

City Research Online

City St George's, University of London

Citation: Ionescu, T. C. (1973). Fundamentals of quasi-digital transduction. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/38380/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

THE CITY UNIVERSITY

DEPARTMENT OF SYSTEMS AND AUTOMATION

FUNDAMENTALS OF QUASI-DIGITAL TRANSDUCTION

A report submitted for the

degree of Doctor of Philosophy

THE CITY UNIVERSITY LIBRARY
ST. JOHN STREET, LONDON, E.C.1.

Traian C. Ionescu

September 1973

CONTENTS

	Aknowledgement	
1	Introduction	1
2	The state of the art in data aquisition systems	4
2.1	Introduction	4
2.2	Requirements imposed on data aquisition systems	5
2.3	A survey of principles used in data aquisition systems	9
2.4	Conclusions	12
3	Quasi-digital signals	15
3.1	Introduction	15
3.2	Definition of the concept of quasi-digital signals	15
3.3	Fundamental characteristics of measurements involving quasi-digital signals	18
3.4	Physical quantities which can be used in generating quasi-digital signals	19
3.5	Structures generating quasi-digital signals	22
3.6	Proposal for a new structure generating quasi-digital signals	31
3.7	The evaluation of the errors of the proposed quasi-digital signal conditioning device	35
3.7.1	The summer	36
3.7.2	The sample and hold devices	45
3.8	Practical results	52
3.9	Conclusions	52
4	Transmission of quasi-digital signals over noisy communication channels	55
4.1	Introduction	55

4.2	A model of the communication channel	59
4.3	The estimation of the p.d.f. by using the method of stochastic approximation	62
4.4	The estimation of the joint probability distribution function of the non-Gaussian noise	71
4.5	The analysis of the communication channel affected by Gaussian noise	75
4.6	Extension to the case of non-Gaussian noise	96
4.7	The analysis of a nonlinear detector for signals corrupted by non-Gaussian noise	98
4.8	Conclusions	107
5	Optimisation of the cost of communication systems	108
5.1	Introduction	108
5.2	Similarities and disimilarities with the problem of depot location	108
5.3	A heuristic algorithm based on the quasi-potential function approach	112
5.4	An algorithm based on a clustering technique	123
5.5	Considerations about the minimisation of the cost of the communication system	131
5.6	Practical results and conclusions	141
6	Final conclusions and lines for future research	144
	Appendices	
A1	The analysis of linear oscillators	147
A2	Some properties of Hermite polynomials	167
A3	About some computational problems	169
A3.1	A method of reducing the store required by the integer programming	169

A3.2 The subroutine for stochastic approximation 172

A3.3 About finding the maxima of the quasi-potential function 175

References 177

AKNOWLEDGEMENT

Sincere gratitude is due to The Romanian Government for its financial support which enabled me to undertake this project.

I would like to thank my supervisor, Professor L. Finkelstein, for his invaluable guidance and support during the course of the work.

Thanks are also due to the members of staff of The Department of Systems and Automation for many stimulating and enlightening discussions and particularly to [redacted].

Finally, the immeasurable debt which I owe to my wife for her untiring support, assistance and patience throughout this work cannot be overemphasised.

Traian C. Ionescu

CHAPTER 1

INTRODUCTION

The availability of on-line digital computers for use as a building element of measurement and control systems necessitates a complete rethinking of the forms of instrumentation. Functions such as signal conditioning, compensation, linearisation, filtering, etc., hitherto carried out by dedicated, special purpose and sometimes not practical hardware, can now be efficiently realised by using a computer. At the same time, digital computers in measurement and control systems need to communicate with other subsystems of the overall system. In particular, they must be connected to sensors.

Up to the present, in spite of significant advances in computer technology, sensor subsystems for use in connection with computers have remained in the form in which they were before the advent of computers, as analogue devices connected to the computer by some form of interface. No systematic studies have been carried out to arrive at systems specifically devised to fit in optimal way into a system centred round a computer.

The aim of the present report is to examine from the first principles the sensor-signal conditioning device-communication channel--interface-computer chain in computer-based measurement and control systems with a view to arriving at conclusions concerning desirable ways to design and analyse the elements of the aforementioned chain. The problem of the industrial process control computer was kept specially in mind.

The motivation to undertake the study arose on the one hand from

industrial needs as seen by the membership and relevant Committee of the SIRA Institute and on the other hand from the interest in The City University in the systemic approaches to instrumentation design (E.R. Carson, A. Pirouzan, 1972). Recent events have confirmed that the study is justified : the IMEKO VI Symposium dealt with means and methods for the instrument-computer interface and a special round table, concerning the problem of interface between the process and the computer was organised. An excellent paper (B.P.Th. Veltman, 1972) analyses in great detail the various aspects of interface. The conclusions mainly reflect the need for a suitable interface (involving both hardware and software) if the capabilities of the digital computer have to be fully used. A conference organised by IEE in September 1973, has the aim of discussing the same topic. The works at IMEKO have shown that the methods and elements used for the interface between the process and the computer are still behind the advances in computer technology. A systemic approach for solving this problem is undertaken in this report. Each element of the chain between the process and the computer is considered and analysed.

The state of the art in data aquisition systems is considered in Chapter 2. The typical requirements imposed to data aquisition systems and possible ways of improving their performance are analysed. The concept of transmitting analogue signals over long and noisy communication channels is challenged and justification for a new approach of this matter is given.

Chapter 3 presents fundamental problems in connection with quasi-digital signals. The notion of quasi-digital signal is postulated. The general properties of quasi-digital signals are discussed. Possible

structures generating quasi-digital signals are examined.

The problem of signal transmission over communication channels is discussed in Chapter 4. The analysis is carried out for channels affected by Gaussian, as well as non-Gaussian noise. A non-Gaussian multivariate probability distribution function is proposed, based on the knowledge of marginal probability distribution functions. The transmission of signals is analysed for the cases of Gaussian and non-Gaussian noise. Also the detection of signals corrupted by noise is examined and possible detector structures are given.

As the cost of the cables connecting the data sources to the computer is unlikely to decrease in the future, Chapter 5 is devoted to the optimisation of the cost of communication system. The leading idea is to use multiplexers placed at suitable locations instead of using an individual cable for each data source. Two heuristic algorithms, which find the optimum position for multiplexers, are proposed. The computer programs implementing them have been tested and the results produced by them are similar. Then the queuing theory for single-server facility is applied for estimating the capacity of multiplexers.

Chapter 6 contains final conclusions and possible lines for future investigation.

The Appendices contain elements relevant to the work which has been done, mainly with reference to the study of an oscillator, using piece-wise linearisation and root locus method, and computational problems encountered during the various stages of the thesis.

CHAPTER 2

The state of the art in data aquisition systems

2.1 Introduction

Although very many industrial automatic control systems use analogue instrumentation, including controllers, on-line digital control is to an increasing extent being used as the standard way of controlling large plants. The reasons for this are :

(i) the flexibility of the computer which is able to implement schemes involving complex control algorithms, feedforward, multivariable control and sequencing ;

(ii) the ability of the computer to optimise the control, including adaptive schemes ;

(iii) the ability to process the measurement information by conditioning, filtering, scaling, linearising, etc. ;

(iv) the ability of the computer to provide good management information.

The advances in digital hardware already discussed make on-line digital computers competitive in terms of cost and reliability with analogue instrumentation used in conventional single loop installation. However, a computer cannot operate at its best if the whole system in which it is connected is not designed having specifically in mind the idea of using a computer. One of the weakest parts of such systems is the interface between the data sources and the computer. The majority of the available transducers have as output either analogue signals, very often of relatively low value (in the range of tens of millivolts), or the variation of the value of a passive circuit element (e.g. mutual

inductance, resistance, etc.). Both these classes of output require conditioning, since the distance between the sensor which has produced them and the computer often exceeds several hundreds of metres. Most signal conditioning devices are designed for being used in connection with remote single loop analogue control instrumentation and therefore produce as output a standard signal (e.g. 2-10 mA), which again is in analogue form. The "classical" solution of realising the process--computer interface consists in transmitting the analogue signal to the computer room, where an analogue to digital (A/D) converter is time--shared by several data sources by using an analogue multiplexer.

This chapter will discuss some of the main requirements imposed on data aquisition systems and some principles used by data aquisition systems which exist at the moment of writing.

2.2 Requirements imposed on data aquisition systems

Several papers (D.P. Allen, 1969 ; Instruments and Control Systems Staff, 1970 ; M.W. Jervis, R.C. Bareham, 1968) describe the requirements imposed on data aquisition systems. In the following the main characteristics which any data aquisition system must satisfy are enumerated and discussed. Problems connected with sensors and transducers have received an excellent treatment in a survey (E.R. Carson, A. Pirouzan, 1972), which also discusses the computer-process interface .

Data sources. The level of the signal source is the major element under consideration. It affects the choice of signal conditioning device, multiplexer and analogue to digital converter. Special care must be shown for low level signals (of the order of mV or tens of mV), as they are easily corrupted by industrial noise. The number of data sources

producing such low level signals, usually thermocouple or bridge output voltages, usually exceeds 30%, as Table 2.1 shows, of the total number

Table 2.1. Typical percentage distribution of data sources classes

i	Parameter	Control	Indication only
1	Temperature	30	30
2	Pressure	20	10
3	Flow	30	40
4	Level	15	10
5	Others	5	10

of points of measurement in an industrial process, therefore it is justified to treat low level signals as a distinct class.

Number of data sources. This parameter determines the number and capacity of multiplexers, also the length of communication channels. Figures of several thousands (Table 2.2) are not unusual for modern chemical plants. In order to cope with future expansions, it is reasonable to prefer a modular structure for multiplexers.

Table 2.2. Data sources for a typical oil refinery.

Process	No. of data sources
Crude and vacuum	150
Delayed coker	150
Fluid catalytic cracker	100
Catalytic reformer	100
Hydrocracker	100
Alkylation	50
Saturates	50
Unsaturates	50
Hydrogen & Hydrotreating	100
	Total 850

Frequency spectrum of data. Shannon's theorem provides the way to calculate the sampling rate such that no loss of information occurs. The sampling rate is individually determined for each point. Its values are taken into account when the multiplexers are chosen and the transmission speed established.

Measurement precision and accuracy. The design of a suitable data aquisition system requires a good knowledge of these parameters. The system components should be in general an order of magnitude more accurate than the desired accuracy of the systems. The combined effect of factors such as type of conversion, resolution, speed, time and temperature drift, etc. has to be considered. One should remember that the sensors accuracy rarely exceeds 0.5%, therefore the overall performance cannot be better than this figure regardless the type of signal conditioning.

Noise. Possibly the most important factor to be considered, the noise can fatally compromise the results of measurement. Generally, the noise is generated by electric fields of motors or power conductors, being particularly violent during transients. The assumption that the distribution of noise is Gaussian is hardly valid in industrial environment. A rigourous analysis of corruption of signals by additive noise must firstly establish a better representation of its distribution.

Geometric lay-out of points. The distance between points of measurement and control room, as well as the spread of points over the site has a considerable impact where corruption by noise and cost of wires is concerned. As the actual cost of cable (about £3 - £8 per metre installed) is likely to increase in future, whilst the plant complexity (therefore the number of points of measurement) has the same tendency, a major saving could be obtained by optimising the cable

lay-out as a function of geometry of points which require measurement and transmission.

On-line computation. Many sensors have nonlinear characteristics in that the output signal is not directly proportional to the value of the measured variable. A computer-based data aquisition system could perform operations such that the result is the linearisation of the measured value. This process, characterised by high accuracy, does however require a particular amount of time which is function of computer speed and therefore affects the choice of computer.

Cost. This is one of the main features of any practical activity. Although the cost of better instrumentation could be much higher than the cost of conventional instrumentation, one has to consider whether or not the improved quality of control does not pay for itself the supplementary cost of more accurate instrumentation.

Reliability. In order to keep the cost of instrumentation at the lowest possible level, the volume of stand-by automatic or even manual equipment should be reduced as far as possible when the process is controlled by computer. This imposes severe conditions concerning the reliability of data aquisition system. Experience has proved that the computer itself is a very reliable component and therefore much care should be taken in order to design a data aquisition system having similar properties.

The aforementioned elements are the main requirements imposed on data aquisition systems. Other elements (display facilities, measurement intervals, manual preset data entry, etc.) can arise in various forms for various practical applications.

The data aquisition system designer should try to answer the question : how should one connect the computer to the process (via the data aquisition

system) such that the information is

- transmitted with minimum loss of accuracy ;
- available when and where is desirable ;
- available at lowest possible cost ;
- transmitted with maximum reliability.

In the next section a brief survey of principles used in existing data aquisition systems will be carried out, analysing the way in which each principle satisfies the requirements mentioned above.

2.3 A survey of principles used in data aquisition systems

There are many data aquisition systems available on the market. Most of them were initially designed to operate in experimental or small installations and came to be used in connection with computers only because no other system was available. As result of this fact, almost all data aquisition systems available have the following features :

- the number of data sources which can be connected to the equipment varies between 10 and 1000 ;
- the signal obtained at the output of the sensor, regardless the level, is transmitted in analogue form over the communication channel. In order to separate in the received signal the useful information, several methods are used (filtering, integration over the noise period, especially when the 50 Hz component induced in long wires has a substantially high level, flying capacitor, etc.). As result of corrupting noise, the signal at the receiver represents the measured parameter with less accuracy as compared with the sensor output signal. Despite all procedures for separating the useful signal from noise, an analogue signal can never be completely free of additive noise ;

- at the receiver the signals are introduced into an analogue multiplexer, using either mechanical switches (reed relays) or semiconductor devices (FET transistors). The multiplexer introduces further limitations in the speed of scanning (for mechanical switches) or nonzero resistance of the switch (for semiconductor devices). Another disadvantage of multiplexers using reed relays consists in the thermoelectric effect appearing at the junction between the two blades in contact. When two relays are used (one for each conductor) the parasitic thermoelectric voltages produced compensate each other ; however, due to the heat radiated by control coils of the relays, it may be possible that the contacts have different temperature and therefore an error appears as result. Due to elastic properties of contact blades, a delay of about 15-20 msec is noticed between the moment of energising the coil and the moment when the contact becomes safe and stable. This obviously has as effect the limitation of the speed of scanning. The semiconductor devices have a nonzero "ON" resistance which is variable not only with the temperature, but also with the current passing through the junction ;
- the multiplexer is followed by a high accuracy system which converts the received analogue signals into digital signals. This system usually consists of a d.c. amplifier (often with self-adjusting gain) and an A/D converter, fast and accurate (up to 15 digits, conversion frequency 10 MHz). Therefore, the computer is fed with quickly and accurately available information. Unfortunately, this information represents an excellent measurement of the multiplexer input signals, but there is no guarantee that the measurand is represented with the same accuracy. The A/D converter can be either of classical type, or based on a relatively new principle (MEDIA) consisting basically of a continuous conversion, using binary Gray code ;

- some of data aquisition systems are designed to be universal (e.g. CAMAC) (H. Bisby, 1971). As a result of the fact that everything is standardised, any item of such a system becomes very expensive ; so one should not be surprised to find out that an empty crate costs over £200 (CAMAC Price List) ;
- the software has various practical realisations and is very much dependent on the type of computer used. Generally the computer ability of processing the signals (by filtering, compensating, linearising, etc.) is used together with the appropriate strategy for data gathering.

An analysis of the aforementioned principles shows that there is still much to be improved in the field of data aquisition systems. The first element to be noticed is the waste of money and lack of accuracy when analogue signals (especially low level) are transmitted individually over communication channels the length of which often exceeds 1 km. Regardless the accuracy of A/D converter used near the computer, the additional noise in the received signal can by no means be removed entirely. According to some estimates (R.L. Aronson, 1971) the cost of wiring is typically 17% of the total control budget for a relatively large centralised system (500 to 1000 data sources). The most optimistic prediction cannot foresee a reduction in the actual cost of wires (£3 - £8 /m of pair installed). On the other hand it is very likely that recent advances in LSI technology will cut down the price of electronic components.

Based on these considerations, an approach which seems to be more economical and to lead to higher accuracy is discussed in the next section.

2.4 Conclusions

The conclusion emerging from a critical analysis of the actual data aquisition systems is that one of the major reasons for not using the computer control of industrial processes on a larger scale is due to the quality of data aquisition systems, still not matching the impressive performance of computers. One should try to solve this problem not merely by looking for a possible way to improve the existing methods, but trying to define the real problem and then trying to find the best way to solve it, regardless the existing solutions. The best solution may be to improve upon existing methods, but on the other hand, a totally different approach may be needed. Rapid advances in technology require a continuous reevaluation of goals and methods to achieve them. If two or three years ago a decision was valid, the same decision may have lost its validity as a consequence of technological developments.

A different pattern of data aquisition system could produce better results at costs not exceeding those of "classical" systems, if not lower. Such a system would have two major features, practically feasible as result of the progress in technology. These features are obtained by :

- connecting the sensor outputs to signal conditioning devices which convert the analogue signal into another form (ideally digital) having much higher immunity to noise ;
- using remote multiplexers (which are now released from tedious task of operating with analogue signals) located such that the cost of wires required to connect the computer to the plant is minimal.

The idea seems to be only too simple and the fact that no existing data aquisition system (except MEDIA, with some reservations) puts it in application is originated by the cost, still high, of A/D converters.

However, as it was mentioned, there are other signals which are as immune to noise as digital signals (when no sophisticated error correcting codes are used). This is the class of "quasi-digital" signals, which will be defined and studied in detail in the next chapter. Generating such signals is an operation as simple as the conversion of a voltage into a standard current signal. Taking into account the improved transmission and multiplexing accuracy, such systems cannot be inferior to "classical" data aquisition systems. Another advantage of such systems consists in the fact that the quasi-digital signal is already "serialised", ready to be transmitted, as opposed to most digital signals, which have to be stored into a register until the process of digitisation is finished and only after this stage is accomplished can they be transmitted.

At the present, the main objection to a system having the features mentioned above is connected with the cost of individual signal conditioning devices. However, one can predict that the tendency in dramatic decreases of prices of integrated circuits will continue in the future. By adding the saving in the cost of wires and the potential supplementary profit obtained as a consequence of using a better, more accurate control system, one can consider that the future belongs to the solution using individual signal conditioning devices, having as output signals with higher noise immunity as compared with analogue signals, the signal multiplexing taking place near the data source. The multiplexers become faster and more accurate, as the signal amplitude is no longer the vital information carrier.

Several problems must be answered in order to fully examine this structure :

- what are the signals at the output of signal conditioning devices ?
- what are the transmission properties of these signals ?

- how should the multiplexers be located such that the cost of wires is minimum ?

The answer to these questions will be given in the next three chapters.

CHAPTER 3

Quasi-digital signals

3.1 Introduction

The need for a systematic approach to the computer-plant interface was emphasised in the previous chapter. One of the main points of such technique consists of an adequate choice of the nature of signals transmitted over communication channels from the data source to the computer room. It is vitally important to use signals with high noise immunity, as the process of separating at the receiver the signal from noise is often complicated and always imperfect. Considerations connected with the cost of hardware do not allow, at least at the present stage, to provide individual A/D converters for each data source. However, there is another class of signals which have superior noise immunity compared with analogue signals. These are quasi-digital signals. It is the aim of this chapter to explain the concept of quasi-digital signal and to give some useful definitions. Also, possible ways and structures of achieving quasi-digital signals are discussed. A signal conditioning device, chiefly designed on economic considerations, is analysed in detail.

3.2 Definition of the concept of quasi-digital signals

One names as "quasi-digital" a signal consisting of a sequence of pulses, which carries the information either in the number of pulses in the sequence, or in the duration of such a sequence. This definition

becomes more explicit by examining the Fig. 3.1. In Fig. 3.1a the analogue signal is converted into a sequence of n pulses of constant frequency, transmitted between the two "flag" pulses, defining the beginning and the end of the message. Fig. 3.1b represents a signal consisting of pulses of variable frequency, the information being recovered either by measuring the time T required to transmit a fixed number of pulses N, or by measuring the period τ of one pulse. It is easy to see

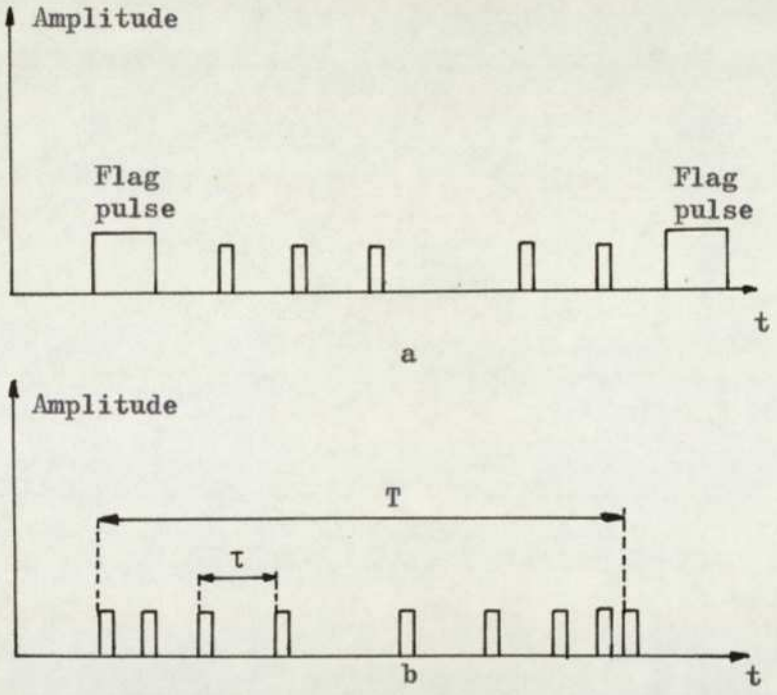


Fig. 3.1

that quasi-digital signals have characteristics which occur both in analogue and digital signals. Although quasi-digital signals would appear as a particular form of digital signals, this is only true as far as the information carrying element (the pulse) is concerned, as the pulses have identical weight in assessing the transmitted information when quasi-digital signals are used, whilst the position of pulses in digital signals is vital for a correct recovery of the message. Therefore loosing or gaining a pulse during the transmission of a

quasi-digital signal is not as dangerous as for digital signals in the same situation.

By examining the two possible aspects of a quasi-digital signal, two possible ways to obtain them can be considered :

a) a variable signal x, which is a function of measurand, is "calibrated" by using a fixed "etalon" c. The result n, always an integer, is a number which satisfies the relation

$$n\,c \leqslant x \leqslant (n+1)\,c \quad ; \tag{3.1}$$

b) a fixed amplitude c is "calibrated" using a variable signal x_i which is a function of measurand. The result n of the calibration is always an integer having the property

$$\sum_{i=1}^{n} x_i \leqslant c \leqslant \sum_{i=1}^{n+1} x_i \quad . \tag{3.2}$$

In the case a) steps of equal amplitude c are generated and summed. The sum is permanently compared with the amplitude x to be measured. When the modulus of the quantity x - nc becomes less than c, the measurement is finished and its result is n. The method is particularly suitable when x is relatively large. The notions of "large" and "small" should be considered in connection with hardware used in the system. As example, if signals in the range of millivolts are measured, they are "small" as compared with the full scale output of an amplifier, normally 10 volts .

In the case b) samples x_i of the measurand, obtained at particular moments of time, are added together and their sum is compared with an etalon c ($c \gg \text{Max } x_i$). The measurement is finished when the difference $\sum_{i=1}^{n} x_i - c$ becomes positive for the first time and its result is n.

At this stage one should analyse some fundamental characteristics of quasi-digital signals as result of measurement.

3.3 Fundamental characteristics of measurements involving quasi-digital signals

Several determinant characteristics of measurements involving quasi-digital signals should be pointed out, as it follows.

Conversion time. Normally one would use sensors with analogue output signal which is converted into quasi-digital signal. It is here where one of main differences between analogue and quasi-digital signals appears. The quasi-digital signals do not allow instantaneous read-out as analogue signals do. As quasi-digital signals are produced during a process involving sampling, the cycle of measurement has a nonzero duration, longer if x is longer for the case a) and the opposite for the case b). Thus there is a time interval between the moment when the information is requested and that when it is available. Therefore the system of measurement should be designed such that the conversion time is adequate for the dynamics of the process. As it will be seen later, this requirement can be almost always fulfilled in practice.

Errors and resolution. Chiefly, the errors are produced by the nonzero value of the calibrating amplitude (fixed for case a), variable for the case b)). The absolute errors $\varepsilon_{abs\ a}$ and $\varepsilon_{abs\ b}$, would be

$$\varepsilon_{abs\ a} = \left| x - c \bmod\left(\frac{x}{c}\right) \right| = x - c \bmod\left(\frac{x}{c}\right) < c \qquad (3.3)$$

and

$$\varepsilon_{abs\ b} = \left| c - \sum_{i=1}^{n} x_i \right| \qquad (3.4)$$

(where mod(y) represents the largest integer less than y) . The relative errors $\varepsilon_{rel\ a}$ and $\varepsilon_{rel\ b}$ can be derived as being :

$$\varepsilon_{\text{rel a}} = 1 - \frac{\text{mod}(\frac{x}{c})}{\frac{x}{c}} \qquad (3.5)$$

and

$$\varepsilon_{\text{rel b}} \simeq \frac{c}{n\, x_0} - 1 \quad , \qquad (3.6)$$

where

$$x_0 = \frac{1}{n} \sum_{i=1}^{n} x_i \quad .$$

In fact the sampling period is chosen such that the measurand samples x_i keep a constant value during the conversion cycle. However, there is no special compulsion for such choice.

By examining the relations given above, one can see that by its very nature the process of quasi-digitisation produces loss of resolution and finally accuracy, chiefly due to the quantisation process. In fact, the relationship between c and x determines the resolution with which an analogue signal can be converted into a quasi-digital signal. Theoretically, a quasi-digital signal can represent the measurand as accurately as desired. This can be achieved by decreasing the amplitude of the etalon in the case a) or by increasing c in the case b). However, technological reasons impose limits which will be discussed later.

3.4 Physical quantities which can be used in generating quasi-digital signals

A realistic analysis of all physical quantities leads to the conclusion that only three classes are suitable for quasi-digitisation, namely :

- TIME
- VOLTAGE
- DISPLACEMENT

The conversion of a signal belonging to any other class into quasi-digital signals either implies an intermediate conversion into one of the above mentioned classes, or it has unacceptable accuracy and resolution and involves complicated instrumentation. Due to these considerations, only the quasi-digitisation of time, voltage and displacement will be considered in this work, the aim of which is to analyse signal conditioning devices having quasi-digital signals as output as part of computer compatible instrumentation. Thus mechanical devices will only be mentioned, without a detailed treatment, being known that such devices involve fine technology and relatively high energies at the output of sensors.

The time quasi-digitisation can be realised basically by using a train of pulses of relatively high frequency, gated by another pulse, having a much longer period (Fig. 3.2). The output of the AND gate reproduces the pulses applied at the input B when a "long" pulse is applied to the input A. This process can produce a number of pulses n which is a function of the measurand : either the measurand is

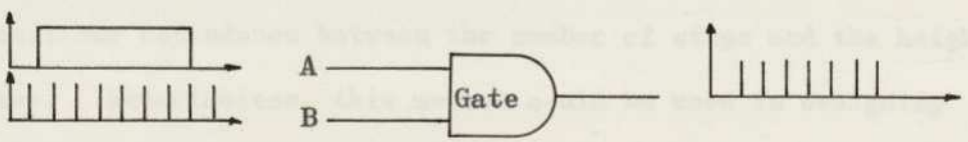


Fig. 3.2

converted into a time duration which is calibrated using n pulses of given, fixed frequency, or the measurand is converted into a frequency and the number of periods n is counted over a given, fixed time interval. The former method is adequate when the measurand is converted into

relatively low frequency signals, whilst the latter is successful when the measurand is converted into high frequency signals.

A similar process takes place when voltages are quasi-digitised. A staircase-shaped voltage is generated and the number of steps required to reach a particular level represents the quasi-digital measure of the measurand (Fig. 3.3).

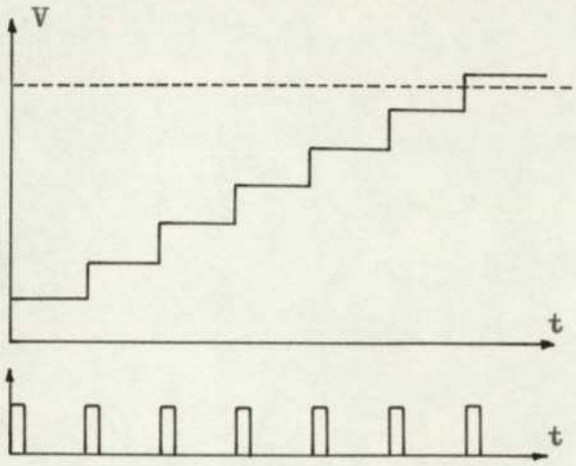


Fig. 3.3

In the case a) the steps have fixed amplitude and the level to be compared with the staircase is a function of measurand, whilst in the case b) the roles are reversed. For the case a) one can recognise one principle used in A/D converters. However, the counter, the final element of such a converter, is not required when quasi-digital signals are used. The second possibility is not used for A/D converters due to the nonlinear dependence between the number of steps and the height of the step. Nevertheless, this method could be used in designing instrumentation destined to be connected to computers, especially when small voltages have to be converted into quasi-digital signals.

The conclusion one can draw after examining the process of quasi-digitisation is that quasi-digital signals could be successfully used in computer compatible instrumentation. The main reason consists in their noise immunity, higher than that of analogue signals and similar

to the noise immunity exhibited by digital signals. A second reason is that the interface between instrumentation using quasi-digital signals and the digital computer is much simpler as compared with analogue instrumentation. Also, multiplexing signals from several data sources over the same communication channel is far easier when quasi-digital signals are employed. A third reason, which will become more obvious in the next sections of this chapter, is supported by the fact that generating quasi-digital signals is almost as simple as converting the output of various sensors into standard current signals (e.g. 4-20 mA).

In the following possible structures generating quasi-digital signals will be briefly analysed and their individual merits will be discussed.

3.5 Structures generating quasi-digital signals

Harmonic and relaxation oscillators. Due to the relatively large number of data sources inside an industrial plant, the designer should do his best in finding the simplest and at the same time the most reliable and accurate form of instrumentation. In choosing quasi-digital instrumentation the first temptation is to use oscillators in which one of the elements of the feedback (e.g. a resistor) is dependent on the measurand, thus the oscillator frequency bearing also a form of relationship with the measured magnitude. Therefore, the first to be analysed is the oscillator. The oscillators have received much attention in the published literature. The analysis of the nonlinear system has attracted many authors and, as result, books such as (N. Minorsky, 1967) (L. Strauss, 1970) deal in great details with theoretical and practical problems concerning oscillators. Many authors still mention the famous

Van der Pol equation which was able to solve the theoretical problems in connection with oscillators having as nonlinear elements electronic valves. However, this equation cannot produce satisfactory results when used for the analysis of oscillators realised with operational amplifiers, uni-junction transistors or more than one nonlinear element. The reason for failure for this cases is obvious : in the Van der Pol equation the characteristic of nonlinear element is approximated by a rational power of the input signal of the nonlinear element ; this approximation produces nice mathematical solutions for the second order differential equation characterising the system, but does not answer questions about systems such as those mentioned above.

A survey of the pertinent literature shows a great variety of types and design of oscillators. The main preoccupation of those who published was to produce stable oscillators. The stability is obtained in various ways : by using diodes (L. Molineux, 1969), by using specialised feedback (L. Mourlam Jr., 1967), or based on the observation that the time constant of resistance-capacitance combination is independent of voltage supply and temperature variations, if good circuit elements are considered (J. Kisslinger, 1968). At the same time many authors tried to produce oscillators having as output a signal of variable frequency. This frequency is a linear function of one parameter of the feedback (D. Meyer, 1968, H.R. Tränkler, 1973). The oscillators designed for this purpose use mainly a Wien bridge, having the period $T = 2\pi\sqrt{C_1C_2R_1R_2}$, and the linearity is obtained by the simultaneous variation of R_1 and R_2, although only R_1 is effectively modified by an external action. Unfortunately, such authors have not considered at all the practical realisability of such designs. Indeed, in order to ensure stable oscillation even in most adverse conditions, as the content of Appendix A1 shows, the gain

of the nonlinear element of the oscillator should have a value above the critical gain. Therefore the expression for T as written above is no longer valid and could lead to errors, as shown in Appendix A1.

Our approach to the problem of finding an universal model of the oscillator starts from the fact that the designer could be assisted by a computer and significant results can be obtained even if the analytical solution is not explicitly known.

Consider the general block diagram of an oscillator, as shown in Fig. 3.4, consisting of a nonlinear amplifier of gain k and a positive feedback, obtained through the feedback network having the transfer function Y(s). Perhaps one could find a function which describes the behaviour of the nonlinear element . But then chances are that the

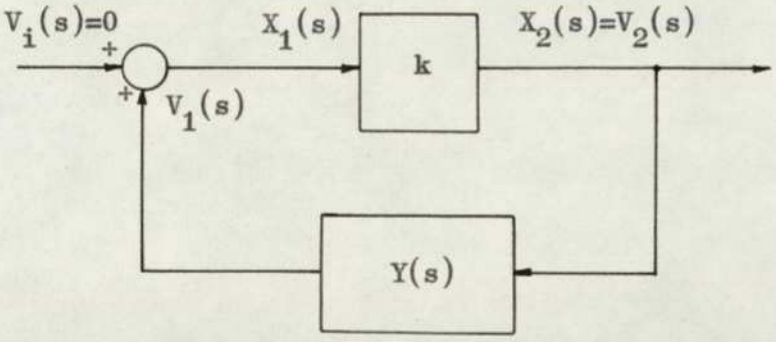


Fig. 3.4

differential equation describing the oscillator is analytically intractable and therefore approximations should be made in order to find the answer. Another method, which makes use of the facilities offered by the digital computer, consists in representing the nonlinear element by a piece-wise linear characteristic, which makes possible the analysis of the waveform obtained at the oscillator output. (The method is treated in great detail in Appendix A1). Consider a general nonlinear

characteristic as shown in Fig. 3.5. Then the oscillator can be represented as in Fig. 3.6 ; the nonlinear element is now replaced by a linear amplifier and a fixed signal, whose gain and amplitude are function of the operation

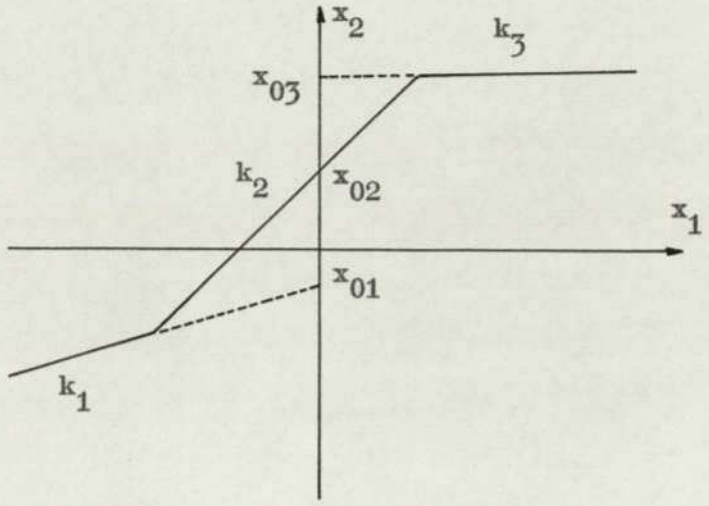


Fig. 3.5

point. Such a representation is supported by evidence obtained in practice. The main circuit elements used in designing oscillators

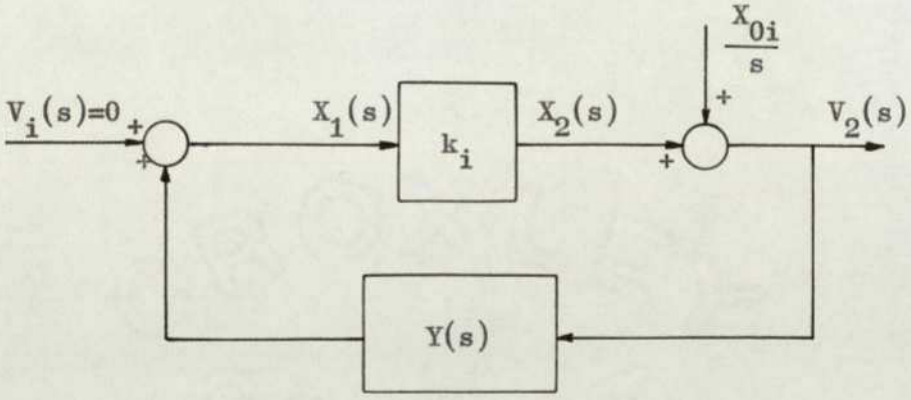


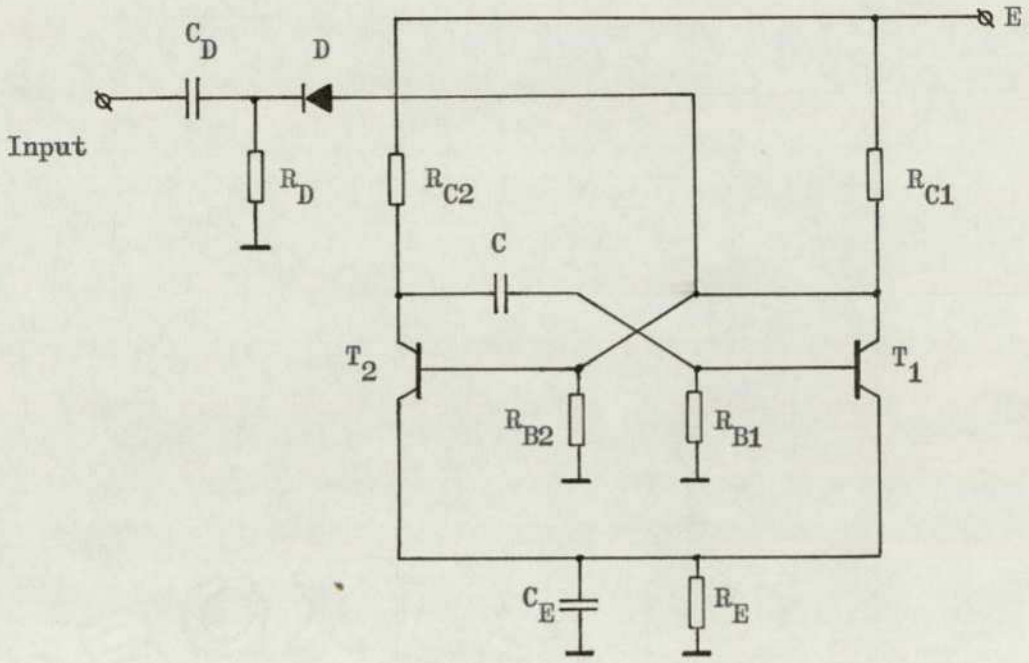
Fig. 3.6

(transistors and operational amplifiers) have characteristics which can be reasonably approximated by three straight lines. As example, a Wien bridge oscillator has been modelled . The computational model produced results in very good concordance with those obtained experimentally, as Appendix A1 shows.

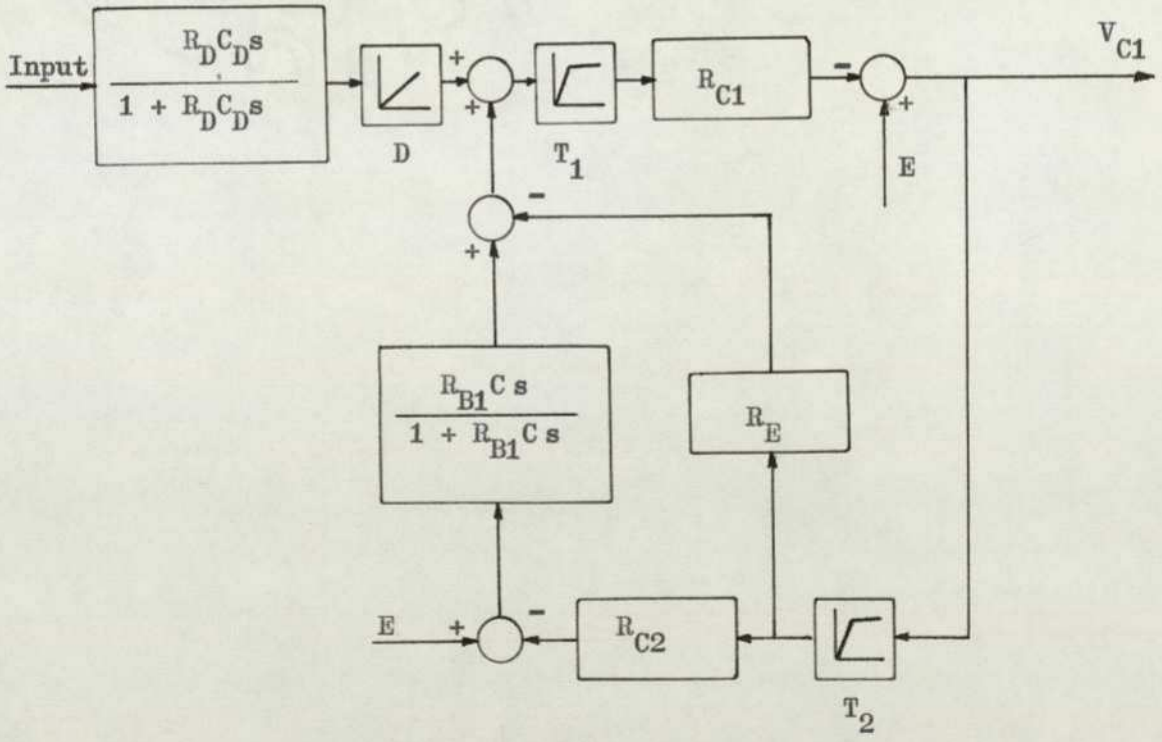
However, the fact that the oscillators are available does not mean that they are suitable for our purpose. One of their disadvantages, a generally nonlinear dependence of the output frequency as function of a variable feedback element could be overcome in systems using computers. The fatal limitations of harmonic and relaxation oscillators consist in their poor sensitivity and in the fact that all feedback elements influence almost in the same proportion the output frequency. Ideally, one should like to have an oscillator having the output very sensitive to one element of the feedback and completely insensitive to all other elements. A study using the root locus method has been carried out in order to investigate the possibility of obtaining such oscillators. As shown in Appendix A1, it is theoretically impossible to design such oscillators.

The same approach can be used when multivibrators or monostable circuits are analysed, although the problem becomes more difficult due to the presence of two nonlinear elements and therefore the computation is more complicated. Fig. 3.7 exemplifies this by showing a monostable circuit (Fig. 3.7a) and its corresponding block diagram (Fig. 3.7b). It is characteristic for monostable circuits that the loop gain is under the critical value and therefore the system does not oscillate continuously, but only responds to an external excitation, applied at the input.

A solution to reduce the sensitivity of the multivibrator output to the variation of environmental conditions has been proposed (A. Pirouzan, 1971). It consists in connecting "in parallel" two multivibrators which are identical in all respects, except the fact that one of them has the resistive sensor included in the feedback, while tho other one has a fixed stable resistor connected in the same position. Therefore, the difference of their periods of oscillation will be less affected by factors such as environmental temperature and voltage supply, but the influence of the


E
C_D
D
Input
R_D
R_{C2}
R_{C1}
C
T_2
T_1
R_{B2}
R_{B1}
C_E
R_E

a


Input
$\frac{R_D C_D s}{1 + R_D C_D s}$
D
T_1
R_{C1}
V_{C1}
E
$\frac{R_{B1} C s}{1 + R_{B1} C s}$
R_E
E
R_{C2}
T_2

b

Fig.3.7

dispersion of the parameters of the components of the two circuits cannot be avoided. Also, synchronisation becomes a major impediment, which can be avoided by using monostable devices triggered by the same pulse. However, the low sensitivity of such devices results in performance below the standard imposed by industrial applications.

Voltage-controlled oscillators. It can be estimated that voltage-controlled oscillators will pay a substantial part on the scene of data aquisition systems in the near future. This is mainly due to the high performance at relatively low cost which can be achieved by these devices. The literature often describes them as voltage to frequency converters. Basically, their principle of operation consists in comparing a linear variable voltage with the reference voltage. Their coincidence is marked by a pulse produced at the output of the converter (F. Bombi, 1970). Other circuits for voltage-controlled oscillators use voltage-variable circuit elements, mainly uni-junction transistors (L.G. Smeins, 1970), the variation of which modifies the time constant of a circuit containing a capacitor which is charged under constant current conditions. The time required for the voltage across the capacitor to reach a given value is thus dependent on the voltage which controls the time constant of the circuit under consideration. Other voltage-controlled oscillators use crystals (M.M. Driscoll, D.J. Healey, 1971). Such circuits have very high stability, but they have a narrow band of tuning (0,2%). Therefore, a good resolution requires operation at high frequencies (above 5 MHz) and this creates problems in connection with the transmission of the output signal over communication channels having limited bandwidth.

In Fig. 3.8 a simplified schematic of a voltage-controlled oscillator, produced by National Semiconductors Ltd., is presented. Q_2 is a voltage controlled source, the collector current of which is a linear

function of the control voltage e_f. Initially Q_5 is OFF and the

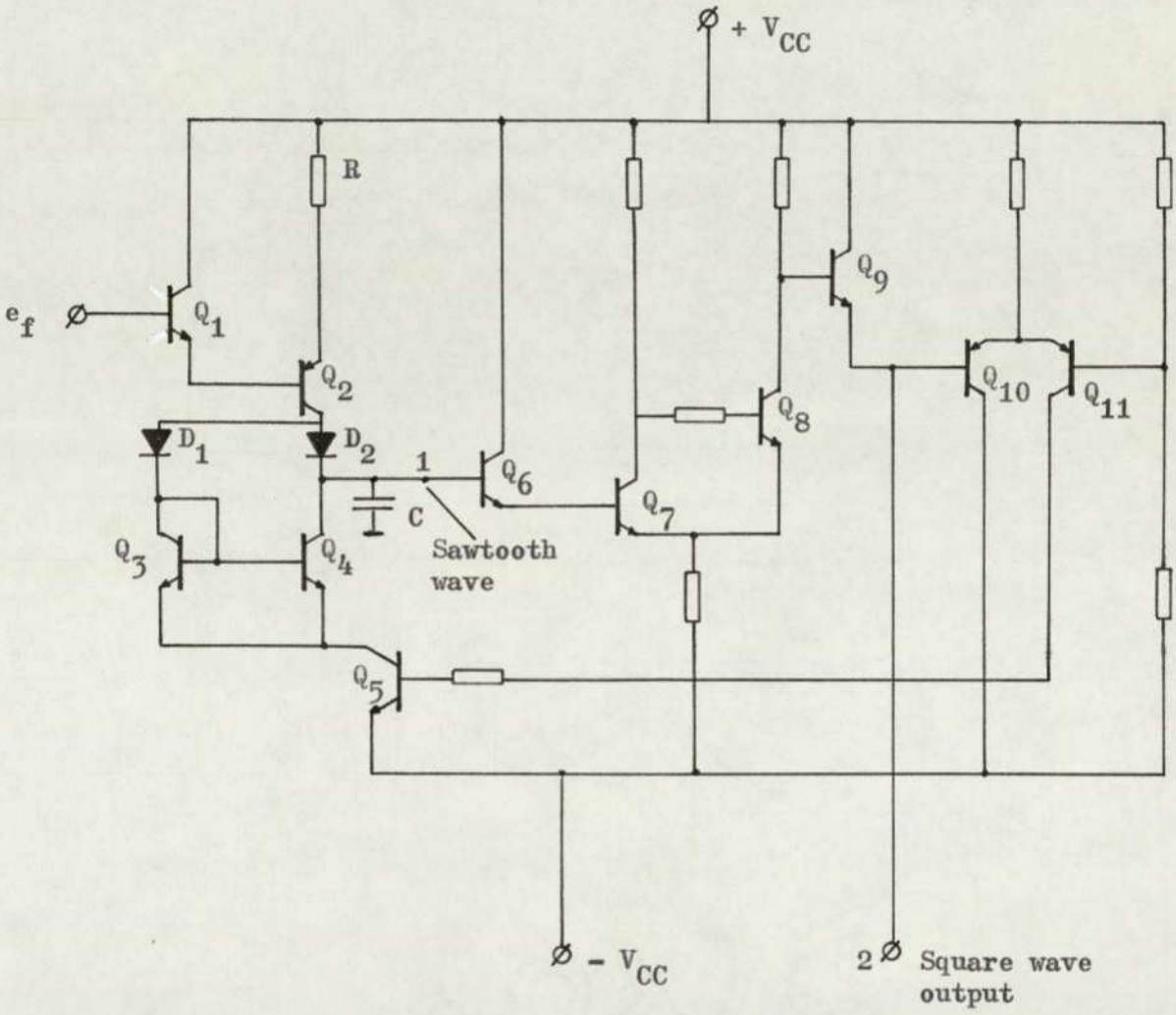


Fig. 3.8

collector current of Q_2 passes through D_2 and charges C in a linear fashion. The voltage across C is therefore a ramp and continues to increase until Q_7 is turned ON ; this turns OFF Q_8, causing Q_9 and Q_{11} to turn ON. This in turn turns ON Q_5. With Q_5 ON, the anode of D_1 is clamped close to - V_{cc} and D_2 stops conducting, since its cathode is more positive than its anode. All of the current supplied by Q_2 is diverted through D_1 and Q_3, which sets up an equal current in Q_4. This current is supplied by the charged capacitor C (which now discharges linearly), causing voltage across it to decrease. This continues until

a lower trip point is reached and Q_2 turns OFF and the cycle repeats. Due to the matching of Q_3 and Q_4, the charge current of C is equal to the discharge current and therefore the duty cycle is very nearly 50%. The waveforms at the points 1 and 2 are shown in Fig. 3.9.

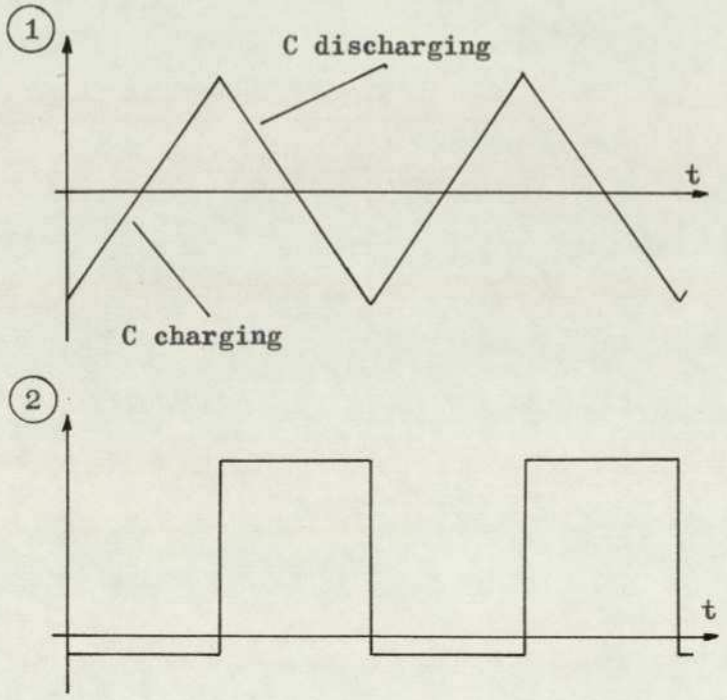


Fig. 3.9

Other versions of voltage-controlled oscillators are designed on the base of multivibrators in which the slew rate of a single capacitor is controlled by a voltage-variable current source, and therefore, the frequency is determined by the control voltage (Motorola).

Regardless the type of circuit used, voltage-controlled oscillators could produce very good performance, especially when used in connection with sensors having low-level output. An example is provided by the request formulated by GEC Elliott for an investigation about possible structures of high performance signal conditioning device, used in converting the output signal of a strain gauge bridge into a quasi-digital signal of high resolution (1 : 10,000), high linearity (0,01%) and good long term stability (0,01%). After carefully analysing the problem, a

voltage-controlled oscillator produced by the Hewlett-Packard Company has been found suitable for this purpose. Of course, devices intentionally designed for this particular application can be considered, but they would also be based on voltage-controlled oscillator principle. Due to advances in large scale integration technology, one can foresee a strong case for designing and building voltage-controlled oscillators as tiny integrated circuits. The versatility of such devices regarding the input and frequency range is another argument in support of use of voltage-controlled oscillators in data aquisition systems of the near future.

3.6 Proposal for a new structure generating quasi-digital signals

In the followings, a new structure which generates quasi-digital signals is proposed, taking into account the requirements imposed to quasi-digital measurement. The block diagram is given in Fig. 3.10. The principle of operation is based on adding together a number n of voltages x (x being the amplitude of the sensor output), until the sum signal exceeds a fixed reference voltage. Therefore the number n represents the quasi-digital measure of the signal x. Assume that in the first instance all voltages V_0, V_1,..., V_6 are zero and the analogue gates G_1 and G_2 are OFF. When V_0 has the value x, the output V_2 of the summer S ($V_2 = V_0 + V_1$) will also have the value x. If a pulse from the switch controller SC switches ON the gate G_1, V_3 also becomes x and so does the output V_4 of the first sample and hold device SH_1. If a new control pulse switches G_1 OFF, V_4 will maintain the value x. By switching ON the gate G_2, the input signal V_5 and therefore the output

signal V_6 of the second sample and hold device SH_2 will also become x.

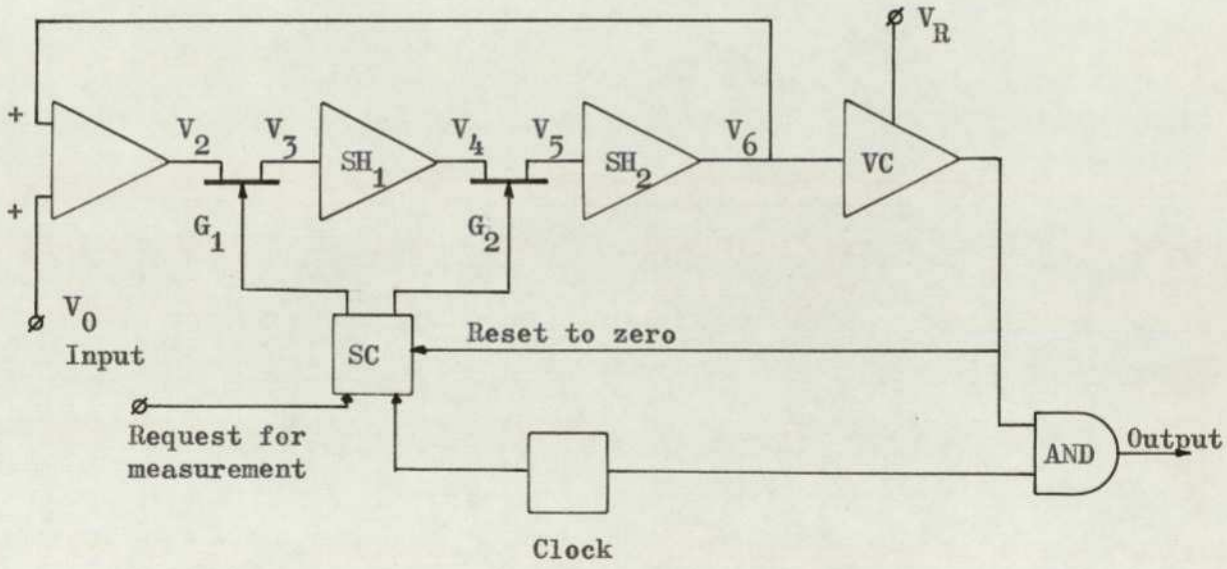


Fig. 3.10

V_6 and then V_1 hold the value x even after G_2 is switched OFF. The output signal V_2 of the summer will now be 2x. When the second cycle, similar to that described above, starts by switching ON the gate G_1, V_3, V_4, V_5, V_6 and V_1 will successively become 2x and, as result, V_2 will have now the value 3x, a.s.o. Thus the value V_6 increases by x after each cycle. Meanwhile, the voltage comparator VC activates the AND gate and pulses from the clock flow through the transmission line. When V_6 exceeds for the first time the reference level V_R, the AND gate is inhibited as the conversion of x into quasi-digital is over. At the same time, SC resets to zero the whole circuit, which is now waiting for a new request for measurement. Fig. 3.11 presents the waveforms of various voltages of the circuit.

From the block diagram presented in Fig. 3.10, associated with the

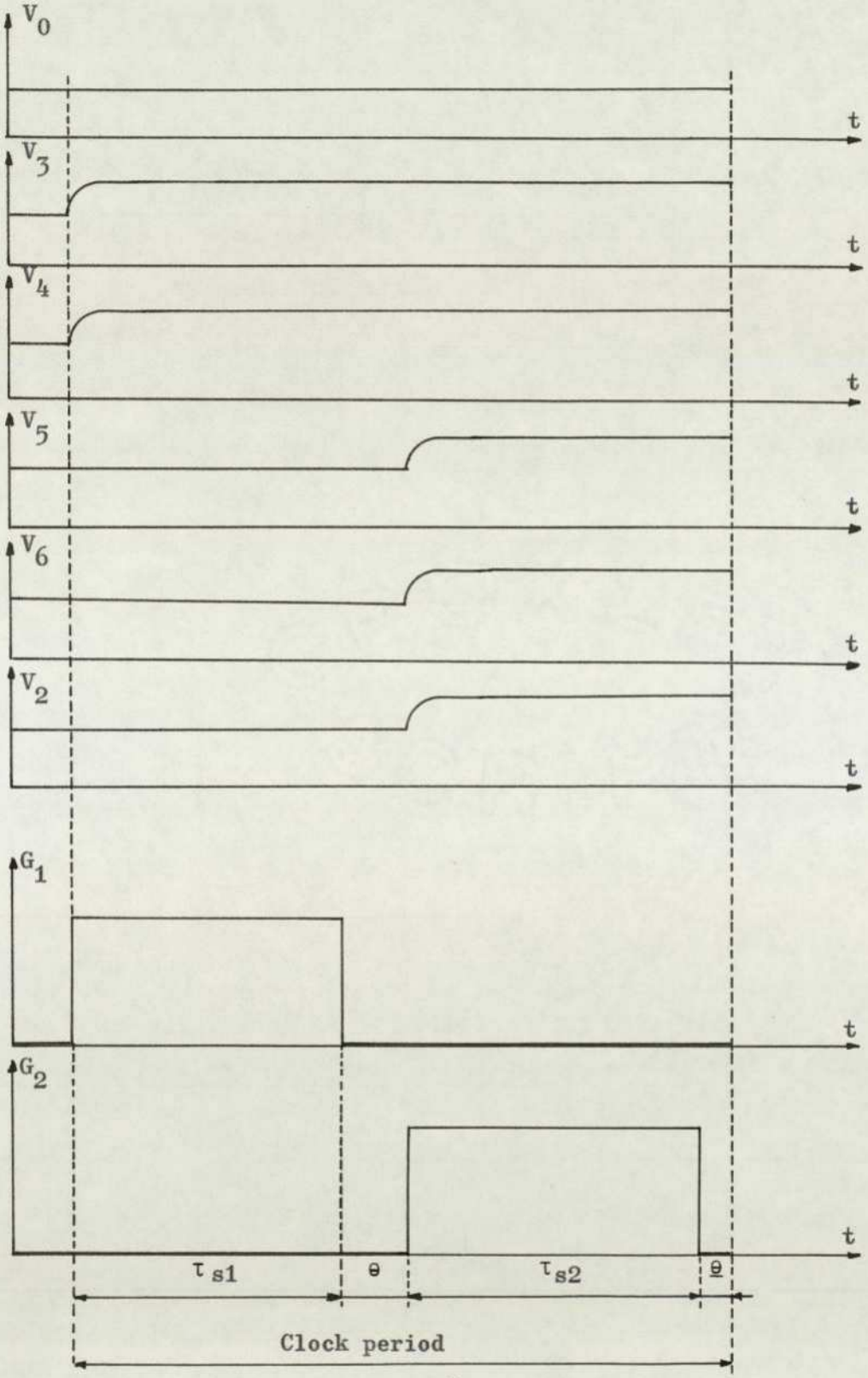


Fig. 3.11

waveform diagrams of Fig. 3.11 one can see that the switch controller should produce one pulse of duration τ_{s1} synchronous with the clock pulse and another one of duration τ_{s2}, delayed by θ with respect to the first. Also it should produce a reset signal at the end of each measurement. One possible block diagram of such a switch is shown in Fig. 3.12.

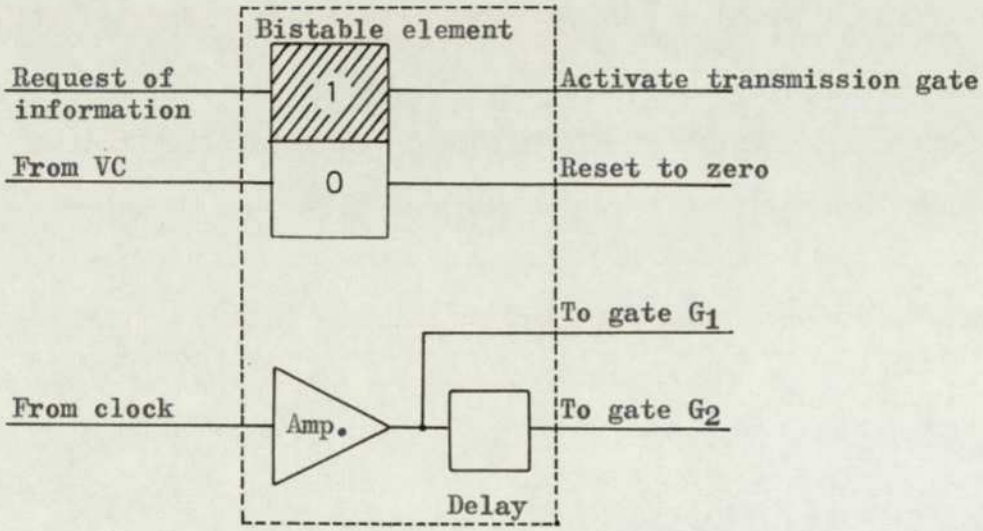


Fig. 3.12

One can see that only one clock, one switch controller and one power supply unit can be used for several signal sources if these are not too widely spread. This is often the practical case, especially in chemical industry and oil refineries, where on the same reaction vessel or distillation column one could find several tens of sensors the output of which requires quasi-digitisation. However, including the clock and switch controller in the local signal conditioner is a relatively simple job, as the number of components is reduced and no severe restrictions are imposed on their operation parameters. The clock frequency is not critical, as only the number of pulses, and not their shape or duration, determines the quasi-digital measure of the input signal.

The aforementioned signal conditioner has the advantage of operating with low level input signals (in the range of those produced by thermocouples, i.e. 20-40 mV), requiring a minimal amount of common-type components. The whole device can be realised as an integrated circuit which would confere to it higher reliability in operation. The performance of such signal conditioner will be discussed in great detail in the next section.

The major disadvantage of the proposed circuit consists in its nonlinear characteristics, which makes it suitable for those applications in which a lineariser is easily and cheaply obtained, therefore in connection with a computer-based control system.

3.7 The evaluation of the errors of the proposed quasi-digital signal conditioning device

The main errors of the proposed quasi-digital signal conditioning device are due to the ambient temperature in which the transducer is working and the nonzero hold time. The temperature variations produce drift of the operational amplifiers used for summing and sample and hold. Supplementary, offset voltages due to power supply voltage variations and noise produce errors which must be taken into account. An analysis of all the errors is made in this section. The analysis is based on the information provided by various manufacturers (Analog Devices, Ancom, Burr-Brown, Motorola, Siliconix, Texas Instruments) in their application manuals.

Although the summer and the sample and hold devices use operational amplifiers of the same type, the analysis of errors is carried out for

each of them due to the different functions performed by the operational amplifiers contained in their structure.

3.7.1 The summer

Errors due to offset voltages. From the point of view of the offset voltages, the circuit diagram shown in Fig. 3.13 has to be considered.

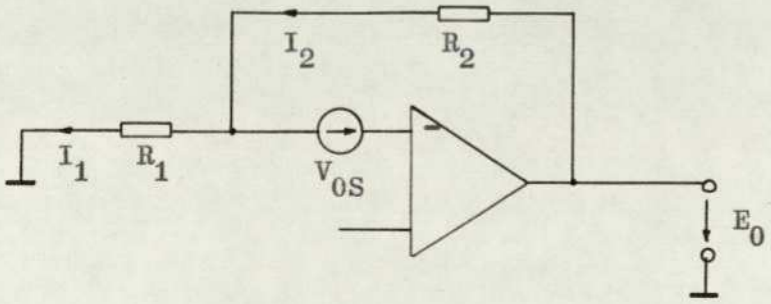


Fig. 3.13

Due to the very high input resistance of the operational amplifiers one can assume that $I_1 = I_2$. As

$$I_1 = \frac{V_{0S}}{R_1} \tag{3.7}$$

and

$$I_2 = \frac{E_0 - V_{0S}}{R_2} \quad , \tag{3.8}$$

then

$$\frac{V_{0S}}{R_1} = \frac{E_0 - V_{0S}}{R_2} \quad , \tag{3.9}$$

which yields :

$$E_0 = \frac{R_1 + R_2}{R_1} \; V_{0S} \quad . \tag{3.10}$$

The usual notation

$$\beta = \frac{R_1}{R_1 + R_2} \qquad (3.11)$$

leads to :

$$E_0 = \frac{1}{\beta} V_{0S} \quad . \qquad (3.12)$$

For a summing amplifier $R_1 = R_2$ and then $\beta = \frac{1}{2}$, therefore

$$E_0 = 2\ V_{0S} \quad . \qquad (3.13)$$

Refered to the maximum voltage $E_{i\ max}$ applied to the input,

$$\varepsilon_{0S} = \frac{200\ V_{0S}}{E_{i\ max}} \quad \% \quad . \qquad (3.14)$$

The offset voltage V_{0S} can be nulled, but an offset voltage drift is produced by variation in the ambient temperature. Although the industrial range of temperature is from - 25°C to + 85°C, the maximum variation in the ambient temperature must be considered as being 60°C (85°C - 25°C), as the nulling of the offset is carried out at the room temperature. The offset voltage due to the variations of the power supply and the offset voltage produced by the ageing of the components should be considered when V_{0S} is calculated.

Errors due to the finite open loop gain A. From Fig. 3.14 one can write :

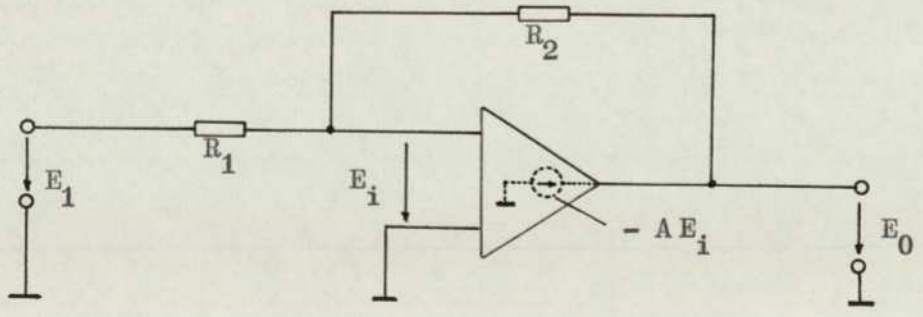


Fig. 3.14

$$\frac{E_1 - E_i}{R_1} = \frac{E_i - E_0}{R_2} \quad . \tag{3.15}$$

If $E_0 = - A E_i$, then

$$\frac{E_1}{R_1} + \frac{E_0}{A R_1} = - \frac{E_0}{A R_2} - \frac{E_0}{R_2} \quad , \tag{3.16}$$

and then, after some transformations :

$$\frac{E_0}{E_1} = \frac{R_2}{R_1} \frac{1}{1 + \frac{1}{A\beta}} \quad , \tag{3.17}$$

$$\beta = \frac{R_1}{R_1 + R_2} \quad . \tag{3.18}$$

Because $A\beta \gg 1$

$$\frac{1}{1 + \frac{1}{A\beta}} = \frac{1 - (\frac{1}{A\beta})^2}{1 + \frac{1}{A\beta}} = 1 - \frac{1}{A\beta} \quad . \tag{3.19}$$

The error due to the finite value of A is :

$$\varepsilon_A = \frac{200}{A} \ \% \tag{3.20}$$

for $\beta = \frac{1}{2}$ (the case of the summing amplifier) .

Errors due to the external resistor network. A further error of $3 \frac{\Delta R}{R} 100$ % is added due to tolerances and variations of the resistors of the input and feedback network (Fig. 3.15). The factor 3 is due to the fact that three resistors are used (2 for the two inputs, one for the feedback) :

$$E_0 = - \left(\frac{R_2}{R_{11}} E_{11} - \frac{R_2}{R_{12}} E_{12} \right) \frac{1}{1 + \frac{1}{A\beta}} \quad . \tag{3.21}$$

Errors due to the bias currents. A circuit showing the influence of the bias currents is shown in Fig. 3.16 . Their contribution in the output voltage is calculated as follows :

$$I_3 = I_1 + I_4 \longrightarrow I_1 = I_3 - I_4$$

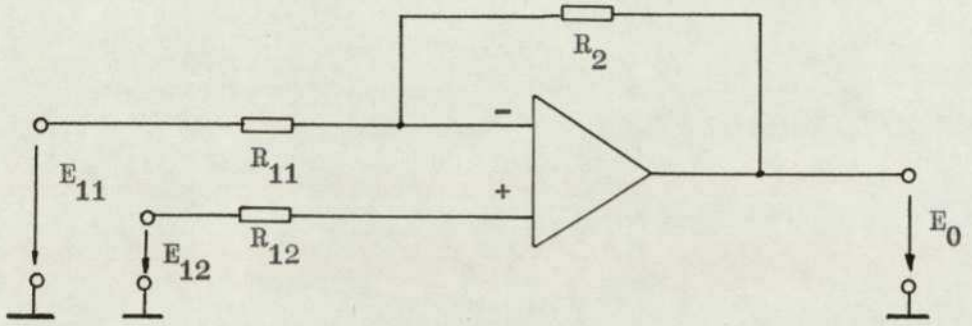


Fig. 3.15

$$E_S = I_2 R_3$$

$$I_3 = \frac{E_S}{R_{11}} \longrightarrow I_3 = I_2 \frac{R_3}{R_{11}}$$

$$I_4 = \frac{E_0 - E_S}{R_2} = \frac{E_0}{R_2} - \frac{E_S}{R_2} = \frac{E_0}{R_2} - \frac{I_2 R_3}{R_2}$$

$$I_1 = I_2 \frac{R_3}{R_{11}} - \frac{E_0}{R_2} + \frac{I_2 R_3}{R_2} \ .$$

Thus

$$E_0 = - R_2 I_1 + \frac{R_2 + R_{11}}{R_2 R_{11}} R_2 R_3 I_2 \ . \qquad (3.22)$$

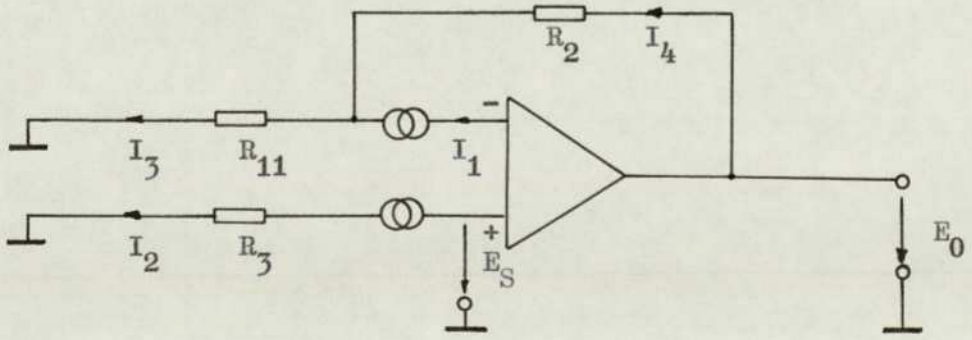


Fig. 3.16

If

$$R_3 = \frac{R_{11} R_2}{R_{11} + R_2} \quad , \tag{3.23}$$

$$E_0 = R_2 (I_2 - I_1) \quad . \tag{3.24}$$

Refering this value to the value of maximum voltage applied to the input :

$$\varepsilon_b = \frac{R_2 (I_2 - I_1)}{E_{i\ max}} 100 \quad \% \quad . \tag{3.25}$$

In most of applications $I_1 = I_2$, then for a proper value of R_3 the effect of bias currents can be neglected. However, the bias currents have a temperature drift and therefore low values for R_2 should be chosen.

Errors due to the finite value of the input resistance of the amplifier. The finite value R_i of the input impedance of the operational amplifier affects the value A of the gain of the amplifier. The following analytical considerations can be made (in connection with Fig. 3.17).

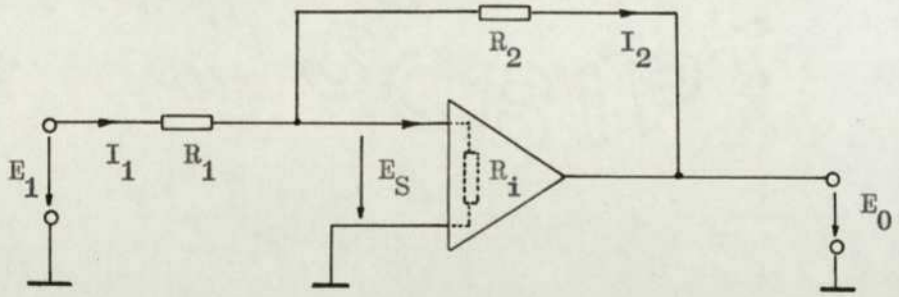


Fig. 3.17

$$I_1 = I_i + I_2 \tag{3.26a}$$

$$I_1 = \frac{E_1 - E_S}{R_1} \tag{3.26b}$$

$$I_2 = \frac{E_S - E_0}{R_2} \tag{3.26c}$$

$$I_i = \frac{E_S}{R_i} \tag{3.26d}$$

$$E_S = \frac{E_0}{A} \quad . \tag{3.26e}$$

Then

$$\frac{E_1 - E_S}{R_1} = \frac{E_S}{R_i} + \frac{E_S - E_0}{R_2}$$

$$\frac{E_1}{R_1} = E_S \left(\frac{1}{R_1} + \frac{1}{R_i} + \frac{1}{R_2} \right) - \frac{E_0}{R_2}$$

and then

$$\frac{E_1}{R_1} = -\frac{E_0}{R_2} \left(\frac{1}{A} + \frac{R_2}{A\,R_1} + \frac{R_2}{A\,R_i} + 1 \right) , \tag{3.27}$$

or

$$\frac{E_0}{E_1} = -\frac{R_2}{R_1} \frac{1}{1 + \frac{1}{A} \cdot \frac{R_2 + R_1}{R_1} \left(1 + \frac{R_2 R_1}{R_2 + R_1} \cdot \frac{1}{R_i} \right)} \quad . \tag{3.28}$$

Notating

$$\frac{R_1 + R_2}{R_1} = \frac{1}{\beta} \quad , \tag{3.29}$$

$$A' = A \frac{1}{1 + \frac{R_1 R_2}{R_1 + R_2} \cdot \frac{1}{R_i}} , \tag{3.30}$$

one can write :

$$\frac{E_0}{E_1} = -\frac{R_2}{R_1} \cdot \frac{1}{1 + \frac{1}{A'\beta}} \quad . \tag{3.31}$$

A' is less than A by a factor of $1 + \frac{R_1 R_2}{R_1 \quad R_2} \cdot \frac{1}{R_i}$.

The value A' should be taken into account when calculations of errors due to the finite value of A are made.

Errors due to the finite value of the output resistance. The nonzero values for the output resistance R_{out} of the operational amplifier lead to further reductions in the value of A. To evaluate this effect the circuit diagram shown in Fig. 3.18 is considered.

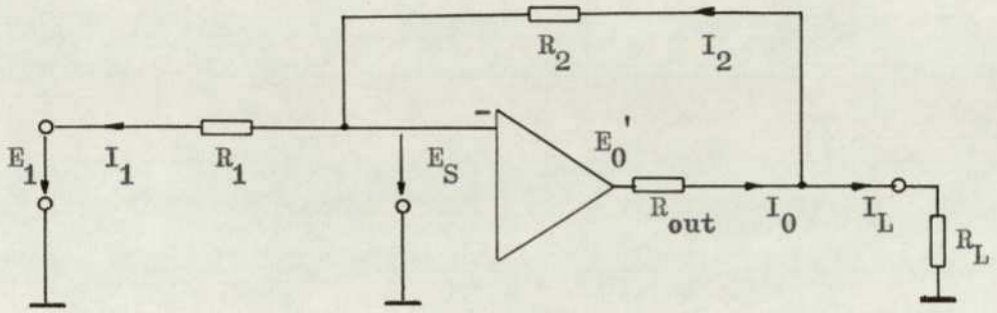


Fig. 3.18

Assuming that a variation ΔE_0 of the output voltage is produced by a variation ΔI_L of the load current I_L flowing through the load resistor R_L, one can write

$$\Delta E_0 = R_{out}' \Delta I_L \quad .$$

Subsequently

$$\Delta E_S = \frac{R_1}{R_1 + R_2} \Delta E_0 = - \frac{\Delta E_0'}{A} \quad . \tag{3.32}$$

Then

$$\Delta E_0' = \Delta E_0 + R_{out} \Delta I_L \quad . \tag{3.33}$$

Replacing terms :

$$\frac{R_1}{R_1 + R_2} \Delta E_0 = \frac{E_0 + R_{out} \Delta I_L}{A} \tag{3.34}$$

$$A\beta \Delta E_0 = \Delta E_0 + R_{out} \Delta I_L \quad , \tag{3.35}$$

then

$$\Delta E_0 = \frac{R_{out}}{A\beta - 1} \Delta I_L \tag{3.36}$$

and therefore

$$R_{out}' = \frac{R_{out}}{A\beta - 1} \simeq \frac{R_{out}}{A\beta} \quad , \tag{3.37}$$

as $A\beta \gg 1$.

This is the output resistance of the closed loop amplifier, which is smaller when A has large values.

Now

$$I_0 = I_L + I_2 = \frac{E_0}{R_L} + \frac{E_0}{R_2}$$

$$E_0' = - A E_S \tag{3.38}$$

$$I_0 = \frac{E_0' - E_0}{R_{out}}$$

$$I_2 \simeq I_1 = \frac{E_1 - E_S}{R_1} \quad .$$

Then

$$E_0 \left(\frac{1}{R_L} + \frac{1}{R_2} \right) = \frac{E_0'}{R_{out}} - \frac{E_0}{R_{out}} \quad , \tag{3.39}$$

$$\frac{E_0}{E_S} = - A \frac{1}{1 + \frac{R_{out}}{R_L} + \frac{R_{out}}{R_2}} \quad . \tag{3.40}$$

This relation shows that R_{out} has a great effect on A, if the load resistance R_L is fairly low. To avoid substantial decreases of A at low values of R_L, a booster amplifier must be used, as shown in Fig. 3.19. The booster amplifier has the gain 1, a high output current and a low output resistance.

Errors due to the input noise. The input noise in operational amplifiers can be described by a noise voltage E_n and two noise currents I_{n1} and I_{n2} associated to each input (Fig. 3.20).

The overall effect of the above mentioned factors on the output is

a noise voltage having the value :

$$E_{0n} = \frac{R_1 + R_2}{R_1} \sqrt{E_n^2 + (R_3 I_{n2})^2 + \left(\frac{R_1 R_2}{R_1 + R_2} I_{n1} \right)^2} \qquad (3.41)$$

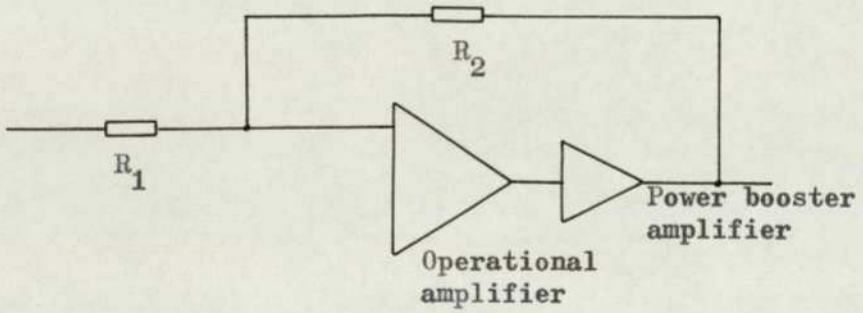


Fig. 3.19

If

$$R_3 = \frac{R_1 R_2}{R_1 + R_2} ,$$

$$I_{n1} = I_{n2} = I_n$$

and

$$\beta = \frac{R_1}{R_1 + R_2} ,$$

$$E_{0n} = \frac{1}{\beta} \sqrt{E_n^2 + 2 (R_3 I_n)^2} . \qquad (3.42)$$

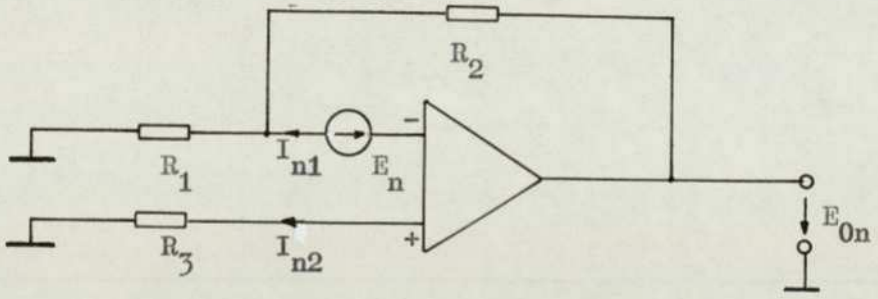


Fig. 3.20

The error produced by the noise will be :

$$\varepsilon_n = \frac{\frac{1}{\beta}\sqrt{E_n^2 + 2\,(R_3 I_n)^2}}{E_{i\,max}}\,100\ \% \ . \qquad (3.43)$$

The noise characteristics are much affected by the frequency and the level of noise is low at lower frequencies.

3.7.2 The sample and hold devices

The main errors of the sample and hold devices are produced by the offset voltage drift, the noise of the operational amplifiers, the finite time for charging the capacitor and the leakage of the charged capacitor (nonzero holding time). These elements will be considered in order to estimate the errors of sample and hold devices used in the transducer.

Errors due to the offset voltage drift. The drift of the operational amplifier was already discussed for the summer. The problem remains the same, except the fact that the equivalent drift voltage is V_{0S} and not 2 V_{0S}, due to the particular connection of the amplifier as voltage follower (Fig. 3.21). The error due to the drift is

$$\varepsilon_d = \frac{V_{0S}}{E_{i\,max}}\,100\ \% \ . \qquad (3.44)$$

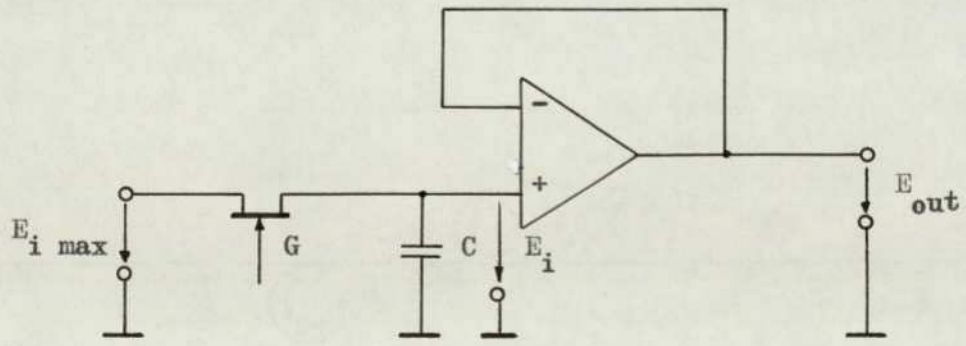


Fig. 3.21

In calculating V_{0S} the effect of temperature, power supply variation and component ageing should be considered.

Errors due to the noise. As it was shown, the noise has as effect a noise voltage having the value

$$E_{0n} = \frac{1}{\beta} \sqrt{E_n^2 + (R_3 I_{n2})^2 + \left(\frac{R_1 R_2}{R_1 + R_2} I_{n1}\right)^2} \quad . \tag{3.45}$$

It can be assumed that $R_2 = 0$, $R_3 = r_{0N}$ (the FET-resistance in the 0N-state), due to the direct feedback connection between the output and the inverting input. Then

$$E_{0n} = \sqrt{E_n^2 + r_{0N}^2 I_n^2} \tag{3.46}$$

and

$$\varepsilon_n = \frac{E_{0n}}{E_{i\,max}} 100 \quad \% \quad . \tag{3.47}$$

Errors due to th finite time t_c of charging the capacitor. To calculate the errors due to the finite time of charging the capacitor the circuit diagram shown in Fig. 3.22 is considered.

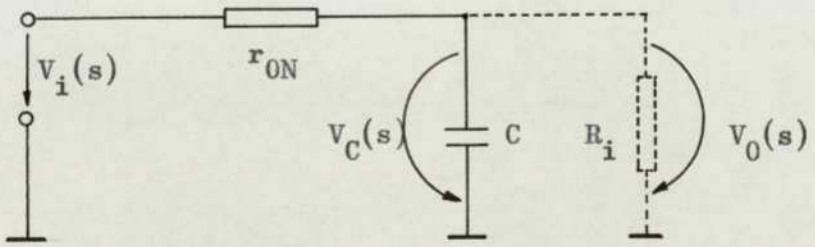


Fig. 3.22

The voltage $V_C(s)$ across the capacitor C can be expressed as

$$V_C(s) = \frac{1/C\,s}{r_{0N} + 1/C\,s} V_i(s) \quad , \tag{3.48}$$

or

$$V_C(s) = V_i(s) \frac{1}{1 + T_c s} \quad , \tag{3.49}$$

in which

$$T_c = r_{ON} C \quad .$$

If $V_i(s)$ is a step voltage

$$V_i(s) = \frac{V}{s}$$

then

$$V_C(s) = V \frac{1}{s (1 + T_c s)} \tag{3.50}$$

and

$$v_C(t) = V (1 - e^{-\frac{t}{T_c}}) \quad . \tag{3.51}$$

The ideal situation is that when $v_C(t) = V$, which takes place when $t \longrightarrow \infty$. If an error of ε_{charge} is imposed, then

$$1 - e^{-\frac{t_c}{T_c}} \geqslant 1 - \varepsilon_{charge} \quad ,$$

or

$$e^{-\frac{t_c}{T_c}} \leqslant \varepsilon_{charge} \quad . \tag{3.52}$$

The time of charging t_c can then be calculated in order to achieve a desired accuracy ε_{charge} :

$$t_c > - T_c \ln(\varepsilon_{charge}) \quad . \tag{3.53}$$

As example, if $\varepsilon_{charge} = 0.0001 = 10^{-4}$:

$$t_c \geqslant 4 T_c \ln 10 \simeq 10 T_c \quad .$$

In this case $V_C \geqslant (1 - \varepsilon_{charge}) V$.

Errors due to the nonzero holding time t_h. Due to the finite input resistance R_i of the operational amplifier, the capacitor C does

remain charged at the initial voltage V_C. To calculate the dynamic behaviour of the circuit, again Fig. 3.22 is refered to.

$$V_0(s) = \frac{R_i}{\frac{1}{Cs} + R_i} V_C(s) \quad . \tag{3.54}$$

In time domain :

$$v_0(t) = V_C \; e^{-\frac{t}{T_d}} \quad , \tag{3.55}$$

in which

$$T_d = R_i C \quad .$$

If the maximum allowed error is $\varepsilon_{discharge}$, then the holding time must satisfy the relation :

$$e^{-\frac{t_h}{T_d}} \geqslant 1 - \varepsilon_{discharge} \quad ,$$

then

$$t_h \leqslant - T_d \ln(1 - \varepsilon_{discharge}) \quad . \tag{3.56}$$

If $\varepsilon_{discharge} = 0.0001$

$$t_h \leqslant 10^{-4} T_d \quad .$$

Errors due to the external resistor network. These errors are negligible, because the zero resistance of the feedback. Under no circumstances is the error larger than 0.001 %.

Errors due to the bias current. In the case of sample and hold circuit it is not possible to compensate the errors due to the bias current as in the case of the summing amplifier. The output voltage produced by bias currents is :

$$E_0 = - R_2 I_1 + \frac{R_2 + R_{11}}{R_{11}} R_3 I_1 \quad . \tag{3.57}$$

With $R_2 = 0$, $R_{11} \rightarrow \infty$, $R_3 = r_{ON}$:

$$E_0 = r_{ON} I_1 \quad . \tag{3.58}$$

The error is reduced by decreasing r_{ON}.

Errors due to the analogue gate. Due to recent developments concerning field effect transistors such devices are now unanimously accepted as suitable for being successfully used as analogue gates. A basic gate is shown in Fig. 3.23. The device is controlled by applying negative pulses at the gate input G. In their absence the device is in "ON" state ; "OFF" state is established when a negative pulse is applied.

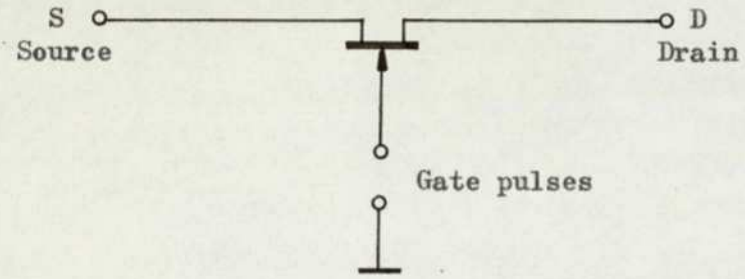


Fig. 3.23

The main features of the gate which should be analysed are the nonzero "ON" resistance between source and drain and feedthrough spikes (both into generator and load) during transient times of gate pulses. In order to get a better insight of the phenomenon, equivalent "ON" and "OFF" circuits are considered and discussed below (Fig. 3.24) (M. Stevens, A. Wood, 1968 ; J. Watson, 1971).

In the "OFF" state leakage currents $I_{DS\ OFF}$, I_{GSO} and I_{BDO} are negligible ($10^{-12}\ I_{DS\ ON}$ to $10^{-15}\ I_{DS\ ON}$) and only at very high temperature their contribution should be taken into account. Although the presence of spikes does not alter the operation and then no drift voltage could appear, the suggested circuit configuration provides a reasonably good protection even against these spikes. Indeed, consider the circuit shown in Fig. 3.25. C_H and R_I represent the holding capacitance and the

input resistance of the buffer used in the sample and hold device.

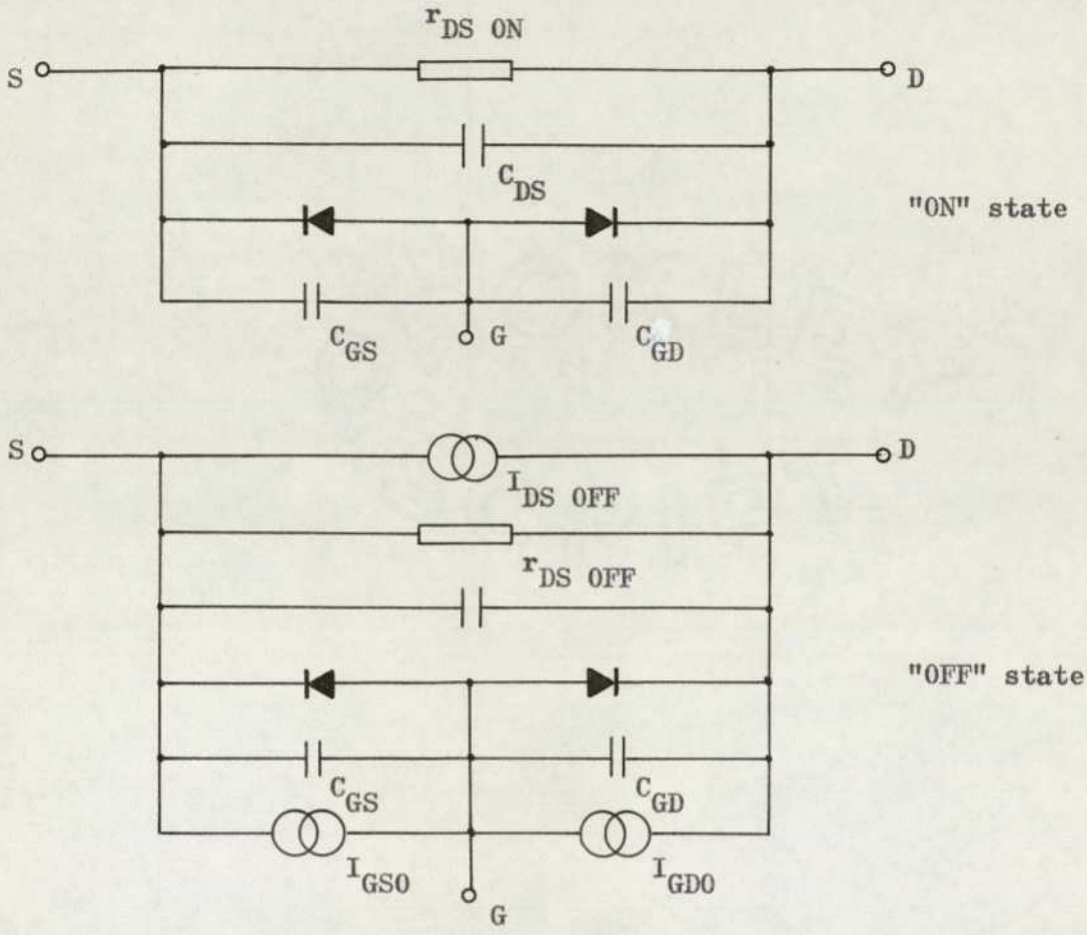


Fig. 3.24

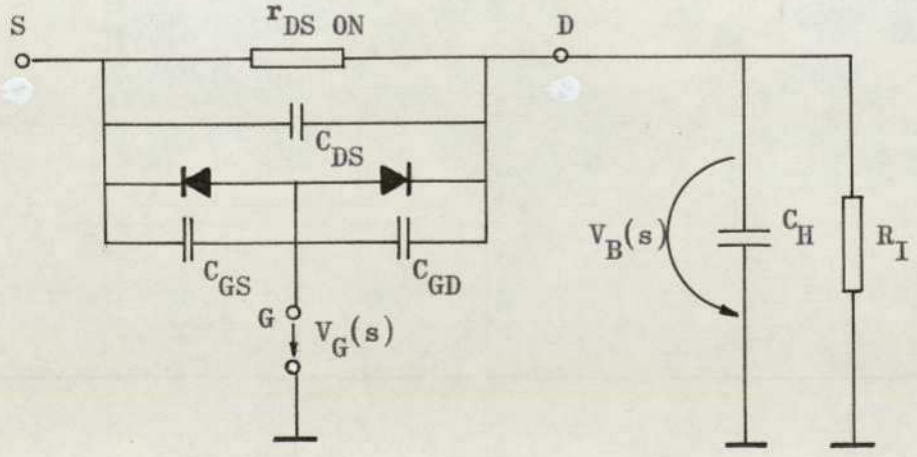


Fig. 3.25

When a step voltage $V_G(s) = \frac{V_0}{s}$ is applied at the gate G, the input voltage

$V_B(s)$ of the buffer is

$$V_B(s) = \frac{V_G}{s} \frac{Z_1(s)}{Z_1(s) + Z_2(s)} ,$$

with

$$Z_1(s) = \frac{\frac{1}{C_H s} R_I}{\frac{1}{C_H s} + R_I} ,$$

$$Z_2(s) = Z_1(s) + \frac{1}{C_{DG} s} .$$

After some manipulations one obtaines :

$$V_I(s) = \frac{V_G}{s} \frac{T' s}{1 + (T + T') s} ,$$

where

$$T = R_I C_H ,$$

$$T' = R_I C_{DG} .$$

In practice $C_{DG} \ll C_H$, hence $T' \ll T$ and thus

$$V_I(s) = \frac{V_G}{s} \frac{T' s}{1 + T s} . \qquad (3.59)$$

In the same domain

$$v_I(t) = \frac{T'}{T} V_G \left(1 - \exp(-\frac{t}{T})\right). \qquad (3.60)$$

It can be seen that although the transient time is relatively long, the amplitude of the spike is very low (in the ratio of C_{DG} and C_H) and therefore the pulses applied to the gate have little influence on the voltage across the capacitor C_H. If the control pulses applied to G have relatively long raising time, the effect of the feedthrough spike is even more diminished.

3.8 Practical results

The proposed structure for a quasi-digital signal conditioning device has been built according to the circuit diagram shown in Fig. 3.26. Initially analogue switches of the type DG 184 (Siliconix) have been used as gates for the sample and hold devices. Their poor performance (long transient time, about 10 times longer than that mentioned in the manufacturer's specification and high value for "ON" resistance) did not permit a relevant experimental analysis of the original conditioning device and only the principle of operation has been checked. At a later stage the DG 184 have been replaced by switches having the circuit diagram as shown in Fig. 3.27 (M. Stevens, A. Wood, 1968). The circuit is under development and although the principle of operation has been checked, the quantitative data obtained so far are not relevant. By using operational amplifiers of the type 741 GTX (Ancom Devices Ltd.) having a drift of $1\,\mu V/^{o}C$, the total estimated error for operation between $-\ 25^{o}C$ and $+\ 85^{o}C$ is less than 1%. This figure can be substantially improved if the signal conditioning device is produced as an integrated circuit and mounted in a temperature controlled capsule, since the main errors are due to the variation of the ambient temperature.

3.9 Conclusions

This chapter has been devoted to the concept of quasi-digital signals and suitable devices to produce them. The advantages offered by a data aquisition system using quasi-digital signals as compared with one using analogue signals are :

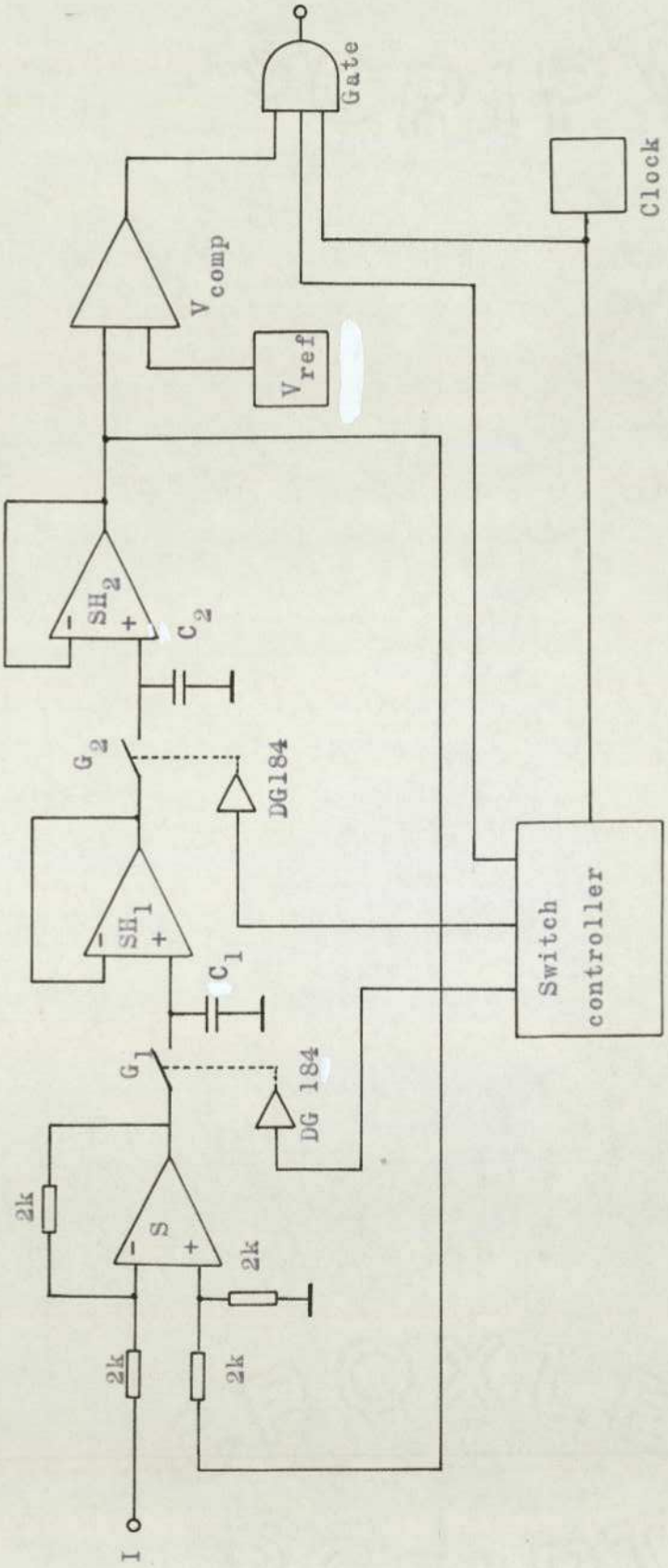

Gate
Clock
V_{comp}
V_{ref}
SH_2
C_2
G_2
DG184
SH_1
C_1
G_1
DG 184
Switch controller
2k
S
2k
2k
2k
I

Fig. 3.26

- the quasi-digital signal conditioning is obtained in a simple manner, similar with standard analogue signal conditioning ;

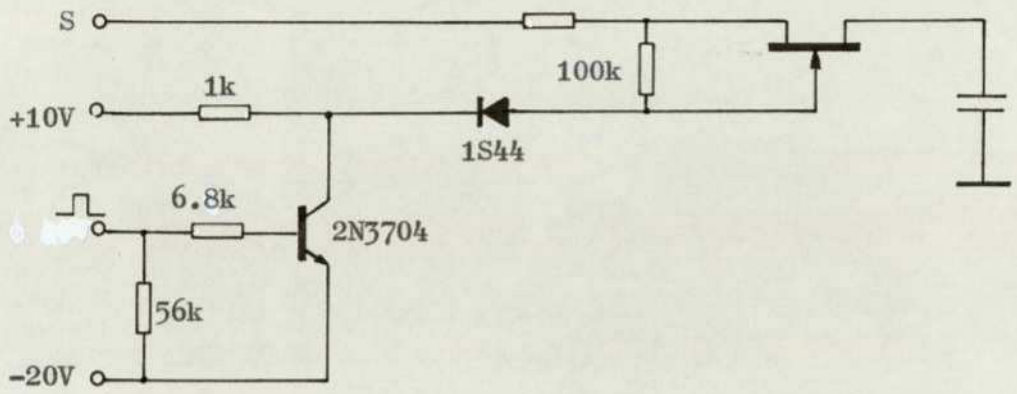


Fig. 3.27

- the use of multiplexers becomes much easier and cheaper ;
- the noise immunity of quasi-digital signals, comparable with that of digital signals, is superior of that achieved by using analogue signals ;
- the interface between the plant and the computer can be realised much easier when quasi-digital signals are used ;
- the accuracy of the process of quasi-digitisation is by no means inferior to that obtained by standard analogue conditioning at comparable costs.

Among possible structures of quasi-digital signal conditioning devices, voltage controlled oscillators look as having the best future, although other specialised structures, having the advantage of low cost, can be imagined.

CHAPTER 4

Transmission of quasi-digital signals over noisy communication channels

4.1 Introduction

The previous chapter has analysed the qualities of the quasi-digital signals from the point of view of representing in an adequate manner the measurand. The advantage of using quasi-digital signals in computer compatible instrumentation have been mentioned. At this stage one should consider the problem of transmitting the quasi-digital signals over the communication channels linking the data sources to the computer. Due to their similarity with digital signals as far as transmission properties are concerned, the analysis will follow the path normally taken for digital signals.

Very much has been written about the transmission of digital signals. Among many books published, one can mention as very good (W.B. Davenport, W.L. Root, 1958), (R.W. Lucky, J. Salz, E.J. Weldon, 1968), (A.J. Viterbi, 1966), (A.D. Whalen, 1971). Apart from exposing general principles, they give good guidance about the way of analysing the problem of signals corrupted by noise. However, since the sources which produce the noise cannot be analysed explicitly and on the basis of the central limit theorem, the analysis is carried out for Gaussian processes only. However, the specialised magasines contain many remarkable contributions in the field. Of very good quality are (J.H. Fennick, I.E.O. Nassel, 1966) and (J.H. Fennick, 1969) which contain surveys of impulse noise on Bell System carrier facilities and telephone lines. Other authors (C.P. Bates,

G.T. Hawley, 1971 ; H.C. Ward, 1967) try to establish models for currents and voltages induced within long transmission cables or for the noise coupled between cable pairs. This way of approaching the problem of noise has produced acceptable results only for particular cases and give only cures instead of finding the causes.

Much attention has been paid to various models of impulsive noise. This name characterises the electrical noise which appears in spikes of high amplitude and short duration, the time interval between spikes being relatively large as compared with the spike duration (P. Mertz, 1961). The authors tried to model impulsive noise either as a combination of standard waveforms (J.D. Moore, 1972) or having a fixed shape and assuming particular distributions (e.g. Poisson) of the moments of occurence of spikes (W.H. Card, P.K. Chanduri, 1965). Impulsive noise has been studied experimentally (M. Kurland, D.A. Molony, 1967) and the results show the influence of the shape and duration of the spike on the average number of errors per transmitted pulse. Also a "safe" threshold is determined such that for values of the signal/noise ratio above this level no errors can occur. However, this affirmation appears exagerated, as the signal/noise ratio can only be estimated with a limited degree of confidence and therefore some errors could occur.

Other authors (P.A. Bello, R. Esposito, 1969) calculate the probabilities of errors due to the impulsive noise based on the assumption that only one impulse might occur during the transmission of one pulse.

The detection of signals corrupted by non-Gaussian noise has been analysed in papers as (A. Bodharamic, 1971), (S.S. Rappaport, L. Kurz, 1966), in which the assumption that the noise belongs to a particular class, being distributed according to a hyperbolic law (R.E. Ziemer, 1967)

in which the noise is assumed Gaussian for low signal/noise ratios and impulsive with Poisson distribution at high ratios.

As conclusion to this survey, one should quote what many authors have said : one of the most outstanding feature of impulsive noise is its variability. Therefore, the first step in examining the electrical noise present in industrial environments is to get a relevant sample of such noise. The next step would be to model this noise such that its variability is taken into account, at the same time provision being made that the Gaussian character of the noise between impulses is reproduced.

Knowledge about the noise in instrument cables operating in plants controlled by computers would be of great help in assessing the transmission and detection properties of quasi-digital signals. Investigations have been carried out in order to obtain relevant information concerning this class of noise. Some data has been provided by Control & Instrumentation, of U.K.A.E.A., refering to electrical noise present in power stations.

Unfortunately, no consistent data about the character and parameters of industrial noise were available at the time of writing (by industrial noise one understands - here and for the rest of this paper - the electrical noise present in channels dedicated to data transmission, operating in industrial environments). However, our belief is that one cannot assume the industrial noise as being purely Gaussian, mainly due to the electrical interference produced by cables carrying strong currents. These cables, supplying energy to motors actuating devices with intermittent operation (valves, pumps, etc.), radiate an electromagnetic field which induce unwanted noise in communication channels, thus corrupting the useful signal. It seems very likely that the main component of this induced noise is centred about the frequency of 50 Hz and therefore it can be eliminated by using a high-pass filter (when signals having the

fundamental frequency well above the mains frequency are transmitted). Nevertheless, spikes of high amplitude and short duration, occuring at random moments of time, are still present in the received signals. Therefore, one cannot achieve the best performance by using a detector tuned merely for separating the noise having Gaussian distribution.

In this chapter a way of estimating the p.d.f. of the non - Gaussian noise, based on the stochastic approximation method, is considered. A test sample is generated, having the p.d.f. given by

$$p(x) = \theta_1 p_G(x) + \theta_2 p_I(x) \ , \qquad (4.1)$$

where $p_G(x)$ and $p_I(x)$ are respectively Gaussian and impulsive p.d.f.s and θ_1 , θ_2 are constants with the property $\theta_1 + \theta_2 = 1$. The sample is used for testing a program which approximates an unknown p.d.f. by a linear combination of known functions. Starting from the estimated non - Gaussian marginal p.d.f.s, a form of multivariate non - Gaussian p.d.f. is proposed and exemplified. Then the transmission of signals over communication channels affected by Gaussian noise is analysed. Special attention is given to the analytic signals, since this is a very convenient representation of signals having narrow bandwidth (I.Selin, 1965) as it is the case of signals produced by voltage - to - frequency converters. Although these devices could have a square wave as output signal, this waveform can be transmitted over a communication channel with limited bandwidth, since only the fundamental frequency carries the information relevant to the measurand. The analysis of the transmission properties and detection of analytic signals is also justified by the fact that a more reliable form of quasi - digital signal is obtained by applying a train of pulses at one input of a gate, which has the second input connected to a source of signal producing a sine wave of constant amplitude and frequency (Fig. 4.1). The procedure is similar to those

used in radar systems. The analysis can be extended to non-Gaussian

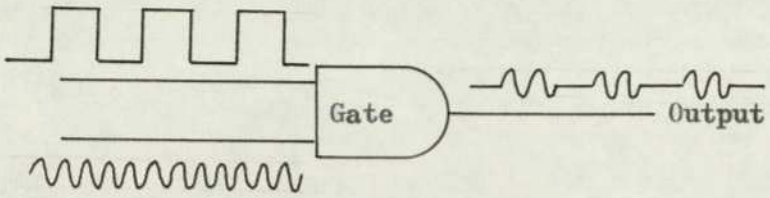


Fig. 4.1

channels, using the proposed multivariate p.d.f. Finally, a one-parameter nonlinear detector is analysed.

4.2 A model of the communication channel

A simplified model of the communication channel is shown in Fig. 4.2. In the particular case of quasi-digital signals the transmitted signal $s(t)$ can have only two values : 1 if a mark was transmitted, 0 if a space was transmitted (the values have been normalised). The noise $n(t)$ is often non-Gaussian, due to various hazards occuring in the industrial environments in which quasi-digital signals are used. The task of the receiver is to decide which of the two signals was transmitted, based on the received signal $z(t) = s(t) + n(t)$. The process which governs the behaviour of

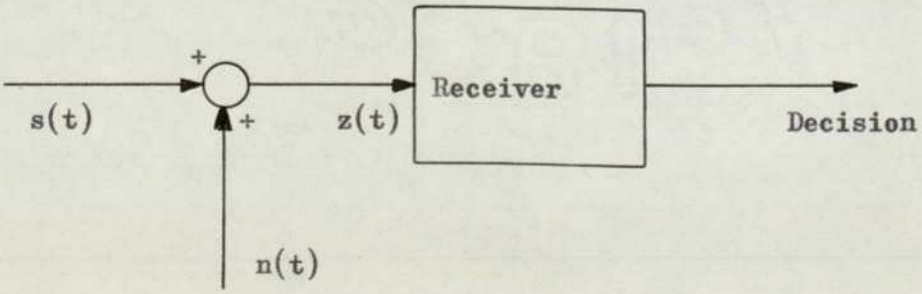


Fig. 4.2

the receiver is presented more explicitly in Fig. 4.3. The signals s_1 and s_0, having the a priori probabilities of occurence p_{a1} and p_{a0} are

transmitted over the communication channel. Due to the noise the points s_1 and s_0 are mapped not into two points of the observation space, but into two regions of this space. Moreover, these regions overlap and therefore the decision process, although deterministic, is based on statistical knowledge about the process. Thus the decision function $D(z)$ maps any point of the observation space into one of the following three points of the decision space : δ_1 if it is more probable that s_1 was transmitted, δ_0 if it is more probable that s_0 was transmitted and δ_{10} if, based on the existing evidence, taking either of the decisions δ_1 or δ_0 involves a too great risk. In fact the decision function calculates the risk involved by taking each of the decision δ_1, δ_0 or δ_{10} and selects

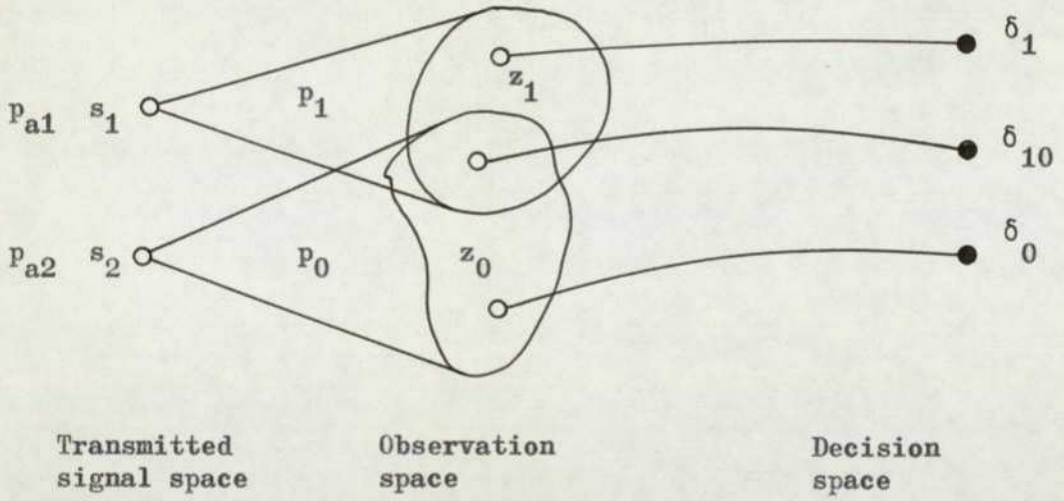


Fig. 4.3

that which minimises the risk. Consider $p(z; s_1)$ and $p(z; s_0)$ as being the transition densities associated with the signals s_1 and s_0 respectively. The cost of the risk involved in taking a decision is given by

$$r = \sum_{i=1}^{N_s} p_{ai} \sum_{j=1}^{N_D} p(\delta_j; s_i)\, L_{ij} , \qquad (4.2)$$

where p_{ai} are the a priori probabilities and L_{ij} are the elements of a loss matrix, which associates costs to taking a decision δ_j when s_i was

transmitted. N_s represents the number of points in the signal space and N_D the number of points in the decision space. For the case of quasi-digital signals $N_s = 2$ and $N_D = 3$. Usually the matrix $\underline{L}$ is chosen such that

$$L_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{otherwise ;} \end{cases} \qquad (4.3)$$

the risk becomes :

$$r_i = \sum_{j=1}^{N_D} p_{ai}\, p(\delta_j ;\, s_i) \qquad (4.4)$$

By analysing all the values of r_i, the detector choses the value of i for which the risk is minimised. The major problem of taking decisions involves a good knowledge of the conditional probabilities $p(\delta_j ;\, s_i)$, as they are essential in determining the likelihoods. The problem is a relatively easy one if these probabilities are Gaussian, but, unfortunately this is often an ideal condition. Due to the nature of processes taking place in an industrial environment the p.d.f. of the noise becomes non-Gaussian and therefore detectors tuned to produce optimum results for noise with normal distribution are no longer optimal under such circumstances. A survey of the published literature (J.H. Fennick, S.I.E.O. Nassel, 1966; A.A. Alexander & R.M. Gryb, 1960) shows that the impulse noise which characterises industrial environments appears as spikes of high amplitude and short duration, the moments of occurence being random. These spikes are induced in communication channels by the electromagnetic fields produced by strong currents, especially during the on-off switching process. The experience has also shown that a realistic description of such noise requires data collected on channels affected by impulsive noise. A theoretical treatment of the problem is impossible in most cases. Therefore an estimation of noise p.d.f. based on noise samples collected

from the channel under observation is the only realistic approach in solving the problem of designing optimal detectors and choosing optimal shapes for the transmitted signals. One of the most used models of non-Gaussian noise is the Gram-Charlier model (W.Feller, 1966), which expresses the p.d.f. as the product between a Gaussian density function and a polynomial expression which requires the knowledge of the moments of high order of the distribution. However, to the best of our knowledge, no model of a multivariate p.d.f. describing the non-Gaussian noise in the general case has been produced so far. As result, problems such as the analysis of the detector performance or detection of analytic signals have been analysed or solved only for the Gaussian state of the channel noise.

In the next section, the p.d.f. of non-Gaussian noise of any form is estimated and then a multivariate non-Gaussian p.d.f. model is proposed and analysed. An example is given.

4.3 The estimation of the p.d.f. by using the method of stochastic approximation

The method of stochastic approximation, described in (J.M.Mendel, K.S.Fu, 1970), is a method having as main aim to minimise a specific criterion function. A standard minimisation routine cannot be applied, since the criterion function is not known explicitly, neither can one observe the exact value of the function for any given setting of the parameters by means of a suitable experiment. The method could be applied even for experiments in which the results are observed only after they have been corrupted by noise, as it uses the stochastic analogues of the steepest descent method in order to obtain the maximum of the function.

The object of the method consists in minimising a deterministic function $J(\underline{\theta})$, where $\underline{\theta}$ is the parameter which can be controlled. The function $J(\underline{\theta})$ is completely unknown, but for every $\underline{\theta}$ a random variable $z(\underline{\theta})$ can be observed, such that

$$E\left(z(\underline{\theta});\ \underline{\theta}\right) = J(\underline{\theta}) \quad . \tag{4.5}$$

Sometimes, one may be able to observe also a noisy value $\underline{y}(\underline{\theta})$ of the gradient function, where $\underline{y}(\underline{\theta})$ satisfies

$$E\left(\underline{y};\ \underline{\theta}\right) = \nabla_{\theta} J(\underline{\theta}) \, . \tag{4.6}$$

One should devise an algorithm to determine $\underline{\theta}^{o}$, the value of $\underline{\theta}$ which minimises $J(\underline{\theta})$ based on the observed values of $z(\underline{\theta})$ or $\underline{y}(\underline{\theta})$. Due to the cost associated to each experiment it is desirable to obtain the satisfactory result with as few observations as possible by processing them in an optimal fashion. At the same time the method should be recursive, such that $\underline{\theta}_k$ being the current estimate of $\underline{\theta}^{o}$ at the k-th step of the iteration, its value can be updated based on its previous value and the additional information provided by $z(\underline{\theta}_k{}^{*})$ or $\underline{y}(\underline{\theta}_k)$.

In some cases the approximation for $\underline{\theta}$ is found by finding the zero of $\nabla_{\theta} J(\underline{\theta})$ and therefore minimising

$$J(\underline{\theta}) = E\left(z(\underline{\theta});\ \underline{\theta}\right) \quad .$$

The problem can be formulated as : find the solution of the equation

$$\nabla_{\theta} J(\underline{\theta}) = \underline{r}(\underline{\theta}) = E\left(\underline{y}(\underline{\theta});\ \underline{\theta}\right) = 0 \quad , \tag{4.7}$$

in which $\underline{r}$ is an unknown function and $\underline{y}(\underline{\theta})$ a random vector. The main assumption made is that for every $\underline{\theta}$ a random vector $\underline{y}(\underline{\theta})$ is observable, where

$$\underline{r}(\underline{\theta}) = E\left(\underline{y}(\underline{\theta});\ \underline{\theta}\right) \tag{4.8}$$

has an unique zero at $\underline{\theta} = \underline{\theta}^{o}$. If $\underline{r}(\underline{\theta})$ were known, the algorithm for finding the zero of $\underline{r}(\underline{\theta})$ would be

$$\underline{\theta}_{k+1} = \underline{\theta}_k - \rho \underline{r}(\underline{\theta}_k) \, . \tag{4.9}$$

As $\underline{r}(\underline{\theta})$ is unknown, its value is replaced function of $\underline{y}(\underline{\theta})$:

$$\underline{\theta}_{k+1} = \underline{\theta}_k - \rho_k \underline{y}(\underline{\theta}_k) \ . \tag{4.10}$$

Therefore, instead of using a fixed gain ρ at every step of the iteration, a variable gain ρ_k is used in order to compensate for the error involved by considering a random function $\underline{y}(\underline{\theta})$ instead of $\underline{r}(\underline{\theta})$. The gain ρ_k should satisfy the conditions :

$$\rho_k \geqslant 0$$

$$\sum_k \rho_k = \infty \tag{4.11}$$

$$\sum_k \rho_k^2 < \infty \ .$$

The algorithm mentioned above has the disadvantage that requires the need of observing values for $\underline{y}(\underline{\theta})$. An algorithm which finds the extremum of $J(\underline{\theta})$ based only on samples of the random variable $z(\underline{\theta})$ is the Kiefer - Wolfowitz algorithm, which estimates the stochastic gradient $\nabla_\theta z(\underline{\theta})$ computed from neighbouring values of the function. The algorithm has the form

$$\underline{\theta}_{k+1} = \underline{\theta}_k - \rho_k \hat{\nabla}_\theta z(\underline{\theta}_k) \ , \tag{4.12}$$

where $\hat{\nabla}_\theta z(\underline{\theta}_k)$ is the estimated gradient of $z(\underline{\theta})$. A possible way of finding it consists of observing 2n random samples of $z(\underline{\theta})$, which are $z(\underline{\theta}_k \pm c_k \underline{e}_j)$, $j = 1,\ldots,n$, $c_1, c_2, \ldots$, is a suitable chosen positive sequence,

$$\underline{e}_i = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{4.13}$$

are n-dimensional vectors having all the components zero except the i-th, which is 1. n represents the dimension of $\underline{\theta}$. The estimated gradient

is then

$$z(\underline{\theta}_k) = \sum_{j=1}^{n} \underline{e}_j \frac{1}{2c_k} \left[z(\underline{\theta}_k + c_k \underline{e}_j) - z(\underline{\theta}_k - c_k \underline{e}_j) \right] . \quad (4.14)$$

The coefficients ρ_k and c_k should fulfil the conditions

$$\begin{aligned} &\sum \rho_k = \infty \\ &\sum \rho_k c_k < \infty \\ &\sum \left(\frac{\rho_k}{c_k}\right)^2 < \infty \\ &\rho_k < d_1 c_k, \quad d_1 \in (0, \infty) . \end{aligned} \quad (4.15)$$

The condition

$$\sum \left(\frac{\rho_k}{c_k}\right)^2 < \infty$$

ensures that the sum of the squares of the correction terms is finite. A set of coefficients ρ_k and c_k having the aforementioned properties could be chosen as

$$\begin{aligned} &\rho_k = \frac{\rho_0}{k} \\ &c_k = \frac{c_0}{k^\alpha}, \quad \alpha \in (0, 0.5), \quad \rho_0, c_0 > 0 . \end{aligned} \quad (4.16)$$

Based on this algorithm a computer program was written for finding a linear approximation of a p.d.f. $p(\underline{x})$ given by

$$\hat{p}(\underline{x}) = \underline{\theta}' \underline{\varphi}(\underline{x}), \quad (4.17)$$

where

$$\underline{\varphi}(\underline{x}) = \begin{bmatrix} \varphi_1(\underline{x}) \\ \vdots \\ \varphi_n(\underline{x}) \end{bmatrix} \quad (4.18)$$

is a vector of known functions and

$$\underline{\theta} = \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_n \end{bmatrix} \tag{4.19}$$

is the vector of unknown coefficients to be determined. The program finds the value $\underline{\theta}^o$ which minimises the criterion

$$J(\underline{\theta}) = E\left(p(\underline{x}) - \underline{\theta}'\underline{\varphi}(\underline{x})\right)^2 . \tag{4.20}$$

The minimising value of $\underline{\theta}$ is

$$\underline{\theta}^o = E\left(\underline{\varphi}(\underline{x})\ \underline{\varphi}'(\underline{x})\right)^{-1} E\left(\underline{\varphi}(\underline{x})\ p(\underline{x})\right) . \tag{4.21}$$

As an observable random vector $\underline{y}(\underline{\theta},\underline{x})$, the expectation of which is the gradient $\nabla_\theta J(\underline{\theta})$, cannot be obtained, another observable vector is considered, having the property

$$E\left(\underline{y}(\theta_k,\underline{x}_k);\ \underline{x}_k\right) = \nabla_\theta J(\underline{\theta})\Big|_{\underline{\theta}=\underline{\theta}_k} + c_k\, g_k(\underline{x}_k) , \tag{4.22}$$

with the bias term vanishing as k approaches infinity. An empirical density function $z_k(\underline{x})$ is considered such that

$$z_k(\underline{x}) = 2^{-m} c_k^{-m} \sum_{e}\left(\prod_{j=1}^{m} e_j\right) F_{k-1}(\underline{x} + c_k\underline{e}) , \tag{4.23}$$

where

$$\underline{e} = \begin{bmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_m \end{bmatrix},\ e_i = +1 \text{ or } -1 , \tag{4.24}$$

$$\underline{x} + c_k\underline{e} = \begin{bmatrix} x_1 + c_k e_1 \\ \cdot \\ \cdot \\ \cdot \\ x_m + c_k e_m \end{bmatrix} \tag{4.25}$$

and $F_k(\underline{x})$ is an empirical distribution function, the derivative of which is $z_k(\underline{x})$. The summation is carried out over all 2^m possible combinations of $\underline{e}$. One empirical density function $F_k(\underline{x})$ could be chosen as being

$$F_k(\underline{x}) = \frac{1}{k-1} \sum_{j=1}^{k-1} \Omega_x(x_j) \ , \qquad (4.26)$$

$$\Omega_x(\underline{y}) = \begin{cases} 1 \text{ if } x_k < y_k \quad \forall\, k = 1,\ldots,m \\ \\ 0 \text{ otherwise} \ . \end{cases}$$

In fact, $F_k(\underline{x})$ represents the ratio between the number of samples which, at the k-th step of the iteration, have all their components less than the corresponding components of $\underline{x}$ and the total number of samples considered up to the k-th step.

The flow-chart of the program is extremely simple and self-explanatory (Fig. 4.4). No such program has been found in any subroutine library and therefore the listing of the program which is based on the algorithm described above is given in Appendix A3. The main feature of this program consists in the minimum number of calculations required to update $\underline{\theta}_k$ at various steps, which makes it suitable for applications necessitating frequent p.d.f. estimation.

The program has been tested by using a test sample belonging to the distribution

$$p(x) = \theta_1\, p_G(x) + \theta_2\, p_I(x) \ , \qquad (4.27)$$

with

$$p_G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad (4.27a)$$

$$p_I(x) = \frac{\lambda^2}{4} \exp\left(-\lambda\sqrt{|x|}\right), \qquad (4.27b)$$

$$\theta_1 = 0.75 \ ,$$

$$\theta_2 = 0.25 \ .$$

The results produced by the program are

$$\theta_1' = 0.729$$

$$\theta_2' = 0.271 \ .$$

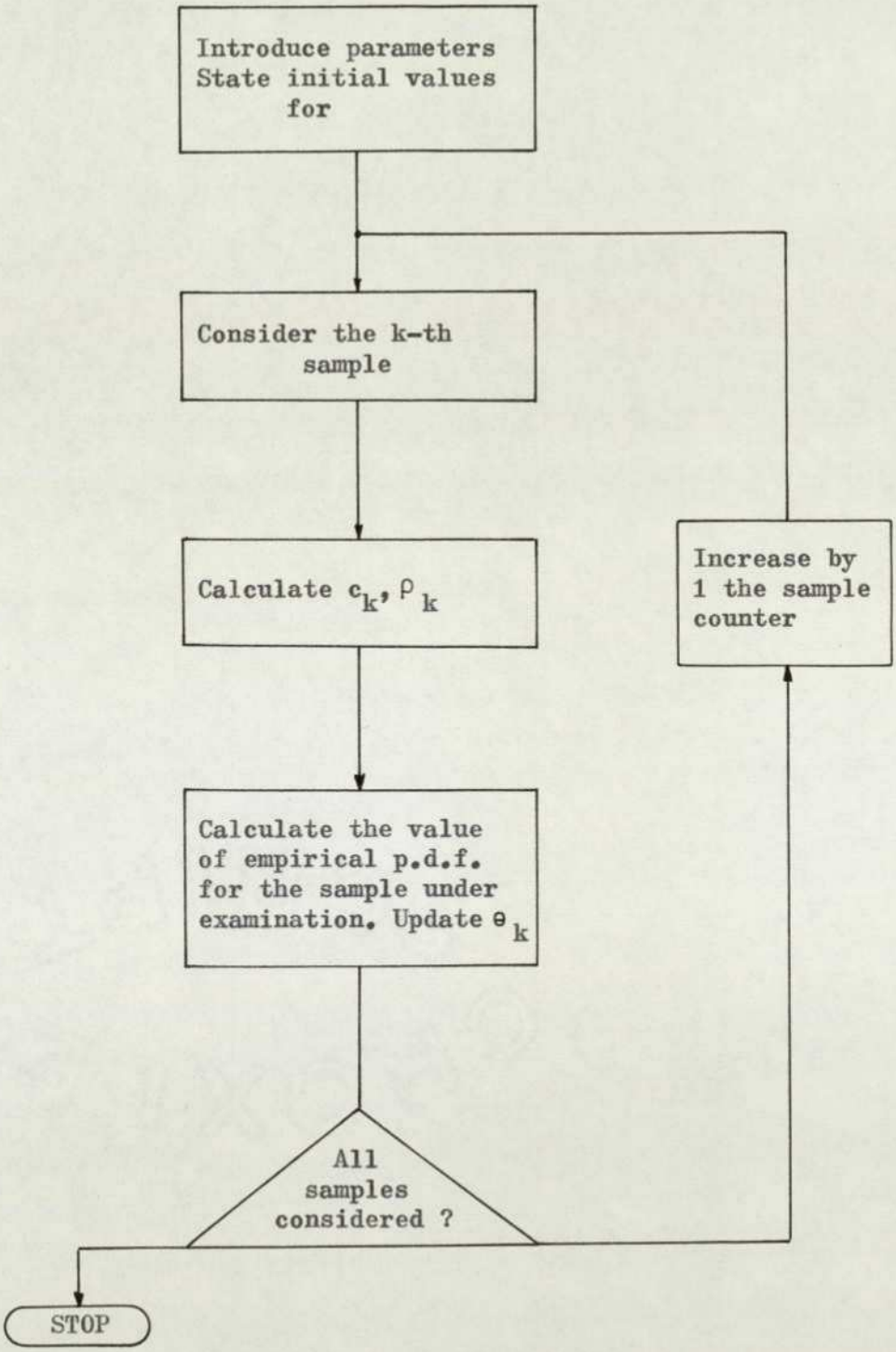

Introduce parameters
State initial values
for
Consider the k-th
sample
Calculate c_k, ρ_k
Calculate the value
of empirical p.d.f.
for the sample under
examination. Update θ_k
All
samples
considered ?
STOP
Increase by
1 the sample
counter

Fig. 4.4

The test sample has been generated by using the algorithm with the flow-chart shown in Fig. 4.5. The algorithm has the advantage of being able to generate samples distributed according to any law. The only elements which have to be supplied are the number of required samples and their desired p.d.f. The principle of operation is illustrated in Fig. 4.5a. Random numbers y_i between 0 and 1 are generated. If $P(x_i)$ is the required probability, the numbers x_i are calculated from the equation $y_i = P(x_i)$.

The results show that the method of stochastic approximation produce very good results, combining a reasonable accuracy with a short computational time. The speed of the algorithm could be increased by using a second order scheme. The method was tested for univariate distribution only, mainly due to the difficulties of generating test samples belonging to multivariate distributions. One can see no reason of failure for such cases since the estimation of the p.d.f. of one variable is a particular case of multivariate distribution estimation.

A last comment should be added about the choice of the function $\underline{\theta}$. For several reasons, Hermite polynomials have been prefered, chiefly due to their property of orthonormality and to the fact that the presence of the factor $\exp(-x^2/2)$ suggests a similarity with Gaussian distribution. If the variable x is replaced by x/σ, where σ is calculated as if the sample had Gaussian distribution, all the coefficients $\underline{\theta}$ but θ_1 become zero when the sample under analysis comes from a purely Gaussian process. This observation is used in the next section, in which a way to estimate the p.d.f. of a multivariate distribution, based on the stochastic approximation of the marginal p.d.f., is proposed.

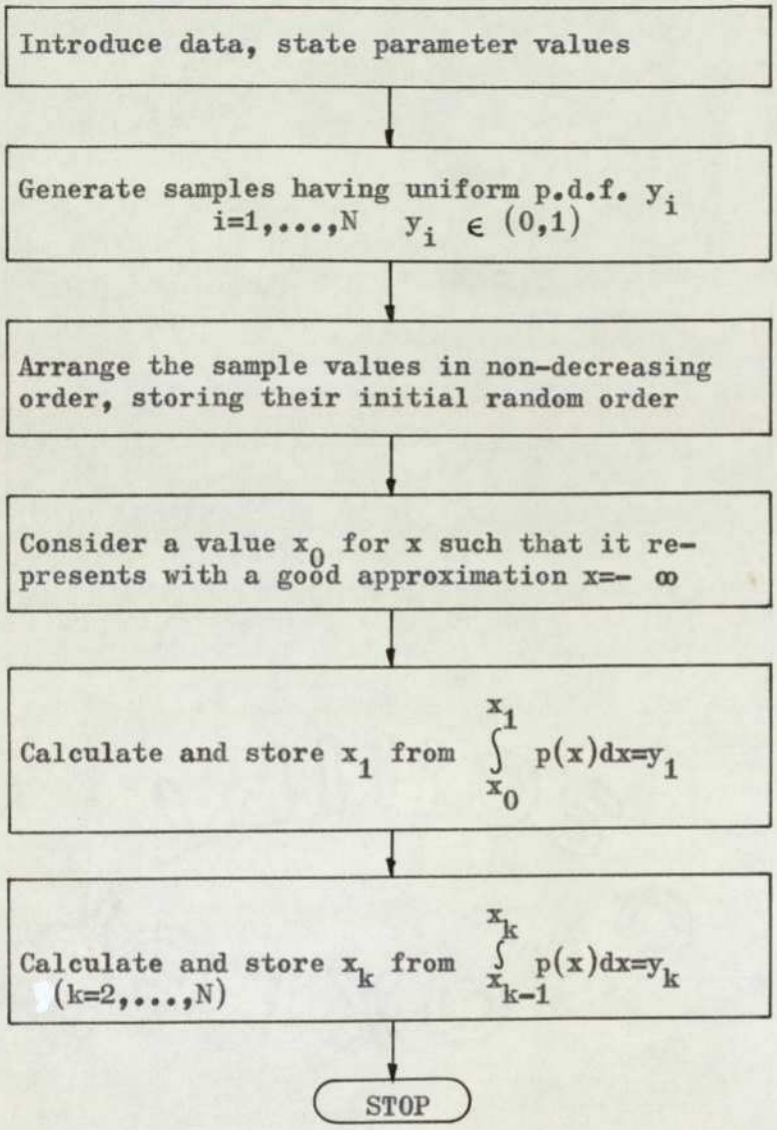


Fig. 4.5

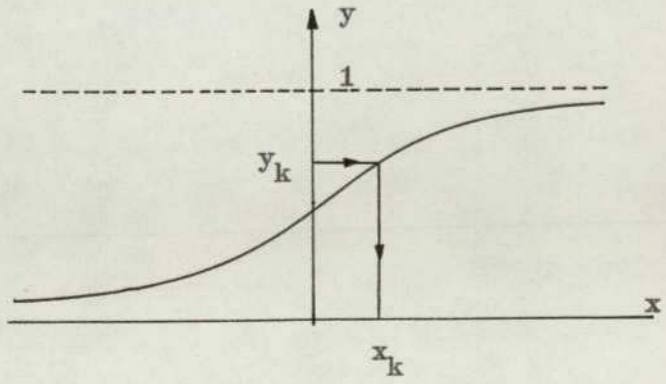


Fig. 4.5a

4.4 The estimation of the joint probability distribution function of the non-Gaussian noise

The next problem after the marginal p.d.f.s have been estimated is to find a mathematical expression for the joint p.d.f. The technique to find such an expression involves merely algebraic calculations and does not use any standard method similar to those of statistical inference. This approach is justified by the fact that, as previously mentioned, the method of stochastic approximation estimates the marginal p.d.f.s as a linear combination of known functions, such that the mean square error between the true density and that as obtained by stochastic approximation is minimised. Therefore, seeking for the expression of the joint p.d.f. a mathematical function having the standard properties of a p.d.f. will implicitly take into account the minimisation process used when the marginal distributions were determined.

Due to reasons explained before, a linear combination of orthonormal functions, built on the base of Hermite polynomials, has been considered in determining the marginal p.d.f.s, which have the form :

$$p(x) = \left(\sum_{i=0}^{n} \theta_{2i} \, H_{2i}(x) \right) \, p_G(x) \, , \qquad (4.28)$$

where θ_{2i} are coefficients determined by stochastic approximation, $H_{2i}(x)$ are Hermite polynomials and $p_G(x)$ is a function having the aspect of a Gaussian probability density function.

The mechanism of determining a joint probability density function, having marginal densities of the form shown above will be exemplified in what follows.

Consider two variables x_1 and x_2. If their p.d.f.s are estimated using two Hermite polynomials, one can write :

$$p(x_1) = \left(\theta_{01} H_0(x_1) + \theta_{21} H_2(x_1) \right) \, p_G(x_1) = (a_1 + b_1 x^2) \, p_G(x_1) \qquad (4.29a)$$

$$p(x_2) = \left(\theta_{02} H_0(x_2) + \theta_{22} H_2(x_2) \right) p_G(x_2) = (a_2 + b_2 x_2^2)\, p_G(x_2) \tag{4.29b}$$

Assume for their joint p.d.f. an expression of the form

$$p(x_1,x_2) = (A + \underline{x}' \underline{D}\, \underline{x})\, p_G(x_1,x_2) , \tag{4.30}$$

where A and the elements of the matrix

$$\underline{D} = \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} \tag{4.31}$$

are unknown coefficients to be determined,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.32}$$

and $p_G(\underline{x})$ a function having the form of a multivariate Gaussian probability density function :

$$p_G(\underline{x}) = \frac{1}{2\pi |C|^{1/2}} \exp\left(-\frac{1}{2} \underline{x}' \underline{C}\, \underline{x} \right) \tag{4.33}$$

with

$$\underline{C} = \begin{bmatrix} c_1^2 & c_{12} \\ c_{12} & c_2^2 \end{bmatrix} \tag{4.34}$$

$|C| = \det(\underline{C})$

The elements c_1^2, c_2^2, c_{12} of the matrix $\underline{C}$ are determined from the available sample as if the joint density is purely Gaussian. The advantage of making this assumption is obvious : when the samples represent a Gaussian process, the coefficients d_1^2, d_2^2 and d_{12} become zero and A becomes 1. In other words, the elements of the matrix $\underline{D}$ represent an empirical measure of the non-Gaussian character of the p.d.f. of the sample under consideration.

$p(x_1,x_2)$ must satisfy two major conditions in order to represent a

multivariate p.d.f., namely :

$$\text{(i)} \quad \int_{-\infty}^{+\infty} p(x_1,x_2)\, dx_1 = p(x_2) \text{ and } \int_{-\infty}^{+\infty} p(x_1,x_2)\, dx_2 = p(x_1) \qquad (4.35)$$

$$\text{(ii)} \quad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x_1,x_2)\, dx_1\, dx_2 = 1 \quad . \qquad (4.36)$$

It should be mentioned that once the condition (i) is satisfied, the condition (ii) is automatically fulfilled, due to the way in which $p(x_1)$ and $p(x_2)$ have been estimated.

After the integrals expressing the condition (i) are calculated, one obtains :

$$p(x_2) = \left((A + \frac{1}{c_1^2} d_1^2) + (d_2^2 - 2\frac{c_{12}}{c_1^2} d_{12} + \frac{c_{12}^2}{c_1^4} d_1^2)\, x_2^2 \right) p_G(x_2) \qquad (4.37a)$$

$$p(x_1) = \left((A + \frac{1}{c_2^2} d_2^2) + (d_1^2 - 2\frac{c_{12}}{c_2^2} d_{12} + \frac{c_{12}^2}{c_2^4} d_2^2)\, x_1^2 \right) p_G(x_1) \quad . \qquad (4.37b)$$

By identification with the expressions for the marginal p.d.f.s, as found by using the method of stochastic approximation, one obtains the following relations :

$$\begin{aligned} A + w_1 &= a_2 \\ A + w_2 &= a_1 \\ c_2^2 w_2 + h_1^2 w_1 - 2\frac{c_2}{c_1} h_1 w_{12} &= b_2 \\ c_1^2 w_1 + h_2^2 w_2 - 2\frac{c_1}{c_2} h_2 w_{12} &= b_1 \quad , \end{aligned} \qquad (4.38)$$

with the notations

$$w_1 = \frac{d_1^2}{c_1^2}$$

$$w_2 = \frac{d_2^2}{c_2^2}$$

$$w_{12} = \frac{d_{12}}{c_1 c_2} \qquad (4.39)$$

$$h_1 = \frac{c_{12}}{c_1}$$

$$h_2 = \frac{c_{12}}{c_2} \quad .$$

One can see that this set of four simultaneous equations produces unique solutions for the unknown parameters A, d_1^2, d_2^2, d_{12}. By solving it, one obtains the following expressions :

$$A = \frac{c_1^2 c_2^2 (h_1 a_2 - h_2 a_1) + h_1 h_2 (c_2^2 h_2 a_2 - c_1^2 h_1 a_1) + (c_1^2 h_2 b_2 - c_2^2 h_1 b_1)}{c_1^2 c_2^2 (h_1 - h_2) + h_1 h_2 (c_1^2 h_1 - c_2^2 h_2)}$$

$$w_1^2 = (a_2 - A)\, c_1^2$$

$$w_2^2 = (a_1 - A)\, c_2^2 \qquad (4.40)$$

$$w_{12} = \frac{c_1^2 c_2^2 (h_1 w_1 + h_2 w_2) + h_1 h_2 (c_1^2 h_1 w_1 + c_2^2 h_2 w_2) - (c_1^2 h_2 b_2 + c_2^2 h_1 b_1)}{4 h_1 h_2}$$

Unfortunately, these expressions, already complicated, become even more complex when the number of variables of the multivariate probability density function increases or the approximation uses more Hermite polynomials. However, for the case of one-parameter nonlinear detector, the performance of which is suitable in most industrial applications, only two variables are required for the expression of the joint p.d.f.

4.5 The analysis of the communication channel affected by Gaussian noise

The main purpose of any analysis of the communication channel is to find the best possible way to separate at the receiver the transmitted signal s(t) from the random noise n(t), permanently present on the channel. In other words, the problem consist of deciding between the two hypotheses, namely :

$$H_0 : z(t) = n(t) \tag{4.41}$$

$$H_1 : z(t) = s(t) + n(t) \ , \tag{4.42}$$

where z(t) is the received signal. One assumes that H_1 is valid when a mark is transmitted and H_0 is valid when a space is transmitted. A possible way of realising the detection process when s(t) is known consists in the use of sampling technique, as developed by Shannon (I. Selin, 1965). One can write :

$$n(t) = \underset{k \to \infty}{\text{l.i.m.}} \sum_{i=-k}^{k} n_i \frac{\sin 2\pi w (t - t_i)}{2\pi w (t - t_i)} \ . \tag{4.43}$$

By

$$\underset{n \to \infty}{\text{l.i.m.}}(x_n) = x$$

one understands the "mean in the limit". Although

$$\lim_{n \to \infty} (x_n) \neq x \ ,$$

the mean of the random variable x_n approaches the value x as n increases without bounds, thus

$$\lim_{n \to \infty} E(|x_n - x|^2) = 0 \ .$$

However, n_i are not always random independent variables and therefore the correlation of the samples must be taken into account if the sampling theorem should be used for finding the probability measure for the

continuous waveform of n(t), expressed as a product of known p.d.f.s. In order to avoid the complications produced by considering the correlation functions, the Karhunen-Loeve expansion will be used. The approach generally follows that recommended by (W.B. Davenport, W.L. Root, 1958; I. Selin, 1965).

The Karhunen-Loeve expansion expresses the noise n(t) as

$$n(t) = \underset{k\to\infty}{\text{l.i.m.}} \sum_{i=1}^{k} n_i \, \sigma_i \, \varphi_i(t) \, , \tag{4.44}$$

where $\varphi_i(t)$ are deterministic functions of time, σ_i are deterministic coefficients independent of time, n_i are random and not function of time, being Gaussian distributed if n(t) is so. A supplementary condition imposes that $\varphi_i(t)$ should be orthonormal over the sampling period [0 , T] :

$$\int_0^T \varphi_i(t) \, \varphi_k^*(t) \, dt = \delta_{ik}, \tag{4.45}$$

δ_{ik} being the Kronecker delta.

If the random coefficients n_i are normalised and uncorrelated (i.e. $E(n_i n_k^*) = \delta_{ik}$), the joint probability can be written as the product of the individual probabilities. In this case, the covariance function of n(t) has the expression

$$R(t_1,t_2) = E(n(t_1) \, n^*(t_2)) = \sum_{k=1}^{\infty} |\sigma_k|^2 \, \varphi_k(t_1) \, \varphi_k^*(t_2) \quad , \tag{4.46}$$

as $E(n_i n_k^*) = \delta_{ik}$. Due to the property of orthonormality, by using the expression of $R(t_1,t_2)$ as in (4.46) :

$$\int_0^T R(t_1,t_2) \, \varphi_i(t_2) \, dt_2 = |\sigma_i|^2 \, \varphi_i(t_1) \tag{4.47}$$

One can see that $\varphi_i(t_1)$ and $|\sigma_i|^2$ are the eigen functions and eigen values of the kernel $R(t_1,t_2)$.

Real and analytic signals. Consider the signal

$$w(t) = u(t) + j\, v(t), \tag{4.48}$$

u(t) and v(t) being real signals. Assume u(t) such that

$$\int_{-\infty}^{+\infty} u^2(t)\, dt < \infty \ .$$

The Fourier transform of u(t) is U(ω) :

$$U(\omega) = \int_{-\infty}^{+\infty} u(t)\, \exp(-j\omega t) dt, \tag{4.49a}$$

with

$$u(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} U(\omega)\, \exp(j\omega t)\, d\omega \ . \tag{4.49b}$$

u(t) being real, $U(-\omega) = U^*(\omega)$. The signal w(t) is called analytic if v(t) is the Hilbert transform of u(t) :

$$v(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u(\tau)}{t-\tau}\, d\tau = \frac{1}{\pi t} * u(t) \quad . \tag{4.50}$$

Therefore

$$V(\omega) = \mathcal{F}\left\{\frac{1}{\pi t} * u(t)\right\} = -\, j\, U(\omega)\, \mathrm{sign}\,\omega \quad , \tag{4.51}$$

as

$$\mathcal{F}\left\{\frac{1}{\pi t}\right\} = -\, j\, \mathrm{sign}\omega \quad .$$

Then the Fourier transform of w(t) is

$$W(\omega) = U(\omega)\, (1 + \mathrm{sign}\omega) \quad , \tag{4.52}$$

or

$$W(\omega) = \begin{cases} 2\, U(\omega) , & \omega \in (0, \infty) \\ U(\omega) & \omega = 0 \\ 0 & \omega \in (-\infty, 0) \end{cases} \tag{4.53}$$

Using the Parseval's theorem :

$$\int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\omega)|^2 \, d\omega \qquad (4.54)$$

one can write

$$\int_{-\infty}^{+\infty} |w(t)|^2 \, dt = \int_{-\infty}^{+\infty} \left(u^2(t) + v^2(t) \right) \, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2 \, d\omega + \frac{1}{2\pi} \int_{-\infty}^{+\infty} |V(\omega)|^2 \, d\omega \quad . \qquad (4.55)$$

Since $U(\omega) = V(\omega)$, the energy is equally distributed between the two components u(t) and v(t). Other properties of w(t) include : w(t) integrable quadratic and u(t) and v(t) linearly independent .

Now, if u(t) is a signal representing a real random process being stationary with zero mean, these properties will be transfered to v(t) by the fact that the Hilbert transform is a linear operator. The correlation function of w(t) is given by

$$E\left(w(t) \, w^*(t-\tau) \right) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} w(t) \, w^*(t-\tau) \, d\tau =$$

$$= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left(u(t) \, u(t-\tau) + v(t) \, v(t-\tau) \right) dt +$$

$$+ j \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left(v(t) \, u(t-\tau) - u(t) \, v(t-\tau) \right) dt =$$

$$= R_u(\tau) + R_v(\tau) + j \left(R_{vu}(\tau) - R_{uv}(\tau) \right) \quad . \qquad (4.56)$$

By using the frequency domain

$$R_u(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} u(t) \, u(t-\tau) \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} U(\omega) \left(U(\omega) \cdot \right.$$

$$\left. \cdot \exp(-j\omega\tau) \right)^* \, d\omega =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2 \exp(j\omega\tau)\, d\omega \quad . \tag{4.57}$$

In a similar manner

$$R_v(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t)\, v(t-\tau)\, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} V(\omega) \left(V(\omega) \exp(-j\omega\tau) \right)^* d\omega =$$

$$= \frac{-j^2}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2 \exp(j\omega\tau)\, \mathrm{sign}^2\omega\, d\omega \,, \tag{4.58}$$

since

$$V(\omega) = -j\, U(\omega)\, \mathrm{sign}\,\omega \,;$$

therefore

$$R_v(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2 \exp(j\omega\tau)\, d\omega \quad . \tag{4.59}$$

Following the same procedure, one can show that

$$R_{uv}(\tau) = -R_{vu}(\tau) \quad . \tag{4.60}$$

Indeed :

$$R_{uv}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(t)\, v(t-\tau)\, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} U(\omega) \left(-j\, U(\omega)\, \mathrm{sign}\,\omega \exp(-j\omega\tau) \right)^* d\omega =$$

$$= \frac{j}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2\, \mathrm{sign}\,\omega \exp(j\omega\tau)\, d\omega$$

$$R_{vu}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t)\, u(t-\tau)\, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} - j\, U(\omega) \left(U(\omega)\, \exp(-j\omega\tau) \right)^* d\omega =$$

$$= -\frac{j}{2\pi} \int_{-\infty}^{+\infty} |U(\omega)|^2\, \mathrm{sign}\omega\, \exp(j\omega\tau)\, d\omega \ .$$

Therefore, the correlation function of the random analytic signal w(t) has the properties :

(i) $R_w^*(\tau) = R_w(-\tau)$ (4.61)

(ii) $R_w(0) = R_u(0) < \infty$ (4.62)

(iii) $R_w(\tau)$ positive definite.

After some properties of the analytic signals have been analysed, the next step consists of producing a set of observables which represent the process over the sampling interval, approach mentioned by (W.B.Davenport, W.L.Root, 1958). For this purpose, Fredholm integrals are used :

$$\int_a^b R(t,t')\, \varphi_i(t')\, dt' = \lambda_i\, \varphi_i(t) \ , \quad t \in [a\ ,\ b] \tag{4.63}$$

$\varphi_i(t)$ being the eigenfunctions and λ_i the eigenvalues associated with the kernel $R(t,t')$. Under these circumstances, $\varphi_i(t)$ form a complete orthonormal basis and therefore

$$\int_a^b \varphi_i(t)\, \varphi_k^*(t)\, dt = \delta_{ik} \ ,$$

where δ_{ik} is the Kronecker delta. According to Mercer's theorem, a set of eigenfunctions and eigenvalues as specified above fully represent the kernel $R(t,t')$:

$$R(t,t') = \sum_{i=1}^{\infty} \lambda_i\, \varphi_i(t)\, \varphi_i^*(t') \ . \tag{4.64}$$

Consider a random process $\underline{z}(t)$

$$\underline{z}(t) = \begin{bmatrix} z_1(t) \\ \cdot \\ \cdot \\ \cdot \\ z_n(t) \end{bmatrix} \tag{4.65}$$

having the correlation function

$$R(t,t') = R^*(t',t) \ .$$

One can prove that the eigenvalues associated to it are real and positive :

$$E\left(\left|\int_a^b z(t)\,\varphi_i^*(t)\,dt\right|^2\right) = E\left(\left|z_i(t)\right|^2\right) = 2\lambda_i = 2\lambda_i^* > 0 \quad . \tag{4.66}$$

This is due to the fact that $z(t)$ is assumed to be integrable square. The sum of the eigenvalues is bounded :

$$\int_a^b \frac{1}{2} E\left(\left|z(t)\right|^2\right) dt = \sum_{i=1}^{\infty} \lambda_i = \frac{1}{2} \sum_{i=1}^{\infty} E\left(\left|z_i(t)\right|^2\right) . \tag{4.67}$$

The following properties could be easily found :

$$E\left(z_i(t)\right) = E \int_a^b z(t)\,\varphi_i^*(t)\,dt \quad = 0 \tag{4.68}$$

$$E\left(z_i^*(t)\right) = E \int_a^b z^*(t)\,\varphi_i(t)\,dt \quad = 0 \tag{4.69}$$

$$E\left(z_i^*(t)\ z_k(t)\right) = \int_a^b\int_a^b \lambda_i(t)\ E\left(z^*(t)\ z(t')\right) \varphi_k^*(t')\,dt\,dt' =$$

$$= \int_a^b\int_a^b \lambda_i(t)\ 2R(t,t')\,\varphi_k^*(t')\,dt\,dt' =$$

$$= 2\int_a^b \varphi_k^*(t')\,\lambda_i\,\varphi_i(t')\,dt' = 2\lambda_i\,\delta_{ik} \quad . \tag{4.70}$$

The last result shows that the samples are uncorrelated. For a stationary random process

$$E\left(z_i(t)\ z_k(t)\right) = E\left(z_i(t)\ z_k^*(t)\right) = 0\ . \tag{4.71}$$

Assuming that

$$z(t) = u(t) + j\ v(t)\ , \tag{4.72}$$

one can see that

$$E\left(u_i(t)\ u_k(t)\right) = E\left(v_i(t)\ v_k(t)\right) = \lambda_i\ \varphi_{ik} \tag{4.73a}$$

and

$$E\left(u_i(t)\ v_k(t)\right) = E\left(u_k(t)\ v_i(t)\right) = 0\ . \tag{4.73b}$$

Consider now a random process $z(t)$ and a set of m filters. The output of the filters will be

$$\underline{f} = \int_a^b z(t)\ \underline{h}^*(b - t)\ dt, \tag{4.74}$$

where

$$\underline{f} = \begin{bmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_m \end{bmatrix} \tag{4.75}$$

$$\underline{h}(t) = \begin{bmatrix} h_1(t) \\ \cdot \\ \cdot \\ \cdot \\ h_m(t) \end{bmatrix}\ . \tag{4.76}$$

$h_k(t)$ is the transfer function of the i-th filter. One can express $h_k(t)$ by using a set of eigenfunctions and eigenvalues :

$$\underline{h}(b - t) = \underset{M\to\infty}{\text{l.i.m.}} \sum_{k=1}^{M} ({}^k\underline{h})\ \varphi_k(t)\ , \tag{4.77}$$

where

$$({}^k h) = \int_a^b h(b - t)\ \varphi_k^*(t)\ dt\ , \tag{4.78}$$

therefore

$$\underline{f} = \underset{M\to\infty}{\text{l.i.m.}} \sum_{k=1}^{M} ({}^{k}h)^* \int_a^b z(t)\, \varphi_k^*(t)\, dt \quad . \tag{4.79}$$

The moments of $\underline{f}$ are given by :

$$E(f_i) = E(f_i^*) = 0 \tag{4.80a}$$

$$E(f_i\, f_k) = 2 \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} \lambda_l\, ({}^{l}h_i)\, ({}^{l}h_k)^* \quad . \tag{4.80b}$$

The expression for the second moment shows that the samples are correlated, as one should expect. One can show that by changing the orthogonal base φ into some other nonorthogonal base, the samples still remain correlated. If the new base is given by

$$\psi_i(t) = \sum_{k=1}^{\infty} a_{ik}\, \varphi_k(t) \quad , \tag{4.81}$$

then

$$f_i = \int_a^b z(t)\, \psi_i^*(t)\, dt = \sum_{k=1}^{\infty} a_{ik}^* \int_a^b z(t)\, \varphi_k^*(t)\, dt \quad . \tag{4.82}$$

Thus

$$E(f_i) = E(f_k) = 0 \quad , \tag{4.83a}$$

$$E(f_i^* f_k) = 2 \sum_{l=1}^{\infty} \lambda_l\, a_{il}\, a_{kl} \quad . \tag{4.83b}$$

Note that a_{il} should be chosen such that

$$\sum_{l=1}^{\infty} |a_{il}|^2 < \infty \quad .$$

Maximum likelihood estimators. Consider now a deterministic signal transmitted over a channel affected by Gaussian noise. Under these circumstances, as the mean vector and the covariance matrix fully specify the Gaussian distribution, the knowledge of $E(\underline{w})$ and $E(\underline{w}^2)$ will characterise the process completely. $\underline{w}$ is an analytic signal for which

M samples are known :

$$\underline{w} = \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_M \end{bmatrix} \quad . \tag{4.84}$$

Consider

$$E(\underline{w}) = \begin{bmatrix} E(w_1) \\ \cdot \\ \cdot \\ \cdot \\ E(w_n) \end{bmatrix} = 0 \tag{4.85}$$

$$E(\underline{w}^2) = \frac{1}{2} \left\| E(w_i^* w_k) \right\| \quad . \tag{4.86}$$

For a signal $\underline{u}(t)$ consisting only of real samples $u_i(t)$ the covariance matrix is symmetric and Hermitian, since :

$$\frac{1}{2} E\left(u_i(t)\, u_k^*(t)\right) = \frac{1}{2} E\left(u_i^*(t)\, u_k(t)\right) = \frac{1}{2} E\left(u_i(t)\, u_k(t)\right) \quad . \tag{4.87}$$

If the transmitted signal $s(t)$ is corrupted by noise $n(t)$ one can write the expression of the received signal $z(t)$ as

$$z(t) = A\, s(t) + \sqrt{N}\, n(t) \ , \tag{4.88}$$

A and N being unknown constants. If one considers a finite number of samples M, the real samples are

$$z_i = \sum_{k=1}^{\infty} a_{ik} \int_a^b \Big(A\, s(t) + N\, n(t) \Big) \varphi_k(t)\, dt \ , \tag{4.89}$$

having as components

$$s_i = \sum_{k=1}^{\infty} a_{ik} \int_a^b s(t)\, \varphi_k(t)\, dt \tag{4.89a}$$

and

$$n_i = \sum_{k=1}^{\infty} a_{ik} \int_a^b n(t)\, \varphi_k(t)\, dt \quad . \tag{4.89b}$$

The sample mean is

$$E\left(\underline{z}(t)\right) = A\ \underline{s}(t) \tag{4.90}$$

The sample covariance matrix is given by

$$E\left\{\left(\underline{z}(t) - E\left(\underline{z}(t)\right)\right)\left(\underline{z}(t) - E\ \underline{z}(t)\right)\right\} = N\ E(\underline{n}\ \underline{n}') =$$

$$= N \left\| \sum_{l=1}^{M} \lambda_l\ a_{il}\ a_{kl} \right\| = N\underline{R}\ , \tag{4.91}$$

R being the covariance matrix of the noise.

The p.d.f. of the received signal is then

$$p(\underline{z};A,N) = \frac{1}{\sqrt{(2\pi)^M\ N^M\ \det R}} \exp\left(-\frac{1}{2N}\ (\underline{z} - A\ \underline{s})'\underline{R}^{-1}(\underline{z} - A\ \underline{s})\right) \quad . \tag{4.92}$$

If the process is analytic

$$z(t) = A\ s(t) + \sqrt{2N}\ n(t) \quad , \tag{4.93}$$

A being a complex unknown constant and N an unknown real constant. The samples are

$$z_i = \sum_{l=1}^{M} a_{il}^* \int_a^b \left(A\ s(t) + \sqrt{2N}\ n(t)\right) \varphi_l^*(t)\ dt\ , \tag{4.94}$$

having as components

$$s_i = \sum_{l=1}^{M} a_{il}^* \int_a^b s(t)\ \varphi_l^*(t)\ dt \tag{4.94a}$$

and

$$n_i = \sum_{l=1}^{M} a_{il}^* \int_a^b n(t)\ \varphi_l^*(t)\ dt\ . \tag{4.94b}$$

The sample mean and the sample covariance matrix are found in the same way as for the real process :

$$E\left(\underline{z}(t)\right) = A\ \underline{s}(t) \tag{4.95a}$$

$$\frac{1}{2} E\left\{\left(\underline{z}(t) - E\left(\underline{z}(t)\right)\right)^* \left(\underline{z}(t) - E\left(\underline{z}(t)\right)\right)\right\} =$$

$$= N\, E(\underline{z}^*\underline{z}') = N \left\| \sum_{l=1}^{M} \lambda_l\, a_{il}\, a_{kl}^* \right\| = N\, \underline{R} \quad . \tag{4.95b}$$

The p.d.f. of the received signal is

$$p(\underline{z};A,N) = \frac{1}{\sqrt{(2\pi)^M\, N^M \det R}} \exp\left(-\frac{1}{2N} (\underline{z} - A\,\underline{s})' \underline{R}^{-1} (\underline{z}^* - A\underline{s}^*) \right) \quad . \tag{4.96}$$

Once the p.d.f. of the received signal has been expressed, any standard procedure can be used for finding estimators for mean and variance. The need for such estimators arises from the necessity of expressing the estimator of the signal/noise ratio (SNR). If A and N can be estimated, then an estimator for SNR is

$$SNR = A^2/N \quad . \tag{4.97}$$

Maximum likelihood estimators will be used for A and N. For real signals the log-likelihood function L is

$$L = \ln p(\underline{z};A,N) = -\frac{1}{2N} (\underline{z} - A\,\underline{s})' \underline{R}^{-1} (\underline{z} - A\,\underline{s}) - \ln \sqrt{(2\pi)^M\, N^M \det R} \ , \tag{4.98}$$

the set of solutions A_e and N_e which maximises it are given by

$$\left.\frac{\partial L}{\partial A}\right|_{\substack{A = A_e \\ N = N_e}} = 0$$

$$\left.\frac{\partial L}{\partial N}\right|_{\substack{A = A_e \\ N = N_e}} = 0 \quad . \tag{4.99}$$

The partial derivative of L with respect to A gives

$$\frac{\partial L}{\partial A} = -\frac{1}{2N} \left(-\underline{s}' \underline{R}^{-1} (\underline{z} - A\,\underline{s}) - (\underline{z}' - A\,\underline{s}') \underline{R}^{-1} \underline{s} \right) \quad . \tag{4.100}$$

Therefore, the equation which gives the value of A_e is :

$$\underline{s}' \underline{R}^{-1} \underline{z} - \underline{s}' \underline{R}^{-1} A_e \underline{s} + \underline{z}' \underline{R}^{-1} \underline{s} - A_e \underline{s}' \underline{R}^{-1} \underline{s} = 0 \quad . \tag{4.101}$$

Due to the symmetry of R

$$A_e = \frac{\underline{z}' \, \underline{R}^{-1} \, \underline{s}}{\underline{s}' \, \underline{R}^{-1} \, \underline{s}} \quad , \tag{4.102}$$

or

$$A_e = \frac{1}{Q} \, \underline{z}' \, \underline{R}^{-1} \, \underline{s} \quad , \tag{4.102a}$$

where Q is the quadratic expression

$$Q = \underline{s}' \, \underline{R}^{-1} \, \underline{s} \quad . \tag{4.102b}$$

The problem is slightly more complicated for N_e. The partial derivative of L with respect to N gives

$$\frac{\partial L}{\partial N} = - \frac{1}{\sqrt{(2\pi)^M \, N^M \, \det R}} \cdot \frac{1}{2\sqrt{(2\pi)^M \, N^M \, \det R}} \, (2\pi)^M \, \det R \, M \, N^{M-1} +$$

$$+ \frac{1}{2N^2} \, (\underline{z}' - A \, \underline{s}') \, R^{-1} \, (\underline{z} - A \, \underline{s}) =$$

$$= - \frac{M}{2N} + \frac{(\underline{z}' - A \, \underline{s}') \, \underline{R}^{-1} \, (\underline{z} - A \, \underline{s})}{2N^2} = 0 \, . \tag{4.103}$$

For $N = N_e$ and $A = A_e$ $\frac{\partial L}{\partial N} = 0$, then

$$N_e = \frac{1}{M} (\underline{z}' - A_e \, \underline{s}') \, \underline{R}^{-1} \, (\underline{z} - A_e \, \underline{s}) \quad . \tag{4.104}$$

In order to find if the estimators are biased, their mean and variance are required. For A_e

$$E(A_e) = E\left\{ \frac{\underline{z}' \, \underline{R}^{-1} \, \underline{s}}{Q} \right\} = \frac{1}{Q} \, E(\underline{z}') \, \underline{R}^{-1} \, \underline{s} = \frac{1}{Q} \, A \, \underline{s} \, \underline{R}^{-1} \, \underline{s} = A \tag{4.105a}$$

$$E\left\{ A_e^{\,2} - E^2(A_e) \right\} = E\left\{ \frac{(\underline{z}' \, \underline{R}^{-1} \, \underline{s})^2}{Q^2} \right\} - A^2 =$$

$$= \frac{1}{Q^2} \, E\left\{ (\underline{z}' \, \underline{R}^{-1} \, \underline{s})^2 \right\} - A^2 = \frac{1}{Q} \left(A^2 Q + N \right) - A^2 = \frac{N}{Q} \quad . \tag{4.105b}$$

The quadratic term Q approaches infinity as b - a increases without bound, therefore A_e is a consistent estimator only for $b - a \longrightarrow \infty$, regardless how many samples are taken within this interval.

The mean and variance of N_e can be calculated in a similar manner:

$$E(N_e) = \frac{1}{M} \, E\left\{ (\underline{z}' - A_e \, \underline{s}') \, \underline{R}^{-1} \, (\underline{z} - A_e \, \underline{s}) \right\} =$$

$$= \frac{1}{M} E(\underline{z}' \underline{R}^{-1} \underline{z} - 2A_e \underline{z}' \underline{R}^{-1} \underline{s} + A_e^2 \underline{s}' \underline{R}^{-1} \underline{s}) =$$

$$= \frac{1}{M} E\left(\underline{z}' \underline{R}^{-1} \underline{z} - \frac{2}{Q}(\underline{z}' \underline{R}^{-1} \underline{s})^2 + \frac{1}{Q^2}(\underline{z}' \underline{R}^{-1} \underline{s})^2 Q\right) =$$

$$= \frac{1}{M} E\left(\underline{z}' \underline{R}^{-1} \underline{z} - \frac{1}{Q}(\underline{z}' \underline{R}^{-1} \underline{s})^2\right) \quad . \qquad (4.106)$$

The moment of $\underline{z}' \underline{R}^{-1} \underline{z}$ can be calculated replacing

$\underline{z} = A \underline{s} + \sqrt{N} \underline{n}$:

$$E(\underline{z}' \underline{R}^{-1} \underline{z}) = E(A^2 Q - 2A\sqrt{N} \underline{s}' \underline{R}^{-1} \underline{n} + N \underline{n}' \underline{R}^{-1} \underline{n}) =$$

$$= A^2 Q - 0 + M N \quad . \qquad (4.107)$$

The moment of the last term is the trace of the noise covariance matrix, therefore has the value M. The moment of $(\underline{z}' \underline{R}^{-1} \underline{s})$ was already calculated, therefore

$$E(N_e) = \frac{1}{M} (A_e^2 Q + M N - A_e^2 Q - N) = \frac{M - 1}{M} N \quad . \qquad (4.108a)$$

Therefore, N_e as it was calculated is a biased estimator of N. An unbiased estimator for N would be

$$N_e = \frac{1}{M - 1} (\underline{z}' - A_e \underline{s}') \underline{R}^{-1} (\underline{z} - A_e \underline{s}) \quad . \qquad (4.108)$$

The variance of N_e is given by

$$E\left(N_e^2 - E^2(N_e)\right) = \frac{1}{M - 1} E\left((\underline{z}' \underline{R}^{-1} \underline{z} - 2A_e \underline{z}' \underline{R}^{-1} \underline{s} + A_e^2 Q)^2\right) - N^2$$

Using the hints given in (A.D.Whalen, 1971) for claculating the moments of expressions such as $(\underline{u} \underline{A} \underline{v})^k$, one can find

$$\text{Var}(N_e) = \frac{2}{M - 1} N^2 \quad . \qquad (4.109)$$

This result shows N_e as a consistent estimator, since its value approaches zero as the number M of taken samples increases without bound.

Similar calculations carried out for samples obtained from analytic signals produce values for the estimators of A and N. However, in this case A is complex and therefore it should be replaced by two real constants :

$$A = r \exp(\varphi) \qquad (4.110)$$

and therefore

$$A_e = r_e \exp(\varphi_e) \quad . \tag{4.111}$$

The same procedure is used in finding the values of r,φ and N which maximise the log-likelihood function, given by :

$$L = - \ln \left((2\pi)^M N^M \det R \right) - \frac{1}{2N} (\underline{z} - A\underline{s})\, \underline{R}^{-1}\, (\underline{z}^* - A\underline{s}) \quad . \tag{4.112}$$

By calculating $\frac{\partial L}{\partial r}$, $\frac{\partial L}{\partial \varphi}$ and $\frac{\partial L}{\partial N}$ one obtains :

$$\begin{aligned}
\frac{\partial L}{\partial r} &= \frac{1}{2N}\ \underline{s}'\underline{R}^{-1}\underline{z}^* \exp(j\varphi) + \underline{z}'\underline{R}^{-1}\underline{s}^* \exp(-j\varphi) - 2\, r\, \underline{s}'\underline{R}^{-1}\underline{s}^* \\
\frac{\partial L}{\partial \varphi} &= \frac{jr}{2N}\ \underline{s}'\underline{R}^{-1}\underline{z}^* \exp(j\varphi) - \underline{z}'\underline{R}^{-1}\underline{s}^* \exp(-j\varphi) \\
\frac{\partial L}{\partial N} &= -\frac{M}{N} + \frac{1}{2N^2} (\underline{z}' - A\underline{s}')\, \underline{R}^{-1}\, (\underline{z}^* - A\underline{s}^*) \quad .
\end{aligned} \tag{4.113}$$

Solving the set of simultaneous equations given by

$$\begin{cases} \dfrac{\partial L}{\partial r} = 0 \\ \dfrac{\partial L}{\partial \varphi} = 0 \\ \dfrac{\partial L}{\partial N} = 0 \end{cases} \tag{4.114}$$

one finds :

$$r_e = \frac{1}{2Q} \left(\underline{s}'\underline{R}^{-1}\underline{z}^* \exp(j\varphi_e) + \underline{z}'\underline{R}^{-1}\underline{s}^* \exp(-j\varphi_e) \right) \tag{4.115}$$

$$\varphi_e = k\pi + \theta \tag{4.116}$$

$$N_e = \frac{1}{2M} (\underline{z}' - A_e\underline{s})\, \underline{R}^{-1}\, (\underline{z}^* - A_e\underline{s}^*) \quad , \tag{4.117}$$

where

$$\underline{z}'\underline{R}^{-1}\underline{s}^* = \rho \exp(j\theta) \tag{4.118}$$

$$Q = \underline{s}'\underline{R}^{-1}\underline{s}^* \, . \tag{4.119}$$

The estimator A_e for A has the mean and variance given by :

$$E(A_e) = A \tag{4.120a}$$

$$E\left(A_e^{\,2} - E^2(A_e) \right) = \frac{2N}{Q} \, . \tag{4.120b}$$

One can see that A_e is again unbiased and its consistency is not affected

by the number of samples taken in the interval (b - a), but by the length of this interval, as the variance of A_e approaches zero when Q, then (b - a) increases without bound.

The expected value of N_e is :

$$E(N_e) = \frac{M - 1}{M} N \quad . \tag{4.121a}$$

Therefore, an unbiased estimator for N_e would be :

$$N_e = \frac{1}{2 (M - 1)} (\underline{z}' - A_e \underline{s}) \underline{R}^{-1} (\underline{z}^* - A_e \underline{s}^*) \quad . \tag{4.121}$$

The variance of the estimator is :

$$Var(N_e) = \frac{1}{M - 1} N^2 \quad , \tag{4.122}$$

therefore the chosen expression for N_e produces a consistent estimator.

The case of incoherent detection of known signals. Finally, consider the case of the incoherent detection of analytic signals, case which can be analysed by making use of the elements calculated so far. One will assume that the signals are completely known, except their phase at the receiver. The p.d.f. of the phase is assumed to be uniform in the interval between 0 and 2π (R.A.Finkler, 1970).

Assume that two possible signals are sent : $s_k(t)$, with k = 0 or k = 1. The received signal will be :

$$z(t) = A\, s_k(t) + \sqrt{2N}\, n(t) \quad . \tag{4.123}$$

Using the Karhunen-Loeve expansion one can produce a set of eigenfunctions $\varphi_i(t)$ and eigenvalues λ_i, starting from the autocorrelation function $R(t - t')$ of the noise (the noise is assumed to be stationary). The eigenfunctions and eigenvalues are given by the integral equation :

$$\int_a^b R(t - t') \underline{\varphi}(t')\, dt' = \lambda \underline{\varphi}(t) \quad . \tag{4.124}$$

The observables at the receiver are

$$z_l = A\, s_{kl} + \sqrt{2N}\, n_l \quad , \tag{4.125}$$

with

$$n_l = \int_a^b n(t)\ \varphi_l^*(t)\ dt \qquad (4.125a)$$

$$s_{kl} = \int_a^b s_k(t)\ \varphi_l^*(t)\ dt\ . \qquad (4.125b)$$

The decision about the transmitted signal is taken based on the a posteori probability, given the received signal $\underline{z}$. Then :

$$p(s_k\ ;\ \underline{z},\ A,\ N) = p(s_k)\ p(\underline{z}\ ;\ s_k,\ A,\ N)/p(\underline{z})\ . \qquad (4.126)$$

The expectation of $\underline{z}$ is :

$$E(\underline{z}) = A\ \underline{s}_k \qquad (4.127a)$$

and its variance is

$$\frac{1}{2}\ E\left\{\left(z^* - E(\underline{z}^*)\right)\left(\underline{z} - E(\underline{z})\right)'\right\} = N\ E(\underline{n}^*\ \underline{n}') = N\ \underline{R} = N\left\|\lambda_i\ \delta_{ik}\right\| \qquad (4.127b)$$

Due to the independence of the observables one can write

$$p(s_k\ ;\ \underline{z},\ A,\ N) = \frac{p(s_k)}{p(\underline{z})}\ \frac{1}{(2\pi)^M \prod_{l=1}^{M} N\lambda_l}\ \exp\left(-\frac{1}{2N}\sum_{l=1}^{M}\frac{1}{\lambda_l}(z_l - A\ s_{kl})\ .\right.$$

$$\left.(z_l^* - A\ s_{kl}^*)\right) =$$

$$= C\ p(s_k)\ \exp\left(-\frac{1}{2N}\sum_{l=1}^{M}\frac{1}{\lambda_l}\left(-As_{kl}z_l^* - A^*s_{kl}z_l + |A|^2 s_{kl}\ s_{kl}^*\right)\right)\ . \qquad (4.128)$$

Considering the new variables given by

$$q_{kl} = \frac{s_{kl}}{l} \qquad (4.129)$$

they should satisfy the relation

$$q_i(t) = \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} q_{il}\ \varphi_l(t)\ , \qquad (4.130)$$

with q_{il} given by

$$q_{il} = \int_a^b q_i(t)\, \varphi_l^*(t)\, dt \quad . \tag{4.131}$$

$q_i(t)$ can be found from

$$s_i(t) = \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} s_{il}\varphi_l(t) = \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} \lambda_l\, q_{il}\varphi_l(t) \tag{4.132}$$

But

$$\lambda_l \varphi_l(t) = \int_a^b R(t - t')\, \varphi_l(t')\, dt' \quad , \tag{4.133}$$

therefore, by inverting the order of the summation and integration :

$$s_i(t) = \int_a^b R(t - t')\, \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} q_{il}\, \varphi_l(t')\, dt' =$$

$$= \int_a^b R(t - t')\, q_i(t')\, dt' \quad . \tag{4.134}$$

If the number M of samples approaches infinity,

$$\underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} z_l^*\, s_{il}/\lambda_l = \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} z_l^* \int_a^b q_i(t)\, \varphi_l^*(t)\, dt =$$

$$= \int_a^b q_i(t)\, \underset{M\to\infty}{\text{l.i.m.}} \sum_{l=1}^{M} z_l^*\varphi_l^*(t)\, dt = \int_a^b q_i(t)\, z^*(t)\, dt \tag{4.135}$$

Therefore

$$p(s_k\, ;\, \underline{z},\, A,\, N) = C\, p(s_k)\, \exp\left(-\frac{1}{2N} |A|^2 \int_a^b q_k\, s_k^*(t)\, dt \right) \cdot$$

$$\exp\left(\frac{1}{N} \mathrm{Re}\left(A \int_a^b q_k(t)\, z^*(t)\, dt \right) \right), \tag{4.136}$$

with

$$A = r \exp(j\varphi) \tag{4.136a}$$

$$\int_a^b q_k(t)\, z^*(t)\, dt = \rho_k \exp(j\theta_k) \tag{4.136b}$$

and

$$p(\varphi) = \frac{1}{2\pi} \quad , \tag{4.136c}$$

the last expression characterising the incoherent detection. After some manipulations, one finds :

$$p(s_k \; ; \; \underline{z}, \; r, \; N) = \frac{C}{2} \exp\left(- \frac{E}{2\, N_0\, N} |A|^2 \right) I_0\left(\frac{\rho_j r}{N}\right) , \tag{4.137}$$

in which $I_0(\cdot)$ is the modified zero order Bessel function,

$$N_0 = \frac{R(t - t')}{2\, \delta(t - t')} \tag{4.138}$$

(assumption which characterises the process with white noise), and

$$E_k = \frac{1}{2} \int_a^b |s_k(t)|^2 \, dt \quad . \tag{4.139}$$

E comes from the assumption that both transmitted signals have equal energy 2E. The a posteriori probability is proportional to ρ_k :

$$p(s_k \; ; \; \underline{z}, \; r, \; N) = C_1\, \rho_k \quad , \tag{4.140}$$

since $I_0(\cdot)$ is a monotonically increasing function. Therefore the detector should perform the function :

$$\left| \int_a^b s_k(t)\, z^*(t)\, dt \right| \quad . \tag{4.141}$$

The schematic diagram of such detector is shown in Fig. 4.6. The conjugate of the received analytic signal is multiplied by $s_0(t)$ and $s_1(t)$ respectively. The switches SW_{11} and SW_{12} are closed at t = a and remain closed until t = b. At this moment, the switches SW_{21} and SW_{22} are closed and the output signals of the two integrators are compared.

The detector decides that s_1 has been transmitted if D is positive or that s_0 has been transmitted if D is negative.

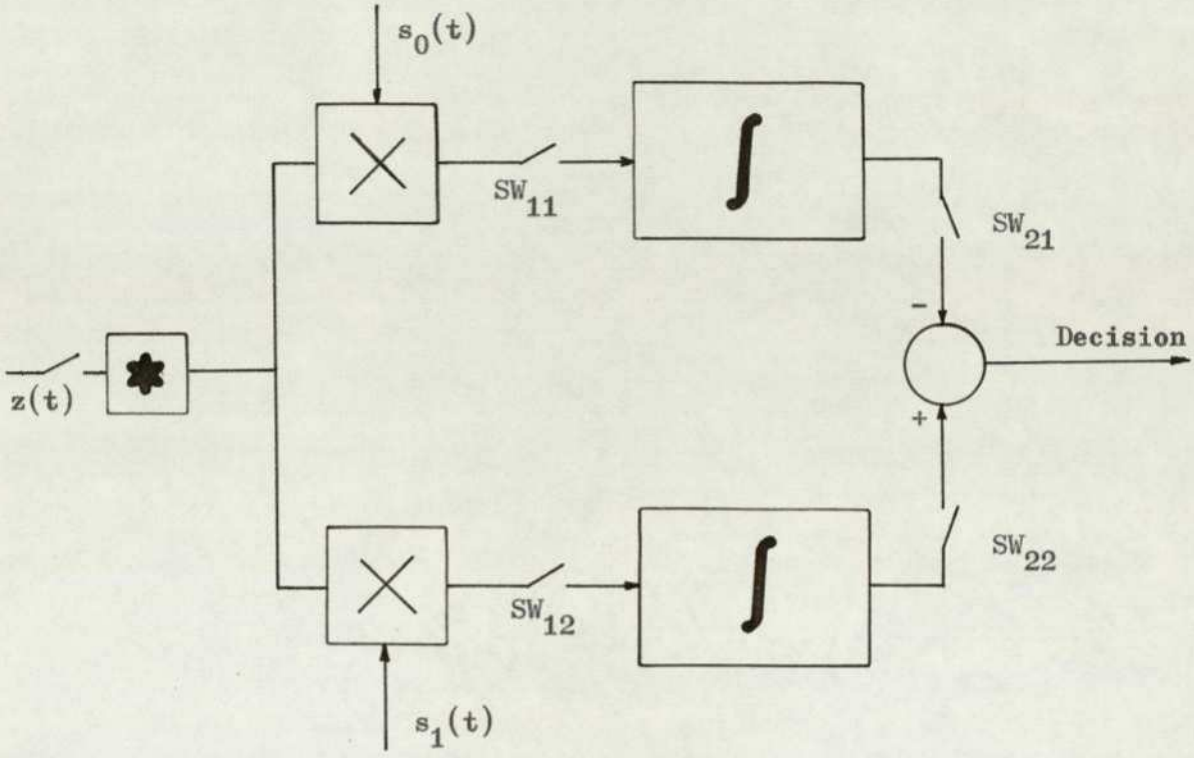


Fig. 4.6

An improved detector structure. The decision process as mentioned above does not take into account the fact that the SNR could be too small and therefore the risk of taking an erroneous decision becomes too high. A detector, having a more complex structure, but leading to better performance, is represented in Fig. 4.7.

The difference between it and the detector shown in Fig. 4.6 is due to the supplementary branch, which performs the estimation of the SNR. One of the signals s_1 or s_0 is selected on the base of maximum likelihood. The selected signal is used for the estimation of SNR, which is compared with a predetermined threshold. The decision taken on the base of the

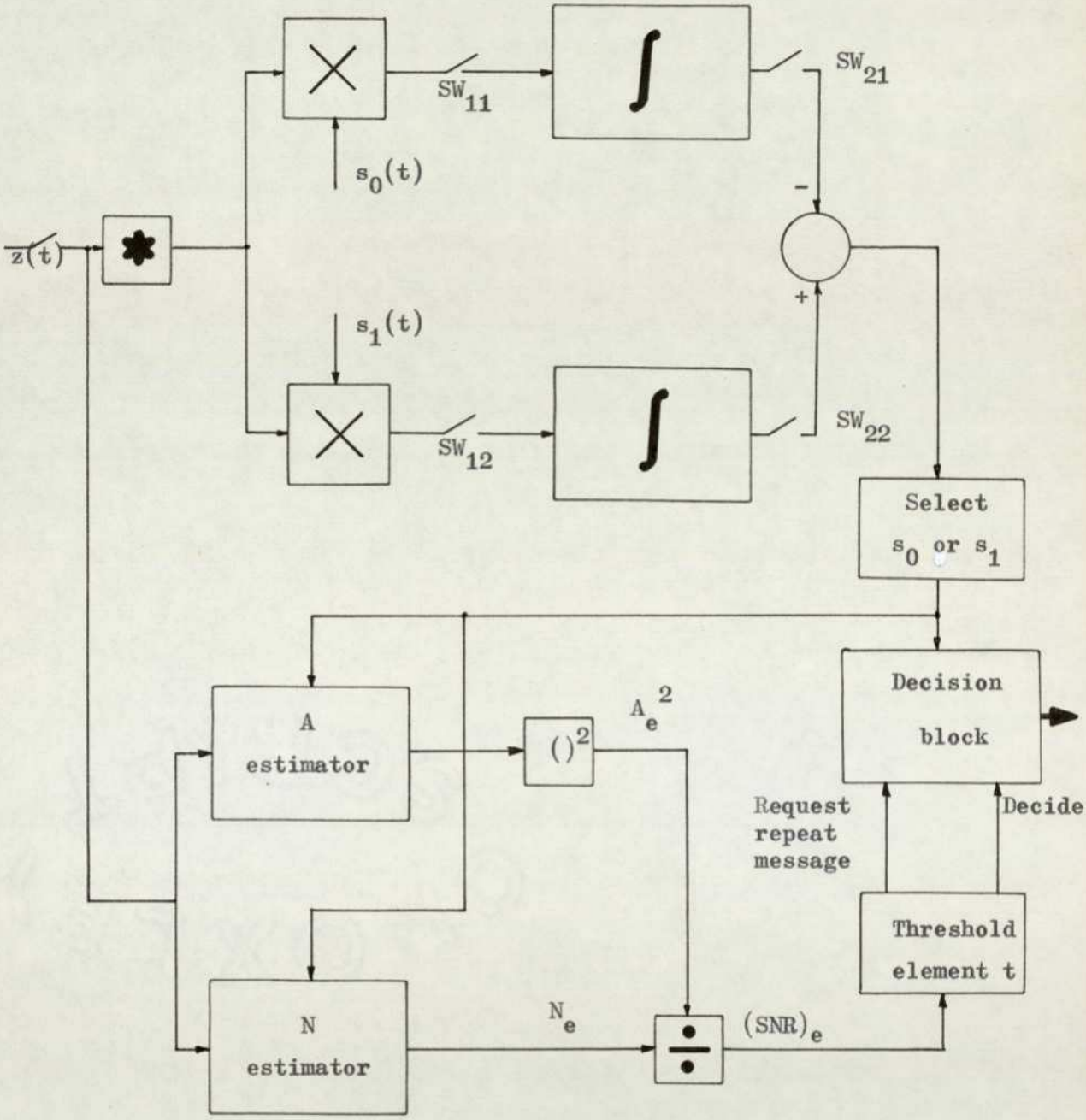

z(t)
$s_0(t)$
$s_1(t)$
SW_{11}
SW_{12}
SW_{21}
SW_{22}
-
+
Select s_0 or s_1
Decision block
A estimator
$()^2$
A_e^2
N estimator
N_e
$(SNR)_e$
Request repeat message
Decide
Threshold element t

Fig. 4.7

maximum likelihood is validated if $(SNR)_e$ exceeds the threshold t. In the opposite case, a request for repeating the message is formulated and sent to the transmitter. The value of t is chosen based on practical considerations (average level of the noise, the signal power, etc).

Finally, it is worth mentioning that the maximum number of samples which should be taken in a time interval T is given by the relation :

$$M_{max} = \frac{1}{2\ B\ T}\ , \qquad (4.142)$$

where B is the equivalent bandwidth of the channel noise. Otherwise, smear-desmear problems are encountered.

4.6 Extension to the case of non-Gaussian noise

In the previous section the transmission and detection of signals received at the end of communication channels affected by Gaussian noise have been analysed. It can be seen that once the p.d.f. of the noise present on the channel can be estimated and expressed in a mathematical form, the analysis of the aforementioned elements can be carried out in the same manner, since all the elements required (the correlation function, the log-likelihood functions, the moments, etc) can be directly calculated. However, the practical solution of the problem becomes more difficult and it is very likely that problems as solving the set of simultaneous equations given by (4.99) or (4.114) could be solved only by using numerical methods.

Consider the noise p.d.f. of the form

$$p(\underline{n}) = (\alpha + Q''(\underline{n}))\ p_G(\underline{n})\ , \qquad (4.143)$$

where

$$\underline{n}(t) = \begin{bmatrix} n_1(t) \\ \cdot \\ \cdot \\ \cdot \\ n_M(t) \end{bmatrix}\ , \qquad (4.143a)$$

or, in a simpler notation, $\underline{n}$ and

$$Q''(\underline{n}) = \underline{n}' \, \underline{D} \, \underline{n} \quad . \tag{4.143b}$$

$\underline{D}$ is a M by M matrix obtained as described in the section 4.3 and $p_G(\underline{n})$ is a function obtained from the available sample as if the distribution was Gaussian. α is also obtained by following the procedure described in the section 4.3. If the signal

$$\underline{z}(t) = A \, \underline{s}(t) + \sqrt{N} \, \underline{n}(t) \tag{4.144}$$

is received, the conditional probability of $\underline{z}$ given A and N is given by :

$$p(\underline{z} \; ; \; A, \; N) = \left(\alpha + \frac{1}{2 \, N} \, Q''(\underline{z} - A\underline{s}) \right) \; p_G \left(\frac{\underline{z} - A\underline{s}}{2\sqrt{N}} \right) \quad . \tag{4.145}$$

The log-likelihood function is then :

$$L = \ln \left(\alpha + \frac{1}{2 \, N} \, Q''(\underline{z} - A\underline{s}) \right) + \ln \; p_G \left(\frac{\underline{z} - A\underline{s}}{2\sqrt{N}} \right) \quad . \tag{4.146}$$

The partial derivatives with respect to A and N of the second term of the expression (4.146) have already been calculated in (4.100) and (4.103). Note

$$L_1 = \ln \left(\alpha + \frac{1}{2 \, N} \, Q''(\underline{z} - A\underline{s}) \right). \tag{4.147}$$

Then

$$\frac{\partial L_1}{\partial A} = \frac{1}{\alpha + \frac{1}{2 \, N} \, Q''(\underline{z} - A\underline{s})} \; \frac{1}{2 \, N} \frac{\partial}{\partial A} \; Q''(\underline{z} - A\underline{s}) \tag{4.148}$$

and

$$\frac{\partial L_1}{\partial N} = \frac{Q''(\underline{z} - A\underline{s})}{\alpha + \frac{1}{2 \, N} \, Q''(\underline{z} - A\underline{s})} \; \frac{-1}{2 \, N^2} \quad . \tag{4.149}$$

As

$$Q''(\underline{z} - A\underline{s}) = (\underline{z} - A\underline{s})' \, \underline{D} \, (\underline{z} - A\underline{s}) \; ,$$

$$\frac{\partial}{\partial A} \, Q''(\underline{z} - A\underline{s}) = 2 \, (A\underline{s}'\underline{D}\underline{s} - \underline{z}'\underline{D}\underline{s}) \tag{4.150}$$

Thus, the set of equations which give the values of A_e and N_e can be written - after some manipulations - as:

$$\frac{A_e\, \underline{s}'\, \underline{D}\, \underline{s} - \underline{z}'\, \underline{D}\, \underline{s}}{\alpha + \frac{1}{2N_e} Q''(\underline{z} - A_e \underline{s})} - (A_e\, \underline{s}'\, \underline{C}\, \underline{s} - \underline{z}'\, \underline{C}\, \underline{s}) = 0 \qquad (4.151)$$

$$\frac{-Q''(\underline{z} - A_e \underline{s})}{\alpha + \frac{1}{2N_e} Q''(\underline{z} - A_e \underline{s})} \frac{1}{N_e} - M + \frac{(\underline{z} - A_e \underline{s})'\, \underline{C}\, (\underline{z} - A_e \underline{s})}{N_e} = 0 \qquad (4.152)$$

N_e can be expressed from (4.151) and, by replacing it in (4.152), one obtains an equation in A_e. Unfortunately the mathematical expressions involved are complicated and therefore the only reasonably practical solution consists in solving the equations (4.151) and (4.152) by using a standard numerical method. After a set of solutions for A_e and N_e were obtained, the signal/noise ratio can be estimated. Due to the large volume of computation involved, it seems that a detector as that shown in Fig. 4.7 should be tuned on-line at moments of time determined on the basis of trial and error. In fact, the values of A_e and N_e should be readjusted if the result of the estimation of the probability density function (which requires far less computation) shows a substantial change in the noise parameters and therefore a reconsideration of the estimator of SNR is desirable.

4.7 The analysis of a nonlinear detector for signals corrupted by non-Gaussian noise

As it has been mentioned before, the non-Gaussian noise present in industrial environments is characterised by spikes of short duration and high amplitude, arising at random moments of time. Usually each spike is followed by a relatively long period of "quietness" ; during this period the noise which corrupts the received signal is merely Gaussian. These considerations show that a detector operating in such conditions

should process in a linear fashion the signals in which only Gaussian noise is present, limiting at the same time the high amplitudes, which are the effect of the corruption of the signal by the impulsive noise. The simplest detector providing these characteristics is represented in Fig. 4.8. z(t) is the received signal and x(t) the signal after being

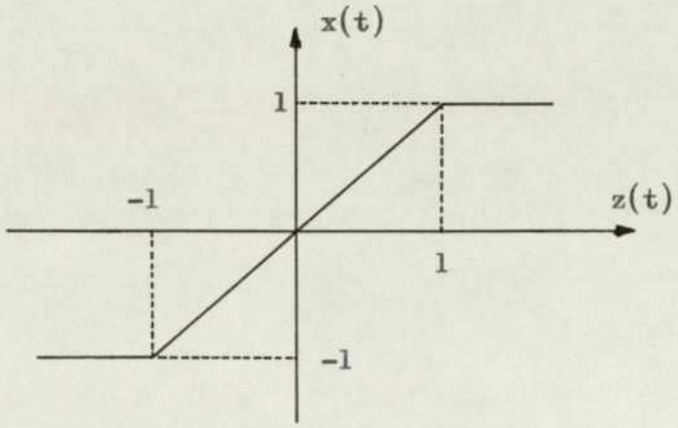


Fig. 4.8

processed by the detector. The signal x(t) is either sampled or integrated, in the latter case the result being a signal y(t) given by

$$y(t) = \int_0^T x(t)\, dt \quad .$$

In order to asses the optimum parameters of the process of transmitting signals (such as the optimum signal shape, the optimum signal power, etc), one should first derive the properties of the detector output signal x(t), especially its probability density function. This problem has been considered in several papers, among them (S.S.Rappaport, L.Kurz, 1966 ; J.Kapp, 1971). However, the way of attacking it not only assumes that the p.d.f. of the noise has a particular form, but also involves extremely lengthy and cumbersome calculations. A different approach, based on the possibility of estimating the noise p.d.f. by using the method of stochastic approximation, is presented in this section. The approach has the advantage of being applicable to any class of noise and

any detector characteristic. The only element required in estimating the probability density function of the variable x is a set of samples for z.

The proposed approach can be briefly described as it follows : given a set of N samples z_k, k = 1,...,N for z, the p.d.f. of z is estimated by using the method of stochastic approximation. After that a set of samples x_k, k = 1,...,N are generated based on z_k and the detector characteristic function f(z). By using again the method of stochastic approximation the p.d.f. of z can be estimated. As it can be seen, this method does not assume a particular form of the noise density function, nor does its principle and realisation change when the detector characteristic is altered. Sometimes the detector could have as characteristic function an expression which is not easily implementable in practical applications, e.g.

$$f(z) = z \exp(- a|z|^{\xi})$$

(J.Kapp, 1971). In such cases their characteristic can be relised by using a linear combination of Hermite polynomials. Such practical realisations involve only amplifiers and analogue multipliers, devices which are readily available at low cost and suitable parameters with respect to the band width and accuracy. An example of obtaining the fourth Hermite polynomial is shown in Fig. 4.9. Therefore one can approximate the detector characteristic function f(z) by a linear combination of Hermite polynomials such that

$$f(z) \simeq f_e(z) = \sum_{k=0}^{N_d} c_k H_k(z) \ . \qquad (4.153)$$

The criterion based on which the coefficients c_k are chosen has the form

$$C = \int_{-\infty}^{+\infty} \Big(f(z) - f_e(z)\Big)^2 p(z)\, dz \ , \qquad (4.154)$$

therefore c_k are chosen such that C is minimised ; in (4.154) p(z) is the p.d.f. of z and this expression was prefered due to the fact that

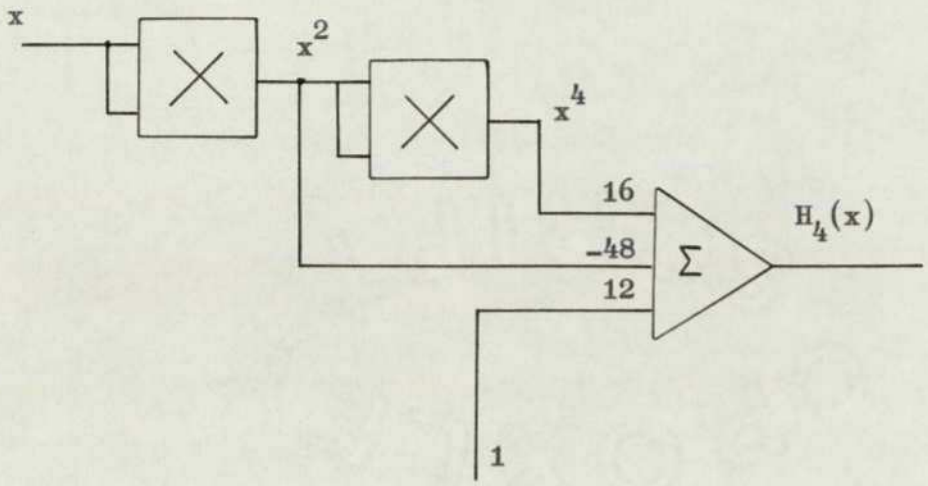


Fig. 4.9

$f_e(z)$ reproduce f(z) with higher accuracy for those values of z which are more probable. Hermite polynomials were chosen in the approximated form of f(z) due to their property of orthogonality, which leads to a very convenient way of obtaining the coefficients c_k. Indeed, C is minimised when

$$\frac{\partial C}{\partial c_k} = 0 \quad ,$$

therefore

$$\frac{\partial}{\partial c_k} \int_{-\infty}^{+\infty} \left(f(z) - \sum_{i=1}^{N_d} c_i \, H_i(z) \right)^2 p(z) \, dz =$$

$$= 2 \int_{-\infty}^{+\infty} H_k(z) \left(f(z) - \sum_{i=1}^{N_d} c_i \, H_i(z) \right) p(z) \, dz =$$

$$= 2 \int_{-\infty}^{+\infty} f(z) \, H_k(z) \, p(z) \, dz - 2 \int_{-\infty}^{+\infty} H_k(z) \sum_{i=1}^{N_d} c_i \, H_i(z) \, p(z) \, dz = 0 \quad . \tag{4.155}$$

N_d represents the number of polynomials used for the approximated detector

characteristic function. $p(z)$ is the probability density function of the received signal z as obtained by using the stochastic approximation method, i.e.

$$p(z) = p_G(z) \sum_{m=0}^{N_p} \theta_m H_m\left(\frac{z}{\sigma}\right) = (b_0 + b_2 z^2 + \ldots + b_{N_m} z^{N_m})\, p_G(z), \tag{4.156}$$

with

$$p_G(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-z^2/2\sigma^2) .$$

The notation N_p refers to the number of polynomials used for the approximated expression for p.d.f.

By expressing $f_e(z)$ as :

$$f_e(z) = \sum_{i=1}^{N_d} c_i H_i(z) = \sum_{i=0}^{N_d} d_i z^i , \tag{4.157}$$

(4.155) reduces to

$$L_k(c_m, M_i) = \int_{-\infty}^{+\infty} f(z) H_k(z) p(z)\, dz , \tag{4.158}$$

$$k = 1,\ldots,N_d$$

$$m = 1,\ldots,N_d$$

$$i = 1,\ldots,N_d N_p$$

where M_i are the i-th moments of the distribution $p_G(z)$ and L_k represents the symbolic form for a linear combination. Therefore, the coefficients c_m, $m = 1,\ldots,N_d$ are easily found by solving the set of simultaneous equations given by (4.158). As example, assume that

$$f_e(z) = \sum_{i=0}^{1} c_{2i+1} H_{2i+1}(z) = c_1 H_1(z) + c_3 H_3(z)$$

$$p(z) = p_G(z) \sum_{k=0}^{2} \theta_{2k} H_{2k}\left(\frac{z}{\sigma}\right) = \left(\theta_0 H_0\left(\frac{z}{\sigma}\right) + \theta_2 H_2\left(\frac{z}{\sigma}\right) + \theta_4 H_4\left(\frac{z}{\sigma}\right)\right) p_G(z)$$

Therefore the choice for odd coefficients only in the expression of $f_e(z)$ is justified by the fact that $f_e(-z) = - f_e(z)$.

$$\frac{\partial C}{\partial c_k} = 0 \ , \quad k = 0,1,2,3 \tag{4.159}$$

implies

$$\int_{-\infty}^{+\infty} H_k(z)\left(c_1 H_1(z) + c_3 H_3(z) \right) \cdot$$

$$\cdot \left(\theta_0 H_0(\frac{z}{\sigma}) + \theta_2 H_2(\frac{z}{\sigma}) + \theta_4 H_4(\frac{z}{\sigma}) \right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-z^2/2\sigma)\, dz =$$

$$= \int_{-\infty}^{+\infty} f(z)\, H_k(z)\, p(z)\, dz \ . \tag{4.160}$$

By expanding the left side of (4.160) one obtains :

$$\begin{aligned} a_{11}c_1 + a_{31}c_3 &= \int_{-\infty}^{+\infty} f(z)\, H_1(z)\, p(z)\, dz \\ a_{13}c_1 + a_{33}c_3 &= \int_{-\infty}^{+\infty} f(z)\, H_3(z)\, p(z)\, dz \quad , \end{aligned} \tag{4.161}$$

with

$$\begin{aligned} a_{11} &= 4(a_0 M_2 + a_2 M_4 + a_4 M_6) \\ a_{31} &= 8\left(-3a_0 M_2 + (2a_0 - 3a_2)M_4 + (2a_2 - 3a_4)M_6 + 2a_4 M_8 \right) \\ a_{13} &= 8\left(-3a_0 M_2 + (2a_0 - 3a_2)M_4 + (2a_2 - 3a_4)M_6 - 3a_4 M_8 \right) \\ a_{33} &= 8\left(9a_0 M_2 - (12a_0 - 9a_2)M_4 + (2a_0 - 12a_2 + 9a_4)M_6 + \right. \\ &\qquad \left. + (4a_2 - 12a_4)M_8 + 4a_4 M_{10} \right) \end{aligned} \tag{4.162}$$

and

$$\theta_0 H_0(\frac{z}{\sigma}) + \theta_2 H_2(\frac{z}{\sigma}) + \theta_4 H_4(\frac{z}{\sigma}) \equiv a_0 + a_2 z^2 + a_4 z^4 \quad . \tag{4.163}$$

The integral of the form

$$I_k = \int_{-\infty}^{+\infty} f(z)\, H_k(z)\, p(z)\, dz \quad , \quad k = 1,2,3$$

can be easily calculated using a numerical technique. Therefore, c_1 and c_3 can be calculated by solving the set of simultaneous equations given by

$$\begin{aligned} a_{11}c_1 + a_{13}c_3 &= I_1 \\ a_{31}c_1 + a_{33}c_3 &= I_3 \quad . \end{aligned} \tag{4.164}$$

Note that the values c_1 and c_3 are function of l, as the integrals I_1 and I_3 are function of l. Therefore, when l has to be chosen in order to maximise an optimum criterion, e.g. SNR, the solution of the system given by (4.158) should be found in a recursive way, using a standard maximisation algorithm.

With the parameters c_i, $i = 1,\ldots,N_d$, found, the p.d.f. of the variable x can be estimated by using the method of stochastic approximation. The estimation is based on the observables x_k of x, directly available if observables z_k of z are known :

$$x_k = \sum_{i=0}^{N_p} c_i\, H_i(z_k) \quad . \tag{4.165}$$

Obviously, $f_e(z)$ should be used instead of $f(z)$ when p.d.f. of x is estimated, since the detector which will be practically implemented has the characteristic $f_e(z)$.

The advantage of the aforementioned procedure consists of the fact that the probability of x can be expressed only as a function of y itself even when the inverse of the function $x = f(z)$ is not mathematically obtainable. In most of practical situations z cannot be explicitly expressed as a function of x and therefore the procedure described in this section should be applied in order to find $p(x)$.

The next step in the detection process is to design a device which, based on $x(t)$, chooses one of the hypotheses (A.D. Whalen, 1971)

$$\begin{cases} H_0 : & z(t) = s_0(t) + n(t) \\ H_1 : & z(t) = s_1(t) + n(t) \end{cases} , \tag{4.166}$$

$s_0(t)$ and $s_1(t)$ being the transmitted signals corresponding to marks and spaces respectively. The optimum decision rule consists of comparing the likelihood ratio with a given threshold λ_0, which will be defined later. First, consider a set of m samples of the received signal. At the moment of time t_k the received signal is

$$z_k = s_{ik} + n_k \tag{4.167}$$

with i = 0 or 1, $z_k = z(t_k)$, $s_{ik} = s_i(t_k)$, $n_k = n(t_k)$.

The likelihood ratio is given by

$$\lambda(\underline{z}) = \frac{p_1(z_1,\ldots,z_m)}{p_0(z_1,\ldots,z_m)} , \tag{4.168}$$

where

$$\underline{z} = \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{bmatrix}$$

and

$$p_i(z_1,\ldots,z_m) , \quad i = 0 \text{ or } 1$$

are the likelihood functions. The decision rule is : choose H_1 if $\lambda(\underline{z}) \geqslant \lambda_0$, otherwise choose H_0. The simplified diagram of the decision mechanism is shown in Fig. 4.10.

Assume that the samples are taken such that they are statistically independent. This can be always realised in practice, since the noise is band-limited. If the samples are taken at time intervals longer than π/Ω, where Ω is the upper limit of the frequency bandwidth, the above mentioned assumption is realised. Assume now that the noise has the p.d.f. given by

$$p(n) = \sum_{l=0}^{N} \theta_{21}\, H_{21}\left(\frac{n}{\sigma}\right) p_G(n)$$

and estimated by using the method of stochastic approximation. Therefore

$$p_i(z_1,\ldots,z_m) = \prod_{j=1}^{m}\left(\sum_{l=0}^{N} \theta_{21}\, H_{21}\left(\frac{z_j - s_{ij}}{\sigma}\right) p_G(z_j - s_{ij})\right),$$

$$i = 1 \text{ or } 0 \qquad (4.169)$$

and then

$$\lambda(\underline{z}) = \frac{\prod_{j=1}^{m}\left(\sum_{l=0}^{N} \theta_{21}\, H_{21}\left(\frac{z_j - s_{1j}}{\sigma}\right) p_G(z_j - s_{1j})\right)}{\prod_{j=1}^{m}\left(\sum_{l=0}^{N} \theta_{21}\, H_{21}\left(\frac{z_j - s_{0j}}{\sigma}\right) p_G(z_j - s_{0j})\right)} =$$

$$= \prod_{j=1}^{m} \frac{\sum_{l=0}^{m} \theta_{21}\, H_{21}\left(\frac{z_j - s_{1j}}{\sigma}\right)}{\sum_{l=0}^{m} \theta_{21}\, H_{21}\left(\frac{z_j - s_{0j}}{\sigma}\right)} \exp\left(- \frac{\left(\sum_{j=1}^{m} (z_j - s_{1j})\right)^2}{2\sigma^2} + \right.$$

$$\left. + \frac{\left(\sum_{j=1}^{m} (z_j - s_{0j})\right)^2}{2\sigma^2} \right) . \qquad (4.170)$$

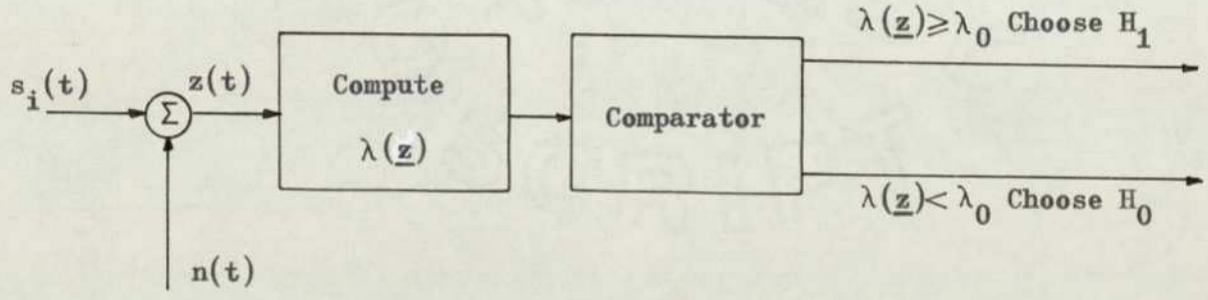


Fig. 4.10

Thus the detector should calculate on the basis of the available sample z_j the value of $\lambda(\underline{z})$, which is then compared with the threshold λ_0. The detector structure is not as simple as for the Gaussian noise, but the

advantage of better performance should be taken into account, especially if the detector is at the end of the line used in time-sharing system by several data sources.

4.8 Conclusions

This chapter has analysed a possible way of describing the probability density function of the noise in its most general way, with special reference to the electrical noise as encountered in communication channels used in industrial environments. Since this noise has a strong Gaussian component, the proposed model for p.d.f. is such that, during the absence of impulsive noise, the Gaussian p.d.f. is exactly represented by the model with particular values of parameters characterising it. This important feature makes possible the successful use of the proposed p.d.f. for the case of near-Gaussian noise. At the same time, no a priori assumption about the distribution of the noise has to be made, due to the universality of the proposed model. One cannot see particular difficulties in implementing such model, although some results, analitically obtainable for the case of Gaussian process, can be obtained with the model only by using numerical methods. This difficulty can be overcome when a computer is used, since the main problem involved in producing the estimated p.d.f. consists in solving a set of algebraic simultaneous equations.

CHAPTER 5

Optimisation of the cost of communication systems

5.1. Introduction

The importance of optimising the cost of wires connecting the data sources to the computer has been pointed out in Chapter 2. This cost could represent as much as 17-20% of the total budget associated to the control equipment and is significant regardless the type of signal transmitted through the communication channel to the computer (H.Simon, 1970). Therefore, a procedure enabling the designer to choose the best location of multiplexers such that the computer is provided with the necessary information at the appropriate moment of time would be of vital concern when a new plant is designed : not only large numbers of data sources can be handled, but also several possibilities for instrumentation lay-out could be considered and therefore the optimum solution could be chosen.

The aim of this chapter is to consider the problem of optimisation of the cost of the communication network, producing suitable algorithms with immediate practical application. It should be mentioned that such algorithms can also be used in other fields, e.g. telephone exchanges location, power substation location, etc. with only minor modifications.

5.2. Similarities and disimilarities with the problem of depot location

The problem of depot location, which is illustrated in Fig. 5.1 (C.D.T.Watson-Gandy, 1971), is one of the favourites of operational

research scientists. The aim of the distribution department is to get the manufactured products from the factory directly to the customer or to an outlet in the proximity of the customer. The practice has shown that an economical method of achieving this is to make use of a system of depots, in which some form of accumulation takes place. In this way, the distributor can make use of bulk shipment from the factories to the depots (named trunking), at which point the goods are portioned down into smaller quantities to be delivered to the local customer. By customer one understands the final point in the delivery chain. Note that the same depot may be supplied from several factories.

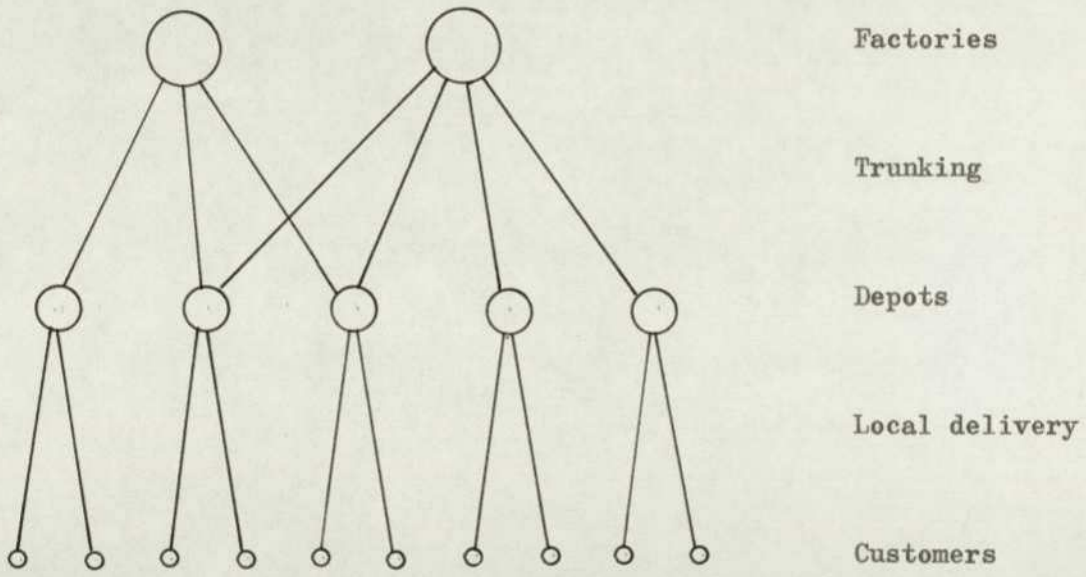


Fig. 5.1

The problem of designing optimal communication networks is in many respects similar to that of depot allocation. Fig. 5.2 shows a typical communication network system, in which data sources are connected to the computer via multiplexers. The question to be answered is : given the fixed locations of data sources, how should one locate the multiplexers and the computer in such a way that the total cost of the communication network is minimised ? In the total cost not only the wires, but

multiplexers are also included.

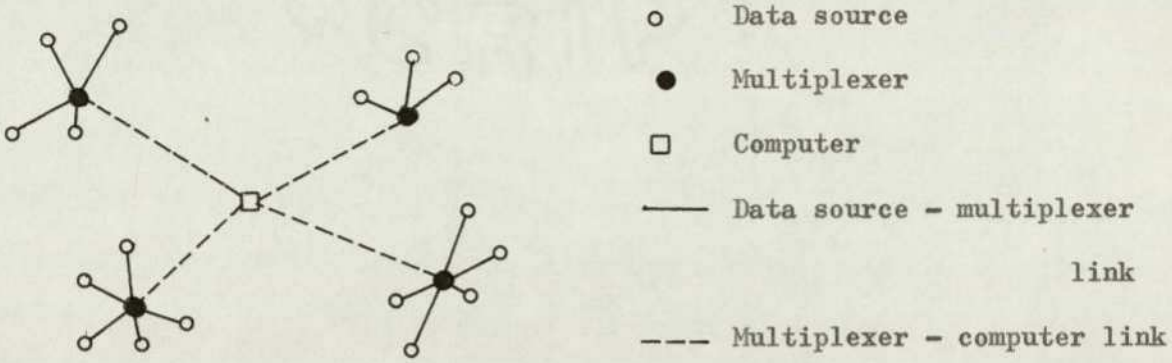


Fig. 5.2

A comparison between the problems of depot location and communication system design shows many similarities between them. The customers and data sources, the depots and multiplexers, the factories and the computer are pairs of similar elements. However, there are some differences between the two problems, namely :

- only one computer is used in the communication system, as opposed to several factories in the depot location problem;
- in emergency, a customer could be supplied from another depot if the normal depot has not the capacity of satisfying its request. A data source is connected to only one multiplexer within the communication system;
- the number of data sources is usually much greater than the number of customers (the notion "customer" is again used to characterise the end point of the transport chain and not the ultimate buyer);

An algorithm generally accepted in solving the depot location problem consists of assigning initial position to the depots according to some guess and to improve these positions such that, by supplying each customer from the "cheapest" depot, the overall delivery cost is minimised. The algorithm is used several times, until no significant improvement from one

run to another is obtained.

The optimisation procedure which will be used for the problem of communication system should answer more complex questions, mainly connected with the restrictions under which the system operates. Among them, one could mention the followings :

- the distance between a data source and the multiplexer which serves it should be limited for reasons determined by noise corruption. A common amplifier, used for all data sources connected to a particular multiplexer, is more economically used if it is located at the same place with the multiplexer;
- the capacity of any multiplexer cannot exceed limits determined by considerations concerning the minimum time interval between two successive requests for information addressed to the same data source. Another factor which imposes a limited multiplexer capacity is generated by conditions of economical operation : the increase of the number of inputs could require power supply units, mechanical constructions, etc., the characteristics of which do not represent acceptable practical solutions;
- there are restrictions imposed on multiplexers and computer locations. These restrictions are not very tough when a new plant is designed, but could become extremely severe when the communication system for an existing plant is redesigned. The main factors contributing to this consist in the existance of some cable paths and of a control room.

All these restrictions should be considered when an algorithm for optimising the communication system is formulated. The most difficult problem in the case of designing a new plant is represented by the initial guess of multiplexer locations. The simple visual inspection of the data sources lay-out could be not only a lengthy but also a misleading procedure, which works only if the data sources are obviously separated. An analysis

of the situation leads to the need of solving a problem, very often encountered in pattern recognition, namely the problem of finding clusters. Of course, the aspect required by the communication system optimisation is simpler, involving clustering based only on geometrical features. In the next two sections two heuristic algorithms will be considered. The ideea of using heuristic algorithms may appear unjustified; better methods such as the integer programming look more attractive. However, the size of the problem makes their use unrealistic. Indeed, the computer program for solving an integer programming problem in connection with a system of thousands of data sources and hundreds of multiplexers requires unusually large store capacity and could lead only to minor improvements as compared with a carefully chosen heuristic algorithm. Despite using special procedures (e.g. the reduction of the size of store when using integer programming programs, described in Appendix A3), the alternative of using integer programming could be ruled out based on computation time consideration. Even using implicit enumeration method (Balas's algorithm), the analysis of a number of possible connections exceeding 10^5 is extremely inefficient. Thus, using heuristic algorithms appears the only solution of practical interest and therefore it will be considered in the followings.

5.3. A heuristic algorithm based on the quasi-potential function approach

The possibility of finding clusters of points was described in a paper (G.A.Butler, 1969) which, unfortunately referes to an algorithm practically unrealisable. The method consists of assigning to each point of the space a field vector which is the resultant of the field

created by each individual point, to which an unit mass is associated. The field vector has a modulus of the form

$$F = \sqrt{\sum_{i=1}^{n} \left(\frac{1}{d_i^{\alpha}}\right)^2} \qquad (5.1)$$

where d_i is the distance between the i-th fixed mass point and the probe, n is the number of fixed masses, which generate the field, and α a positive exponent. The magnitude of α determines the number of field nodes, which increases as α becomes greater. A probe consisting also of an unit of mass is moved in steps following the direction of the field until the position of zero field is found. This position is boxed by sensing the step in which all the field components change sign. When a minimum is boxed the step of search is reduced and the procedure continues in a similar manner until the size of the step corresponds to a desired accuracy. After a field node was found, the search for another node starts from another position and basically the same procedure is used. The method used for avoiding to find nodes already found is to alter the field components in the probe point directed towards established nodes by a fraction which increases as the probe approaches the node; when the probe coincides with a point situated at an infinitesimally small distance from the node the field component directed towards it is zero. In this way, the author claims, the probe "flows" through the established nodes and then other nodes can be found. The procedure continues until no node can be found in a given number of steps counted from the beginning of the search from the last initial position. This kind of approach has the advantage of considering simultaneously the influence of all points to be divided into clusters, which could lead to much better results than those obtained by using procedures which take into account only some of the points (e.g. those "near" to the point suspected to be the centre of a cluster).

An analysis of this algorithm shows that it never works in practice, although the paper gives some results obtained after a "long and patient effort of a programmer". The analysis uses the situation described by Fig. 5.3 . The probe point P_k is characterised by a field vector F_k, having the components F_{kx} and F_{ky} directed along the coordinate directions (a two-dimensional problem is considered). The new position P_1 (in the absence of the node N) is obtained by displacing P_k along each coordinate direction with a step length s. The displacement signs are given by the directions of the field components F_{kx} and F_{ky} along the coordinate directions. Due to the presence of the node N, the component F_{kN} of the field directed towards the node N is altered so that it becomes $F_N' = cF_N$.

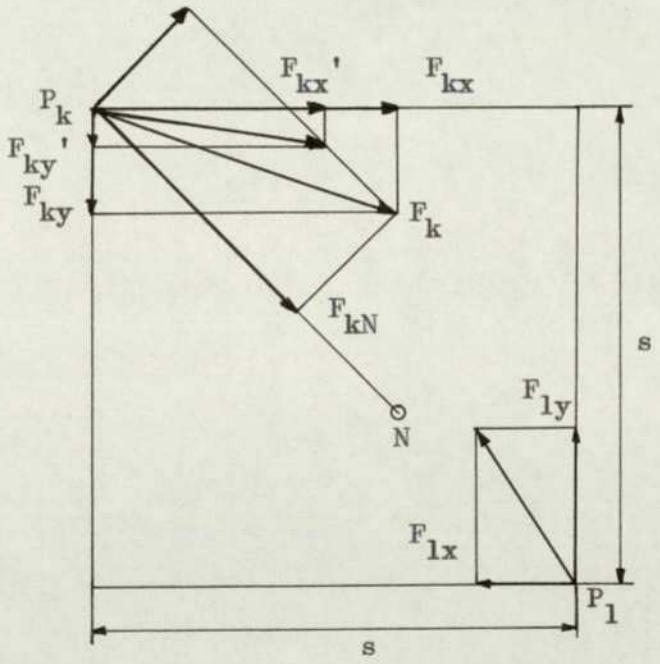


Fig. 5.3

The author suggests for c values in the range $[0\ ,\ 1]$. Therefore F_N' has the same sign as F_N and consequently the components F_{kx}' and F_{ky}' of the modified field have the same sign as F_{kx} and F_{ky}. Thus, despite the field alteration, P_k still moves into P_1 , where the components F_{1x} and F_{1y} of the field F_1 have opposite signs with respect to F_{kx} and F_{ky}, as

F_{1N} is directed towards N. Even after altering F_{1N} the situation remains unchanged and finding N again as a node cannot be avoided. The problem cannot be solved by considering negative c, as this implies the artificial creation of non-existing nodes.

The technique of the aforementioned algorithm, should it have worked, of establishing cluster centres by considering all the points at any stage can be used in designing another algorithm, the principle of which will be explained below.

Consider a set of electric charges of the same sign and of the same magnitude with a fixed geometrical lay-out. One could attach to any point P of the space a quasi-potential function V_P given by

$$V_P = \sum_{i=1}^{n} \frac{1}{1 + \left(\frac{d_i}{D}\right)^{\alpha}} \qquad (5.2)$$

where n is the number of the electric charges, d_i is the distance between the point P and the i-th charge, D is a scaling factor and α an exponent. The way of choosing D and α will be explained later. The quasi-potential function V_P has local maxima near the regions of the space containing high concentrations of charges (i.e. a high number of charges, as their magnitude is the same). Therefore, finding cluster centres is equivalent to maximising the quasi-potential function V_P. The number of maxima is determined by D and α . When D is constant, one can visualise that for $\alpha = 0$ the potential function is a constant and therefore has no extremum points. For α large, the quasi-potential function will have local maxima corresponding to each charge. For a resonable value of α one can find a desired number of maxima of quasi-potential function. On the other hand, when α is constant various values of D could produce changes in the aspect of the quasi-potential function : very large values for D will have as

effect a constant value for V_p, and small values of D will produce many local maxima, each of them associated to one of the charges creating the quasi-potential. It is wise, for computational reasons, to adjust the number of maxima of V_p by changing D and holding $\alpha = 2$. The justification is obvious : given the coordinates of the charges, finding a power of the normalised distance $\frac{d_i}{D}$ can be achieved only by calculating first the square $\left(\frac{d_i}{D}\right)^2$ of the same normalised distance. Therefore D will be the variable parameter and $\alpha = 2$ in the adopted algorithm. One major hurdle remains : if the quasi-potential function has multiple extrema, how can one use a conventional maximisation routine in order to find them all ? The solution adopted for solving this problem consists in generating random initial guesses and using a hill climbing algorithm for finding the extrema; if the currently found maximum coincides with one previously found, the new one is deleted and the search resumes in the same manner. The process of finding maxima of V_p is considered terminated when the last m found extrema are among those previously found. If the number N of found extrema is not satisfactory (too small), D is decreased and the procedure is repeated. The flow-chart of this algorithm is given in Fig. 5.4 . Details about the computer program which implements the algorithm are provided in Appendix A3. The program was checked for two sets of sources, one of them with very neat separation between clusters, the other one with clusters interfacing. The quasi-potential levels are plotted in Fig. 5.6 and 5.8 respectively, using a computer-driven graph plotter. The crosses mark the quasi-potential maxima as found by the described algorithm. Fig. 5.5 and 5.7 represent the lay-out of the two sets of data sources considered.

The next problem is how to use the cluster centres found in the way described above. The method used for the communication system

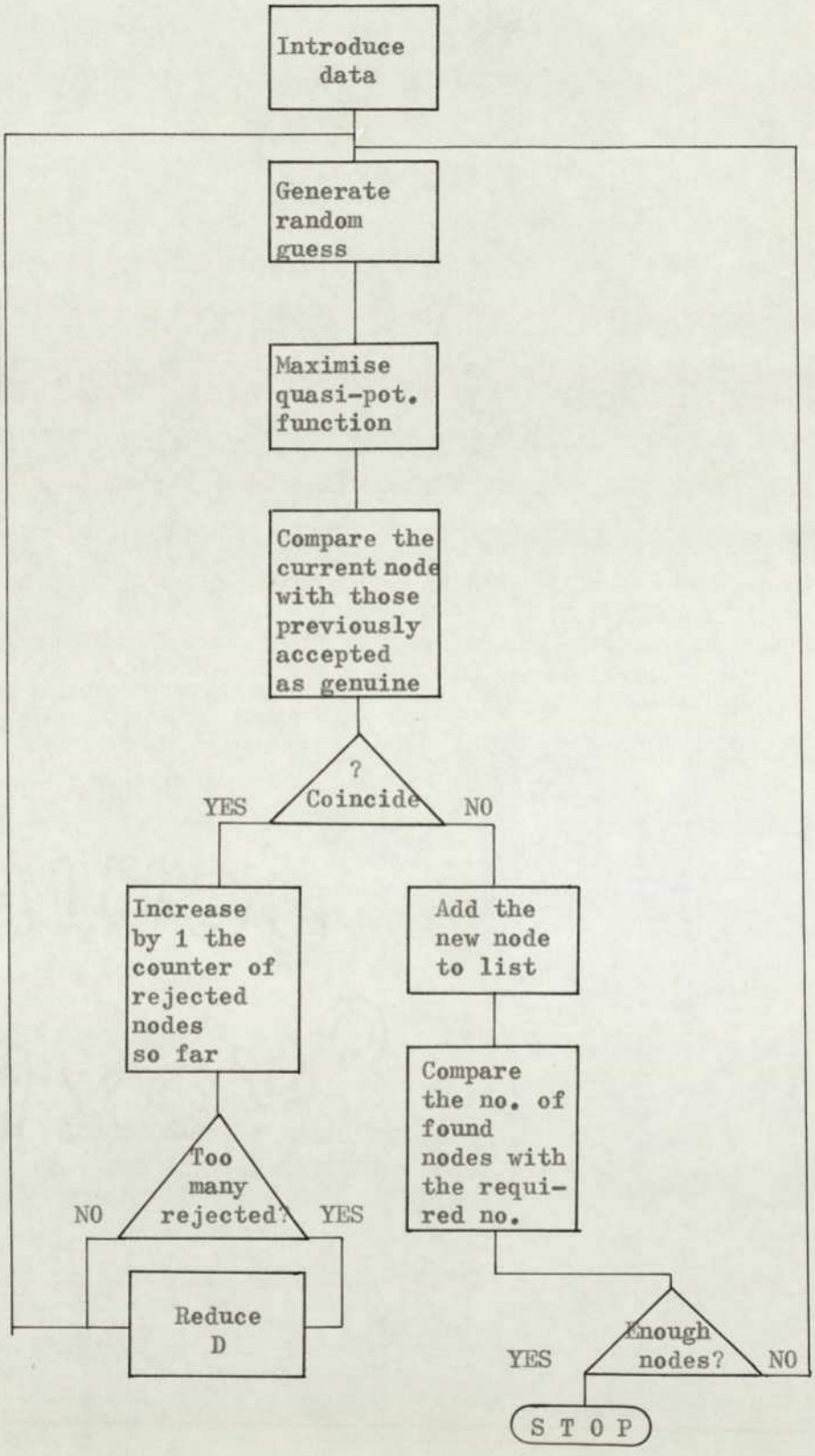

Introduce data
Generate random guess
Maximise quasi-pot. function
Compare the current node with those previously accepted as genuine
? Coincide
YES
NO
Increase by 1 the counter of rejected nodes so far
Add the new node to list
Too many rejected?
NO
YES
Compare the no. of found nodes with the required no.
Reduce D
Enough nodes?
YES
NO
S T O P

Fig. 5.4

Fig. 5.5

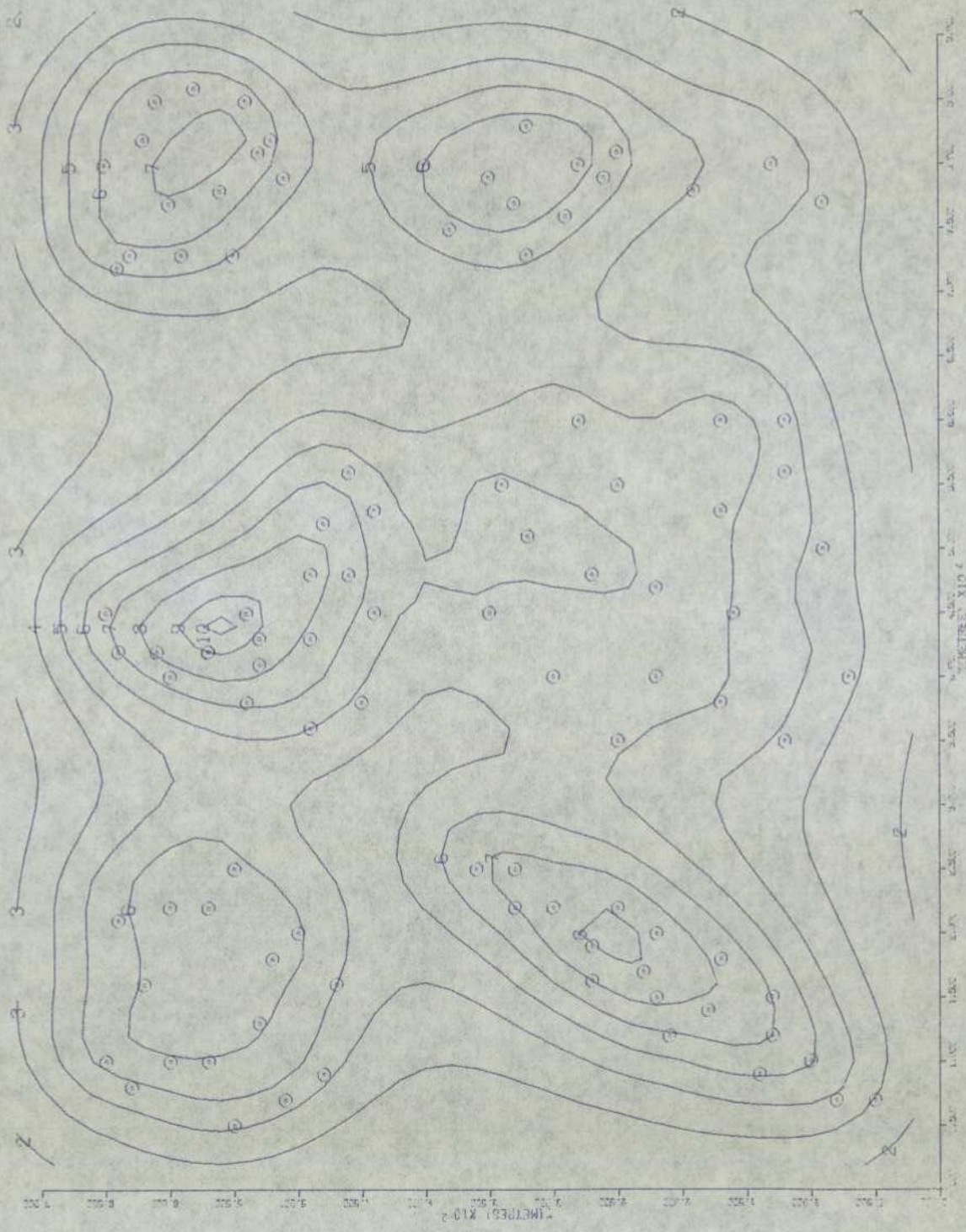


Fig. 5.6

Fig. 5.7

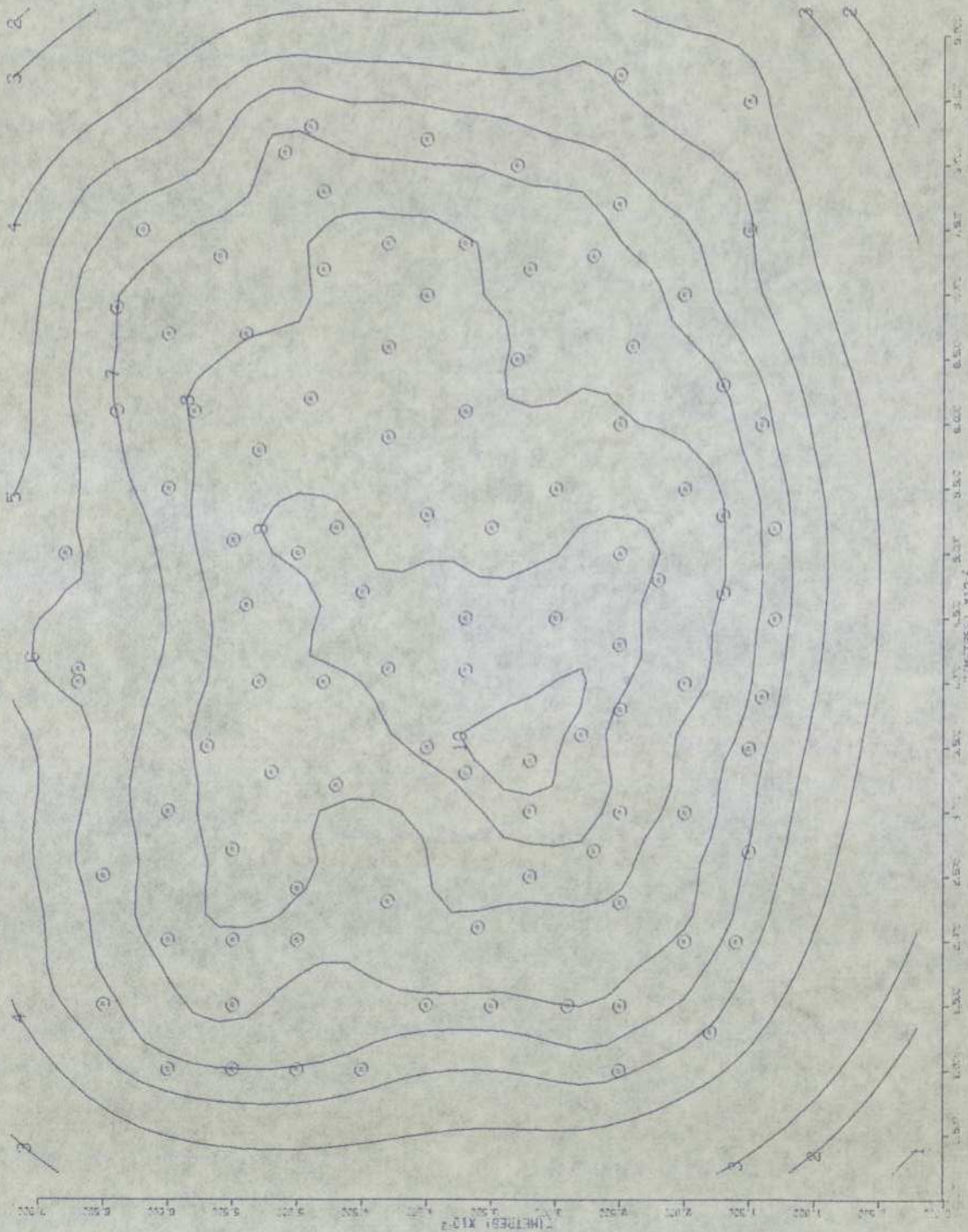


Fig.5.8

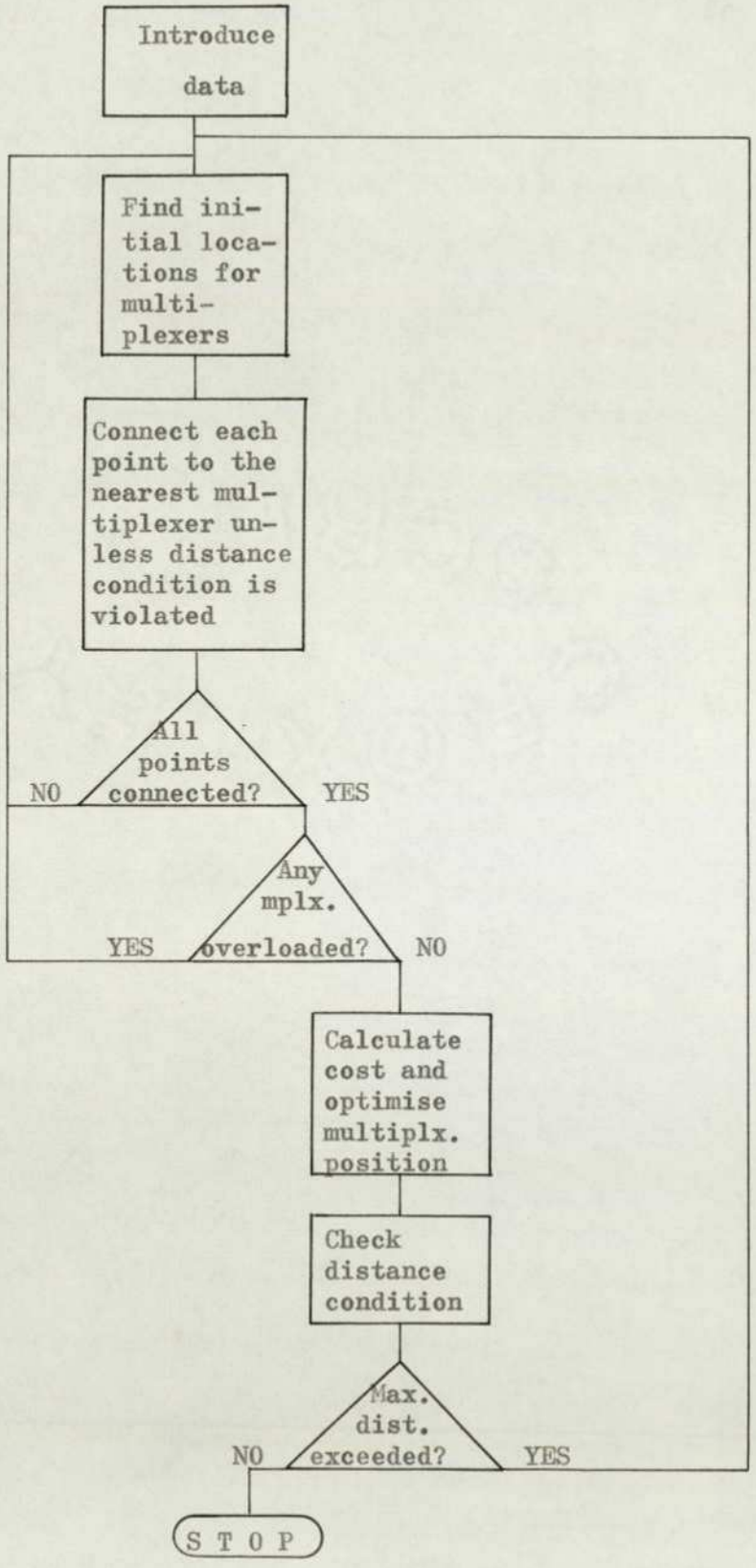

Introduce data
Find initial locations for multiplexers
Connect each point to the nearest multiplexer unless distance condition is violated
All points connected?
NO
YES
Any mplx. overloaded?
YES
NO
Calculate cost and optimise multiplx. position
Check distance condition
Max. dist. exceeded?
NO
YES
S T O P

Fig. 5.9

optimisation appears now obvious : replace the charges by data sources and potential maxima by multiplexers. The locations of the latter were determined such that they would coincide with centres of high aglomeration of the former. An appropriate heuristic algorithm is illustrated in the flow-chart of Fig. 5.9 . Initial locations for multiplexers are generated using the potential approach previously described. Then each point is connected to the nearest multiplexer, unless the point - nearest multiplexer distance exceeds a given value (as mentioned before, this value is function of the signal conditioning device used at the output of the sensor and of the noise level). After all points were considered, the algorithm checks if any point remained unconnected to any multiplexer. If the answer is yes, new initial locations for multiplexers are generated. The multiplexer loading is also checked and again new locations are generated, should any multiplexer be overloaded. As stated before, overloading could lead to unacceptable long scanning times and to difficulties in practical realization of multiplexers. When a set of initial multiplexer locations is found such that none of these restrictions are violated, the cost of wires connecting data sources to multiplexers, as well as the cost of the latter, is calculated and minimised by using a conventional optimisation routine, which shifts locally the multiplexers until an optimum is achieved. The distance condition is again checked and if satisfied, the problem is over; if not, a new initial guess is generated, and the procedure is repeated.

The whole procedure could be applied several times and the optimum optimorum selected accordingly. This is facilitated by the relative short computation time of each version, combined with a modest requirement for computer store.

The method has been tested and the results are summarised below.

Fig. 5.6 shows the aspect of the quasi-potential function contours for a set of hundred data sources which are relatively well separated into clusters (Fig. 5.5). Fig. 5.8 shows the same contours for a set of data sources for which clusters cannot be obviously obtained. The geometrical lay-out of the points generating them is shown in Fig. 5.7 . In Fig. 5.5 and 5.7 the data sources are represented by circles, whilst the position of maxima of quasi-potential are symbolised by crosses. All the aforementioned drawings have been obtained by using an ICL graph plotter.

The next section will describe another method of allocating suitable positions for multiplexers.

5.4. An algorithm based on a clustering technique

The ideea of using a clustering technique in finding optimum locations for multiplexers in a data communication system is suggested by the very nature of the problem. It is obvious that these devices should be placed where the density of data sources is maximum, in other words, somewhere inside the possible clusters one can isolate. However, this problem is far less complicated than those currently encountered in pattern recognition, mainly because the only answer one would like to obtain is where the points are concentrated and not their possible pertinence to one class or another. Secondly, the final aim of finding clusters in pattern recognition problems is to estimate probability densities of each of a number of classes, in order to obtain a statistically optimum clasification of an input into one of these classes. Obviously, this is a much more complex problem and could be too expensive for the case of communication system optimisation.

The method proposed in this section starts from the ideea of

clustering points in the n-dimentional space by using hyper-ellipsoids, the axes of which are determined by the variance of position of the clustered points (G.Sebestyen, J.Edie,1966). The technique can be summarised as it follows : the axes of the hyper-ellipsoidis have as length the variance of the p.d.f. (assumed Gaussian) of points belonging to the cluster defined by the considered hyper-ellipsoid. The first point position vector is considered as the first "typical sample". The density associated to it (assumed Gaussian) is calculated and stored. The second learning vector is used to generate a second cell, similar to the first, if it falls sufficiently outside the first cell. If the second vector falls inside the first cell, the centre of that cell is shifted to the mean of the two vectors and the shape and the size of the cell, then the probability density is recalculated as more information about the process becomes available. If the second vector falls outside the first cell, but not too far, it is stored temporarily to be reused after all the points were considered in the same manner as that described above. The use of so called "guard zones" (regions outside established clusters, of dimensions suitable chosen) prevents clusters of overlapping and therefore of producing cases in which the ambiguity does not allow a proper decision to be taken. The method makes use of several approximations and arbitrary chosen parameters : the points are assumed to be locally Gaussian - distributed, and the notions of "inside" or "outside" the guard zone established as function of a threshold, the value of which is determined somewhat experimentally.

The idea on which the technique described above is based could be useful in specific problems such as the optimisation of the cost of the communication system. One possible method serving this purpose is described below.

Consider the data sources arranged in a random sequence. A cell, consisting of a circle of radius R, is associated to the first data source, such that the centre of the circle coincides with the data source position. The distance between the second data source and the centre of the established cell is considered. If this distance exceeds R, a second cell, centred on the second data source, is established. In the opposite case the second data source is associated to the first cell and its centre is shifted to the mean position of the points considered. The process continues in the same manner until all the data sources are examined. Obviously, when a new point is considered, the distance between it and all the cells centres are calculated in the order in which the cells were established, and the point is assigned to the first cell for which the distance between the cell centre and the considered point does not exceed R. A good value for R is the maximum distance one can allow between a data source and the multiplexer to which it is connected. However, this does not guarantee that at the end of the computational process, all the data sources attached to a particular cell still remain inside the same cell, in other words there is no guarantee that the restriction of a maximum distance is still not violated. If such a violation occurs, this could be an indication that an insufficient number of cells have been generated, as there would be data sources too far from any multiplexer. Therefore if R is only a fraction of the maximum permitted data source multiplexer distance, this unwanted thing could be avoided. By trial and error it was found that $R = 0.9\ D_{max}$ is a satisfactory choice for both considered cases (clusters well separated and overlapping clusters).

One can see that such a method is strongly influenced by the order in which the data sources are considered. As result, if two circles overlap all the data sources lying in the common region will be assigned

to the circle which had the chance of being established first. This obvious disadvantage could be eliminated if the procedure of assigning each data source to the circle the centre of which is nearest is used. The results are even better if the data sources are considered in several random sequences, due to the fact that the computation takes a relatively short time. A program accomplishing these functions has been written. Fig. 5.12 presents its flow-chart. Again, the distributions of data sources shown in Fig. 5.5 and Fig. 5.7 were used. Fig. 5.10 and 5.11 show one set of circles as found by the mentioned program. One should mentionm that in the case of well-separated clusters, the centres of a set of circles, as well as the number of data sources associated to them, did not differ too much as compared with other sets of the same number of circles. The consistency of this result shows that the clustering technique used is effective for this case. When the clusters of data sources "overlap", the method again proves to be useful. In opposition to the pattern recognition problems, the centres of circles indicate zones of high concentrations of data sources and they may overlap without producing too much inconvenience. The second step of the method (assigning each data source to the nearest centre) corrects the imperfections introduced by considering the points in a given sequence.

It should be remembered also that the cluster centres are required only to serve as initial guess position for multiplexers in a cost optimisation procedure. Once each data source was connected to a multiplexer, the position of the latter is modified such that the cost of cables realising the connections is minimised. This procedure is applied for each multiplexer and although one cannot prove theoretically that this is the optimum method of minimising the cost of the communication system, one could assume that the results de not differ considerably from the

Fig. 5.10

Fig. 5.11

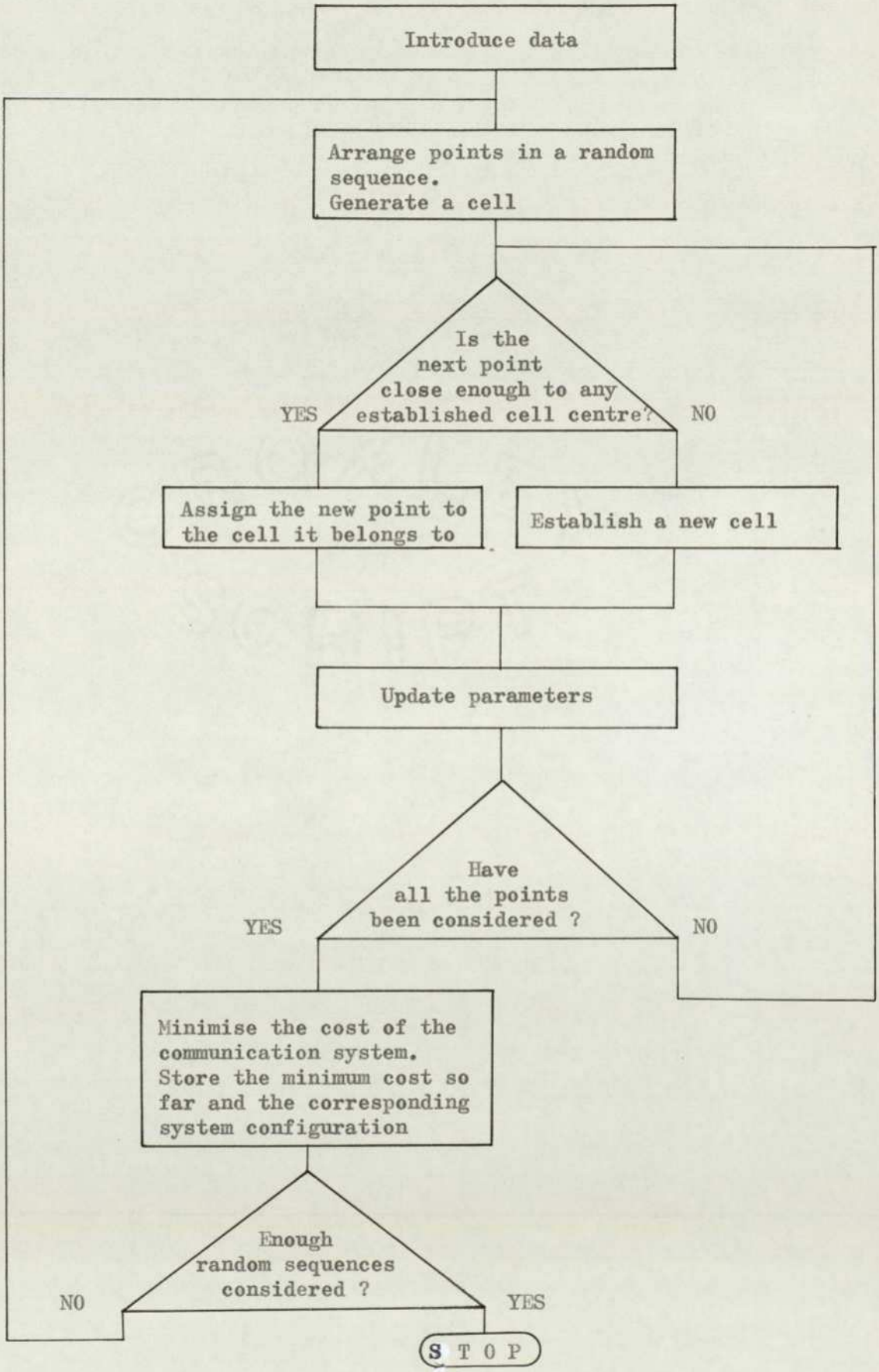

Introduce data
Arrange points in a random sequence.
Generate a cell
Is the next point close enough to any established cell centre?
YES
NO
Assign the new point to the cell it belongs to
Establish a new cell
Update parameters
Have all the points been considered ?
YES
NO
Minimise the cost of the communication system. Store the minimum cost so far and the corresponding system configuration
Enough random sequences considered ?
NO
YES
STOP

Fig. 5.12

theoretical absolute minimum. The flow-chart of the algorithm used in writing the program is shown in Fig. 5.12 .

Finally, a method somewhat similar to that described above can be mentioned. The method consists in finding the distance between all possible pairs of data sources and to start a cluster by using the points separated by the minimum distance. The cluster has the centre in the mean of the points positions. From the points which remain that which is nearest to the cluster centre is considered. If the mean of these three points does not drift away from the cluster centre with more than an empirically chosen threshold, the new point is assigned to the existing cluster. A new cluster is formed in the opposite case. The method, although conceptually simple, involves a large amount of computation and this amount increases dramatically with the number of points, and therefore is considered unsuitable for application involving several thousands of points.

5.5. Considerations about minimisation of the cost of the communication system

The previous two sections have described mainly heuristic algorithms for finding appropriate locations for multiplexers as the first step in the minimisation of the cost of the communication system. These locations are used by a routine, which connects each data source to the nearest multiplexer. After this stage each individual system, composed of a multiplexer and the data sources connected to it, is considered and a conventional minimisation routine is used for finding the position of the multiplexer for which the cost of cables connecting it to each data source is minimised. For this purpose a subroutine based on Powell's

algorithm was used.

However, a more detailed analysis shows that there are some restrictions which have to be considered before and during the cost minimisation. One of these restrictions consists of making sure that none of the multiplexers is overloaded (forced to accept more inputs than the number it is able to deal with). If such cases occur, a satisfactory solution could consider the use of two more multiplexers having the same location. The advantage of such a way of solving the problem consists of the saving in the cost of cable paths between the multiplexer and the computer (which could be several hundreds of metres or even kilometres) and also in the cost of installing, housing and maintaining the communication equipment. However, alternative solutions, involving an increased number of multiplexers should be considered and the best geometrical lay-out chosen. The second problem which requires attention is in connection with the eventuality that, during the cost minimisation operation, the multiplexer could drift away too far from some of the data sources connected to it. This fact is more unlikely to happen if the algorithm based on use of circles is employed in finding the multiplexers positions. The last, but nonetheless important fact to be aware of, consists of ensuring that any data source could transmit signals at time intervals non exceeding those determined by the dynamics of the controlled process. Therefore the number of data sources connected to a multiplexer, as well as the data sources themselves, must be chosen such that each data source can be accessed and activated at time intervals not exceeding given values. The exact solution of this problem is not simple at all. One possible way to approach it could be described as it follows. Assume that a data source connected to a multiplexer requires the use of communication facilities for a time interval t_s, and this request occurs at

time intervals not exceeding T_i. The time t_s is mainly determined by the length of the message to be transmitted. When quasi-digital signals represented by a number of pulses in a sequence are used, t_s varies not only from one data source to another, but also for the same data source, as the measured signal varies in time. t_s has the same values for all data sources when a quasi-digital signal conditioning device such as a voltage to frequency converter is used and t_s is chosen sufficiently long to allow an accurate conversion of the analogue signal into a quasi-digital one. Allowance should be made for t_s to include the duration of interrogation signal from the computer. The time T_i is chiefly a function of the process dynamics, and its values could be in the ratio of 1 : 100 if high speed and low speed processes have their significant measurands transmitted to the computer by using the same multiplexer. This being the case, the problem of determining the optimum configuration of the communication system becomes a problem of queue theory. In practical applications such problems are solved based on statistical properties of the message exchanges between the computer and the data sources (J. Martin, 1972). A possible way to conduct the solving of this question is presented below.

This specific problem of communication system design can be formulated as "single-server queuing" problem, the aim of which is to produce the approximate average size of queues and the average time of waiting in them. Also, the probability that the queue exceeds a particular size can be estimated. Based on the knowledge of the elements mentioned above, the response time can be estimated and therefore one can estimate if the communication system is able to provide the required information within an acceptable time interval from the moment of request. Consider the case of a queue with a single server, in which

the server is the multiplexer and the messages to be handled are the signals produced by the sensors connected to it. The result of measurement is available at the multiplexer at any moment from any data source. However, the transmission of each data requires a particular duration of time, often dependent of the magnitude of the measurand. Therefore the data sources should be connected to multiplexers such that the probability that a message is transmitted to the computer at time intervals longer than a prescribed value does not exceed a given value (ideally zero). In Fig. 5.13 such a situation is presented. By "service" one means the capability of transmitting the message, and by " facility" the way in which the service is done, i.e. by using a multiplexer. The mean E(n) of the number of arrivals per second should be interpreted as the mean number of results of measurement which have to be transmitted in any second. The mean service time $E(t_s)$ represents the average time required to transmit one message, the standard deviation of which is notated by σ_{ts}. Assume that the number of signals waiting

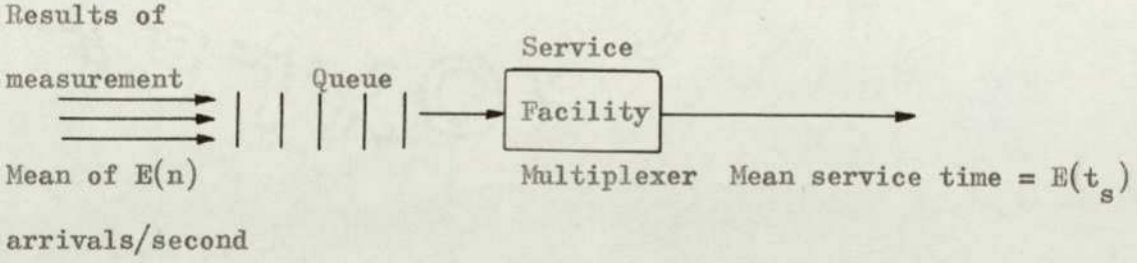


Fig. 5.13

for being transmitted is w and the total number of data sources connected to the multiplexer is q. In fact w represents the number of messages from data sources which must be transmitted under ideal operating conditions in the next millisecond or so. Assume that t_w is the time a data source waits before the message it has produced is transmitted and t_q is the total time for waiting and being transmitted, i.e.

$$t_q = t_w + t_s \quad . \tag{5.3}$$

Therefore

$$E(t_q) = E(t_w) + E(t_s) \tag{5.4}$$

and

$$Var(t_q) = Var(t_w) + Var(t_s) \quad , \tag{5.5}$$

since the duration of waiting for a particular message is independent of its length. Define a facility utilisation r as

$$r = E(n)\ E(t_s) \quad . \tag{5.6}$$

r represents the degree in which the facility is used and its magnitude provides information about the duration of waiting in a queue for a message before being processed. The steady state is of interest, when

$$E(w) = E(n)\ E(t_w) \tag{5.7}$$

and

$$E(q) = E(n)\ E(t_q). \tag{5.8}$$

By replacing (5.4) in (5.8) one obtains

$$E(q) = E(n)\ E(t_w) + E(n)\ E(t_s) \quad ,$$

or, using (5.6)

$$E(q) = E(n)\ E(t_w) + r \quad . \tag{5.9}$$

A theorem, developed by Khintchine and Polloczek, can be used in the case of single-server queuing theory. This theorem can be expressed as

$$E(w) = \frac{r^2}{2(1 - r)} \left[1 + \left(\frac{\sigma_{ts}}{E(t_s)} \right) \right]^2 \quad . \tag{5.10}$$

This result helps in estimating the queue-size. It should be mentioned that the formula (5.10) gives good results for any distribution of service time and any dispatching discipline. By dispatch discipline one understands the way in which the next message to be transmitted is chosen. Obviously, the dispatch discipline is established by the

computer, which indicates to the multiplexer which data source should transmit the message, but in many cases the multiplexer has a preset cycle of scanning. Even in such cases the theorem (5.10) can be applied. By replacing E(w) from (5.10) in (5.9), (5.4) and (5.6) one obtains :

$$E(q) = r + \frac{r^2}{2(1-r)}\left[1 + \left(\frac{\sigma_{ts}}{E(t_s)}\right)\right]^2 \qquad (5.11)$$

$$E(t_w) = \frac{r\,E(t_s)}{2(1-r)}\left[1 + \left(\frac{\sigma_{ts}}{E(t_s)}\right)\right]^2 \qquad (5.12)$$

$$E(t_q) = \frac{r\,E(t_s)}{2(1-r)}\left[1 + \left(\frac{\sigma_{ts}}{E(t_s)}\right)\right]^2 + E(t_s) \quad . \qquad (5.13)$$

The term

$$\frac{1}{2}\left[1 + \left(\frac{\sigma_{ts}}{E(t_s)}\right)\right]^2$$

describes the mean queue characteristics and is dependent on the dispersion of the service times. When the service times are constant, $\sigma_{ts} = 0$ and therefore

$$E(q) = r + \frac{r^2}{2(1-r)} \qquad (5.14)$$

and

$$E(t_q) = E(t_s)\left[1 + \frac{r}{2(1-r)}\right] . \qquad (5.15)$$

When t_s are exponentially distributed $\sigma_{ts} = E(t_s)$, thus :

$$E(q) = \frac{r}{1-r} \qquad (5.16)$$

and

$$E(t_q) = \frac{E(t_s)}{1-r} \qquad (5.17)$$

The practice has shown that these are extreme situations, most service times having values between these two cases, either because the service times cannot be absolutely constant, due to various lengths of the messages which are transmitted, or as result of the fact that the dispersion of t_s is not as wide as $E(t_s)$. As the design should consider the worst possible case, exponentially distributed t_s should be assumed as a guarantee for satisfactory operation. Also, the relations (5.16) and (5.17) have a simple form which makes the calculations easier. Fig. 5.14 shows a graph of $E(q)$ as function of r for the extreme cases $\sigma_{ts} = 0$ (the lower curve) and $\sigma_{ts} = E(t_s)$ (the upper curve). In practice σ_{ts} has the values corresponding to the shaded area. Fig. 5.15 shows the graph of $E(t_q)$ as a function of r, again for the same cases. In order

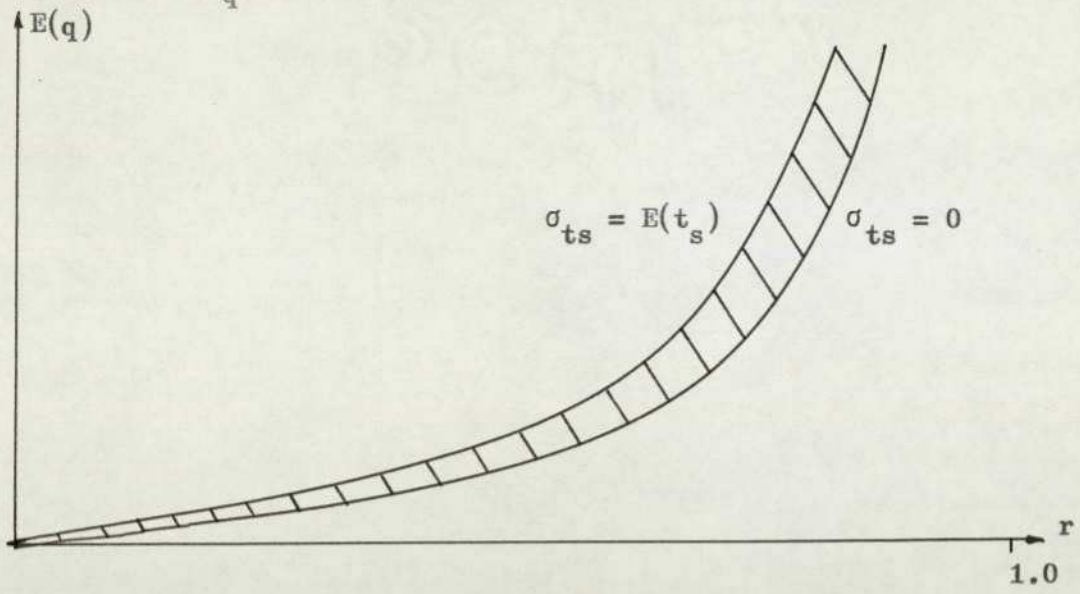


Fig. 5.14

to use for σ_{ts} values of practical significance, this parameter should be estimated before starting the design of the queue handling. With $E(t_q)$ and $E(t_s)$ estimated, the mean $E(t_w)$ of the waiting time t_w of a message before being transmitted is given by

$$E(t_w) = E(t_q) - E(t_s) \quad . \tag{5.18}$$

One can see from Fig. 5.15 and Fig. 5.14 that for values of r near 100% the size of the waiting queues becomes very large ; in other words,

as the device which provides the facility is used near its full capacity, one should expect long waiting queues and long waiting times before a message is processed. It can also be seen that the accurate knowledge of σ_{ts} has little importance for low values of r ; however, it is desirable to have for σ_{ts} an estimate as near to the true value as possible for low and high values of r. The designer should take into account the fact that an occasionally long message, which requires a long time for being processed, will produce much longer queues when the multiplexer is used near its full capacity.

Based on the same theorem of Khintchine and Polloczek, mentioned before, the standard deviation of the queue size can be evaluated (J.Martin, 1972). The result is shown in Fig. 5.16, in which the standard

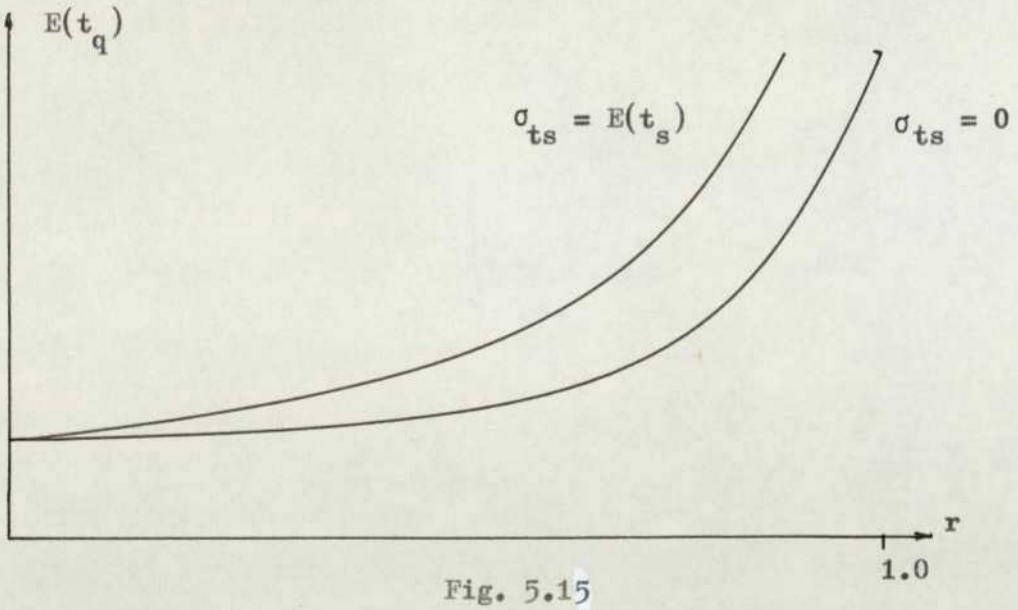


Fig. 5.15

deviation of the queuing times is plotted against the facility utilisation. The lower curve is plotted for constant service times ($\sigma_{ts} = 0$) and the upper curve for exponentially distributed service times ($\sigma_{ts} = E(t_s)$). Again , one can see that the latter is the worst case, as σ_{ts} increases more rapidly with r for the case of exponentially distributed service times.

The probability of queuing time t_q being greater than a given time

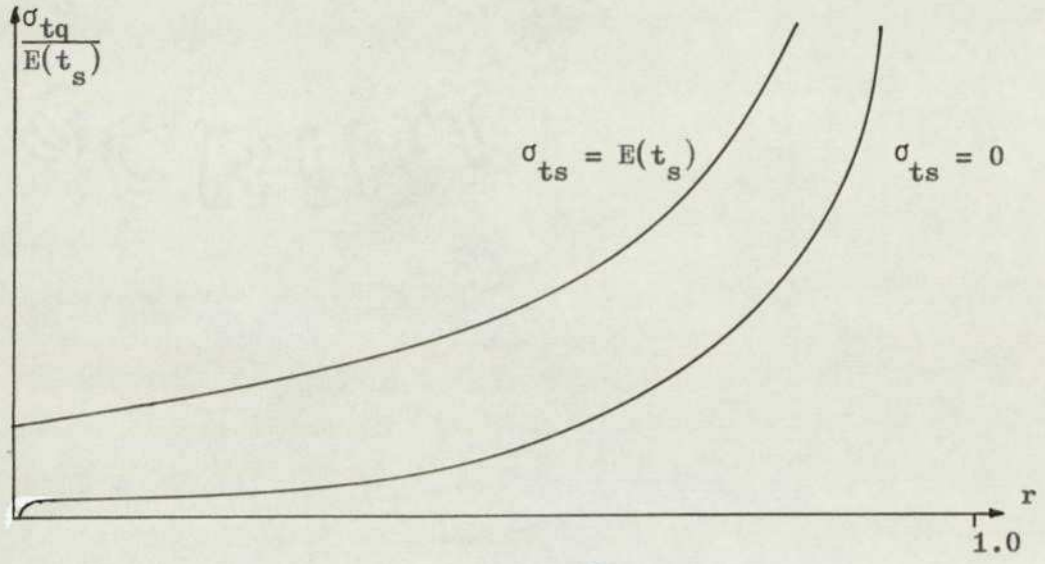


Fig. 5.16

T for the case of exponentially distributed service times is given by :

$$P(t_q > T) = \exp(- (1 - r)^T / E(t_s)) \quad . \qquad (5.19)$$

In Fig. 5.17 $P(t_q > T)$ is plotted against r. Obviously, P becomes smaller as T increases. For $T > t_s$, when constant service times are assumed, P = 0.

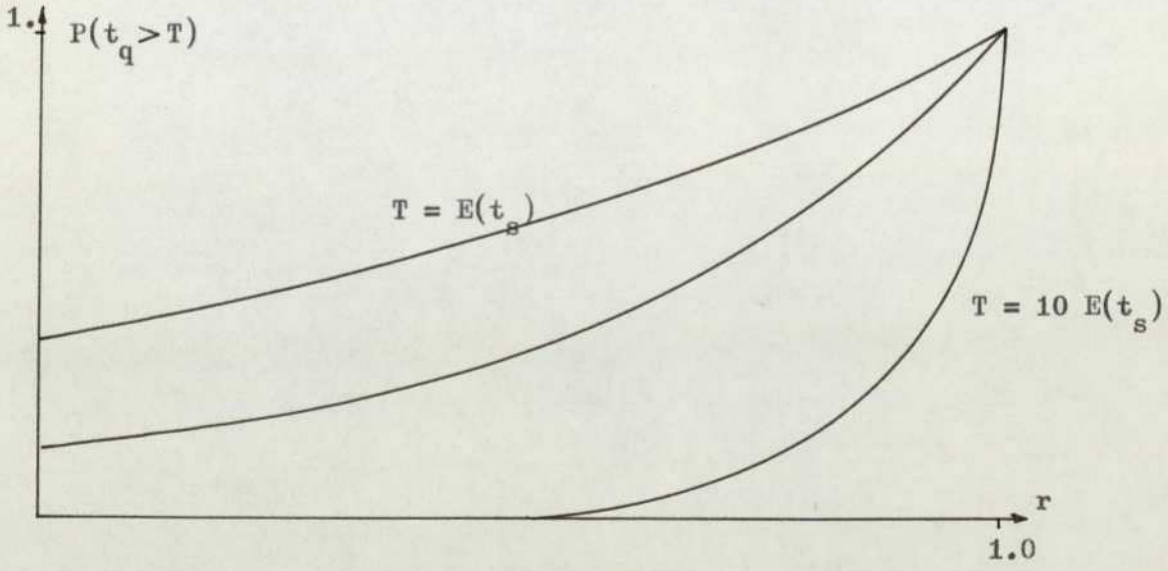


Fig. 5.17

The concepts mentioned above have to be considered when the required capacity of the multiplexer is calculated.

As one can easily see, the theory of queues could lead to unwanted

complications in the design of a communication system for the purpose of being used in a computer - controlled industrial process. Practical considerations show that, whilst t_s could have values of the order of 10 - 20 milliseconds, T_i usually takes values in the range from several seconds to tens of seconds or even minutes. This is the case of big plants, spread over large areas, where the problem of the minimisation of the cost of the communication system becomes vital. Such plants normally involve low speed processes (e.g. chemical plants, oil refineries, etc) and the assumption about large values of T_i as compared with t_s is correct. In this case, at least 100 - 200 data sources could be connected to the same multiplexer, if this becomes necessary. However, the queuing theory is invaluable when very large systems are designed and therefore the optimum use of the hardware is an essential element in increasing the profitability of the system. In fact, the conclusions obtained by using the queuing theory are always considered by the software designers and this is reflected in the manner of designing the appropriate hardware.

A final word should be said about the fact that the theoretically optimum cable path does not always correspond to that which can be implemented in practice. One could often encounter the problem of being forced to follow particular cable paths, especially when obstacles should be avoided or safety rules observed. In such situations, the experience of the designer has an important contribution in solving the problem, but the theoretically optimum location of the multiplexers is invaluable in deciding where these devices should be installed. Of course, one can imagine complicated algorithms for finding the solution, similar to those used in the design of printed circuit boards, but we have the feeling that such complications could save a particular amount from the cost of wires, compensating this by excessively increasing the cost of the design.

5.6 Practical results and conclusions

In this chapter an algorithm which minimises the cost of the communication system has been developed. The main concept at the base of this algorithm is that multiplexers should be used when the number of data sources becomes reasonably high (of the order of hundreds or thousands). Two heuristic algorithms, which provide initial guesses for the location of the multiplexers have been suggested and programs implementing them have been written. Although both these algorithms produce similar results, we consider the second one, which clusters the data by using circles of fixed radius, more adequate for solving the problem in hand. This affirmation is supported by the fact that the restriction that each data source should be connected to a multiplexer by a cable the length of which does not exceed a given value is satisfied almost any time the mentioned algorithm is run. This restriction could be violated if the case illustrated in Fig.5.8 occurs. Assume that after the first data source has been assigned as the centre of the circle (represented by dotted line), the next n-1 data sources have positions such that the centre of the circle is shifted progressively towards the aglomeration of the n - 1 data sources and the final circle (represented by full line) does not contain any longer the first point. If such cases occur, the final check made about the data source-multiplexer distances shows that the distance restriction is violated and therefore the solution as a whole is rejected and the procedure resumed with data sources arranged in a new random sequence.

A computer program implementing the algorithm has been run for two data sources geometric lay-outs, as shown in Fig. 5.5 and 5.7. The results are summarised in Table 5.1.

Table 5.1

Type of data sources lay-out	No. of multiplexers	Cost before optimisation	Cost after optimisation
Clusters well separated	6	147,056.51	114,303.57
- " -	7	155,726.51	108,973.70
- " -	8	146,935.19	108,992.57
- " -	9	171,249.49	107,355.26
Clusters overlapping	6	134,493.56	94,138.01
- " -	7	137,765.63	97,178.39
- " -	8	139,628.31	93,605.15
- " -	9	139,030.05	95,568.14

One can see that the optimisation is effective : the cost before minimisation is greater by 30% - 70% than the cost after minimisation for the considered cases, although the way in which the initial positions of multiplexers are found is somewhat optimal in itself. The fact that the solution employing a smaller number of multiplexers leads to greater costs should not surprise. Two extreme solutions could be imagined : one which involves only one multiplexer for all data sources, the other which involves a multiplexer for each data source. The cost of the former is £180,587.28 for well separated clusters of data sources and £152,734.86 for overlapping clusters of data sources. The prices for cables have been chosen in agreement with the information available (H. Simon, 1970) and the results of the discussions at the round table organised at IMEKO VI , 1973, Dresden. Both these solutions have have approximately the same cost (the cost of fixed units of each multiplexer is negligible as compared with the cost of cables). Therefore the optimum solution, if such solution exists, should involve a number of multiplexers between one and the number of data sources. The more

advantageous solution is that which requires a greater number of multiplexers, since the waiting time of the transmitted message becomes shorter. The fact that the run of the computer program takes a relatively short time (less than 3.5 seconds for the analysis of 50 different random sequences of data sources and four cost minimisations on a CDC 7600 machine) and small store (less than 7 kwords for 100 data sources and probably 10 kwords or so for 1000 data sources), several possible configurations of the data communication system can be produced and then the optimum configuration can be selected.

The practical results show the validity of the method. The systems considered here are very small (only 100 data sources) and therefore the saving in the cost of wires does not appear as significant as it is in the case of large systems, consisting of thousands of data sources. However, the percentage saving (40%) is a good indication about the usefulness of the proposed algorithm. An outstanding feature of it consists in its versatility, which makes it applicable to any system of data communication, regardless the type of transmitted signal (analogue, digital or quasi-digital), the nature of the transmitted signals (messages or electric power) or the type of data sources and multiplexers (consumers and power stations, customers and telephone exchanges, data transmission systems, etc)

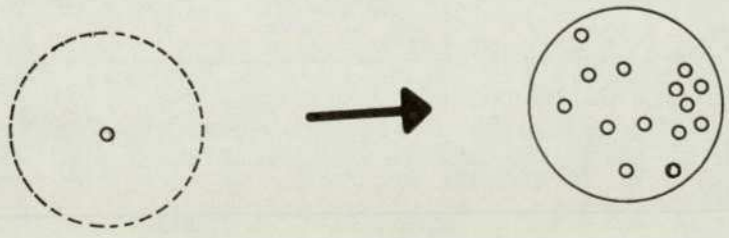


Fig. 5.18

CHAPTER 6

Final conclusions and lines for future research

This report has analysed from a systemic point of view problems of the process-computer interface. Each element of the chain between the data sources and the computer input has been examined. The study shows that the interface is still behind the advances in computer technology and thus the computer cannot be used at its full capabilities, since the information about the controlled process is not available in the best form yet. It has been emphasised that the conventional analogue signal conditioning devices will be gradually replaced by other forms of conditioning (digital or quasi-digital) as these will become cheaper due to advances in the technology of large scale integration.

Possible structures of quasi-digital conditioning devices have been analysed as an alternative for A/D converters as well as for classical analogue conditioning. The harmonic oscillators have been discarded after a detailed analysis which has shown their incapability of satisfying the requirements imposed for such instrumentation. It seems that, apart from specialised structures specifically designed as quasi-digital conditioning devices, the voltage-controlled oscillators will be used more and more in computer-based process control schemes.

As the output of the signal conditioning devices mentioned above has to be transmitted over long communication channels, the report has analysed possible ways of detecting the transmitted signals with minimum probability of error. Such analysis cannot assume the noise in communication channels operating in industrial environments as being merely Gaussian, therefore a possible way of expressing the probability

density function of the noise has been proposed. The advantage of the solution adopted is given by its versatility. Indeed, the proposed p.d.f. does not assume any particular mathematical form (as Poisson distribution, Cauchy distribution, etc.) and its parameters can be found based only on the knowledge of a noise sample obtained from the process under examination. The proposed probability density function does not produce spectacular theoretical results such as those obtained for Gaussian distribution. However, the use of the computer permits the calculations of all parameters required by the use of standard numerical methods. The implementation of the hardware used for detection of the transmitted signals is straightforward and is not too complicated as compared with the Gaussian process, but leads to better results.

As the size of the plants controlled by computer is usually large and the number of data sources often exceeds a thousand, it has been considered that only the use of multiplexers represents a realistic solution for the structure of the communication system. Therefore, finding the optimum locations for these devices is a problem the answer to which should be given. Since the design of a new plant often requires the consideration of several geometric lay-outs, the method of optimising the cost of the communication system should be effective and very efficient. Due to the size of the problem, the use of classical methods such as integer programming would be an aberation. Therefore, heuristic methods have been looked for. Two heuristic algorithms have been proposed and tested and the results are extremely satisfactory. The computer programs implementing them are very efficient and require small store.

This report has not fully solved by any means the problem of computer compatible instrumentation. There are some particular points which require attention in any future research, such as :

- an investigation from technological and cost point of view of possible structures of quasi-digital signal conditioning devices. This investigation should consider the elements involved in the mass production of such devices (structure, cost, etc.) in tight connection with technological requirements (conditions of operations, parameters, etc.). Reliability considerations show that the conditioning devices should be thought and realised as integrated circuits ;
- a detailed practical analysis of the electrical noise present in industrial environment, conducted for a large sample of noise, collected from various sites. Such analysis will show how complicated the estimated probability density function should be in order to achieve given standard performance after detection. At the same time the procedure of estimating the p.d.f. needs to be implemented as a computer program characterised by short computation time.

The results obtained so far show that the quasi-digital instrumentation has a very strong case in improving the performance of computer based control schemes and will become more and more used in the future in industrial applications.

APPENDIX A1

The analysis of linear oscillators

A careful consideration of the methods available has shown that a combination between piece-wise linearisation and root locus could provide the theoretical background for modelling any type of oscillator. This is based on the fact that the root locus method can produce good results when used for the analysis of linear systems and all nonlinear hardware elements used in oscillators (operational amplifiers and transistors) have a characteristic which can be easily approximated by three linear regions. This is shown in Fig. A1.1 for an operational amplifier and Fig. A1.2 for a transistor.

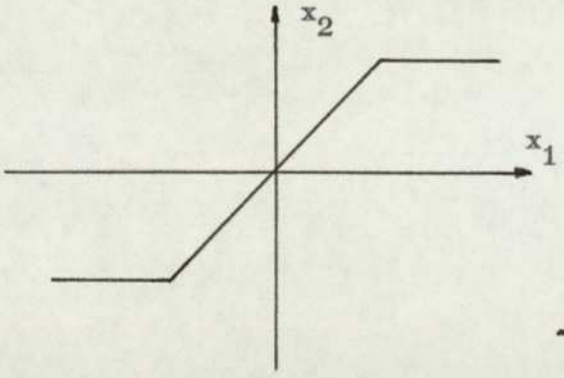


Fig. A1.1

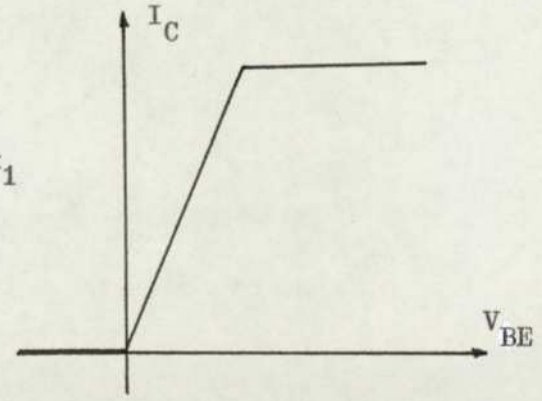


Fig. A1.2

An oscillator is a system consisting of a nonlinear amplifier with positive feedback (Fig. A1.3). The feedback network having the transfer function Y(s) is usually linear. The system starts to oscillate for a value k_c of the gain k named critical gain. If the gain is maintained at the value k_c the system oscillates with a frequency determined only by the feedback network. However, this ideal situation cannot be easily implemented in practice and thus gains above the critical value k_c should be used in order to ensure the operation of the oscillator even in worst

conditions.

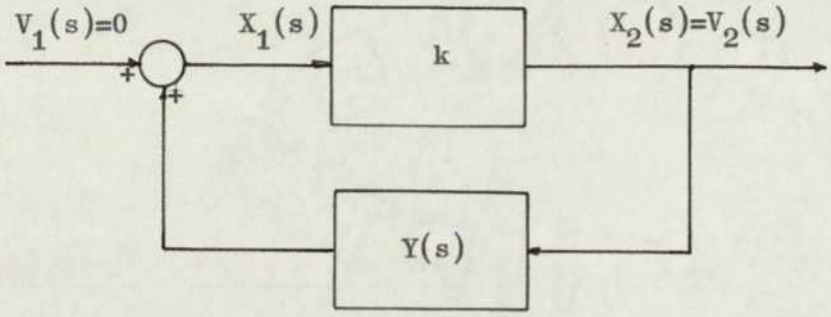


Fig. A1.3

Assume the general nonlinear characteristic approximated by three straight line regions (A,B and C), as shown in Fig. A1.4. Then the oscillator can be represented in the general form shown in Fig. A1.5.

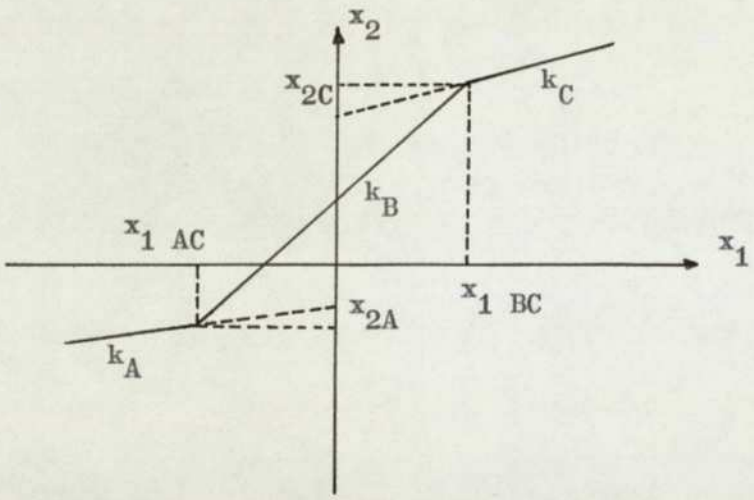


Fig. A1.4

The subscript n refers to either of the regions A, B or C. For the particular case $k_n = 0$ the oscillator becomes an open-loop system; in other words $X_2(s)$ remains constant, although $X_1(s)$ varies.

The output signal $v_2(t)$ can be obtained if the following procedure is used : assume that the operation point of the nonlinear element is in the region B. Thus $v_2(t)$ can be expressed as function of k_B, Y(s) and initial conditions. $v_1(t)$ can also be expressed. The point of operation remains in the region B until either

$$x_1(t_{BA}) = x_{1\ AB} \tag{A1.1}$$

or

$$x_1(t_{BC}) = x_{1\ BC} \tag{A1.2}$$

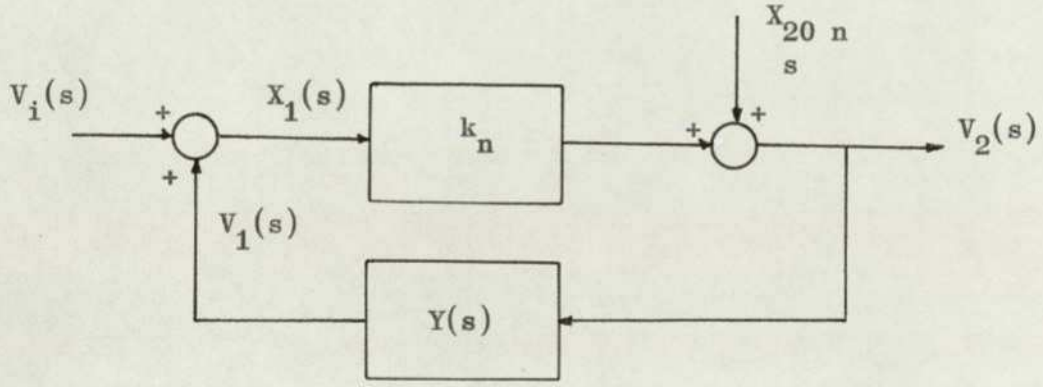


Fig. A1.5

By t_{IJ} the time required for transition from the region I to the region J was notated. When I = J t_{IJ} represents the time during the which the system operates in the region I.

The evolution of the system can be determined by solving the equations (A1.1) and (A1.2) : the operation point will move in the region A if $t_{BA} > t_{BC}$ or in the region C if $t_{BA} < t_{BC}$. Therefore the gain of the nonlinear element becomes k_c and, since all initial conditions and parameters are known, one can calculate the time duration t_C for which the system operates in the region C by solving the equation

$$x_1(t_C) = x_{1\ BC} \tag{A1.3}$$

Then after a time t_C the operation point moves again into the region B and, due to the positive feedback, it will be pushed towards the region A. This takes place after a time t_{BA} given by

$$x_1(t_{CA}) = x_{1\ AB} \tag{A1.4}$$

(measured from the moment at which the system began to operate again in the region B). The procedure continues and the time durations t_A and t_{AC} (characterising the operation of the system in the regions A and B

respectively) can be calculated. The period T of oscillation is then given by

$$T = t_C + t_{CA} + t_A + t_{AC} \quad . \tag{A1.5}$$

As example, a Wien bridge oscillator, having the circuit diagram as shown in Fig. A1.6, will be analysed. The time constant T and the gain k are given by :

$$\begin{aligned} T &= R\,C \\ k &= (R_1 + R_2)/R_2 \quad . \end{aligned} \tag{A1.6}$$

The nonlinear element, consisting of an operational amplifier, has a symmetric characteristic as shown in Fig. A1.7. Assume that at the initial moment the system operates in the region B. One can write the expressions for input and output voltages of various elements of the oscillator :

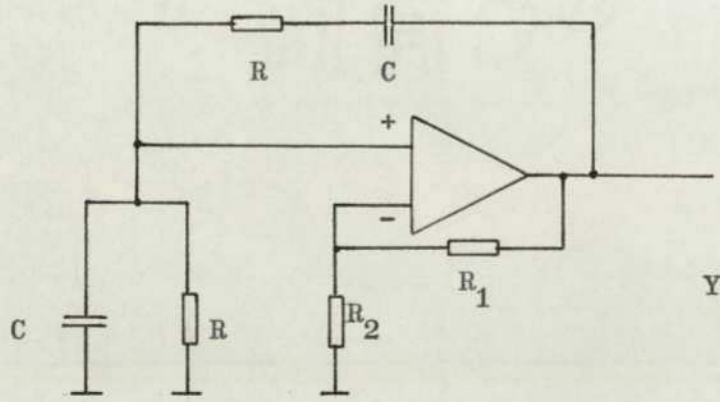
$$V_2(s) = k\,X_1(s) \tag{A1.7}$$

$$X_1(s) = V_1(s) \tag{A1.8}$$

$$V_1(s) = \frac{Ts}{1 + 3Ts + T^2s^2}\,V_2(s) \tag{A1.9}$$

thus

$$V_2(s)\left(1 + (3 - k)Ts + T^2s^2\right) = 0 \tag{A1.10}$$



$$Y(s) = \frac{Ts}{1 + 3Ts + T^2s^2}$$

Fig. A1.6

If the initial conditions are considered, (A1.10) becomes :

$$V_2(s)\left(1 + (3 - k)Ts + T^2s^2\right) - (3 - k)T\,V_{02} - T^2s\,V_{02} - T^2V_{02}' = 0 \qquad (A1.11)$$

where

$$V_{20} = V_2(0)$$
$$V_{20}' = \left.\frac{dV_2(t)}{dt}\right|_{t=0} \qquad (A1.12)$$

Assume $V_{20}' = 0$; then

$$V_2(s) = V_{02}\,T\,\frac{(3 - k) + Ts}{1 + (3 - k)Ts + T^2s^2} \qquad (A1.13)$$

and thus

$$v_2(t) = V_{02}\,\frac{\exp(-qt)}{p}\,\sin\left(pt + \tan^{-1}\frac{p}{q}\right), \qquad (A1.14)$$

with

$$p = \frac{1}{2}\left(6k - 5 - k^2\right) \; ; \quad q = \frac{1}{2T}(3 - k) \; .$$

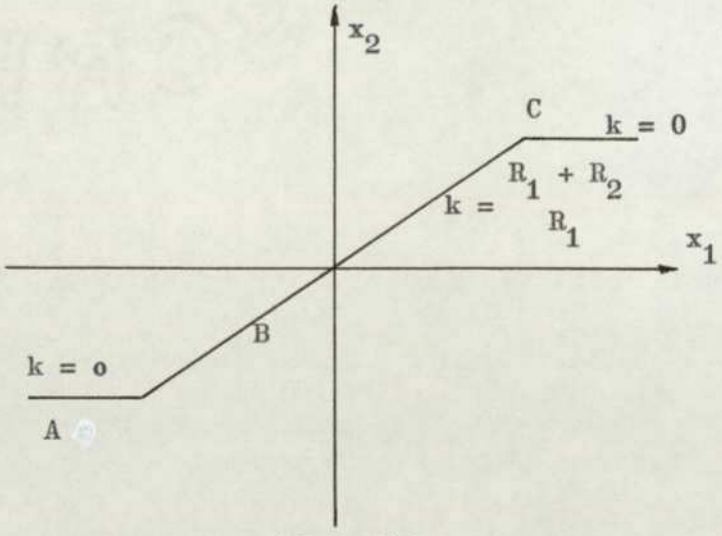


Fig. A1.7

Three cases could be considered :

i) $k < 3$. In this case $\lim_{t\to\infty} v_2(t) = 0$, therefore the output $v_2(t)$ is a damped sine wave.

ii) $k = 3$. The output signal $v_2(t)$ becomes

$$v_2(t) = V_{02}\,\sin\frac{t}{T}2\pi \qquad (A1.15)$$

(A1.15) shows that for the critical value of the gain (k = 3) the oscillator output is a sine wave, the amplitude of which is determined only by the initial condition V_{02}. The period of oscillation is a function of the feedback elements only. It can be seen that any perturbation in the system will alter the amplitude of the output signal.

iii) $k > 3$. One can see that $\lim_{t\to\infty} v_2(t) \to \pm\infty$. Due to the characteristic of the nonlinear element $v_2(t)$ does not exceed the saturation value V_{2sat}. When the operational amplifier enters the saturation region, its gain becomes zero and this state is maintained as long as $v_1(t) > \frac{V_{2sat}}{k}$. The oscillator has the behaviour of an open loop system (Fig. A1.8). The voltage at the input of the operational amplifier is described by

$$V_1(s) = \frac{V_{2sat}}{s} \frac{Ts}{1 + 3Ts + T^2s^2} \quad . \qquad \text{(A1.16)}$$

The period of oscillation can be determined by using the relations

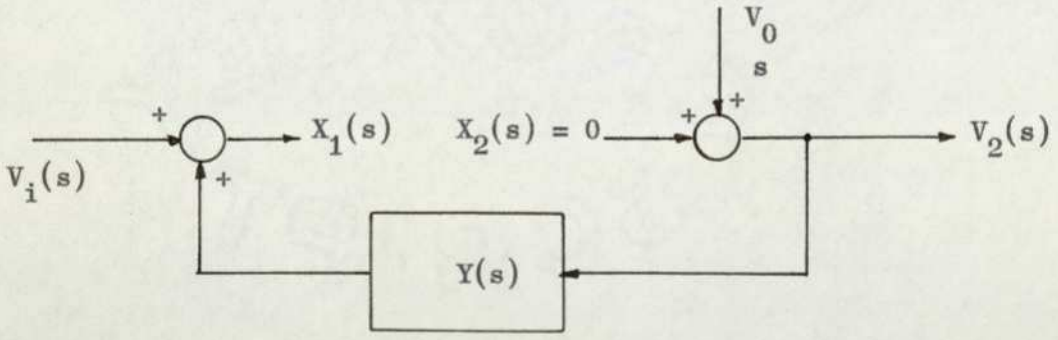


Fig. A1.8

(A1.14) and (A1.16). Assume that the time origin coincides with the moment at which the operational amplifier output enters the saturation. Therefore,

$$v_1(0) = \frac{V_{2sat}}{k} = V_{01}$$

$$\left.\frac{dv_1(t)}{dt}\right|_{t=0} = \frac{1}{k} \left.\frac{dv_2(t)}{dt}\right|_{t=0} = V_{01}' \qquad \text{(A1.17)}$$

When the initial conditions are considered, (A1.16) becomes :

$$V_1(s)\ (1 + 3Ts + T^2s^2) - 3TV_{01} - T^2V_{01}s - T^2V_{01}' = TV_{2sat}$$

or

$$V_1(s) = \frac{V_{2sat}}{k}\, T^2\, \frac{s + r}{1 + 3Ts + T^2s^2} \quad , \tag{A1.18}$$

where

$$r = \frac{V_{01}'}{V_{01}} + \frac{k + 3}{T} \quad .$$

The inverse Laplace transform of $V_1(s)$ is

$$v_1(t) = \frac{TV_{2sat}}{k\sqrt{5}} \Big(w_1 \exp(- w_1 t) - w_2 \exp(- w_2 t) + r \Big(\exp(- w_2 t) - \exp(- w_1 t) \Big) \Big) \quad , \tag{A1.19}$$

thus

$$v_1(t) = \frac{V_{2sat}}{k} \frac{T}{\sqrt{5}} \Big((w_1 - r) \exp(- w_1 t) - (w_2 - r) \exp(- w_2 t) \Big) \quad . \tag{A1.20}$$

w_1 and w_2 are given by

$$w_1 = \frac{1}{2T} (3 + \sqrt{5})$$

$$w_2 = \frac{1}{2T} (3 - \sqrt{5}) \quad .$$

The system will remain in saturation for a time t_{sat} given by :

$$v_1(t_{sat}) = \frac{V_{2sat}}{k} \quad . \tag{A1.21}$$

When the operational amplifier changes its mode of operation from saturation into the region of gain k, the initial conditions for the latter mode of operation are obtained from the equations

$$v_2(0) = V_{2sat} \tag{A1.22a}$$

$$\left. \frac{dv_2(t)}{dt} \right|_{t=0} = 0 \quad , \tag{A1.22b}$$

since $v_2(t)$ remains unchanged during the saturation. The relations (A1.14), (A1.22a) and (A1.22b) can be used for finding $v_2(t)$:

$$v_2(t) = V_{2sat} \frac{\exp(-qt)}{p} \sin(pt + \tan^{-1} \frac{p}{q}) \quad . \qquad (A1.23)$$

The system operates in the region of $k = 3$ for a time duration t_1 given by the smallest root of the equation

$$v_2(t) = V_{2sat} \quad . \qquad (A1.23a)$$

Due to the symmetric characteristic of the nonlinear element, the period T_o of oscillation can be calculated :

$$T_o = 2 (t_{sat} + t_1) \quad . \qquad (A1.24)$$

Fig. A1.9 shows the variation in time of the signals $x_1(t)$ and $x_2(t)$. The same signals have been obtained experimentally by using a Wien bridge oscillator. A good concordance between the theoretical and experimental results has been noticed. For $k = 5$ the computer program modelling the oscillator has produced the value $T_o = 2.722$ for the period, against the value of 2.76 found experimentally. The error, less than 1.5%, is probably due to the inaccuracy of the time base of the graph plotter used in the experiment. The experimental results have also shown the marked influence of the gain k on the period of oscillation. Table A1.1 summarises these results.

Table A1.1

k	T_k	T/T_3 100
3.0	2.16	0
3.5	2.34	8.3
5.0	2.76	27.7

One can see that for 16% increase of k as compared with the critical value, the period of oscillation has an increase of 8.3%. Therefore,

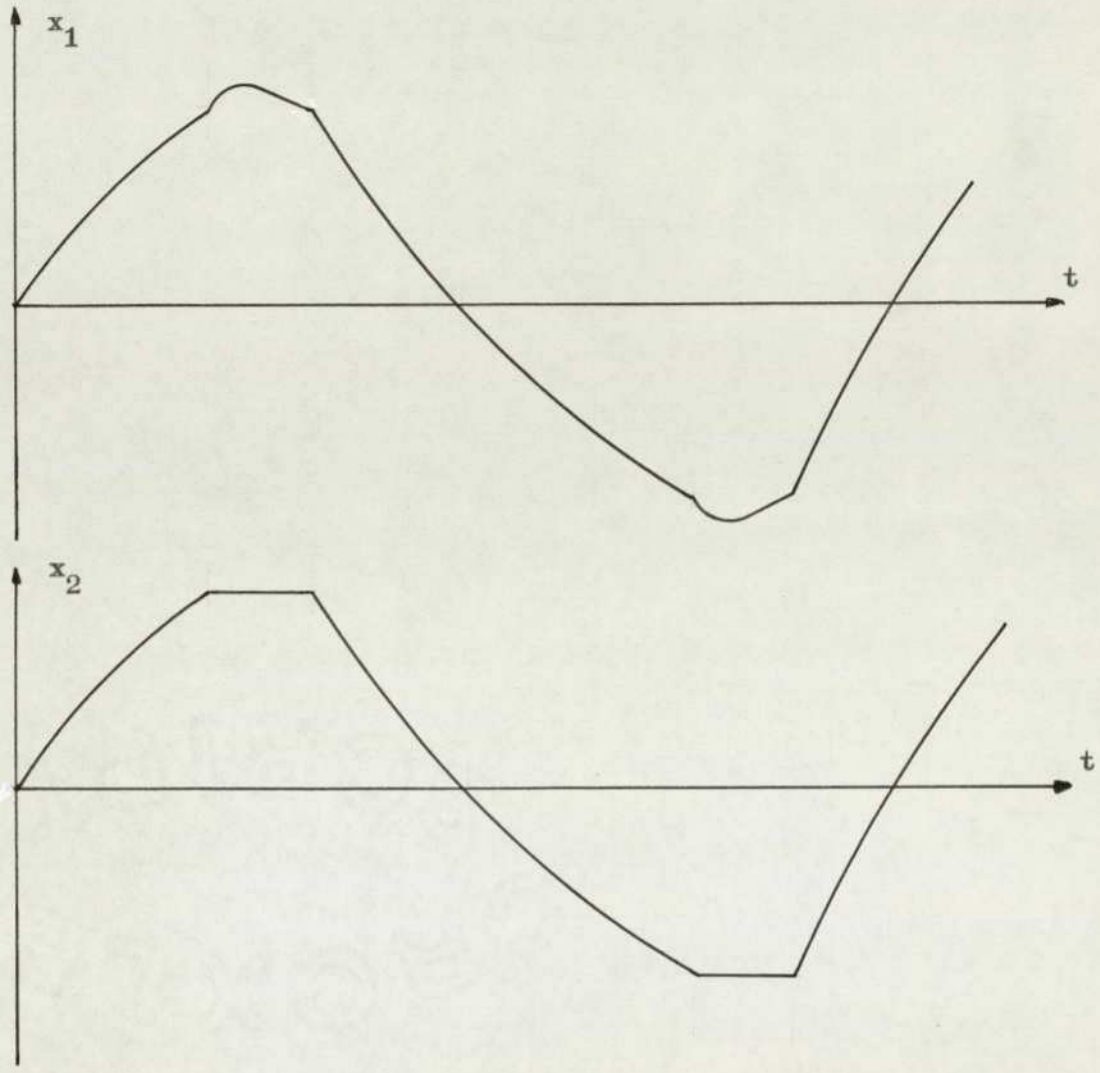


Fig. A1.9

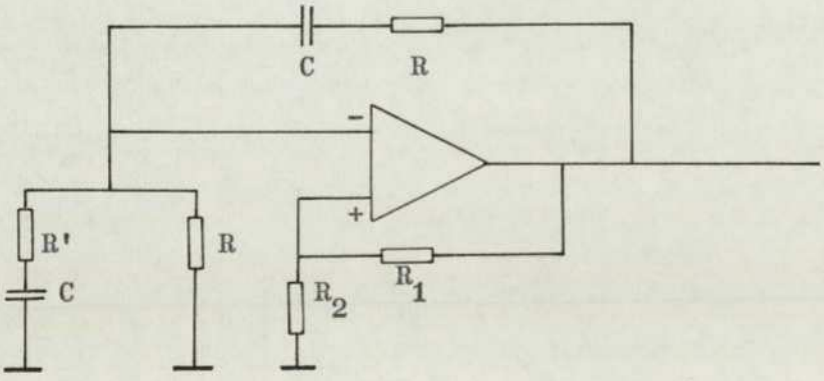


Fig. A1.10

practical considerations do not allow the linear dependence between T and $R = R_1 = R_2$, and thus the principles exposed in (D. Meyer, 1968) and (H.R. Tränkler, 1973) are inappropriate for practical implementation.

The use of the root locus method. Consider an oscillator having a second order characteristic equation of the form

$$1 - k\frac{a_1 s^2 + a_2 s + a_3}{b_1 s^2 + b_2 s + b_3} = 0 \quad . \tag{A1.25}$$

The equation of the root locus is given by :

$$(b_1 - k\,a_1)\,s^2 + (b_2 - k\,a_2)\,s + (b_3 - k\,a_3) = 0 \quad . \tag{A1.26}$$

For $s = \sigma + j\omega$, (A1.26) becomes

$$(b_1 - k\,a_1)\,(\sigma^2 - \omega^2) + (b_2 - k\,a_2) + (b_3 - k\,a_3) + \\ + j\left(2\sigma\omega\,(b_1 - k\,a_1) + (b_2 - k\,a_2)\omega\right) = 0 \quad .$$

Thus

$$(b_1 - k\,a_1)\,(\sigma^2 - \omega^2) + (b_2 - k\,a_2) + (b_3 - k\,a_3) = 0 \tag{A1.27a}$$

and

$$\omega\left(2\sigma(b_1 - k\,a_1) + (b_2 - k\,a_2)\right) = 0 \quad . \tag{A1.27b}$$

By eliminating k between (A1.27a) and (A1.27b) one obtains :

$$\omega = 0 \tag{A1.28a}$$

and

$$(a_1 b_2 - a_2 b_1)\,(\sigma^2 + \omega^2) - 2\,(a_3 b_1 - a_1 b_3)\sigma + (a_2 b_3 - a_3 b_2) = 0 \quad . \tag{A1.28b}$$

(A1.28a) shows that part of the locus coincides with the real axis, whilst (A1.28b) represents the equation of a circle, with the centre of coordinates

$$\left(\frac{a_3 b_1 - a_1 b_3}{a_1 b_2 - a_2 b_1}\ ,\ 0\right)$$

and crossing the imaginary axis at

$$\omega = \pm \sqrt{\frac{a_3 b_2 - a_2 b_3}{a_1 b_2 - a_2 b_1}} \quad . \tag{A1.29}$$

Assume that the coefficients a_i and b_i, i = 1,2,3, satisfy the relationship

$$a_1 b_2 - a_2 b_1 = 0 \quad . \tag{A1.30}$$

Under such circumstances, the circle becomes a straight vertical line which crosses the imaginary axis at $\omega = \pm \infty$. A method of designing sensitive second order oscillators can be developed. If only b_2 is variable, a_1, a_2 and b_1 can be chosen such that

$$a_1 b_{2\,\min} - a_2 b_1 = \eta \tag{A1.30a}$$

where η is a small quantity and $b_{2\,\min}$ is the minimum value b_2 can take. As result, for small percentage variations of b_2 ω will change between large limits. The dependence $\omega = f(b_2)$ is nonlinear, but a high sensitivity is achieved. One possible high sensitivity oscillator has the circuit diagram as shown in Fig. A1.10. The variable element is R'. The transfer function Y(s) of the feedback network is

$$Y(s) = \frac{T\,(T + T')\,s^2 + (2\,T + T')\,s}{(T'^2 + 2\,T\,T')\,s^2 + (3\,T + T')\,s + 1}, \tag{A1.31}$$

with

$$\begin{aligned} T &= R\,C \\ T' &= R'\,C \\ k &= \frac{R_1 + R_2}{R_2} \quad . \end{aligned} \tag{A1.32}$$

k is the gain factor of the operational amplifier.

For

$$k = 3 + \frac{T'}{T} \tag{A1.33}$$

the system will oscillate with the frequency

$$\omega = \frac{\sqrt{2 + T'/T}}{\sqrt{T^2 - T\,T' - T'^2}} \qquad (A1.34)$$

The system will have high sensitivity if the condition :

$$T\,(T + T')\,(3\,T + T') - (T'^2 + 2\,T\,T')\,(2\,T + T') \leqslant \eta \qquad (A1.35)$$

obtained from (A1.30a), is fulfilled.

As the second order oscillators cannot offer adequate performance (the relation (A1.35) is easily violated when T is modified as result of enviromental conditions, ageing, etc), one should find whether higher order oscillators, having more sophisticated feedback networks, can achieve better performance. For reasons which will appear obvious later, a sixth order feedback is first considered :

$$Y(s) = \frac{a_1 s^6 + a_2 s^5 + \ldots + a_7}{b_1 s^6 + b_2 s^5 + \ldots + b_7} \quad ; \qquad (A1.36)$$

the characteristic equation of the system is

$$1 + k\,Y(s) = c_1 s^6 + c_2 s^5 + \ldots + c_7 = 0 \qquad (A1.37)$$

with

$$c_i = b_i - k\,a_i \quad , \quad i = 1,\ldots,7 \quad .$$

The frequency of oscillation can be determined by replacing in (A1.37) $s = \sigma + j\omega$ and $\sigma = 0$. One finds :

$$(-c_1\omega^6 + c_3\omega^4 - c_5\omega^2 + c_7) + j\,(c_2\omega^5 - c_4\omega^3 + c_6\omega) = 0 \quad , \qquad (A1.38)$$

or

$$c_1\omega^6 - c_3\omega^4 + c_5\omega^2 - c_7 = 0 \qquad (A1.38a)$$

$$\omega\,(c_2\omega^4 - c_4\omega^2 + c_6) = 0 \qquad (A1.38b)$$

The analysis is considerably simplified, without loss of generality, if the assumption that c_1, c_3, c_5 and c_7 are not function of k is made. Under such circumstances ω is determined from (A1.38a) and the critical value of the gain from (A1.38b). A measure of the sensitivity of the

system can be obtained from (A1.38a). Note

$$w = \omega^2 \ . \tag{A1.39}$$

If the variable parameter which produces changes in the frequency of oscillation ω is x, the sensitivity can be characterised as $\frac{d\omega}{\omega}$. From (A1.39)

$$dw = 2\,\omega\, d\omega \ . \tag{A1.40}$$

Therefore, if (A1.40) is divided by ω^2, one obtains :

$$\frac{dw}{w} = 2\,\frac{d\omega}{\omega} \tag{A1.41}$$

and then, at a given frequency, one can consider $\frac{dw}{w}$ a measure of sensitivity instead of $\frac{d\omega}{\omega}$. Thus, from (A1.38) and (A1.41) :

$$\frac{dw}{w} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial w}}\ \frac{x}{w}\,\frac{dx}{x} \ , \tag{A1.42}$$

or

$$\frac{dw}{w} = -\frac{x}{w}\,\frac{d_1 w^3 - d_3 w^2 + d_5 w - d_7}{3\,c_1 w^2 - 2\,c_3 w + c_5}\,\frac{dx}{x} \ , \tag{A1.43}$$

where

$$f(x,w) = c_1(x)\,w^3 - c_3(x)\,w^2 + c_5(x)\,w - c_7 \qquad \text{from (A1.38a)} \ ,$$

$$d_i = \frac{dc_i(x)}{dx} \ , \ i = 1,3,5,7$$

One can choose $c_1(x) = 1$, therefore $d_1 = 0$. (A1.43) becomes :

$$\frac{dw}{w} = \frac{x}{w}\,\frac{d_3 w^2 - d_5 w + d_7}{3\,w^2 - 2\,c_3 w + c_5}\,\frac{dx}{x} \ . \tag{A1.44}$$

In order to achieve a high sensitivity, the following condition must be fulfilled :

$$\left| 3\,w^2 - 2\,c_3 w + c_5 \right| \leqslant \eta \ ; \tag{A1.45}$$

also the ratio x/w should be sufficiently large. Thus, it follows that the equation (A1.38a) should have a double root w_0 if the conditions (A1.45) and (A1.38a) are to be satisfied. Then (A1.38a) can be written as :

$$(w - w_0)^2 (w - w_1) = 0 \quad ,$$

therefore (A1.44) becomes

$$\frac{dw}{w} = \frac{x}{w} \frac{w_1' (w - w_1)}{3 (w - w_0) (w - w_2)} \frac{dx}{x} \quad , \qquad \text{(A1.46)}$$

in which

$$w_1' = \frac{\partial w_1}{\partial x} \ ,$$

and w_2 is the second root of the derivative of (A1.38a). If

$$|w - w_0| \leqslant \varepsilon \quad ,$$

ε being a small quantity,

$$\frac{dw}{w} = x \frac{w_1' (w - w_1)}{3 (w - w_2)} \frac{1}{\varepsilon} \frac{dx}{x} \quad . \qquad \text{(A1.47)}$$

Therefore a high sensitivity is achieved, but this sensitivity is dependent on the frequency of oscillation. In order to ensure that an acceptable high sensitivity is obtained when w varies between its extreme values, w_{min} and w_{max}, the parameters must be chosen such that

$$w_1 \leqslant w_{min} < w_{max} \leqslant w_0 \ . \qquad \text{(A1.48)}$$

In this case the dependence of the sensitivity on w has the aspect shown in Fig. A1.11

Now the systems of 3-rd, 4-th and 5-th order can be discussed. For a 3-rd order system (A1.37) becomes

$$c_1 s^3 + c_2 s^2 + c_3 s + c_4 = 0 \qquad \text{(A1.49)}$$

Then (A1.38a), (A1.38b) and (A1.43) can be written as

$$c_2 \omega^2 - c_4 = 0 \qquad \text{(A1.50a)}$$

$$\omega\,(c_1\omega^2 - c_3) = 0 \qquad \text{(A1.50b)}$$

$$\frac{dw}{w} = -\frac{x}{w}\,\frac{d_2 w - d_4}{2\,c_2 w}\,\frac{dx}{x} \quad . \qquad \text{(A1.51)}$$

For a 4-th order system the same relations become :

$$c_1 s^4 + c_2 s^3 + c_3 s^2 + c_4 s + c_5 = 0 \qquad \text{(A1.52)}$$

$$c_1\omega^4 - c_3\omega^2 + c_5 = 0 \qquad \text{(A1.53a)}$$

$$(c_2\omega^2 - c_4) = 0 \qquad \text{(A1.53b)}$$

$$\frac{dw}{w} = -\frac{x}{w}\,\frac{d_1 w^2 - d_3 w + d_5}{2\,c_1 w - c_3}\,\frac{dx}{x} \qquad \text{(A1.54)}$$

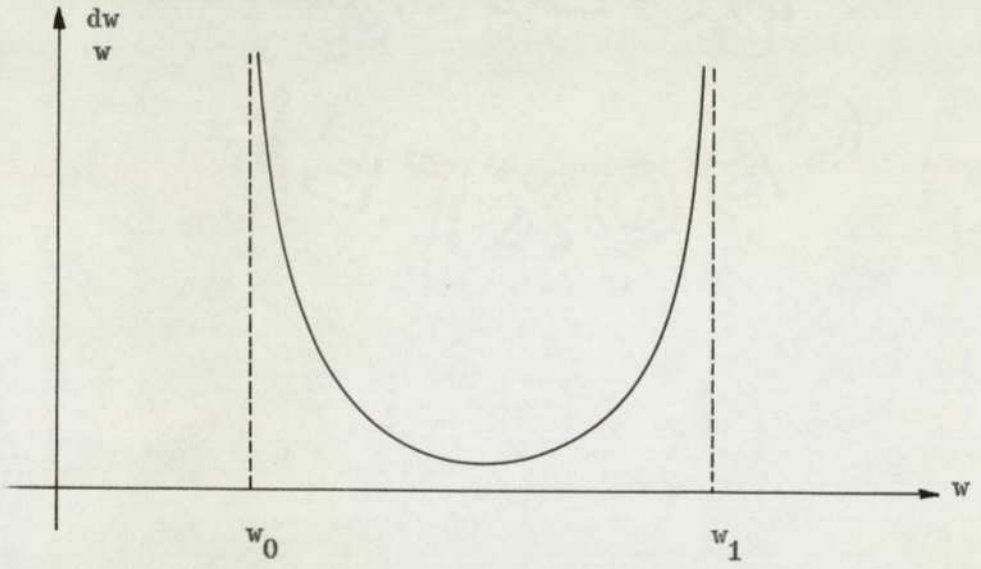


Fig. A1.11

For a 5-th order system the relations (A1.37), (A1.38a), (A1.38b) and (A1.43) have the form

$$c_1 s^5 + c_2 s^4 + \ldots + c_6 = 0 \qquad \text{(A1.55)}$$

$$c_2\omega^4 - c_4\omega^2 + c_6 = 0 \qquad \text{(A1.56a)}$$

$$(c_1\omega^4 - c_3\omega^2 + c_5) = 0 \qquad \text{(A1.56b)}$$

$$\frac{dw}{w} = -\frac{x}{w}\,\frac{d_2 w^2 - d_4 w^2 + d_6}{2\,c_2 w - c_4}\,\frac{dx}{x} \quad . \qquad \text{(A1.57)}$$

The analysis of (A1.51) shows that the sensitivity of the third order system cannot be adjusted, since the condition $2\,c_2 w \leqslant \eta$ implies an

unstable feedback network. At the same time, (A1.50a) cannot have a nonzero double root. The case of fourth and fifth order system is similar : high sensitivity is achieved for

$$\left|2\ c_1 w - c_3\right| \leqslant \eta \qquad \text{(A1.58)}$$

and

$$\left|2\ c_2 w - c_4\right| \leqslant \eta \quad . \qquad \text{(A1.59)}$$

The maximum sensitivity is achieved when the equations (A1.53a) and (A1.56a) have a double root, obtained from (A1.58) and (A1.59) respectively. Assume $\eta = 0$ and replace w as obtained from (A1.58) and (A1.59) in (A1.54) and (A1.57) respectively. One obtains :

$$\frac{dw}{w} = \frac{x}{w}\frac{d_3}{2}\frac{dx}{x} \qquad \text{(A1.60)}$$

and

$$\frac{dw}{w} = \frac{x}{w}\frac{d_4}{2}\frac{dx}{x} \quad , \qquad \text{(A1.61)}$$

therefore the sensitivity cannot be controlled.

The conclusion one can derive from the above considerations is that the simplest system the sensitivity of which can be made reasonably high for a relatively wide range of frequencies is the sixth order system. The practical realisability of such high sensitivity systems will be discussed below. Considerations based on the root locus theory lead to a root locus as shown in Fig. A1.12. Such system must have two roots in the positive semiplan, therefore the feedback network is unstable and the idea of realising this system should be abandoned. It seems highly impractical to consider feedback networks of orders even higher than six. Even more, it can be proved that no system, regardless its order, can achieve high sensitivity as postulated before.

Consider an eighth order system, having the root locus as plotted in Fig. A1.13. The parameters shown on the plot must satisfy the

conditions

$$\begin{aligned} \omega_0 &< \omega_1 \\ \omega_2 &< \omega_0 \\ k_0 &< k_1 \\ k_2 &> k_1 \\ \omega_0 &\text{ - double root} \end{aligned} \tag{A1.62}$$

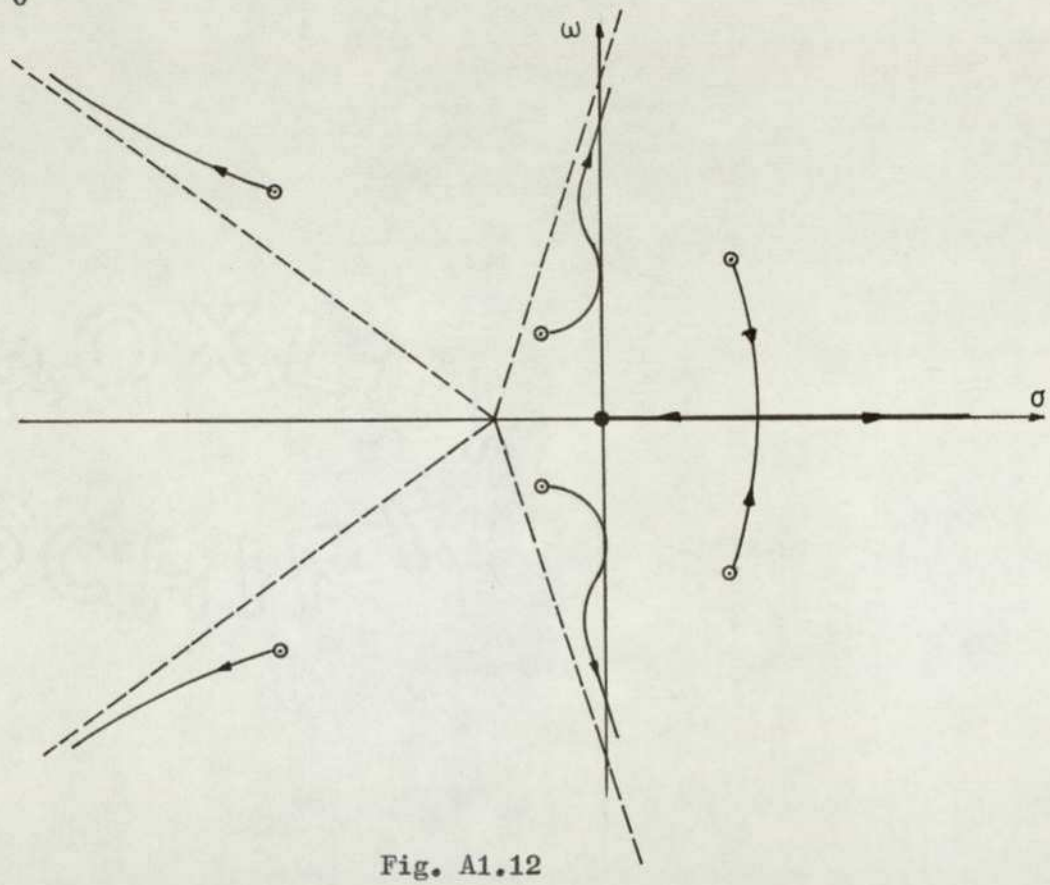


Fig. A1.12

In other words, if the transfer function is

$$Y(s) = \frac{a_8 s}{b_1 s^8 + \ldots + b_9} \tag{A1.63}$$

and the characteristic equation is

$$\mathrm{Re}\left(f(\omega,k)\right) + j\ \mathrm{Im}\left(f(\omega,k)\right) = 0 \tag{A1.64}$$

with

$$\mathrm{Re}\left(f(\omega,k)\right) = b_1\omega^8 - b_3\omega^6 + \ldots + b_9 = 0 \tag{A1.64a}$$

$$\mathrm{Im}\left(f(\omega,k)\right) = \omega\,(b_2\omega^6 - b_4\omega^4 + \ldots + b_8 + k\,a_8) = 0\ . \tag{A1.64b}$$

The conditions (A1.62) can be written as :

$$
\begin{aligned}
&\mathrm{Re}\left(f(\omega_0,k_0)\right) = 0\\
&\frac{\partial}{\partial\omega}\mathrm{Re}\left(f(\omega,k)\right)\Big|_{\substack{\omega=\omega_0\\ k=k_0}} = 0\\
&\mathrm{Im}\left(f(\omega_0,k_0)\right) = 0\\
&\frac{\partial}{\partial\omega}\mathrm{Im}\left(f(\omega,k)\right)\Big|_{\substack{\omega=\omega_0\\ k=k_0}} = 0\\
&\mathrm{Re}\left(f(\omega_1,k_1)\right) = 0\\
&\mathrm{Im}\left(f(\omega_1,k_1)\right) = 0\\
&\mathrm{Re}\left(f(\omega_2,k_2)\right) = 0\\
&\mathrm{Im}\left(f(\omega_2,k_2)\right) = 0
\end{aligned}
\qquad \text{(A1.65)}
$$

Assume $b_1 = 1$, $a_8 = 1$, $w = \omega^2$. The relations (A1.65) produce the set of simultaneous equations :

$$
\begin{aligned}
&w_0^4 - b_3 w_0^3 + b_5 w_0^2 - b_7 w_0 + b_9 = 0\\
&4\, w_0^3 - 3\, b_3 w_0^2 + 2\, b_5 w_0 - b_7 = 0\\
&b_2 w_0^3 - b_4 w_0^2 + b_0 w_0 - b_8 + k_0 = 0\\
&3\, b_2 w_0^2 - 2\, b_4 w_0 + b_6 = 0\\
&w_1^4 - b_3 w_1^3 + b_5 w_1^2 - b_7 w_1 + b_9 = 0\\
&b_2 w_1^3 - b_4 w_1^2 + b_0 w_1 - b_8 + k_1 = 0\\
&w_2^4 - b_3 w_2^3 + b_5 w_2^2 - b_7 w_2 + b_9 = 0\\
&b_2 w_2^3 - b_4 w_2^2 + b_6 w_2 - b_8 + k_2 = 0
\end{aligned}
\qquad \text{(A1.66)}
$$

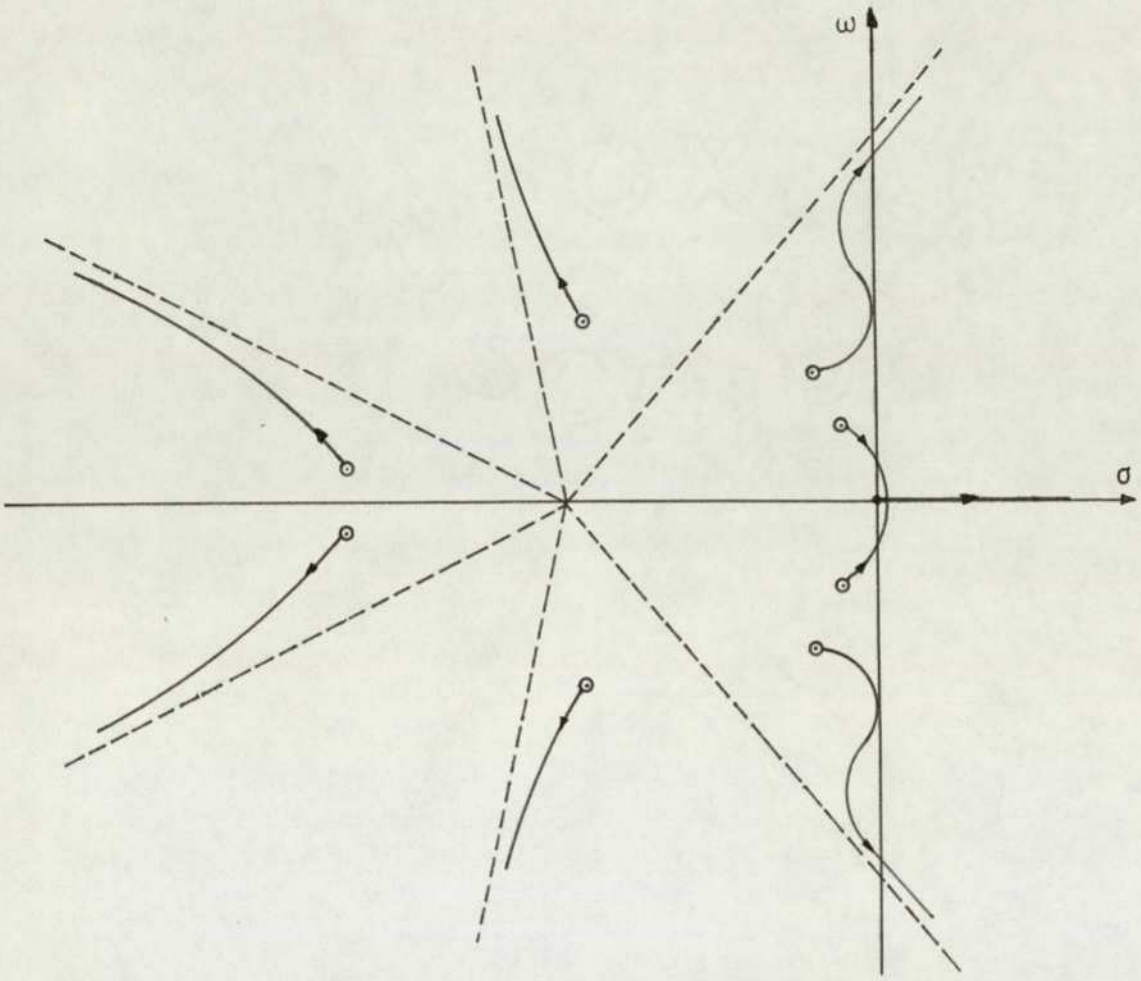


Fig. A1.13

After some manipulations, the set (A1.66) reduces to

$$
\begin{aligned}
(w_1 + 2\,w_0)\,b_3 - b_5 &= w_1^2 + 2\,w_1 w_0 + w_0^2 \\
(w_2 + 2\,w_0)\,b_3 - b_5 &= w_2^2 + 2\,w_2 w_0 + w_0^2 \\
(w_1 + 2\,w_0)\,b_2 - b_4 &= -\frac{k_1 - k_0}{(w_1 - w_0)^2} \\
(w_2 + 2\,w_0)\,b_2 - b_4 &= -\frac{k_2 - k_0}{(w_2 - w_0)^2} \quad .
\end{aligned}
\tag{A1.67}
$$

From (A1.62) it is obvious that the set of simultaneous equations (A1.67) has no real solution, since, regardless of w_0, w_1 and w_2, $k_1 - k_0$ and $k_2 - k_0$ are positive and therefore b_3 and b_5 have signs opposite to those

of b_2 and b_4.

One can conclude that the class of oscillators studied here cannot be used as devices producing quasi-digital signals ; the second order systems have a sensitivity which can be adjusted at a high level, but it decreases rapidly as the frequency changes. The systems of higher order have the retically a wider range of frequencies for which the sensitivity remains high, but are practically unrealisable.

APPENDIX A2

Some properties of Hermite polynomials

The aim of this appendix is to summarise the properties of Hermite polynomials which have been used at various stages of this work.

The Hermite polynomials (R. Courant D. Hilbert, 1953) $H_n(x)$ are the orthogonal polynomials in the interval $(-\infty,+\infty)$ having the weight function

$$w(x) = \exp(-x^2). \tag{A2.1}$$

The generating function $g(x,t)$ is given by :

$$g(x,t) = \exp(-t^2 + 2tx) = \exp(x^2)\exp\left(-(t-x)^2\right) =$$

$$= \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n \quad . \tag{A2.2}$$

Therefore, the Hermite polynomials are defined as

$$H_n(x) = \frac{\partial^n g(x,t)}{\partial t^n}\bigg|_{t=0} = (-1)^n \exp(x^2) \frac{d^n \exp(-x^2)}{dx^n} \; . \tag{A2.3}$$

From (A2.3) it follows that

$$H_n'(x) = 2n\, H_{n-1}(x) \; . \tag{A2.4}$$

Also

$$H_n(x) - 2x\, H_{n-1}(x) + 2(n-1)\, H_{n-2}(x) = 0 \; . \tag{A2.5}$$

Based on (A2.5) first Hermite polynomials are :

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x \\ H_4(x) &= 16x^4 - 48x^2 + 12 \; . \end{aligned} \tag{A2.6}$$

One can see that

$$H_{2k}(-x) = H_{2k}(x) \tag{A2.7a}$$

and

$$H_{2k+1}(-x) = -H_{2k+1}(x) \ . \qquad \text{(A2.7b)}$$

It can be shown (using A2.3) that the Hermite polynomials are orthogonal :

$$\int_{-\infty}^{+\infty} H_m(x)\, H_n(x)\, \exp(-x^2)\, dx = \begin{cases} 0 & \text{for } m \neq n \\ 2^n\, n!\sqrt{\pi} & \text{for } m = n \end{cases} \qquad \text{(A2.8)}$$

Therefore, a set of functions $f_k(x)$ given by

$$f_k(x) = \frac{H_k(x)\, \exp(-x^2/2)}{\sqrt{2^k\, k!\, \sqrt{\pi}}} \qquad \text{(A2.9)}$$

represent a set of orthonormal functions, with the property

$$\int_{-\infty}^{+\infty} f_m(x)\, f_n(x)\, dx = \delta_{mn} \ , \qquad \text{(A2.10)}$$

where δ_{mn} is the Kronecker delta.

APPENDIX A3

About some computational problems

A3.1 A method of reducing the store required by the integer programming

At one stage of this work the possibility of using the integer programming in view of minimising the cost of the communication system has been considered. In fact a particular case of integer programming, the so-called "zero-one programming" has been examined. The method consists in minimising the cost function of the form

$$C = \sum_{i=1}^{m} \sum_{j=1}^{n} l_{ij} \, c \, \lambda_{ij} \quad , \qquad (A3.1)$$

where l_{ij} represents the distance between the i-th multiplexer and the j-th data source, c is the specific cost of the cable connecting the data sources and the multiplexers and λ_{ij} is an integer variable which has the value 1 if the j-th data source is connected to the i-th multiplexer and 0 otherwise. The algorithm solving this problem is based on the so-called "implicit enumeration", which has been formulated by Balas and which has received much attention in various works in the field of operational research (T. Ibaraki, 1971 ; C.C. Petersen, 1967 ; C.A. Trauth, R.E. Woolsey, 1969). The problem of minimisation is solved by considering each possible value of λ_{ij} for all i and all j. However, some of the combinations of λ_{ij} (named solutions) are not explicitly considered since, if one can show that they will not produce a value for the cost function below the minimum value up to that particular stage (E. Balas, 1965). This procedure of considering some solutions implicitly enumerated gave the name to the algorithm (SIAM Review, 1965).

Despite some improvements of the algorithm (F. Glover, 1965), the problems involving more than 100 variables require a computational time beyond any practical values (C.A. Trauth, R.E. Woolsey, 1969). A quick estimate shows that if the problem involves 2,000 data sources and 50 multiplexers, figures which are perfectly realistic, the computer store required for the variables λ_{ij} reaches the considerable figure of 100 kwords. Note that, apart from λ_{ij}, other variables have to be stored and therefore the problem of compiling the program becomes very difficult to handle. A possible way to reduce in a substantial manner the store required is given below.

One of the features provided by the version of extended FORTRAN produced by CDC is the so-called "masking" procedure. Two variables A and B (integers or real numbers), processed by the function

C = A.AND.B

give as result a variable C ; the bits of the word C are the logical AND of the corresponding bits of the words A and B (Fig. A3.1). Since the

A	0	1	0	1	·	·	·
B	0	0	1	1	·	·	·
C	0	0	0	1	·	·	·

Fig. A3.1

variables λ_{ij} have only the values 0 or 1, it is unreasonable to store each λ_{ij} in a full store location, therefore using all the 60 bits of a word (the case of a CDC 7600 machine) to store an information which can be contained only in 1/60 of the same word. The idea of reducing the

volume of store necessitated by the problems of zero-one programming appears now obvious : store each λ_{ij} as a bit of a word and not as a word in itself. A subroutine which handles the bits of a word is given below. The word is one element of the array of integers X(M,N). The subroutine can either read the bit existing in the I-th row and the J-th column of the field determined by X (Fig. A3.2), or write a given bit in the position characterised by I and J. These two options are controlled by the value of the integer INDEX : for INDEX=1 the bit existing in the position

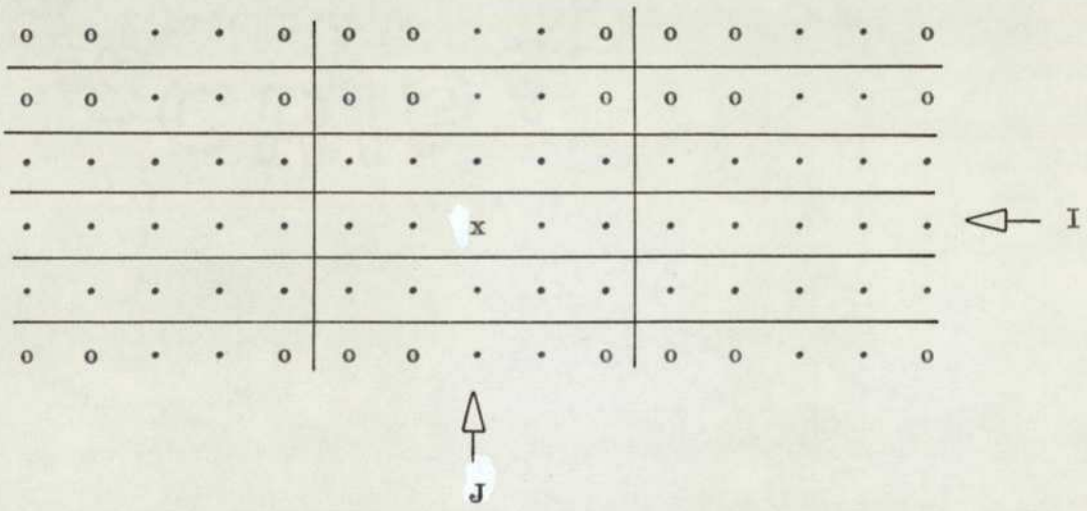


Fig. A3.2

determined by I and J is read and the result is returned in BIT ; for INDEX=2 the variable BIT becomes an input parameter of the subroutine and its value (0 or 1) is written in the position determined by I and J. In Fig A3.2 the elements of the array are separated by continuous lines ; the position determined by I and J is marked by x. Note that there is no need for specifying the particular element of the array X which contains the bit of interest, since the subroutine is provided with the possibility of finding it. In practical applications only 48 bits of each word can be used, since the other 12 are reserved for the sign and the exponent of the stored value and are accessed with more difficulty. The subroutine is

written in FORTRAN (the extended version for CDC 7600 machine). It uses 45 bits from each word. The subroutine has been tested and it has been found as being very efficient.

```
      SUBROUTINE BIN(X,M,N,I,J,BIT,INDEX)
      INTEGER BIT,X(M,N)
      K=(J+45)/46
      L=46*K-J
      MASK=2**L
      INT=X(I,K).AND.MASK
      IBIT=INT/MASK
      GO TO (1,2),INDEX
1     BIT=IBIT
      RETURN
2     IF(BIT.EQ.IBIT) RETURN
      X(I,K)=X(I,K)+MASK*(-1)**IBIT
      RETURN
      END
```

A3.2 The subroutine for stochastic approximation

The subroutine, named STAPP, has as parameters the number NS of samples and the number NF of functions used in the process of the estimation of the p.d.f. The amplitude X of samples and the coefficients THETA of the functions in the linear combination which approximates the p.d.f. are in the common block B1. The function values are calculated by a subroutine named FUNC, which has as parameters an integer showing the number of the function, a real variable which is the abscisa and another real variable in which the value of the function is returned.

```
      SUBROUTINE STAPP(NS,NF)
C     PURPOSE:TO ESTIMATE AN UNKNOWN P.D.F.BY USING A LINEAR
C     COMBINATION OF KNOWN FUNCTIONS
C     X(5000) - THE SAMPLE BASED ON WHICH THE P.D.F. IS ESTIMATED
C     THETA(10) - THE COEFFICIENTS OF THE LINEAR COMBINATION
C     F(10) - ARRAY CONTAINING RETURNED VALUES OF FUNCTIONS USED IN
C                                                    ESTIMATION
C     NS - NO. OF SAMPLES USED FOR ESTIMATION
C     NF - NO. OF FUNCTIONS USED FOR ESTIMATION
      REAL X(5000),THETA(10),F(10)
      COMMON /B1/X,THETA
C
C     STATE PARAMETER VALUES
C
      A1=1.0
      A2=2.0
      C1=1.0
      DO 1 K=1,NF
      THETA(K)=1.0/FLOAT(NF)
 1    CONTINUE
      ALPHA=1.0
      BETA=0.5
      DO 5 K=1,NS
C
C     CALCULATE RO AND C
C
      AK=FLOAT(K)
      RO=A1/(A2+AK**ALPHA)
      C=C1/AK**BETA
C
C     CALCULATE Z
C
      XPL=XK+C
      XMN=XK-C
      CALL CDF(XPL,K,NS,FKPL)
      CALL CDF(XMN,K,NS,FKMN)
      Z=(FKPL-FKMN)*0.5/C
```

```
C
C     CALCULATE Y
C
      SCALAR=Z
      DO 2 K1=1,NF
C
C     K1 DETERMINES WHICH FUNCTION IS CONSIDERED
C
      CALL FUNC(K1,X(K),F(K1))
      SCALAR=SCALAR-F(K1)*THETA(K1)
 2    CONTINUE
C
C     CALCULATE THE NEW VALUE OF THETA
C
      STH=0.0
      DO 3 K2=1,NF
      Y=F(K2)*SCALAR
      THETA(K2)=R0*Y*THETA(K2)
      STH=STH+THETA(K2)
 3    CONTINUE
      DO 4 K3=1,NF
      THETA(K3)=THETA(K3)/STH
 4    CONTINUE
 5    CONTINUE
      RETURN
      END

      SUBROUTINE CDF(SAMPLE,K,KM,FK)
      COMMON /B1/X(5000),DUMMY(10)
      IFK=0
      KS=K-KM
      IF(KS.LE.0) KS=1
      IF(K.EQ.1) GO TO 2
      DO 1 I=KS,K
      IF(SAMPLE.GT.X(I)) IFK=IFK+1
 1    CONTINUE
      AIFK=IFK
      AKKS=K-KS
```

```
      FK=AIFK/AKKS
      RETURN
2     FK=IFK
      RETURN
      END
```

A possible way of writing the subroutine FUNC is given below. The number of different functions is 5.

```
      SUBROUTINE FUNC(N,X,Y)
      GO TO (1,2,3,4,5),N
1     Y=.....
      RETURN
2     Y=.....
      RETURN
      .....
      .....
5     Y=.....
      RETURN
```

A3.3 About finding the maxima of the quasi-potential function

The method of finding clusters based on maximising a quasi-potential function has been described in Chapter 5. A program implementing it has been written and tested. Some points should be mentioned in connection with this program :

- the random initial positions used by the maximisation routine have been generated by using a random number generator. The period of the generator being extremely large, there has been no need for checking if the current random initial position coincides with any former initial position ;
- the number of unsuccessful attempts to find maxima has been chosen as criterion of stopping the program. An attempt is considered unsuccessful when the found maximum is too close to one already found. Two maxima

have been considered as being "too close" if the distance between them did not exceed 0.5 of the maximum permitted distance between the data source and the multiplexer ;

- although the program has produced good results, the method of quasi-potential functions is not as good as that based on the use of circles, although it could produce much better results when used in pattern recognition.

REFERENCES

ALEXANDER, A.A., GRYB, R.M. "Capabilities of the telephone network for data transmission". The Bell Syst. Tech. Journ., Vol. 39, May 1960, pp 431-76

ALLEN, D.P. "How to chose data aquisition systems". Control Engineering, Vol. 16, No. 11, Nov. 1969, pp 102-6

ARONSON, R.L. "Line-sharing systems for plant monitoring and control". Control Engineering, Vol. 18, No. 1, Jan. 1971, pp57-76

BALAS, E. "An additive algorithm for solving linear programs with zero-one variables". Operations Research, Vol. 13, No. 4, July-Aug. 1965, pp 517-46

BATES,C.P., HAWLEY, G.T. "A model for currents and voltages induced within long transmission cables by an electromagnetic wave". IEEE Trans. Electromagn. Compat. (USA), Vol. EMC-13, No. 4, Nov. 1971, pp 18-31

BELLO, P.A., ESPOSITO, R. "A new method for calculating probabilities of errors due to impulsive noise". IEEE Trans. on Commun. Technol., Vol. COM - 17, No. 3, June 1969, pp 368-79

BISBY, H. "Modular nuclear instrumentation. Development of modular systems for nuclear instrumentation and data handling". Nuclear Engineering International, Vol. 16, No. 178, 1971, pp 321-328

BODHARAMIC, A. "Optimal signal detection for low frequency data transmission in the presence of non-Gaussian noise". Univ. of Maryland, 1971

BOMBI, F. "High-performance voltage to frequency convertor has improved linearity". Electron. Eng. (GB), Vol. 42, No. 514, Dec. 1970, pp 61-4

BUTLER, G.A. "A vector field approach to cluster analysis". Pattern

Recognition, Pergamon Press, Vol. I, 1969, pp 291-99

CARSON, E.R., PIROUZAN, A. "Computer compatible instrumentation : a case for quasi-digital transducers". Research memorandum. The City University, London, 1972

COURANT, R., HILBERT, D. "Methods of mathematical physics". Vol. 1, Intrescience Publishers Inc. NY, 1953

DAVENPORT, W.B., ROOT, W.L. "An introductional to the theory of random signals and noise". Mc. Graw-Hill, NY, 1958

DRISCOLL, M.M., HEALEY, D.J. "Voltage controlled crystal oscillators". IEEE Trans. Electron. Devices (USA), Vol. ED -18, No. 8, Aug. 1971, pp 528-35

FELLER, W. "An introduction to probability theory and its applications". J. Wiley & Sons, NY, 1966

FENNICK, J.H. "Amplitude distributions of telephone channel noise and a model for impulse noise". The Bell Syst. Tech. Journ., Vol. 48, Dec. 1969, pp 3243-63

FENNICK, J.H., NASSEL, I.E.O. "The 1963 survey of impulse noise on Bell System Carrier Facilities". IEEE Trans. on Commun. Technol., Vol. COM - 14, No. 4, Aug.1966, pp 520-5

FINKLER, R.A. "The estimation of S/N ratio for real and analytic signals and Gaussian noise". G. Washington Univ., 1970

GLOVER, F. "A multiphase dual algorithm for the zero-one integer programming problem". Operations Research Vol. 13, No. 6, Nov-Dec. 1965, pp 879-919

IBARAKI, T. "Complementary programming". Operations Research, Vol. 19, Sept.-Oct. 1971, pp 1523-8

JERVIS, M.W., BAREHAM, R.C. "Instruments-computer interface problems in CEGB nuclear power stations". IEE Conference on industrial measurement

techniques for on-line computers. London, June 1968, pp 9-13

KAPP, J. "Performance of suboptimum detectors and signal selection in Gaussian and impulsive noise". New York University, 1970

KISSLINGER, J. "Pulse generator is supply and temperature independent". Electronics, Vol. 41, No. 8, May 1968, pp 106

KURLAND, M., MOLONY, D.A. "Observations on the effects of pulse noise in digital data transmission systems". IEEE Trans. on Commun. Technol. Vol. COM - 15, No. 4, Aug. 1967, pp 552-6

LUCKY, R.W., SALZ, J., WELDON JR., E.J. "Principles of data communication". Mc. Graw-Hill Book Co., NY, 1968

MARTIN, J. "Systems analysis for data transmission". Prentice-Hall, Englewood Cliffs, New Jersey, 1972

MENDEL, J.M., FU, K.S. "Adaptive, learning and pattern recognition systems, theory and applications". Academic Press, NY, 1970

MERTZ, P. "Model of impulsive noise for data transmission". IRE Trans. on Commun. Syst., Vol. CS - 9, June 1961, pp 130-7

MEYER, D. "A process for converting quantities to be measured into frequency". Philips Technical Journal, Vol. 29, No. 6, 1968, pp 189-96

MINORSKY, N. "Theorie des oscillations". Gauthier-Villars, 1967

MOLINEUX, L. "Op amp sine wave uses diodes for stability". Electronics, Vol. 41, No. 23, Nov. 1968, pp 121

MOORE, J.D. "Characterisation of impulse noise and analysis of its effects upon correlation receivers". Univ. Alabama, 1972

MOURLAM JR., L. "Feedback loop stabilises FET oscillator". Electronics, Vol. 40, No. 17, Sept. 1967, pp 97

PETERSON, C.C. "Computational experience with variants of the Balas algorithm applied to the selection of R & D projects". Management Science, Vol. 13, No. 9, May 1967, pp 736-45

PIROUZAN, A. "A sensor system with pulse number output". Measurement & Control, Vol. 4, No. 12, 1971, pp 207-210

RAPPAPORT, S.S., KURZ, L. "An optimal nonlinear detector for digital data transmission through non-Gaussian channels". IEEE Trans. on Commun. Technol., Vol. COM -14, No. 3, June 1966, pp 266-74

SEBESTYEN, G., EDIE, J. "An algorithm for non-parametric pattern recognition". IEEE Trans. on Electron. Comp., Vol. EC - 15, No. 6, Dec. 1966, pp 908-915

SELIN, I. "Detection theory". RAND Corporation, R-436-PR, June 1965

* * * "Integer programming by implicit enumeration and Balas' method". SIAM Review, Vol. 9, No. 2, Apr. 1967, pp 178-90

SIMON, H. "Multiplex systems save multibucks in refinery and chemical plants". ISA - 70, Silver Jubilee, Philadelphia, 1970

SMEINS, L.G. "Stable unijunction VCO needs no critical components". Electronics (USA), Vol. 43, No. 21, Oct. 1970, pp 100

STEVENS, M., WOOD, A. "The FET as a chopper". Appl. Rep. B 60, Texas Instruments, 1968

STRAUSS, L. "Wave generation and shaping". Second edition, Mc Graw-Hill Book Co., NY, 1970

TRANKLER, H.R. "Ein linear harmonisher Messoscillator". ETZ-b, Bd. 25, H. 9, 1973, pp 220-22

TRAUTH JR., C.A., WOOLSEY, R.E. "Integer linear programming : a study in computational efficiency". Management Science, Vol. 15, No. 9, May 1969, pp 481-93

VELTMAN, B.P.TH. "The inevitable interface ". Survey lecture for the 5-th IFAC Conference, Paris, 1972

VITERBI, A.J. "Principles of coherent communication". Mc Graw-Hill Book Co., NY, 1966

WARD, H.C. "Considerations affecting the mitigation of noise coupled between cable pairs". IEEE Trans. on Commun. Technol., Vol. COM - 15, No. 4, Aug. 1967, pp 648-52

WATSON, J. "An introduction to FET". Siliconix, 1971

WHALEN, A.D. "Detection of signals in noise". Academic Press, NY, 1971

ZIEMER, R.E. "Error probabilities due to additive combinations of Gaussian and impulsive noise". IEEE Trans. on Commun. Technol., Vol. COM - 15, No. 3, June 1967, pp 471-4