1969, p. 421

33. Aaron, M.R., Fleischmann, J.S., McDonald, R.A., Protonotarios, E.N., "Response of Delta Modulation to Gaussian Signals", Bell Syst. Tech. Journal, vol. 48, no. 5, May-June 1969.

34. van de Weg, H., "Quantizing Noise of a Single Integration Delta Modulation Systems", Philips Research Reports, vol. 8, no. 5, October 1953

35. O'Neal, J.B., "Delta Modulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and T.V. Input Signals", Bell Syst. Tech. Journal, vol. 45, no. 1, January 1966, pp 117-141

36. Protonotarios , E.N., "Slope Overload Noise in Differential Pulse Code Modulation," Bell Syst. Tech. Journal, vol. 46, no. 9, November 1967, pp. 2119-2162

37. Goodman, D.J., "Delta Modulation Granular Quantizing Noise", Bell Syst. Tech. Journal, vol. 48, no. 5, May-June 1969

38. Wang, P.P., "Idle Channel Noise of Delta Modulation", IEEE Trans. Communication Technology, vol. COM-16, no. 5, October 1968, pp. 737-742

39. Kikkert, C.J., "Measurement of Quantization Distortion for Random Inputs", Proc. IREE Aust., vol. 32, no. 12, December 1971

40. Fowler, A., A General Purpose Delta Codec, P.M.G. Research Lab. Melbourne , (private communication) 1969

41. Brownell, R., Unpublished M. Eng. Sc. Thesis, University of Tasmania 1969

42. Rader and Gold, "Digital Filter Design Techniques in the Frequency Domain", Proc. IEEE, vol. 55, no. 2, February 1967.

43. Rader, "On Digital Filtering", IEEE Trans., vol. AU-16, no. 3, September 1968

44. Shinners, S.M., "Control System Design", Wiley, 1964, p. 362