FAULT DETECTION USING

TRANSFER FUNCTION TECHNIQUES

/

BY

NEIL E PATERSON

A thesis submitted to the faculty of Science of

the University of Edinburgh for the degree of

Doctor of Philosophy

Department of Electrical Engineering

1987

ABSTRACT

A new method for identifying the parameters of a continuous system has been developed. This method the Modified Modulating Function Method (MMFM), is a development of Shinbrot's Modulating Function Method. The auto and cross correlation functions are used in place of the input and output signals in the identification algorithm. This enables pre-estimation checks to be performed on the data, and also enables important information about the system to be determined prior to identification. A fast correlation algorithm was developed to perform the large number of correlations needed. This employed Fell's Skip algorithm in a Relay correlator. The MMFM was developed as a fault detection system to be applied to a Diesel Alternator set at Stornoway power station. Tests were carried out on simulated systems to determine the viability of the system before on-line tests at Stornoway power station.

A review of system identification has been carried out, along with a review of fault detection methods using transfer function techniques. This reveals a need for an accurate method for determining the continuous transfer function. The Modified Modulating Function Method is a possible solution to this problem.

For Helen

ACKNOWLEDGEMENTS

CONTENTS

TITLE PAGE

ABSTRACT

DECLARATION OF ORIGINALITY

ACKNOWLEDGEMENTS

GLOSSARY

LIST OF CONTENTS

CHAPTER ONE

INTRODUCTION

Fault detection using transfer function techniques has largely been used to detect and diagnose faults in analogue electronic circuits. Only recently has there been an attempt to extend these techniques for use on mechanical systems. This research work has been devoted to the application of fault detection using transfer function techniques to a large diesel engine used by the North of Scotland Hydro-Electric Board (NSHEB) for the generation of electricity in their Stornoway power station.

Diesel engines are expensive to maintain and repair. In addition, rising fuel oil prices over the last decade have increased operation costs still further. Manufacturers have made great efforts to improve the efficiency of their engines to offset these costs. Even so the diesel engine market is under threat from other power sources. Railways are being electrified, small rural and island power stations are being connected to grid systems. Even the shipping market, where the diesel engine has achieved all but total dominance as the main propulsion unit, is in a world wide slump. This means the diesel manufacturer, to

survive, must reduce the capital cost of his plant and the operational costs by lowering maintenance and fuel costs.

One possible method of achieving this is to employ a condition based maintenance scheme, where the maintenance requirements of the engine are determined by its condition. This reduces unnecessary maintenance and also allows the maintenance and repair of the engine to be scheduled to suit the plant duty. This combined with an ability to reduce the number of unplanned outages allows for a reduction in the plant margin, and also a reduction in the spares inventory held, thus a large saving in the capital cost of the plant can be made.

The necessary instrumentation for this condition monitoring system can be expensive and requires many access points around the engine. In this research work it is intended to reduce the cost of implementing a condition based maintenance scheme by using a transfer function technique to detect faults or a loss of performance in the engine. This method would require only two access points on the engine thus reducing instrumentation costs considerably. For this, a means of estimating the parameters of a linear lumped parameter continuous transfer function was required. The Modified Modulating Function Method was developed

for this purpose.

1-1 Condition Monitoring.

Condition monitoring (CM) of large mechanical systems is playing an ever increasing role in the safe and economic operation of these systems. When operating a condition monitoring regime the objective is to extract sufficient information from the system to be able to estimate the condition or health of that system, and to make maintenance decisions based on that information. Condition monitoring very largely supersedes the need for planned preventative maintenance (PPM). PPM can result in components being adjusted or replaced that need not be changed, also a component may fail before the next scheduled maintenance period, thus resulting in an unplanned outage. This can lead to unnecessary maintenance costs and system down time. This regular maintenance also introduces the possibility of human error when maintenance tasks are carried out. Therefore it would be better to have a system which only implemented system maintenance, when required. Condition monitoring is such a system.

Condition monitoring aims to reduce costs by increasing the reliability of the plant, and reducing maintenance costs. This is achieved by monitoring

3

various parameters of the system, such as temperatures, pressures and vibration at various critical points around the plant. This information represents a signature of the plant. It has been shown [1] that these parameters will change significantly before a failure situation arises in the plant. The faulty signature can be compared with the healthy signature and some decision arrived at as to the cause of the incipient failure, and the lead time to failure. This can be achieved by a thorough knowledge of the plant and its various failure mechanisms, or by comparison with past failure events. The latter course requires a long build up time to acquire sufficient data. Having assessed the likely cause of the incipient failure and the likely time to failure, it will then be possible to order the replacement parts and organise the maintenance staff to implement the necessary maintenance at a time suitable for the operation of the plant.

In this way it should be possible to reduce the number of maintenance staff necessary to repair unplanned plant outages and also reduce the inventory of spares necessary to cover these events, as they can now be planned and the parts ordered in advance. As a result of this, machine downtime will be reduced, and much more productive time can be gained from each machine. Hence the plant margin, the extra machines necessary to

perform the plant function while repairs are carried out on other machines, will be reduced, thus lowering the capital cost of the plant. As a result of the plant being continuously monitored, its performance should not be allowed to drift far from the optimum, and hence it is expected that fuel costs will be reduced.

Condition monitoring (CM) systems require a lot of expensive transducers and signal conditioning equipment, along with the necessary computer equipment, user knowledge and experience. The CM package can be expensive. Examples of the cost benefits and penalties of a CM system are given in [1]. It was felt that there was a need for a low cost on-line device which could operate without user interaction and hence be suitable for remote locations. The device would use a minimum of test points, input and output, and would infer information about the internal state of the plant from these two information sources. This approach is usually called Performance Monitoring (PM), where the relationship between the input and the output of the system is monitored.

In this research work it was decided to investigate the PM approach by monitoring the transfer function of the plant. This particular relationship between the input and output, if properly used, can give information

about the condition of the plant, and offers a potentially large saving in transducers and signal conditioning equipment. User experience is not a necessary requirement as the diagnostic capability of the device is programmed into it, during the design of the PM system, from a prior knowledge of the plant.

The disadvantage of this system is that it may be insensitive to some serious component faults which may not affect the performance of the plant greatly. For instance, if a bearing were to wear, the power output of an engine may not be significantly reduced until the bearing fails catastrophically. Hence the PM system would not in this case detect incipient failure, whereas a CM system employing a temperature or vibration transducer on the bearing may well detect this fault. It is however expected that the PM system would detect incipient faults related to the performance of the engine, such as fouling of fuel injectors and turbo-chargers.

1-2 Diesel Engines and their Applications.

Rudolph Diesel in 1893 patented a power cycle which has formed the basis of the diesel engine throughout the years. In this cycle a piston compressed air in a cylinder until its temperature was raised sufficiently

to ignite a fuel. This was followed by a constant pressure combustion and an adiabatic expansion of the gas. As Germany had no natural oil resources but plenty of coal this cycle was originally designed to use coal dust as the fuel. Coal dust proved to be an unsuitable fuel due to the need for a powerful compressor to overcome the cylinder pressure when injecting this fuel, so oil was used in its place.

The diesel engine relies on the heat of the compressed air igniting the fuel, which is injected into the compressed cylinder just before top dead centre. A high compression ratio is required to raise the air to this ignition temperature and it is this feature which gives the diesel its high thermal efficiency. Also diesel fuel has a higher calorific value than petrol, 11% higher, which also improves the fuel consumption figures. The efficiency of the engine depends on good combustion of the fuel, which is dependent upon air and fuel mixing evenly throughout the cylinder. This is achieved in three ways, which can also be used to classify the engines. Slow and medium speed diesels rely on the fuel injection distributing the fuel in a fine vapour throughout the cylinder. These engines tend to be large 2 mega watts (MW) and above, and are used in ships and power stations. An intermediate size of engine, used in trucks, buses and small boats, employs a

modest amount of air swirl (turbulence) caused by the induction stroke to mix the injected fuel with the air. In small high speed engines, of the type being fitted to modern cars, the mixture of fuel and air is achieved by a high degree of turbulence. The engines that are of interest in this work are the large slow and medium speed engines, used mainly at constant load over a long time span.

In this bracket of large engine there are two types, four and two stroke. The very large engines 24MW tend to be slow speed two stroke engines. Sultzer is a manufacturer of this type of engine. The medium speed engines are mostly four stroke, and a great many manufacturers of this type of engine exist, of which Mirrlees Blackstone is one. All the large engines use turbo-chargers to increase the air charge in the cylinder and increase the power output from a given size of engine.

Over the years manufacturers have used devices like the turbo-charger to increase the efficiency of the diesel, until it can now obtain a thermal efficiency of 40% making it the most efficient practical thermal cycle available. With the escalating fuel prices this has made the diesel engine the most popular ship propulsion unit. In addition to the very large number of diesels

used for marine application, diesels have been used in the generation of electricity in rural areas or developing countries. In a developing country a diesel generator can be placed close to the capital city providing the early needs of the city and larger towns. Once the demand for electricity rises, a larger steam plant can be built and it may be possible to remove the diesel to a more remote site, thus forming the basis of an electric grid. In developed countries the diesel generator can still play an important role in supplying electricity to remote or island locations which cannot be economically connected to the national grid system, and cannot support a larger steam plant. Usually 30 MW is considered the economic maximum capacity of a diesel plant, although there have been instances of much larger stations, for example 92 MW in Taiwan. Diesels are ideal for low load work, being small and compact they can be located near to the towns they supply. In this type of work diesels are used throughout the world, in developed as well as developing countries.

Although a diesel engine offers the most thermally efficient means of converting fuel into energy; the Stirling cycle is more efficient but not practical. The cost of diesel fuel compared to coal, and the economies of scale achieved in a large steam plant, make the diesel generation of electricity an expensive

exercise. Also the complex nature of a diesel engine means that the maintenance costs are high.

Mirrlees Blackstone diesels at Stornoway power station run for 3000 hours between routine servicing, and require a major overhaul every 12000 hours. This workload on a station such as Stornoway, where there may be nine such engines, requires a large number of skilled fitters to carry out maintenance and repair any unexpected failures. Added to this expenditure is the cost of carrying a large inventory of spare parts to repair and service the engines, and the capital cost of providing backup plant to cover unplanned outages.

Pollution can be a problem with diesel engines. Although the engines do not produce the amount of toxic gas that a petrol engine does, the heavy fumes produced by a diesel are under suspicion as being carcinogenic. It is therefore advisable to reduce the amount of black smoke emission from diesels by ensuring that the diesel engine runs efficiently.

Throughout the ninteen seventies the diesel engine market expanded. In the wake of the oil crisis, ship owners were replacing gas turbine plant with the cheaper to run diesels. This expansion has been halted due to the world wide recession and the industry has now been

left with an over capacity due to the continual expansion in the late seventies. The traditional manufacturing bases in Europe and the U.S.A are now under the threat of strong competition from the Far East. Although there has been consolidation aimed at righting this over capacity [2], the diesel market today is still a very competitive place. To survive, the diesel manufacturer must offer a product that has economic advantages over its competitors. The reduction of running costs can play its part in achieving this advantage. Some form of condition monitoring capability is one approach to the reduction of maintenance, fuel and capital costs.

1-3 Diesel Condition Monitoring

The variety of types of diesel and the variety of uses they are put to is quite enormous, but diesels still suffer common problems due to poor design or misuse. Fagerland [3] has studied the common failures of diesels used for ship propulsion these are, burnt exhaust valves; bearing failures; cracks in the cylinder head, liner, piston, turbocharger and beadplate; high lubrication oil consumption. Chambers [4] discusses various diesel failures that have resulted in very large insurance claims. These have included bearing failure due to oil contamination which led to a claim for £100,000 for the replacement of the crankshaft.

11

Failures in large medium speed diesels can be very expensive, another reason for the interest in a condition monitoring maintenance regime, which may detect the fault before catastrophic failure.

For condition monitoring of a diesel engine, various parameters such as temperature and pressure are monitored from key points around the engine. These are monitored on-line by a computer, Fig (1-1), and it may be possible to predict imminent failure in the engine by trend analysis on the various parameters [5]. It is important to monitor the parts of the engine that are known to be susceptible to failure. Konstantinos [6] has proposed a list of parameters to be monitored.

1) Instantaneous performance parameters.
2) Crankshaft bearing condition.
3) Top piston ring wear.
4) Exhaust valve wear.
5) Turbo-charger system behaviour.
6) Fuel injection system.
7) Cylinder pressure.
8) Temperature of critical components.
9) Vibration.
10) Oil and cooling water analysis.

This list of parameters is broadly similar to the parameters measured by Fagerland et al [7] and Langballe et al [8]. The instantaneous performance parameters are the various temperatures, pressure and flow rates in the engine and its various subsystems. These will show a deterioration in the condition of the engine and enable
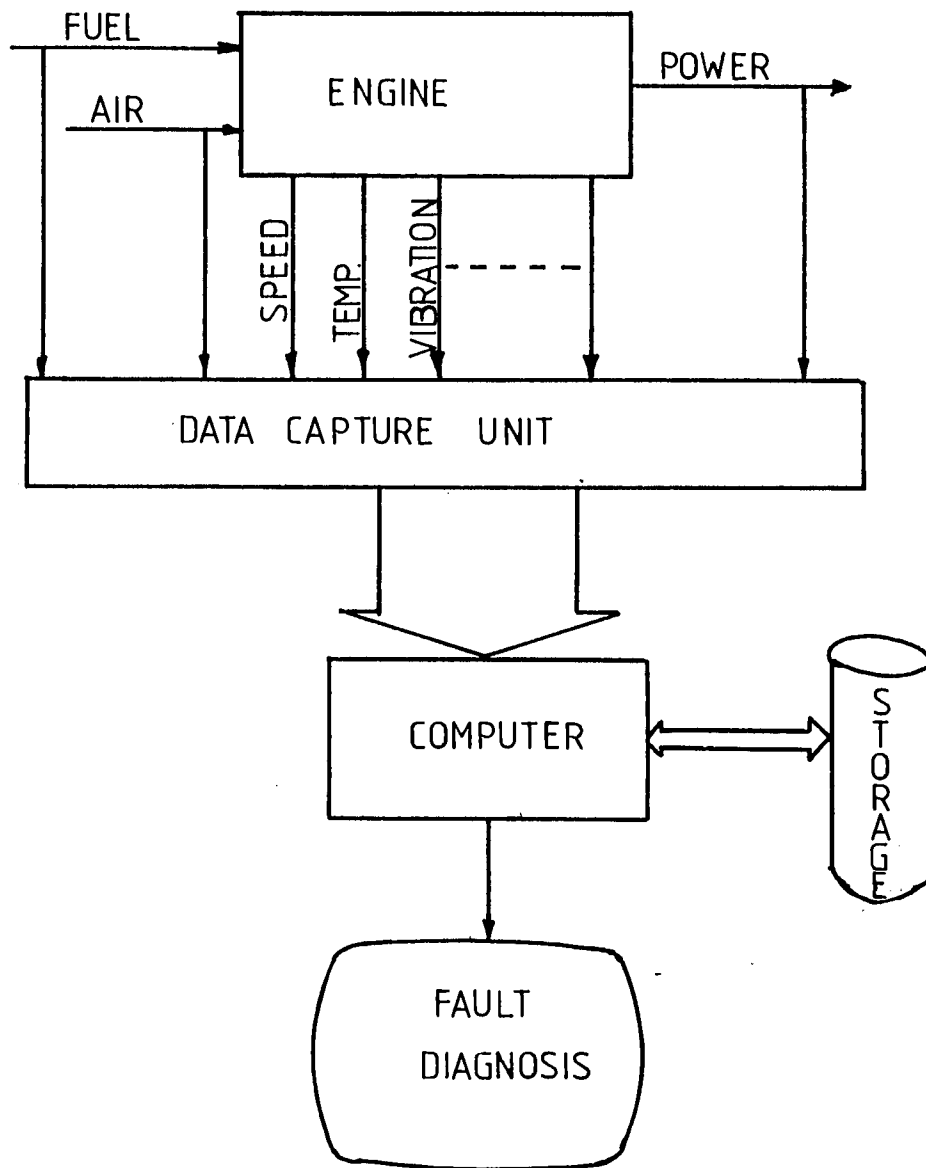
12

FIGURE 1-1
COMPUTER MONITORING SYSTEM

accurate diagnosis of the fault.

The remainder of the list concerns the detection of faults in more specific areas of the engine or in the case of vibration and oil analysis, by the use of well know CM tools.

The only method generally available to monitor the condition of the crankshaft bearing is to monitor the bearing temperature. If wear occurs in a bearing, friction will cause a build up of heat, which can be detected by a thermocouple located in the bearing shell or by measuring the temperature of the lubrication oil flowing out from the bearing.

Top piston ring wear is a common fault in all reciprocating engines. The piston rings usually deteriorate before other components in the piston-connecting rod assembly, therefore it is advantageous to monitor this parameter of the engine. A method of monitoring the wear in the top piston ring has been developed for large diesels [6]. This uses a non-contacting sensor set into the cylinder liner to detect the magnetic properties of the piston ring. As the piston ring wears, its magnetic property changes and this can be used to determine the condition of the piston ring.

Exhaust valve wear is a serious problem on diesels and is usually caused by overloading of the engine or poor combustion. This results in higher exhaust gas temperatures than is permissible, causing damage to the exhaust valves. The solution is to monitor the temperature of the exhaust gas and when this exceeds a specific limit, trigger an alarm.

Fouling of the Turbo Charger system is the biggest problem. Deposits of carbon from the exhaust gas build up on the turbine side of the Turbo Charger reducing its efficiency. Monitoring the pressure of the air after the compressor will give a good indication of any drop in the Turbo Charger efficiency.

The fuel injection process must be functioning correctly for proper combustion to be obtained. A badly adjusted or malfunctioning fuel injection system will lead to poor fuel consumption and pollution. It is therefore possible to deduce the condition of the fuel injection system from the exhaust gas temperature and cylinder pressure.

Monitoring the pressure in each cylinder of a diesel can give a good indication of poor injector condition, poor spray pattern, excess fuel or bad ignition timing.

The pressure can be measured either by a pressure sensor or a strain gauge on one or more of the cylinder head bolts.

The temperature or thermal load of critical components can be measured by the location of thermocouples in the components. The cylinder wall temperature and piston crown temperature are of most interest.

Vibration spectra is a classical condition monitoring tool. Vibration of components such as bearings can be measured using vibration transducers; a signature for a normal component can be obtained which subsequent signatures can be compared against, and a fault declared if the vibration amplitude at certain critical frequencies exceeds a safe limit [1].

Lubrication monitoring is another classic approach to CM. The lubrication oil can be monitored for particles of new debris [1]. The amount of debris in the oil will continually increase at a constant rate between oil changes. Any increase in the rate or size of these particles can indicate an incipient failure in some part of the machine. The detection of these particles can be achieved using ferrographic techniques, which can detect ferrite materials, or spectral analysis techniques. A diagnosis of the fault can be made from a knowedge of

the location of various materials in the engine. These techniques are usually off-line and the lead time between sample and fault reporting is critical.

A condition monitoring system can make use of all or some of these methods. Signals from the various transducers are usually fed into a computer which can make comparisons with the normal operation of the engine. If any reading is found to be abnormal, the computer can diagnose the fault using some form of pattern recognition or fault tree analysis. Prediction of failure can be made by some form of trend analysis, using one of the numerical techniques available, such as least squares method [5].

1-4 The Programme of Research

The objective of this research work was to investigate a low cost on-line machinery health monitoring system which could be operated at remote locations without any operator intervention. It was the intention that this system would use a minimum of test points and still provide a useful amount of information about the health and performance of the machine being monitored. The method which seemed to fit this requirement was that of monitoring the parameters of the small signal differential equation or transfer function, which would

17

require only two test points. The parameters of the transfer function would be monitored for any deviation from the normal, which may be due to an incipient failure, or loss of performance on the machine.

For this purpose a method of system identification was developed which was thought robust enough to detect bad data and operate without skilled intervention, after the initial setting up procedure. This method, the Modified Modulating Function Method of linear system identification, was developed from the classical approach to continuous system identification proposed by Shinbrot [56].

The development of this method, and the software to implement it, has been the major activity during this research. Once developed this method was tested using software and hardware simulated systems. All the software for the method was written on an HP85 computer in either BASIC or Assembler. The HP85 was used to control the tests through the IEEE 488 Bus as well as perform the identification on the system. The portability of the HP85 was also required for on-site testing at Stornoway Power Station.

One of the base load diesel engines in the station was chosen for the tests, several tests were carried out on

this engine to establish the transfer function. This was obtained using the Modified Modulating Function Method and then checked against that obtained by a commercial spectrum analyser.

1-5 Thesis Format

Chapter two reviews the more popular techniques of system identification and parameter estimation. Particular attention is given to the methods of continuous model system identification. Also reviewed in this chapter are the methods of fault detection based on transfer function techniques.

Chapter three presents a theoretical development of the Modified Modulating Function Method, and the correlation techniques upon which it depends. Modelling of the diesel engine and the problems encountered are also discussed in this chapter.

In chapter four the experimental system used and the software developed for the various experiments and software simulations is described.

Chapter five presents the results obtained from the extensive laboratory testing of the MMFM through the use of software and hardware simulated systems. Also

presented are the results obtained during the on-site testing at Stornoway Power Station.

Chapter six contains the conclusions drawn from this work and recommendations for further work.

CHAPTER TWO

SYSTEM IDENTIFICATION AND FAULT DETECTION

Fault detection using transfer function techniques is a combination of two fields of study, System Identification and Fault Isolation. In this chapter the literature from these fields is reviewed. The first half of the chapter deals with the field of System Identification and Parameter Estimation. This field has undergone considerable expansion in the last twenty five years and now covers a wide variety of different techniques and methods. A limited review of the field as a whole, with a closer inspection of the techniques of particular interest to this project, is presented here. The second half of this chapter reviews the literature on the use of transfer functions for fault detection. This area relies heavily on techniques developed in the previous section and it is therefore appropriate that these two fields are reviewed together.

2-1 System Identification

System identification and parameter estimation has been greatly developed over the last twenty five years or so. The classical approach to system identification, frequency response testing, has been largely surpassed

by the discrete methods of parameter estimation, which now dominate the literature. This is due to the introduction of computers which are capable of processing large amounts of data quickly. The discrete methods allow parametric models which are easily implemented on computer, and can therefore be used in simulation studies, controller design or in the case of self tuning, in the controller itself. In this, they have a great advantage over frequency response techniques which may partly explain the considerable interest in discrete techniques. The difference in approach between the classical and the discrete methods highlights one of the main categorisations in system identification. There are two main categories, non-parametric and parametric identification.

The non-parametric identification is the classical technique in system identification. This method results in a graphical representation of the system. The graphical model is generally not suitable for direct implementation on a computer. The non-parametric methods include frequency response, step response and impulse response techniques. These methods are still of great interest and in general use, and have also been used in fault detection studies for a number of years.

Parametric identification results in a mathematical

model of the system which can be easily implemented on a digital computer. Parametric identification can be split into two approaches, discrete or continuous. Discrete methods use difference equations to model the system and the transfer functions operate in the Z - plane. Continuous methods use differential equations as a model of the system, the transfer functions of these methods operate in the S - plane.

As the field of system identification is so large and varied a complete literature survey would prove impossible to carry out here. Instead we shall split the review into three sections, and look closely only at the techniques that are of particular interest in fault detection. The first of these three sections will cover the non-parametric techniques, the second section will deal with the parametric methods and the third section will review the literature of a specific technique, the Modulating Function Method, around which this research work is based.

For fault detection it is thought better to use the continuous model of the system rather than the discrete model [83]. This will enable deviations in the parameters of the model to be traced back to the system, and enable the fault to be located. This is because continuous models are directly derived from the physical

23

laws that govern the system, and the process coefficients such as length, mass, temperature etc. will make themselves apparent in the coefficients of the model. Therefore any deviations in the process coefficients caused by a fault, may be detected in the model coefficients.

For this reason more attention will be given to continuous model parameter estimation than discrete parameter estimation in the parametric section. This is despite the fact that most of the effort in system identification has gone into discrete methods of parameter estimation.

In this review we shall limit ourselves to the problem of modelling a linear, single input, single output, lumped parameter system. For a review of non-linear work the reader is directed to the recent review by Billings [9]. Distributed systems have been reviewed by Lubrusky [10].

2-2 Non-Parametric Identification

Non-Parametric Identification of a system results in a graphical representation of the system. These graphical models can take the form of frequency response magnitude and phase curves, impulse response and step response plots. Therefore the system is described by

24

how it reacts to a given input rather than being described by a mathematical equation relating the input to the output of the system.

This is the classical approach to the identification problem, and has been used successfully for many years. Perhaps these methods should be turned to first when trying to identify the system, so that a picture of its response can be obtained. These methods have a great advantage over parametric identification in that no model of the system is needed before identification can commence. This means that there will be no modelling errors or prior assumptions made about the system that may turn out to be erroneous and invalidate the identification. Being simple and easy to perform, a step response could be performed before any parameter estimation and a suitable model structure determined from this [11].

Although they are important and reveal considerable information about the system, non-parametric models are not easy to use on a computer. Hence there has been much interest in deriving parametric models from these graphical models. Rake [11] describes methods of obtaining parametric models from non-parametric models, also Payne [12] describes a method for curve fitting the frequency response. There are other examples in the

literature, these two being given as examples. There are two main approaches to non-parametric identification. The time domain methods, which are graphical representations of the system transient response, and frequency domain methods which represent the steady state response of the system to a periodic signal. Kwiatkowski [13] compared the two approaches to identification and concluded that the choice of method depended upon the system to be identified.

2-2-1 Time Domain Methods

Time domain methods use non-periodic inputs to excite a transient response from the system. The classical approach is to test a system with a step, ramp or pulse input [11,14]. For these tests a large signal to noise ratio is required, otherwise the output response may not be accurately recognised. This can be achieved by using a larger input signal, but this might lead to the use of a test signal that is too large, causing damage to the plant or product, or pushing the system into a non linear region and hence invalidating the result. This problem can be overcome by repeating the test and averaging the result. As long as the noise is stationary with a zero mean, it will reduce with successive averages, thus a small test signal can be used. However there is the disadvantage that a long

examination time is required and the system might drift
during the test.

A system is completely described by its weighting
function h(t),

$$y(\tau)= \int_{o}^{t} h(t)\ u(\tau-t)dt \qquad (2.1)$$

where u and y are system input and output, respectivly.
The obvious method of obtaining h(t) is to use as an
input, the Dirac impulse function $\delta(t)$ which has the
property of infinite height and zero width and unit
area. If this is used in equation (2.1) it will result
in $y(\tau)=h(\tau)$. This kind of test signal is impossible
to generate in practice, although in seismic work it is
often possible to devise a test signal that is a close
approximation to an impulse. A more practical general
approach is to use a step input, the weighting function
can then be obtained from the derivative of the step
response. However the derivative of a noisy response is
usually meaningless, and noise free step responses
require a considerable number of averages. Therefore if
h(t) is required an alternative method of obtaining it
must be used.

The correlation method [15,16] is the most popular

method of achieving the weighting function without the
need for an impulse test signal.

The relationship between the input u(t), and output
y(t), signals is given by the convolution integral (2.1)
A similar relationship exists for the auto and cross
correlation functions,

$$r_{yu}(\tau) = \int_{o}^{\tau} h(t) r_{uu}(\tau - t) dt$$

where

$$r_{yu}(\tau) = \frac{1}{2T} \int_{-T}^{T} y(t) u(t - \tau) dt \qquad (2.2)$$

From this relationship it can be seen that if the auto
correlation of the input signal can be made to look like
an impulse, then the cross correlation between the input
and output signals will resemble the impulse response or
weighting function. The auto and cross correlation can
be performed on the natural noise in the system, but
this is rarely suitable. Very long integration times
would have to be used on the correlation integral (2.2)
to obtain a good approximation. In addition the
bandwidth of the noise may not be sufficient to excite
all the modes in the system, and the auto correlation

function would not approximate an impulse if the noise bandwidth was narrow.

Instead a noise signal with sufficient bandwidth would be used to excite the system. The most popular noise signal is the Pseudo Random Binary Sequence (PRBS). The use of this signal has many advantages. The test is relatively quick (defined by the sequence length) and because the signal is of a low level, typically 2% of full scale output, it can be carried out on-line with no disruption to plant or process. The PRBS signal is a two level signal, fig. 2.1, which is easy to generate using a shift register and an exclusive - OR gate, fig 2.2. The bandwidth of the PRBS noise is controlled by selecting the clock rate. The auto correlation function of the PRBS signal, shown fig. 2.3, is a narrow spike of finite height and width twice the clock period. There is a small offset, which will result in a small bias in the cross correlation but this is negligible when a long sequence is used. The triangular nature of the auto correlation will result in a weighting function that is slightly distorted, but again this can be designed to have a negligible effect. Because the PRBS signal is Pseudo-Random its auto correlation function repeats at the sequence length N. Great care must be taken to ensure that the cross correlation function has settled before the auto correlation function repeats, ie the
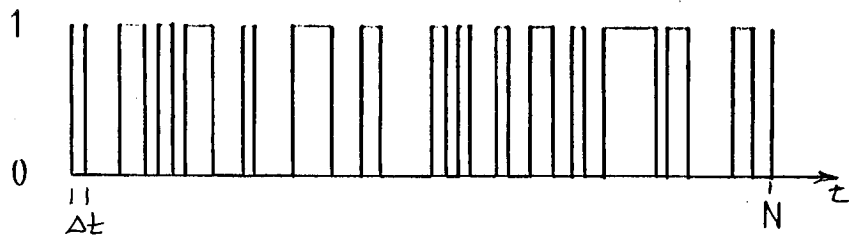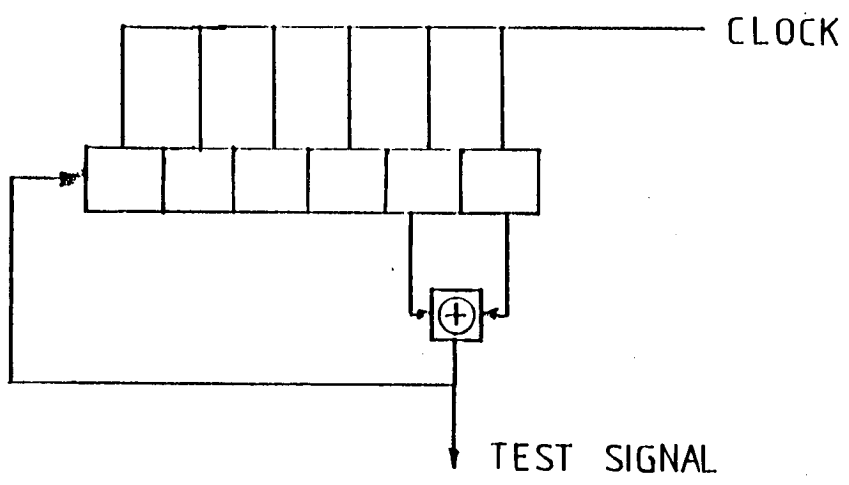
29

FIGURE 2-1

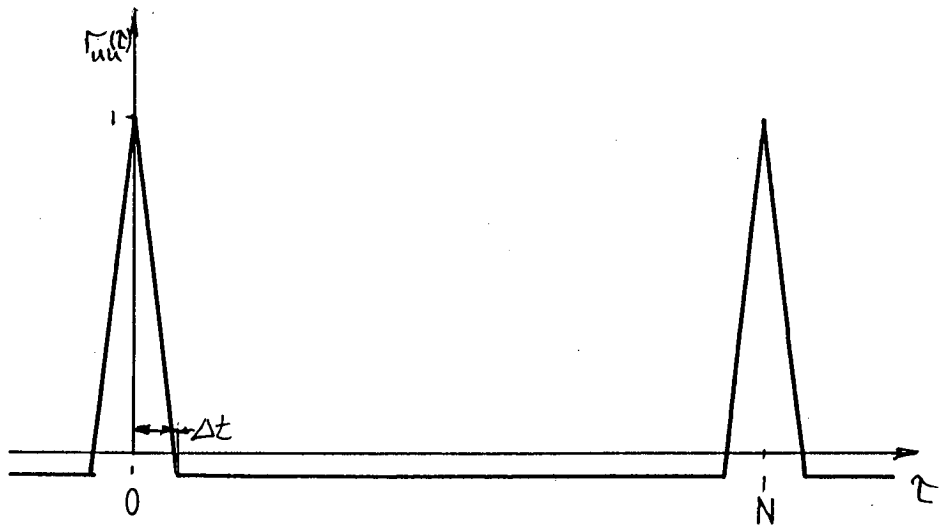PRBS SIGNAL

FIGURE 2-2

PRBS GENERATOR
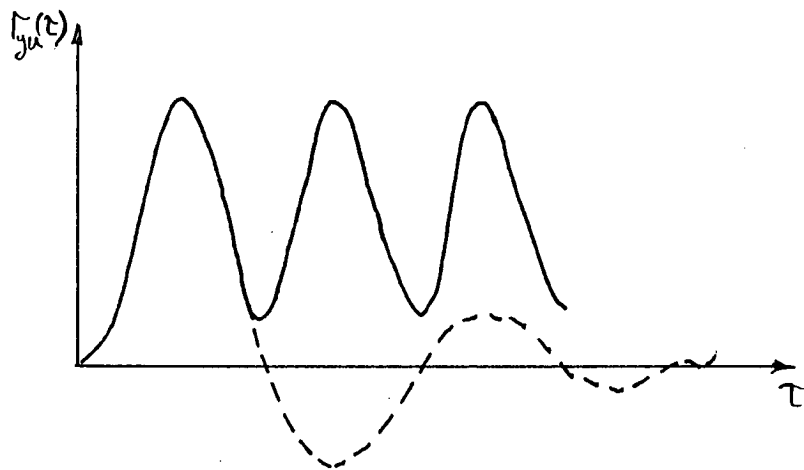
FIGURE 2-3
AUTO CORRELATION OF
PRBS NOISE

FIGURE 2-4
THE EFFECT OF A SHORT PRBS
SEQUENCE LENGTH ON THE CROSS
CORRELATION

sequence period, must be less than the 2% settling time T_s. The effect on the cross correlation function by having a short sequence length is shown in figure 2.4.

Inevitably the signals measured will be corrupted with noise, fig 2.5. The input PRBS signal is assumed to be noise free and any variance in the auto correlation signal will be due to truncation of the integration time from $\pm\infty$ to $\pm T$. The cross correlation variance will partly be due to integration time but spurious system noise will be more of an influence. The variance problems can be overcome by averaging the auto and cross correlation function estimates.

Correlation has been widely used as a system identification tool [15]. Applications were limited initially by the lack of computer power to perform the correlations necessary. Even so on-line identification of distillation columns [17,18], and other chemical plant [19] has been achieved. Godfrey [19] has used this method on nuclear power plant, and shows an interesting range of other applications by other authors. More recently Flower & Windlett [20,21] have used these techniques to identify the dynamics of a Diesel engine. Lang et al [22] has used the technique on a voltage regulator.
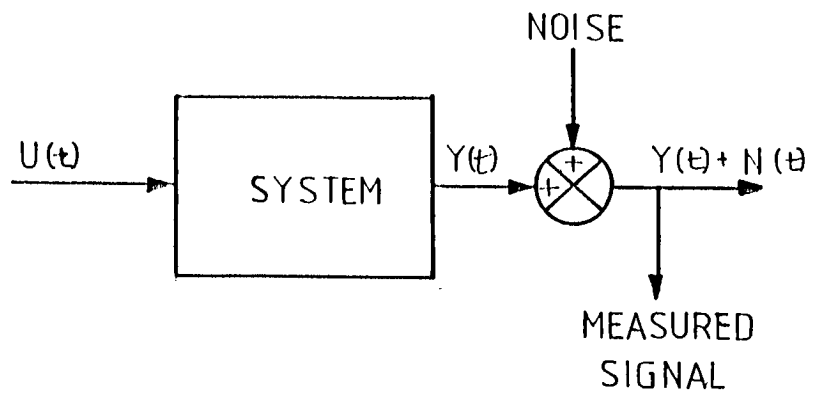
32

FIGURE 2-5

INFLUENCE OF NOISE

2-2-2 Frequency Domain Methods.

Probably the most popular method of system identification is that of transfer function analysis. This is the method of injecting sine waves into a system at different frequencies and measuring the steady state response of the system, i.e. the magnitude and phase shift of the output signal. These measurements give the frequency response plot. Athough well understood and a very useful tool, this method needs very long experiment times. Low frequency test signals can have long periods and will require many averages to overcome noise problems. It would therefore be better if instead of trying to identify each point on the frequency response curve individually, all the points on the curve could be determined simultaneously. This can be achieved using the Discrete Fourier Transform (DFT). The DFT requires a large number of multiplications, N^2, for a N point DFT. This makes the algorithm less attractive for identification work as it is slow to implement. Cooley and Tukey [23] proposed a fast algorithm, the Fast Fourier Transform (FFT) to reduce the number of multiplications to $N.\log_2 N$. This discovery led to an enormous expansion in this field, and now commercial spectrum analysers are available.

The FFT is used to calculate the discrete Fourier

sequences $u_i(k)$ and $y_i(k)$ of the input $u(t)$ and output $y(t)$ signals. From these the auto and cross periodgram $P_{uu}(W)$ and $P_{yu}(W)$ is calculated. These are averaged to produce the auto and cross spectra, $S_{uu}(W)$ $S_{yu}(jW)$. The transfer function $H(jw)$ is then obtained from,

$$H(jw) = \frac{S_{yu}(j\omega)}{S_{uu}(j\omega)}$$

Welstead [16] decribes DFT techniques and their problems (eg windowing). He also gives real applications of the techniques. Ljung and Glover [24] compare the spectral estimation techniques with the discrete parametric techniques now popular in system identification. They conclude that spectral analysis and parameter estimation are not competitive techniques, but that they should be used to complement one another.

2-3 Parametric Identification

Parametric identification has become the most widely researched identification technique over the past twenty years or so. This has been due, principally due to the need for a model that can be used in computer simulation to aid system and controller design. Parametric models are mathematical relations which define the system response for a given input.

Parametric models are amenable to computer implementation and as such, have a great advantage over non-parametric models.

Parametric identification can be categorised into two approaches. Discrete and continuous methods. It is thought that continuous models are more useful for fault detection purposes [83] and we shall, therefore, concentrate on this approach. There appears to be no unified approach to the problem. There are a large number of techniques, and variations on techniques, in both discrete and continuous methods. The main methods of interest in discrete identification will be mentioned, and a comprehensive review of continuous methods will be presented.

Although there is no unified approach one problem all methods suffer from, is modelling. A good model of the structure of the system is needed before identification of the parameters of the model is carried out. A brief review of the techniques of modelling is given first.

2-3-1 System Modelling

A model of a system can be obtained either by theoretical modelling of the system, using the physical laws that govern the system to produce a mathematical

36

model, or by experimentation. A combination of the two is also a possibility. A theoretical model of the system is obtained if the system cannot be tested because it is still in the design stage, or if for safety reasons no experimentation is permitted. In this case, the physical laws of conservation of mass, energy and momentum are applied to the system and a mathematical relation, or set of relations is obtained. This relation describes the system's response to a given input.

If experimentation is permitted and a system is so complex that the coefficients of the model cannot be determined easily using theoretical techniques, system identification and parameter estimation techniques may be used. For these methods the structure of the model should be known. This enables a known number of parameters to be identified. Thus system identification requires a certain amount of theoretical modelling to be carried out in advance.

Models can be classified broadly into four categories depending on the types of mathematical relation the model is based on. It may belong to any one or a combination of the following categories:

    a) Steady state or dynamic

    b) Linear or Non-linear

c) Lumped or Distributed

d) Continuous or Discrete

a) A steady state model is usually based on single algebraic equations. Flow through a pipe could be modelled this way, but if surges in the flow or transients existed, and were of interest, a dynamic model would be required. This would normally be a differential equation.

b) A dynamic model of a system may be linear or non-linear. In general, if a system is non-linear it is better to linearise the model around an operating point. Linear models are more easily identified than non-linear models. Provided any test signal is kept small, and the system does not drift, this should be a valid simplification. This is well described by Fasol and Jorgl [25].

c) Similarly systems that are described by partial differential equations, distributed systems, are more easily identified if a lumped parameter model, a differential equation is used. Kubresly [10] has surveyed the different methods of distributed system parameter estimation, but ordinary differential equations are better understood in the parameter estimation context.

38

d) Continuous modelling is the classical approach and theoretical modelling results in a continuous model. This can be made discrete if a discrete estimation method is to be used. If we are restricted to SISO models then the black box transfer function model, derived from the system differential equation is often used. A typical transfer relation is shown below,

$$\frac{Y(s)}{U(s)} = G(s) = \frac{B(s)}{A(s)}$$

Where A and B are polynomials in s, and U and Y are the system input and output respectively. Hence for example,

$$A(s) = a_1 s + a_2 s^2 + \cdots + a_n s^n$$

For discrete systems a similar is obtained ie,

$$Y(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} U(z^{-1})$$

Where z^{-1} is the backward shift operator and

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n}$$

Note that this is a common, but inexact, use of functional notation ie $s \neq z^{-1}$.

2-3-2 Discrete Techniques.

Discrete techniques of system identification are by far the most popular approach to the problem. There is a very large collection of literature with very many techniques, and variations on techniques, being reported. As continuous methods of identifcation are of more interest in fault detection, and the number of discrete techniques available would prohibit a comprehensive review, only a brief review of the main discrete parameter estimation techniques will be presented here. For brevity, state estimation will not be reviewed, neither will recursive versions, which have been developed for most of the discrete methods.

The methods reviewed here are either all extensions of the least squares method, or techniques derived to overcome the major problem of least squares estimation. It is therefore appropriate to start with least squares estimation and introduce the other methods as the problem becomes apparent.

The least squares technique was first developed by Gauss for use in astronomical experiments. Lately it has become popular as a means of system identification.

The least squares method estimates the parameters of a model of the system, using successive samples of the input/output records of the system. If we assume that the system can be modelled by the z-transform function,

$$A(z^{-1}) \, Y(z^{-1}) = B(z^{-1}) \, U(z^{-1}) \qquad (2.3)$$

where z^{-1} is the backward shift operator. Equation (2.3) can be written,

$$y_k = -a_1 y_{k-1} - \cdots - a_n y_{k-n} + b_0 U_k + \cdots b_m U_{k-m}$$

This is more compactly expressed in vector form,

$$y_k = \underline{z}_k^T \underline{\theta}$$

where $\underline{z}^T$ is the transpose of $\underline{z}$.

and
$$\underline{z}_k^T = \left[-y_{k-1}, \cdots, -y_{k-n}, \, U_k, \cdots, U_{k-m} \right]$$

and
$$\underline{\theta}^T = \left[a_1, \cdots, a_n, \, b_0, \cdots, b_m \right]$$

The least squares estimate of the parameter vector, $\hat{\underline{\theta}}$, is that estimate of $\underline{\theta}$ which minimises the sums of the squares of the measurement errors, or residuals. The

41

residual is usually formed from the equation error, although other error models exist, such as output error, figure 2.6.

The residual, e_k is defined as,

$$e_k = y_k - z_k^T \cdot \hat{\underline{\theta}}$$

For N samples of the input u_k and the output y_k ie k=1,2,...,N., N should be very much larger than the number of parameters to be estimated (m+n+1) [26].

From this a cost function $J(\theta)$, which is a function of the parameters, can be formed. The best parameter estimates are obtained when the residuals are small and hence the cost function is a minimum. This is achieved by differentiating the cost function with respect to θ and equating to zero to find the minimum. The cost function is defined by,
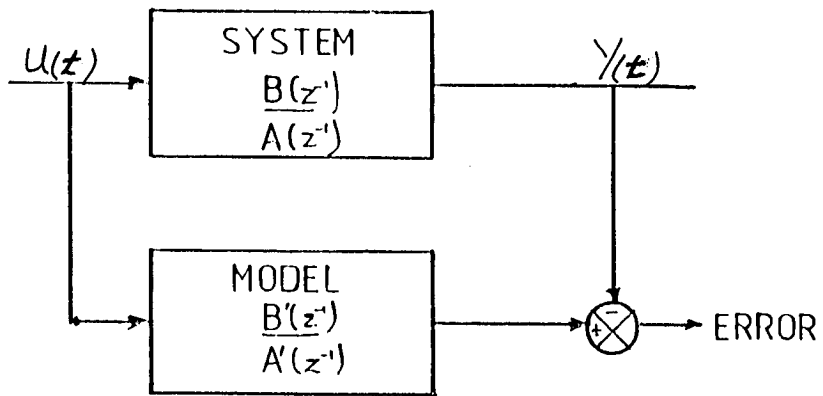
$$J(\theta) = \sum_{k=1}^{N} e_k^2$$

Differentiating and setting the result to zero leads to a set of linear equations known as the normal equations from which the parameter estimate θ can be obtained ie,

42
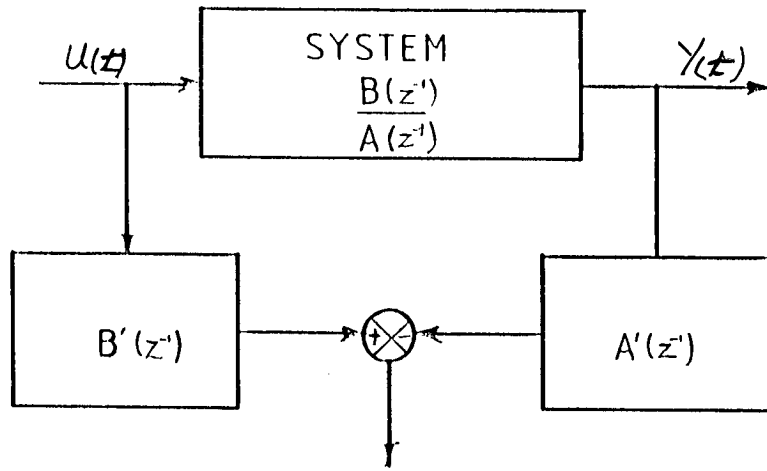
EQUATION ERROR

OUTPUT ERROR

FIGURE 2-6

43

$$\left(\sum_{k=1}^{N} z_k z_k^T \right) \hat{\Theta} = \sum_{k=1}^{N} z_k y_k \qquad (2.4)$$

This is the deterministic approach to least squares estimation. Real systems however are not deterministic, noise contamination of the input/output signals means that a stocastic approach must be used. Equation (2.3) can be re-expressed,

$$Y(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} U(z^{-1}) + \mathcal{E}(z^{-1}) \qquad (2.5)$$

$\mathcal{E}(z^{-1})$ is the effect of all the noise on the system. It is known that the estimate of $\underline{\theta}$ will only be unbiassed if the noise $\mathcal{E}(z^{-1})$ is white and has zero mean. Otherwise the estimate $\hat{\underline{\theta}}$ will be biased, and by an unknown amount. If the residual e_k is not a sequence of random variables, but there exists some dependence between successive residuals, then the residuals are said to be correlated and the parameter estimates will be biased.

This is the major problem of system identification. Most of the methods developed are aimed at overcoming this problem, usually by including in the system model a model of the noise. We shall look at four popular techniques, these are

a) Generalised least squares

b) Instrumental variables

c) Maximum likelihood

d) Correlation with least squares

a) The Generalised Least Squares method of system identification tries to overcome the problem of bias by replacing the white noise term $\mathcal{E}(z^{-1})$ in equation (2.5) by a coloured noise term $C(z^{-1})$. This is an attempt to model the noise by a filter driven by white noise. This coloured noise is a sequence of correlated random variables defined by,

$$C(z^{-1}) = \frac{\mathcal{E}(z^{-1})}{F(z^{-1})}$$

Equation (2.5) now becomes

$$Y(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} U(z^{-1}) + \frac{\mathcal{E}(z^{-1})}{F(z^{-1})} \qquad (2.6)$$

$\mathcal{E}(z^{-1})$ is uncorrelated white noise with zero mean. The problem of estimation can now be dealt with by using the least squares method with modified input/output signals, as shown in figure 2.7. Then new output is,
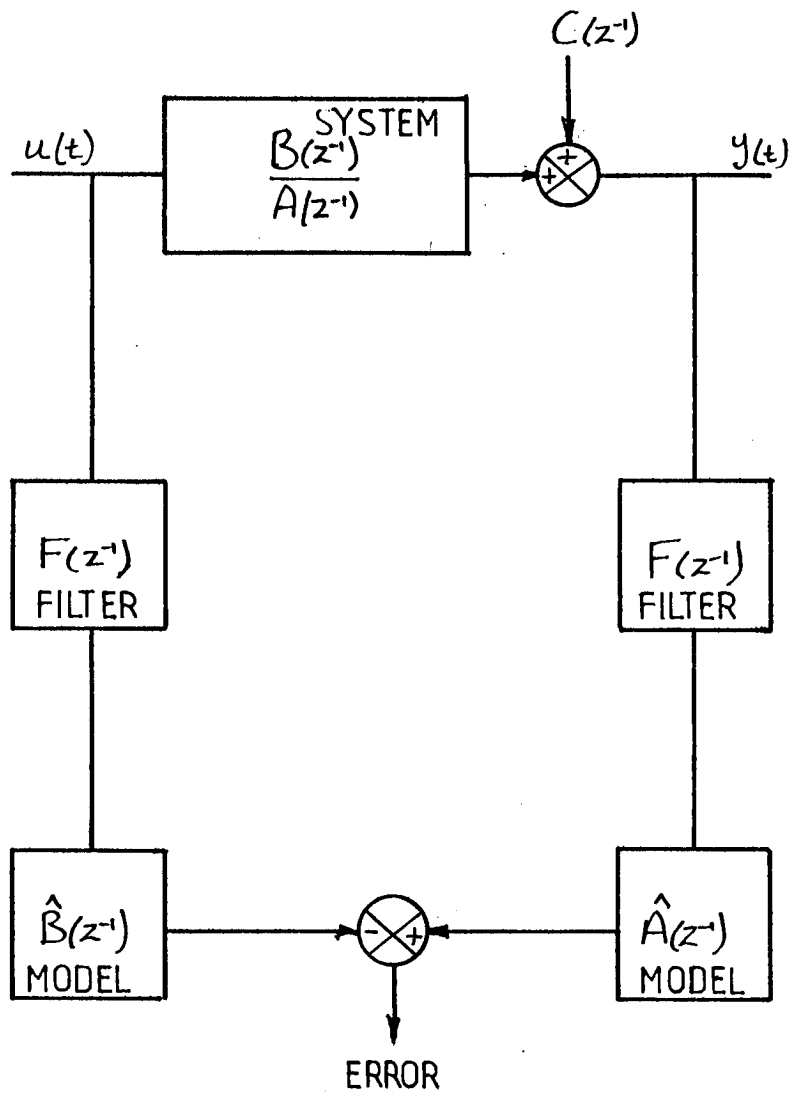
$$F(z^{-1}) \, Y(z^{-1})$$

45

FIGURE   2·7

GENERALISED   LEAST   SQUARES   METHOD

and new input is,

$$F(z^{-1}) \, U(z^{-1})$$

The problem arises in determining $F(z^{-1})$. An iterative procedure is generally used to determine the polynomial $F(z^{-1})$ [27].

1) Set $F(z^{-1})=1$

2) Use new input/output signals $F(z^{-1})U(z^{-1})$ and $F(z^{-1})Y(z^{-1})$
   for parameter estimation

3) Use 2) in least squares estimate of $\underline{\theta}$

4) Use $\underline{\hat{\theta}}$ to form residual $e_k = A(z^{-1})Y(z^{-1}) - B(z^{-1})U(z^{-1})$

5) Model this biased residual with $F(z^{-1})e_k = \mathcal{E}(z^{-1})$

6) Have the parameters of $F(z^{-1})$ converged.

   No - goto 2) use latest estimate of $F(z^{-1})$.

   Yes - stop

b) Instrumental Variables is another method used to overcome the problem of biased estimates. For this an additional variable is introduced, the instrumental variable. This is chosen so that it is strongly correlated with the noise free system input and output signals, but not correlated with the noise that corrupts the input/output signals. These variables are used in the cross products in the co-variance matrix, where, as in the least squares approach, the noise product is squared. In the instrumental variable method the noise

is rejected, because the instrumental variable is independent of the noise, therefore the noise is correlated out.

For the instrumental variable the least squares estimator (2.4) is modified,

$$\left(\sum_{k=1}^{N} \hat{\underline{x}}_k \underline{z}_k^T \right) \hat{\underline{\theta}} = \left(\sum_{k=1}^{N} \hat{\underline{x}}_k y_k \right)$$

Where $\hat{\underline{x}}_k$ is the instrumental variable vector defined as,

$$\hat{\underline{x}}_k = \left[-\hat{x}_{k-1}, \cdots , -\hat{x}_{k-n}, u_{k-1}, \cdots , u_{k-m} \right]^T$$

This instrumental variable is chosen so as to be highly correlated with the noise free vector $\underline{x}_k$

$$\underline{x}_k = \left[-x_{k-1}, \cdots , -x_{k-n}, u_{k-1}, \cdots , u_{k-m} \right]^T$$

Where x_{k-1} is the noise free output, y_{k-1} being the corrupted output see figure 2.8. If $\hat{\underline{x}}_k$ is available the parameter estimates $\hat{\underline{\theta}}$ will be unbiased. The problem is that $\hat{\underline{x}}_k$ is seldom known and unless it is correlated strongly with $\underline{x}_k$, and independent of the noise $\mathcal{E}_k$, a good estimate of $\underline{\theta}$ will not be achieved. Young [28] discusses the various possible methods of generating an instrumental variable. The use of an auxiliary model is a popular
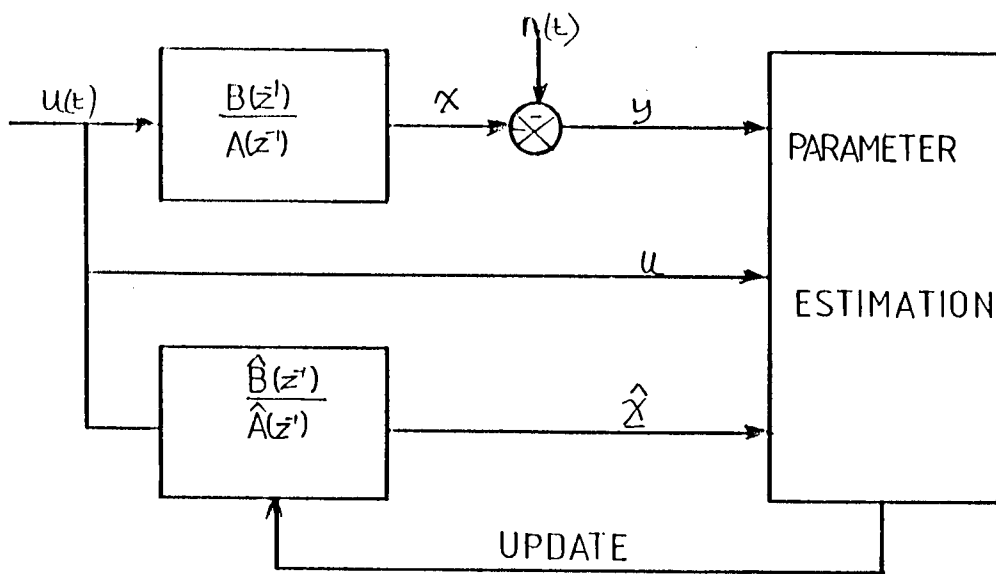
FIGURE 2 - 8
INSTRUMENTAL VARIABLE
METHOD

approach. This is a model which is updated with each successive estimate of the system until the parameter estimates converge, figure 2.8.

c) The maximum likelihood method [30] deals with the problem of correlated residuals in quite a different way from the two previous approaches. The maximum likelihood estimate of the parameters is obtained by maximising a function of the observations and the parameters, the likelihood function. This function is the probability density function of the observed data. It is necessary to know the shape of the probability density function to implement this function.

d) Correlation with least squares. From the least squares estimator equation (2.4) it is found that the elements of $z^T.z$ are essentially auto and cross correlation of the input and output signals [33]. Therefore if the auto and cross correlations are calculated, they can be used in the estimation of the parameters. This is the method of correlation with least squares [31]

If the model order is not known and may also have a time delay term present then the procedures looked at above suffer from an additional problem. The model order and time delay must be determined using an

50

iterative technique [31]. Using this technique the model order and time delay is increased and a least squares parameter estimate performed in each iteration until the loss function is minimised. This involves a great deal of computational effort. One way of overcoming this problem is to identify a non-parametric model first, from which the model order and time delay can be determined, and then use least squares estimation on this model [32].

Once a model of the system has been obtained a verification procedure should be carried out to ensure that the model order is correct, and that there is no longer any bias due to correlated residuals. Isermann [34] suggests a loss function test to determine the correct model order. The system is repeatedly modelled with increasing order of model, and a loss function formed which is dependent on the model order.

$$V(m) = \int_{-\infty}^{\infty} e^2(m) \, dm$$

The order of the system is obtained when the loss function V(m) is a minimum.

Isermann [34] also suggests that the residuals can be tested for whiteness, by performing an auto correlation

on them. Any colouring of the residual will show as a spread in the auto correlation function.

As a final check on the model, the system and model output should be compared, if any significant deviation is detected, modifications can be made to the model. The model should be checked for different input signals, eg steps and ramps and different amplitude input signals. All of these methods have received much attention in the past, the Astrom and Eykhoff [33] survey paper covers many of the methods mentioned here, and a great deal besides. This paper, although fifteen years old, is still a useful and comprehensive survey.

Isermann et al [32] has compared six identification methods. A comparison of the variances of the parameter estimates and the impulse response for the six estimates was made. These were carried out using two different signal to noise levels. This study was carried out on a simulated system and hence the structure was known already. It is interesting to note that the methods that perform best, and which had the smallest deviations from the true values, were those with better noise models. The instrumental variable method performed well for example and, as would be expected, the least squares method did not perform as well.

Many of the studies which have taken place have been simulation studies, comparatively few have taken place on real systems. This may be due in part to the expense and availability of testing real systems, although it seems possible that methods which work well in the laboratory do not perform well on real systems identification. Another reason is the model order. Strejc [35] suggests that the estimation of systems with order greater than three is not practicable. One reason for this is the loss of information due to sampling , but the main reason is contaminating and unmeasurable noise, which real systems have in abundance. This limit on the estimation procedure is obviously a handicap when it comes to identifying real systems. However, low order models of quite complex systems are acceptable in some circumstances [35].

When performing identification on real systems the design of the experiment is important. It is wise to gather as much information about the system before identification begins. This should enable selection of sample time and measurement time. Isermann [34] suggests 6 to 15 samples per 95% settling time for PID control algorithms. If the sample time chosen is too large then the dynamic behaviour is not described accurately, and if the sample rate is too small, ill conditioned matrices result, because the difference

equations become approximately linearly dependant [34]. The selection of the input signal is also very important. This should be chosen so that the system is persistently excited, i.e. the whole dynamic range of the system is excited. PRBS signals or band limited noise are suitable for this purpose.

Recently there has been a number of studies carried out on real plant. Kallstrom [36] has investigated ships' steering dynamics as has Trankel [37] who uses a maximum likelihood method. McDyer [38] has used least squares estimation to identify a low order model for three different types of generating plant, gas turbines, steam plant and hydro plant. These simple models achieved reasonably good results. Hogg et al [39] has also investigated generating plant; least squares and instrumental variable methods were used.

2-3-3 Continuous Techniques

Although most of the effort in system identification has been centred on the discrete techniques, there has been a small but significant parallel development in continuous system identification methods. Continuous methods estimate the parameters of the differential equation which results in a s-plane transfer function. There exists a great many analytical techniques to

handle continuous models. For fault detection continuous models are of particular importance because not only is detection of the fault important, but it is also necessary to diagnose and locate the fault in the system as well. For this purpose it is necessary to have a model that can be related directly to the system through the physical laws that govern the system. Continuous models are the closest to the physics of the system and any deviation in a model parameter can be related to a malfunction in a system component [83]. Discrete models are further removed from the direct relationship between the system and the model by the sampling process and therefore the parameters cannot be so easily related to system failures.

If we consider a SISO lumped parameter linear system which is represented by the differential equation

$$\overset{n}{y}(t) + a_{n-1}\overset{n-1}{y}(t) + \cdots + a_0 y(t) = b_{m-1}\overset{m-1}{u}(t) + \cdots + b_0 u(t) \quad (2\cdot 7)$$

where, $\quad \overset{n}{y}(t) = \dfrac{d^n y}{dt^n}$

This can be represented as a transfer function,

$$Y(s) = \frac{B(s)}{A(s)} U(s) \qquad (2.8)$$

$$A(s) = s^n + a_1 s^{n-1} + \cdots + a_n$$

$$B(s) = b_1 s^{m-1} + \cdots + b_m$$

This will be the basic model used here. Another popular representation is the state space model but for brevity only the transfer function model will be considered.

There are two approaches to the problem of estimation of the continuous system parameters. One approach, the indirect method, first estimates the discrete model of the system and from this a continuous model is obtained [43]. The other approach is direct, in that the parameters of equation (2,7) are calculated directly using one of the techniques available. It is these methods that are of most interest here.

One method is to differentiate the input and output signals, substituting these derivatives in the differential equation. This method is generally disregarded because of the problems of differentiating noisy signals. However the first class of methods reviewed here attempts to obtain the derivatives by

using state variable filters. The second class of
methods repeatedly integrates the differential equation
so that an integral equation is formed. The third class
avoids the differentiation by means of a Modulating
function. This is looked at in detail in the next
section.

In the first class of methods, the discrete methods
which were studied in the previous section can be
applied to continuous models. Young [40] gives an
interesting survey of these and the other methods
designed to overcome the problem of the derivatives of
noisy signals. The least square method illustrates the
problem. From equation (2.8) an equation error or
residual can be formed, Figure 2.9

$$e(t) = \mathcal{L}^{-1}\left[A'(s)\,Y(s)\right] - \mathcal{L}^{-1}\left[B'(s)\,U(s)\right]$$

A'(s) indicates the model rather than the system A(s).
To form this residual, derivatives of the input and
output signals must be formed, ie

$$e(t) = \overset{n}{y}(t) + a_1'\overset{n-1}{y}(t) + \cdots + a_n' - b'\overset{m-1}{u}(t) - \cdots - b_m'$$

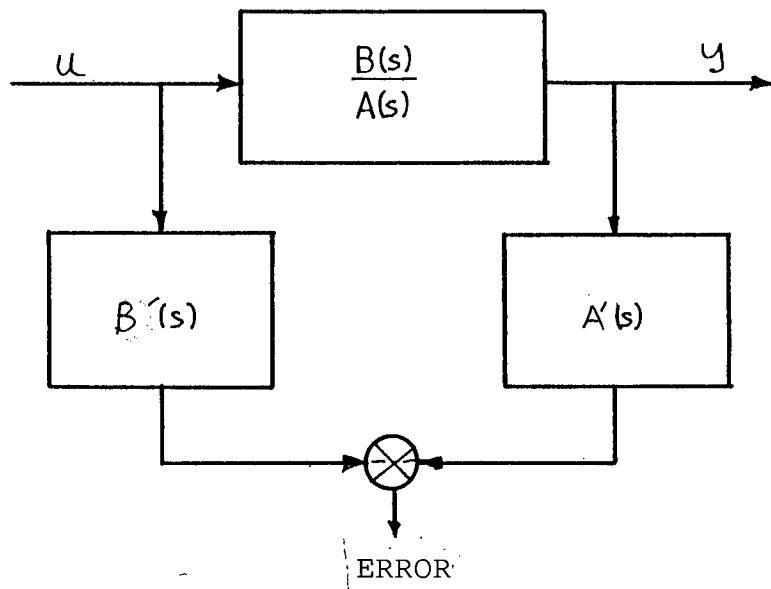This would be very difficult in the case where the

57

FIGURE 2-9
EQUATION ERROR OF A
CONTINUOUS SYSTEM

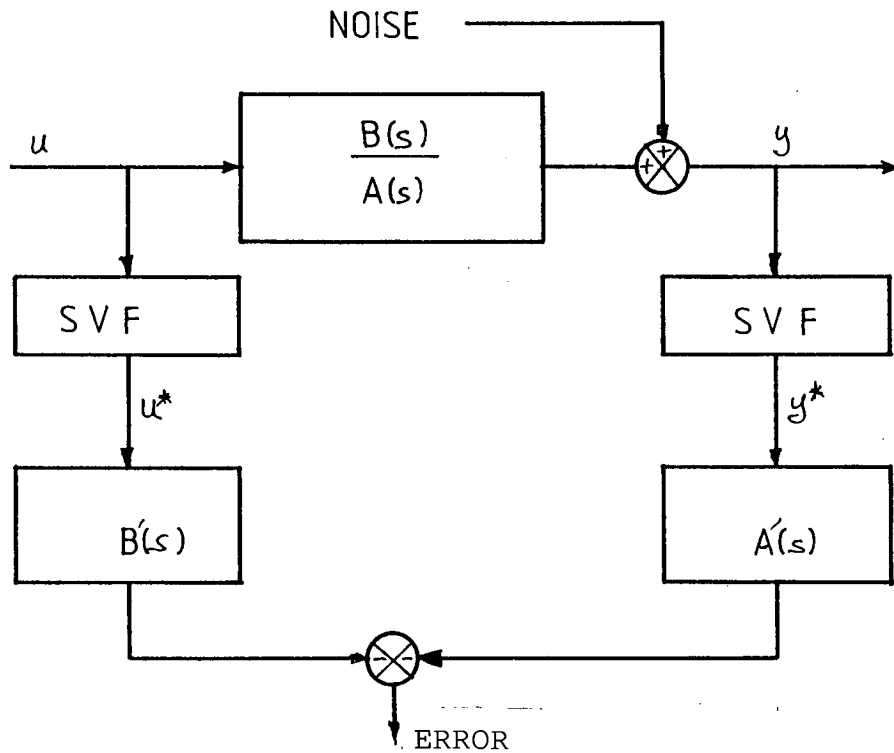FIGURE 2-10
THE USE OF STATE VARIABLE FILTERS
IN THE ESTIMATION OF A CONTINUOUS
SYSTEM

58

signals y(t) and u(t) were not deterministic, but
contaminated with noise, as the noise would be amplified
by the differentiation. To avoid this problem a
generalised equation error is formed in place of the
equation error. In this method the derivatives are
obtained by the use of a State Variable Filter, which
filters the input and output signals and provides
filtered time derivations of the input/output signals
[40], figure 2.10

The cost function J is defined not as a sum of the
squares of the residuals, but as an integral,

$$J = \int_{T_1}^{T_2} e^2(t)dt$$

Where T_1 to T_2 is the sample window. As in the discrete
case, minimising this cost function will provide an
estimate of the parameters.

The Instrumental Variable method can be applied in the
continuous case. This overcomes the problem of bias due
to correlated residuals. Another signal is introduced,
the instrumental variable, which is highly correlated
with the noise free input/output signals but independent
of the noise. Young [40] proposes an auxiliary model
which is updated by the parameter estimates to produce a

noise free input/output vector, figure 2.11 This is similar to the discrete approach, but again the derivatives of the noisy signals are required and therefore a state variable filter is included. As the auxiliary model output $\hat{x}$ becomes highly correlated with y, the auxiliary model will match the system, and an estimate of the parameters will be achieved.

In the second class of methods the problem of derivatives of the input/output signals is overcome in other ways. Diamensis [41] used repeated integrations of the input/output signals to obtain a set of linear equations in which the only unknowns are the parameters. Consider equation (2.7) this can be integrated n+m times.

$$\sum_{i=0}^{n} a_i \underbrace{\int_{0}^{t_k} \cdots \int_{0}^{t_k} y(t) \, dt^{n+1-i}}_{n+1-i} = \sum_{j=0}^{m} b_j \underbrace{\int_{0}^{t_k} \cdots \int_{0}^{t_k} u(t) \, dt^{m+1-j}}_{m+1-j} \qquad (2.9)$$

A set of linear equations can be formed by use of different integration times t_k; k=1,2,...m+n. This is perhaps better illustrated using a first order system. Consider,
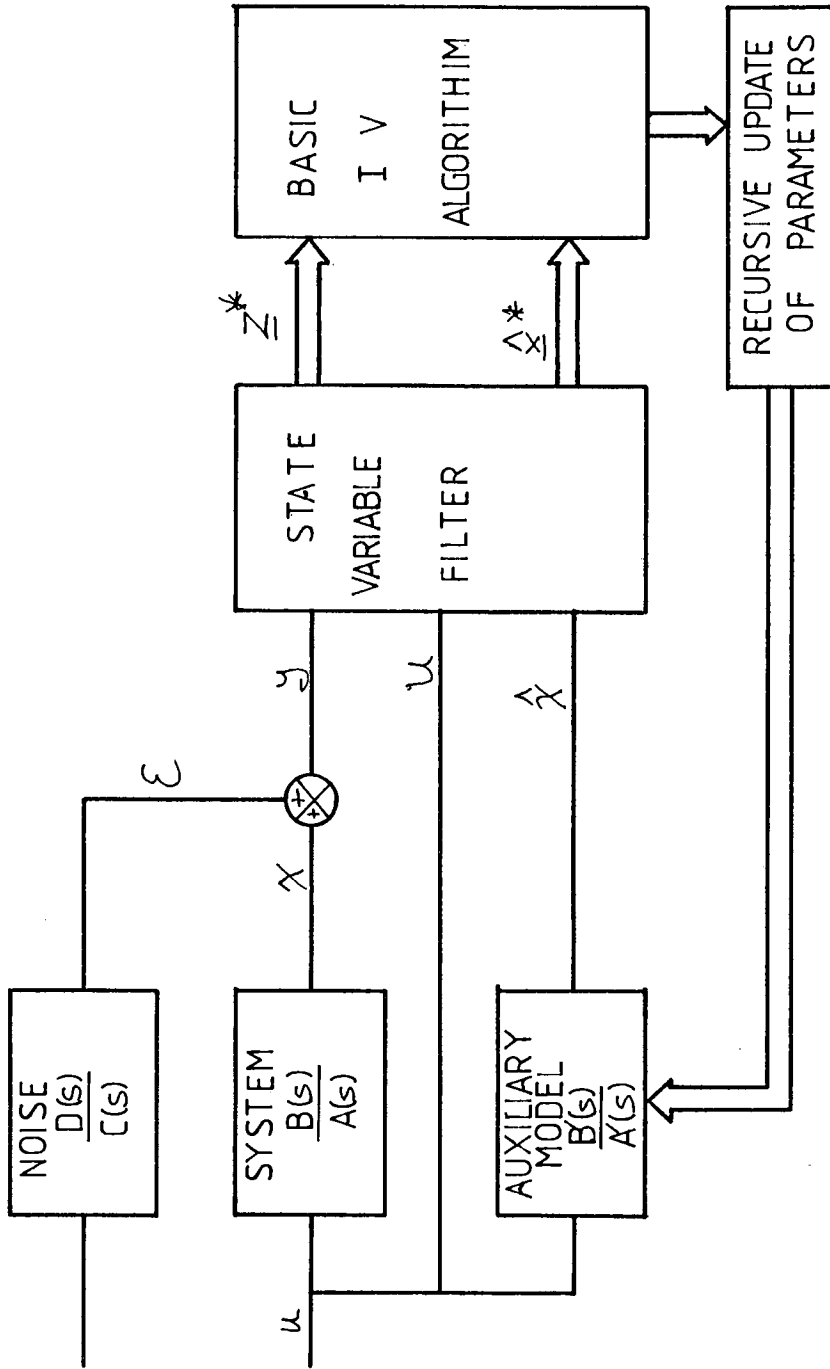
FIGURE 2-11

CONTINUOUS INSTRUMENTAL VARIABLES

$$b\frac{dy(t)}{dt} + y(t) = a\,u(t)$$

Integrating from t=0 to t_1 gives,

$$b\int_0^{t_1} y(t)dt + \int_0^{t_1}\int_0^{t_1} y(t)\,dt^2 = a\int_0^{t_1}\int_0^{t_1} u(t)\,dt^2$$

Integrating from t=0 to t_2 gives,

$$b\int_0^{t_2} y(t)dt + \int_0^{t_2}\int_0^{t_2} y(t)\,dt^2 = a\int_0^{t_2}\int_0^{t_2} u(t)\,dt^2$$

This can be solved, using standard matrix techniques for b and a. This method uses the input/output signals and there is no requirement for a test signal, although it should be noted that the input signal should have a bandwidth wide enough to excite all the modes in the system. The use of repeated integrals can lead to problems of compounding errors, in high order systems.

To overcome the problems of having to perform multiple integrations Garnett and Eisenberg [42] introduced an integral transform, which reduces the multiple integrations to a single integration, and hence overcomes the problem of compounded errors.

Another solution to the integral equation is to set the integration limits over a shorter time interval and approximate the signal with a known function. Sinha [44] has compared three such methods, each based on a successively more complex approximation to the signal. These three methods are the Block Pulse, Trapezoidal and Cubic Spline methods.

The basis of the Sinha method is to integrate the differential equation N times, substituting the estimates of the signals into the integral equation. An equation error is formed and used to solve for the unknown parameters by using Least Squares technique. The integration interval is over a sample period [KT,(K+1)T]. For the first order system described above:

$$Y_{k+1} - Y_k + a_1 \int_{kT}^{(k+1)T} y\, dt + a_0 \int_{kT}^{(k+1)T} \int_{kT}^{(k+1)T} y\, dt_1\, dt_2 = b \int_{kT}^{(k+1)T} u\, dt$$

To solve this integral equation an estimate of the input/output signals over the interval KT to (K+1)T is needed. In the Block Pulse method [45,46] the signal is approximated by a constant equal to the mean value of

63

the function over the integration interval, figure 2.12. Hence,

$$y(t) = \hat{y}_k = \frac{1}{2}\left(y_k + y_{k+1}\right)$$

The trapezoidal method [47] improves the estimate of the signal and hence the accuracy of the solution to the integral equation by modelling the signal with a ramp between the two sample points as shown on figure 2.13. In this case

$$y(t) = \frac{1}{T}\left[\left\{(k+1)T - t\right\}y_k + (t - kT)\,y_{k+1}\right]$$

This increases the computational burden necessary to estimate the parameters, but an improvement in accuracy compared with the Block Pulse method can be expected.

The Cubic Spline method [44] carries this improved estimate of the signal further, by modelling the signal with a third order polynomial as shown on figure 2.14. In this case

$$y_n(t) = \frac{t_n - t_{n-1}}{h}\,y_n + \frac{t_{n-1} - t_{n-2}}{h}\,y_n + (at+b)(t_n - t_{n-2})(t - t_{n-1})$$
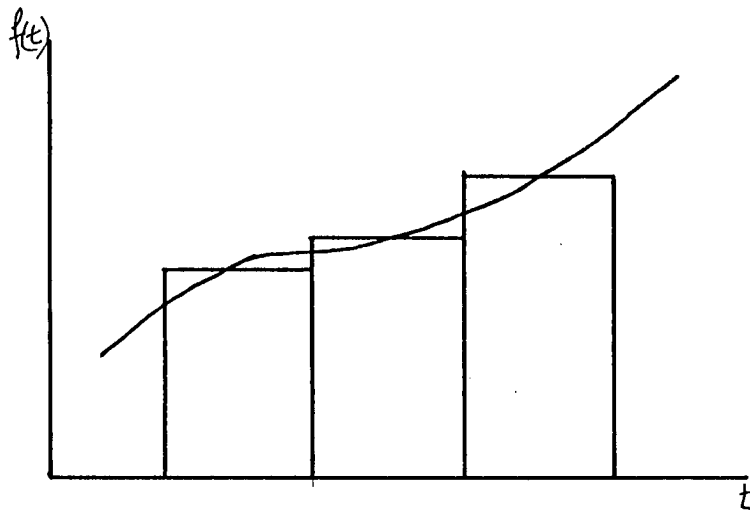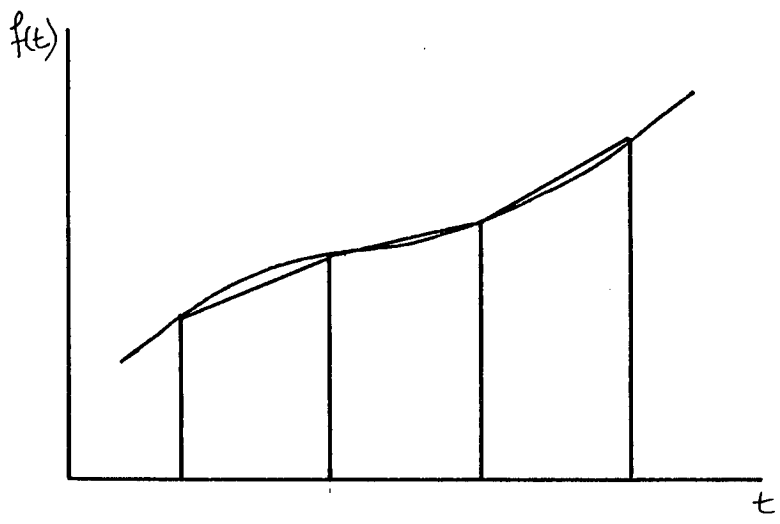
FIGURE 2-12
BLOCK PULSE METHOD

FIGURE 2-13
TRAPEZOIDAL METHOD

where h is the sampling interval.

The repeated integrations have a smoothing effect on the noise resulting in a good parameter estimate [47]. The Least Squares method is used unless the noise level is high, when a more sophisticated technique such as the Instrumental Variables or Maximum Likelihood method, should be used.

Sinha [44] has compared the three methods and has found that the accuracy increases with an increase in complexity but that the Trapezoidal method achieves results that are acceptable without a high level of computation. For noisy signals the sample period should be reduced to cope with this additional problem. The higher the order of model estimated, the greater the number of integrals to be evaluated. It has been found that the more complex functions handle these integrals more successfully [47].

Walsh functions [48] can be used in a similar fashion to the above methods. Again an integral equation can be formed over a time interval [0,T]. The input/output signals are estimated by Walsh series, and these series substituted in the integral equation.

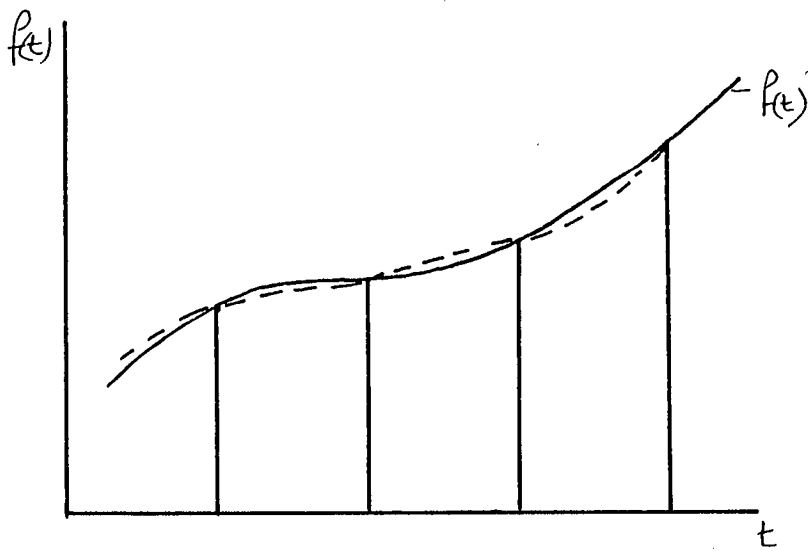A signal can be approximated by a Walsh series in a similar fashion to the Fourier series,
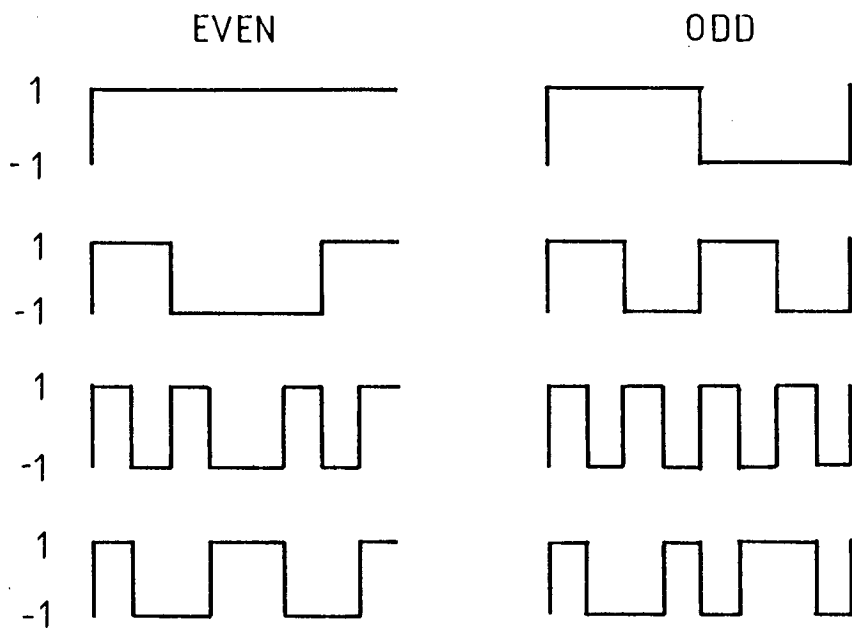
FIGURE   2-14

CUBIC SPLINE METHOD

FIGURE 2-15

WALSH FUNCTIONS

67

$$f(t) = \sum_{n=0}^{\infty} C_n \, \psi_n(t) \qquad (2.10)$$

The coefficients (C_n) of the series can be determined using,

$$C_n = \int_0^1 f(t) \, \psi_n(t) \, dt \qquad n = 0, 1, 2 \qquad (2.11)$$

Where ψ are the Walsh functions. The first eight Walsh functions are shown in figure 2.15. The integral of a Walsh function can be expressed in terms of a series of Walsh functions.

$$\int_0^t \psi_0(t) \, dt \simeq \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{1}{8} & 0 \end{bmatrix} \begin{bmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix}$$

Chen [50] extends this to define a P matrix which translates integration into multiplication.

$$\int_0^t \underline{\psi}_{(4)}(t) \, dt \simeq \underline{P}_{(4 \times 4)} \, \underline{\psi}_{(4)}$$

$$
\begin{bmatrix} \int_0^t \psi_0 \\ \int_0^t \psi_1 \\ \int_0^t \psi_2 \\ \int_0^t \psi_3 \end{bmatrix}
\cong
\begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{1}{8} & 0 \\ \frac{1}{4} & 0 & 0 & -\frac{1}{8} \\ \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\times
\begin{bmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix}
$$

The parameter estimates of the transfer function are obtained by repeatedly integrating the differential equation in a similar manner as used in the previous methods. The input/output signals are expanded into Walsh series using equations 2.10 and 2.11. These Walsh series are substituted in the integral equation (2.9), and the integrals are simplified using the P matrix. Hence

$$
y(t) = C' \psi(t) \qquad \text{and} \qquad U(t) = h' \psi(t)
$$

gives

$$
C'[I + a_1 P + a_2 P^2 + \cdots + a_n P^n]\psi(t) =
$$

$$
h'[b_1 P + b_2 P^2 + \cdots + b_n P^n]\psi(t)
$$

This equation can be solved for the m+n+1 unknown parameters by sampling the input/output signals m+n+1 times, and forming a loss function which is solved for the unknown parameters a,b using the least squares technique.

Corrington [49] has used Walsh functions to solve differential equations. A similar approach employing a P matrix is used to that of [50,51] to determine the sytem output to a given input. Bohn [52,53] has developed a technique which eliminates some of the redundances in the P matrix and improves computation. Rao [54] has used Walsh functions to estimate the parameters of systems with time lags by employing a loss function to determine the unknown lag. Bohn [55] has also developed a technique to handle pure time lags using Walsh functions.

Continuous methods of system identification all suffer from the same problem, how to avoid the use of derivatives of the input and output signals. The continuous methods discussed here attempt to overcome this by the use of State Variable Filters, or repeated integration of the differential equation. These introduce further arithmetic complications and can, in the case of high order systems, result in very complex and clumsy integral equations which would be prone to

truncation errors, and errors introduced by the
approximations to the system signals.

2-4 Modulating Function Method

The methods of continuous model parameter estimation
reviewed in the previous section all tried to overcome
the problem of having to use the derivatives of the
input/output signals. One method of overcoming this
difficulty which has so far not been discussed is the
Modulating Function Method (MFM). As this research work
is based on a development of the MFM it is appropriate
to review the development of this technique
independently from the other continuous methods.

The Modulating Function Method could almost be
described as the classical solution to the problem of
obtaining the parameters of a continuous model. The
method, which was first proposed by Shinbrot [56,57], is
conceptually simple. It overcomes the problem of
derivatives of noisy signals and only uses the
measurable input/output signals of the system. There is
no repeated integration of the signals to form integral
equations, thus no sampling errors are accumulated.
Since the method is one shot it alleviates the need for
the use of Least Squares techniques.

Shinbrot´s method forms the bases of the MFM´s

71

proposed by other authors. These other techniques only usually vary in the application or the choice of modulating function.

Shinbrot proposed that the system equation be multiplied by a modulating function which would have the special property that the function and its derivatives are zero, at the beginning and end of the sample window [0,T].

$$\emptyset[0,T] = \frac{d}{dt}\emptyset[0,T]=0 \qquad\qquad (2.12)$$

Given this, if the modulating function $\emptyset(t)$ is multiplied through the differential equation, and the resulting equation then integrated over the sample time window [0,T], an equation is formed in which no reference to the derivatives of u,y are made, but only to the derivatives of the modulating function.

A second order differential equation is used here to illustrate the method.
Consider

$$\ddot{y}(t) + a\dot{y}(t) + by(t) = Cu(t)$$

multiply by a modulating function

$$\emptyset(t)\ddot{y}(t) + a\emptyset(t)\dot{y}(t) + b\emptyset(t)y(t) = C\emptyset(t)u(t)$$

integrate from t=0 to T

$$\int_0^T \phi(t)\ddot{y}(t)\,dt + a\int_0^T \phi(t)\dot{y}(t)\,dt + b\int_0^T \phi(t)y(t)\,dt = c\int_0^T \phi(t)u(t)\,dt$$

integration by parts gives

$$\int_0^T \phi(t)\ddot{y}(t)\,dt = \left[\phi(t)\dot{y}(t)\right]_0^T - \left[\dot{\phi}(t)y(t)\right]_0^T + \int_0^T \ddot{\phi}(t)y(t)\,dt$$

using equation 2.12 gives,

$$\int_0^T \ddot{\phi}(t)y(t)\,dt - a\int_0^T \dot{\phi}(t)y(t)\,dt + b\int_0^T \phi(t)y(t)\,dt = c\int_0^T \phi(t)u(t)\,dt$$

The result is an equation whose unknowns are the system parameters, everything else is known or easily measured. Thus the equation can be solved for the unknowns a,b and c using successive modulating functions ϕ_n, n=1,..N. A block diagram of the method is shown in figure 2.16. Shinbrot has extended this method to include non-linear systems, where the parameters are modelled by polynomials [56].

The choice of $\phi(t)$ is crucial, the end point conditions (2.12) must be met for the method to work. Shinbrot has chosen a trigonometric modulating function, see figure 2.17,
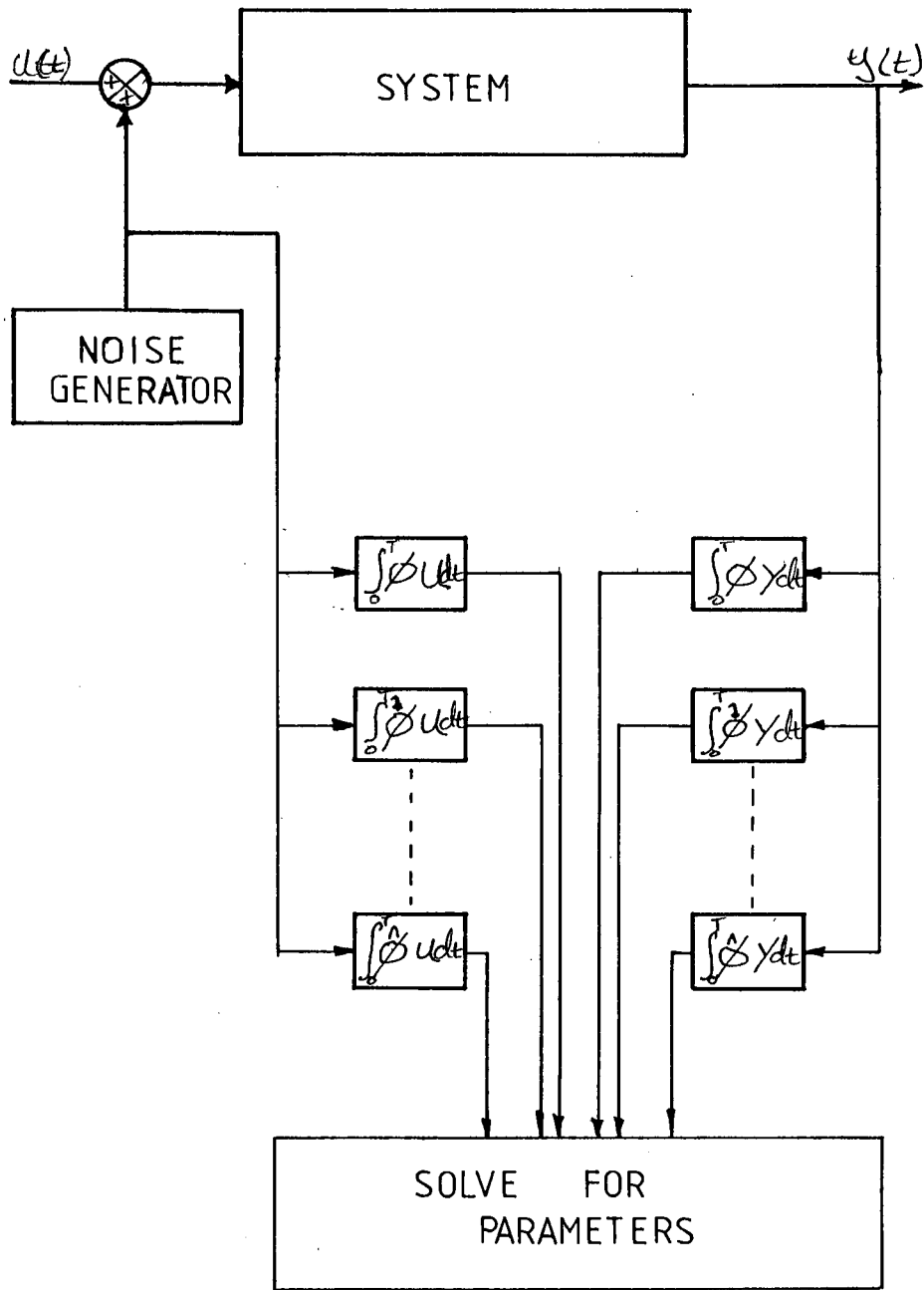
$$\phi_n = \sin^2 w_n t$$

which has the property

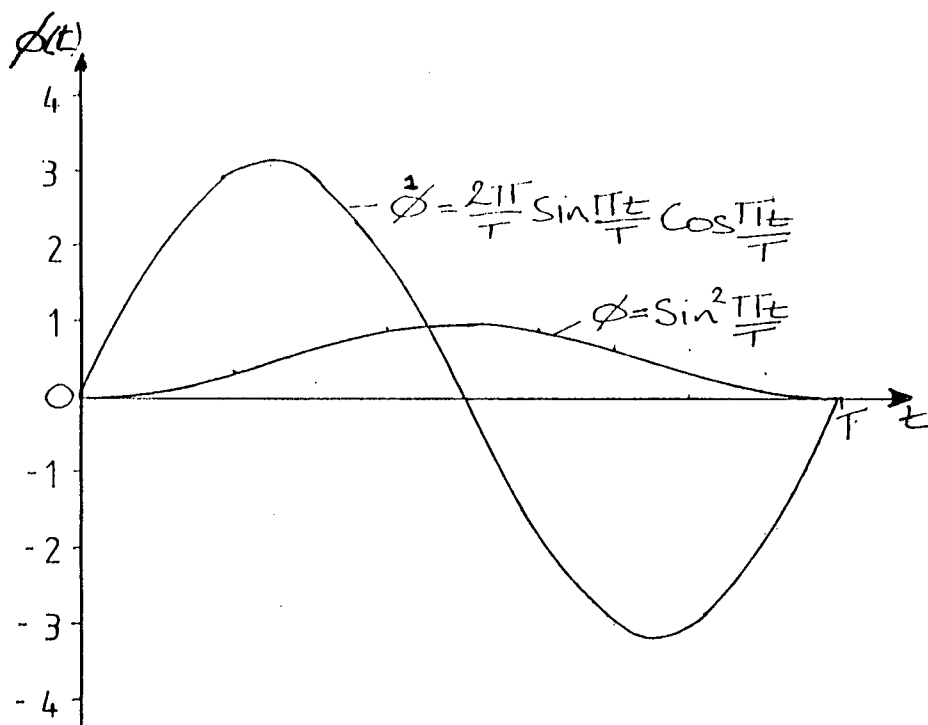FIGURE 2-16
BLOCK DIAGRAM OF THE
MODULATING FUNCTION METHOD

FIGURE 2-17
TRIGONOMETRIC MODULATING
FUNCTIONS

75

$$\phi_n[0,T] = \frac{d \phi_n[0,T]}{dt} = 0$$

The frequency w_n has to be chosen carefully to ensure that the boundary conditions are satisfied, ie $w_n = n\pi/T$.

Loeb and Cahen [58,59] have developed the modulating function method further by extending the modulating function end point conditions,ie

$$\phi[0,T]=0$$
$$\frac{d}{dt}\phi[0,T]=0$$
$$\vdots$$
$$\overset{\wedge}{\phi}[0,T]=0$$

$$(2.13)$$

Where $\overset{\wedge}{\phi}$ is the nth derivative of ϕ

This allows the parameters of any order of system to be determined, where Shinbrot was limited to 2nd order because his modulating function end point condition was limited to the first derivative. A solution for the nth order differential equation can now be obtained.

$$(-1)^n a_n \int_0^T \phi(t)\, y(t)\, dt + (-1)^{n-1} a_{n-1} \int_0^T \phi^{n-1}(t)\, y(t)\, dt + \cdots$$

$$\cdots + a_0 \int_0^T \phi(t)\, y(t)\, dt = \int_0^T \phi(t)\, u(t)\, dt \qquad (2.14)$$

A set of simultaneous equations can be formed using successive samples of u and y, eg $[t_1$ to $t_2]$, $[t_3$ to $t_4]$, $[t_5$ to $t_6]$ or different modulating functions ϕ_1, ϕ_2, ϕ_3 In this case,

$$
\begin{bmatrix}
(-1)^n \int_0^T \phi_1^n y\, dt & \cdots & (-1)^0 \int_0^T \phi_1 y\, dt \\
\vdots & & \vdots \\
\vdots & & \vdots \\
(-1)^n \int_0^T \phi_m^n y\, dt & \cdots & (-1)^0 \int_0^T \phi_m y\, dt
\end{bmatrix}
\begin{bmatrix}
a_n \\
\vdots \\
\vdots \\
a_0
\end{bmatrix}
=
\begin{bmatrix}
\int_0^T \phi_1 u\, dt \\
\vdots \\
\vdots \\
\int_0^T \phi_m u\, dt
\end{bmatrix}
$$

In order that matrix invertion can be performed the modulating functions should be orthogonal [58], ie

$$\int_{T_1}^{T_2} \phi_1(t)\, \phi_2(t)\, dt = 0$$

Loeb has considered the possibility of a modulating function based on the exponential function described by Schwartz [60],

$$\phi(t) = A(t) \exp\left(\frac{B(t)}{(t-a)(b-t)}\right)$$

This satisfies the boundary condition (2.13) for t<a and t>b where a and b are arbitrary. The functions A(t) and B(t) and their derivatives are continuous for a<t<b.

Takaya [61], following on from Loeb's work has proposed the use of modulating functions based on Hermite functions. These satisfy the boundary condition (2.13) and indeed prove to be most versatile and useful functions. These modulating functions can be applied to any order of system and are the modulating functions used in this research work. The Hermite function is described by,

$$H_n(r) = (-1)^n \exp\left(\frac{r^2}{2}\right) \frac{d^n \exp\left(-\frac{r^2}{2}\right)}{d r^n}$$

Hermite functions do not alone satisfy the orthogonal property of the modulating functions. To achieve this criteria the Hermite function must have a weighting function defined by,

$$\int_{-\infty}^{\infty} \exp(-r^2/2)H_n(r)H_m(r)dr$$

$$= n!\sqrt{2\pi} \quad \text{if } m=n$$

$$= 0 \text{ otherwise}$$

Thus Takaya proposed the use of weighted Hermite functions as modulating functions, where the higher derivatives of the modulating functions are obtained by using higher order Hermite functions,

$$\hat{\phi}(r) = (-1)^n \frac{\exp\left(-r^2/2\right) H_n(r)}{\sqrt{2\pi}}$$

Substituting (t-T/2) for r results in modulating functions that satisfy equation (2.13). The first four Hermite modulating functions are shown on figure 2.18

Maletinsky [62,63] has proposed a modulating function method based on spline type modulating functions. these modulating functions are based on repeated integrations of Dirac delta functions. Maletinsky obtains the estimates of the parameters using a recursive least squares technique rather than the one shot simultaneous equations used by Shinbrot and Loeb. This is achieved by forming an equation error from the modulated system differential equation (fig 2.19). Repeated equation
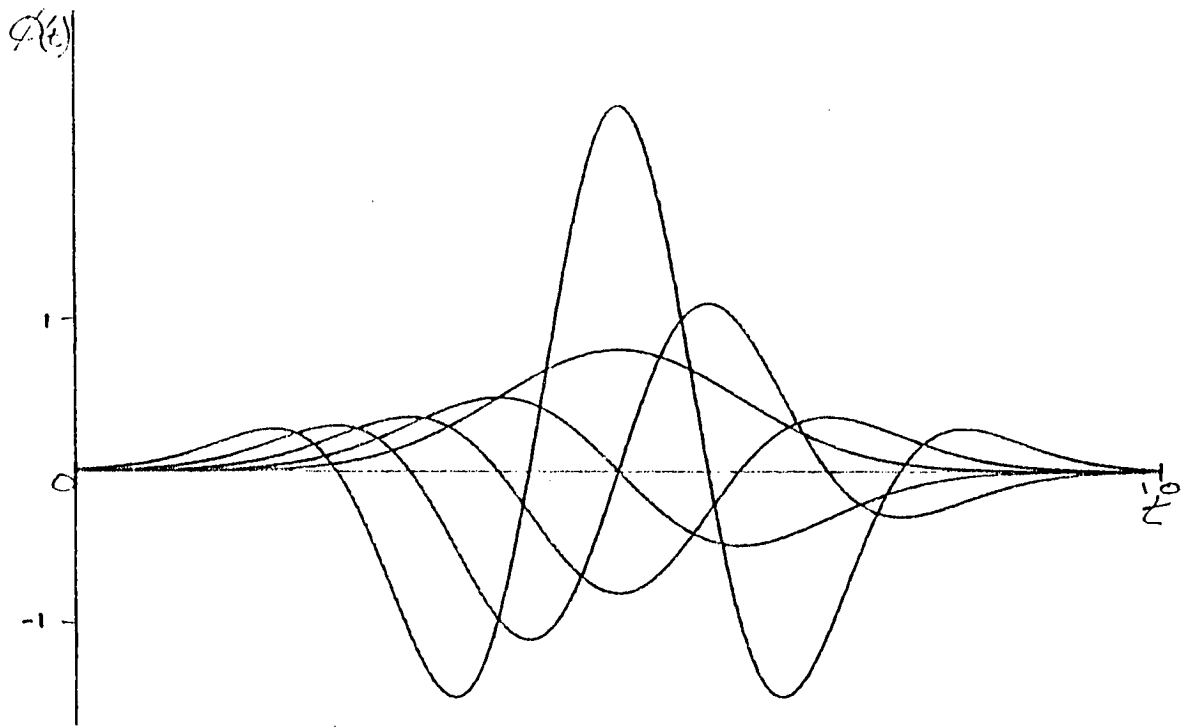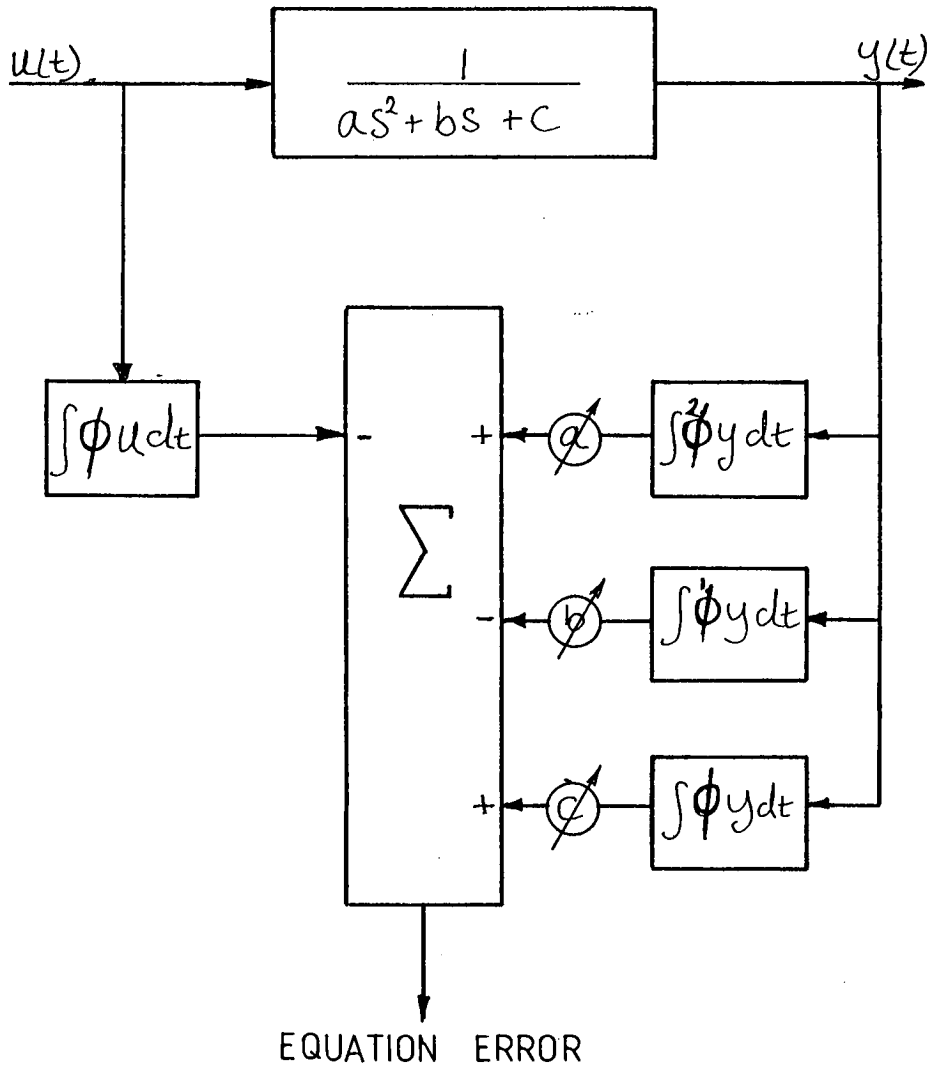
79

FIGURE 2-18
HERMITE MODULATING FUNCTIONS

FIGURE   2 -19
CUBIC SPLINE MODULATING
FUNCTION   METHOD

errors are formed using different samples of the input/output signals, this is then solved for the parameters using the least squares technique. Maletinsky recognises the problem of bias due to noise and suggests the use of the Instrumental Variables technique to overcome this problem. To ensure accuracy of the parameter estimates the input signal should excite the whole bandwidth of the system and the order of the system should also be known.

The spline type modulating functions proposed by Maletinsky are based on repeated integrations of the Dirac Delta function, (figure 2.20). For a known order of system, n, a weighted sum of Dirac delta functions can be formed from the following equation

$$\phi_n(t) = \sum_{i=0}^{n} (-1)^i C_n^i \, \delta(iT - t)$$

Where $C_n^i = \dfrac{n!}{i!(n-i)}$ and $C_n^n = C_n^0 = 1$

Where n is the order of the system

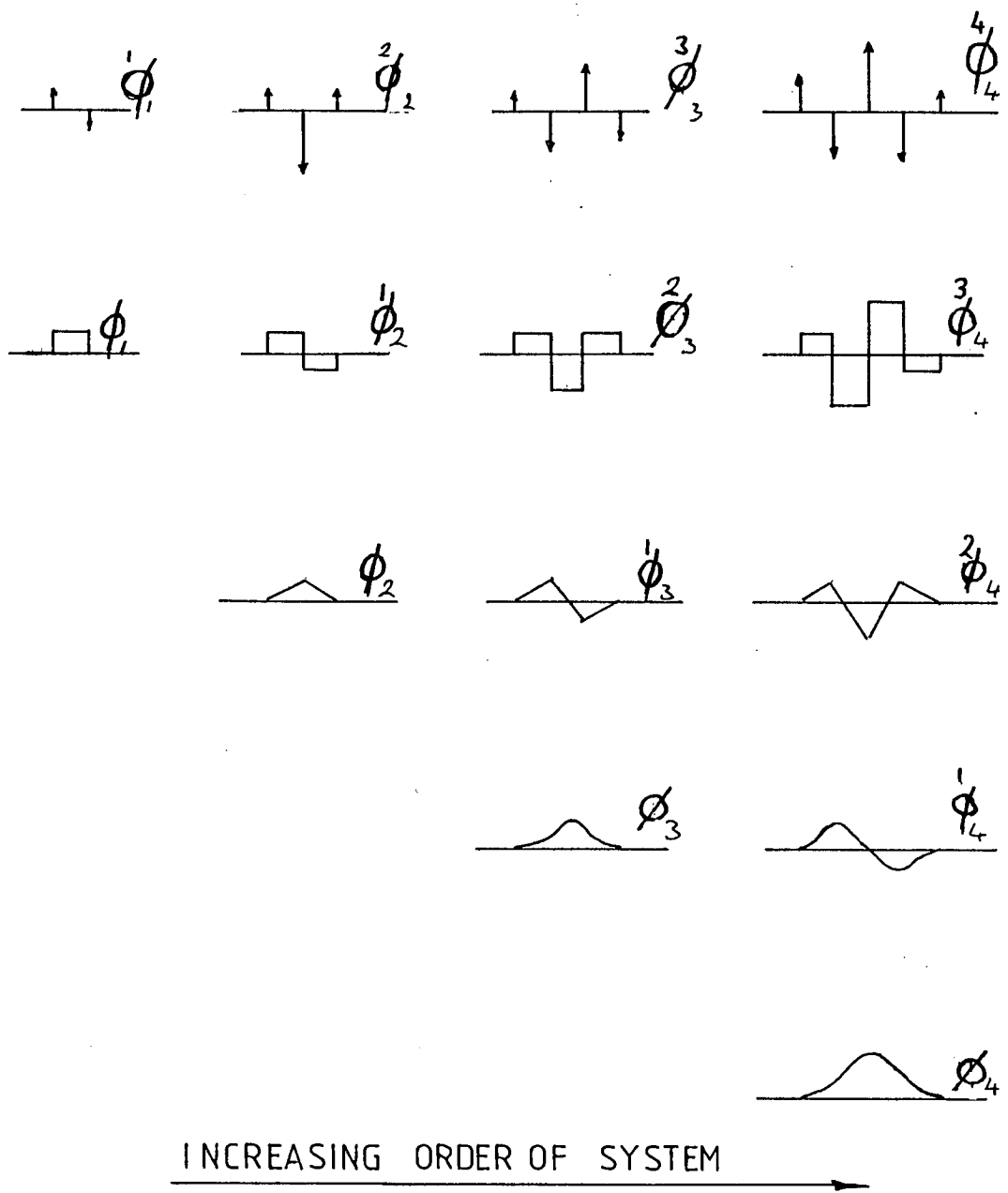The integral modulating function equation (2.14) can now be restated using.

INCREASING ORDER OF SYSTEM

FIGURE 2-20
SPLINE TYPE MODULATING FUNCTIONS

$$\int_0^{nT} \phi_n(t) \, x(t) \, dt = \sum_{i=0}^{n} (-1)^i \left(c_n^i \, x(iT) \right)$$

Belorusets [64] uses a similar approach to Shinbrot in obtaining modulating functions. For a system with two unknown parameters the sample window is split into two sub-intervals and modulating functions are formed on each of these

$$\phi_1(t) = \sin^2(2\pi t/T) \quad t \in [0, T/2]$$

$$\phi_2(t) \sin^2[(2\pi(t-T/2))/T] \quad t \in [T/2, T]$$

$\phi_2(t)$ is obtained by translating $\phi_1(t)$ into the sub-interval T/2 to T. Shortening the sample window makes the set of linear simultaneous equations ill conditioned therefore another form of modulating function was considered by Belorusets. A polynomial modulating function was proposed, figure 2-21

$$\phi_1(t) = t^2(t-T)^2$$

$$\phi_2(t) = t^3(t-T)^2 \quad t \in [0, T]$$

The problem with polynomial modulating functions is that they only satisfy the boundary conditions (2.12) for the modulating function and its first derivative, as with Shinbrot.

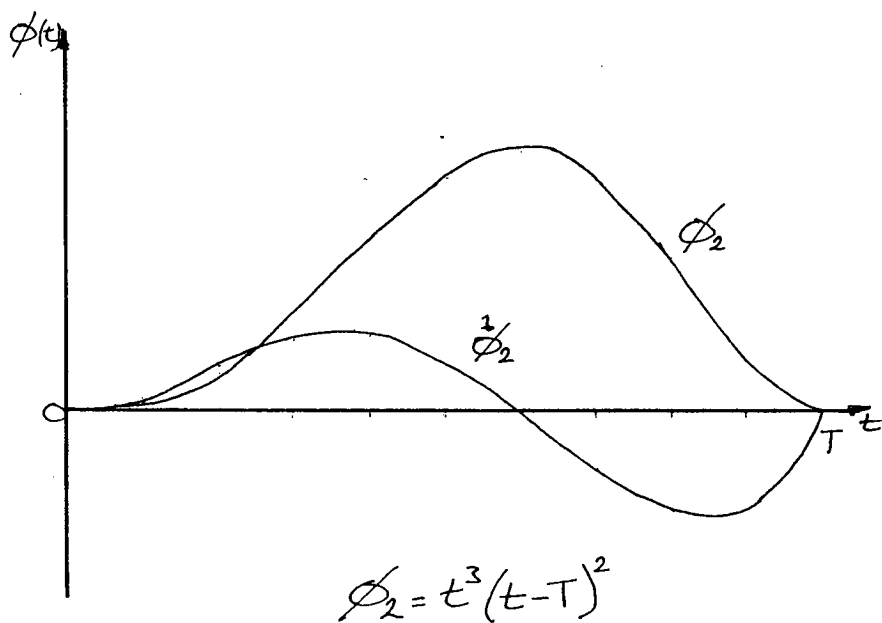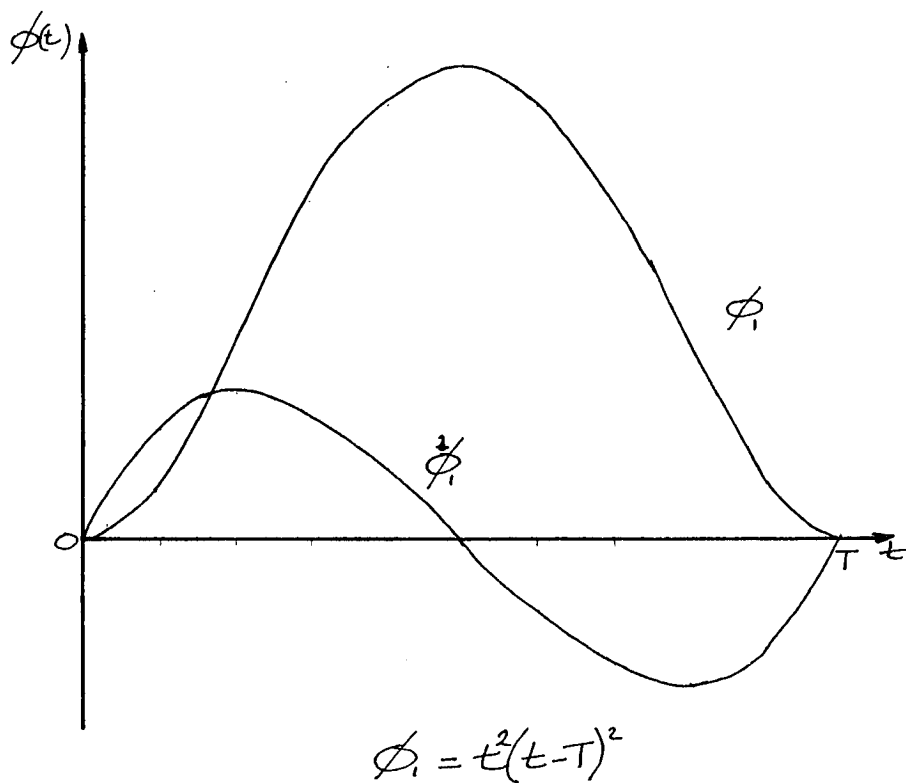$$\phi_1 = t^2(t-T)^2$$

$$\phi_2 = t^3(t-T)^2$$

FIGURE 2-21

POLYNOMIAL MODULATING FUNCTIONS

Shinbrot suggested the use of carefully chosen sinusoids as modulating functions and Pearson [65] has extended this proposal by using the Fast Fourier Transform (FFT) to compute the modulating functions. Pearson's modulating functions are a summation of cosines for the odd modulating functions, and a series of sine functions for the even ordered modulating functions.

$$\phi_{C,K}(t) = \sum_{j=0}^{\frac{n+1}{2}-1} a_{k,j} \cos(k+j)\omega_o t, \quad \omega_o = \frac{2\pi}{T}$$

$$\phi_{S,K}(t) = \sum_{j=0}^{\frac{n}{2}-1} b_{k,j} \sin(k+j)\omega_o t, \quad \omega_o = \frac{2\pi}{T}$$

where $a_{k,j}$ and $b_{k,j}$ are chosen so that the end conditions 2.22 of the modulating function are met.

Then $\phi_{C,K}^{i}(0) = \phi_{C,K}^{i}(T) = 0, \quad i = 0, 2, 4, \cdots (2(\frac{n-1}{2}))$

And $\phi_{S,K}^{i}(0) = \phi_{S,K}^{i}(T) = 0, \quad i = 1, 3, 5 \cdots (2(\frac{n}{2})-1)$

Using these modulating functions enabled Pearson to produce a set of linear equations which are solved by the least squares technique.

Pearson has extended this work to periodically time varying models, and also to non-linear systems [66]. Pearson claims that a large computational saving is afforded by using this method, because of the use of the FFT algorithm. This eliminates the need to perform the numerical integration of the input and output signals.

The Poisson Moment functional approach has been explored by various authors, Loeb has used modulating functions derived by Shwartz which are based on Poisson functions [67]. More recently Saha [68] and Fairman [69] have developed modulating function techniques based on Poisson functions.

The basis of the Poisson functional approach is that the higher derivatives of a signal can be expressed in terms of itself through the use of Poisson functions. This is achieved by modelling the input/output signals by an exponentially weighted series of impulses,

$$f(t) = \sum_{i=0}^{\infty} M_i \{f(t)\} e^{-c(t-t_o)} \delta^{(i)}(t-t_o)$$

$$M_i(f(t)) = \int_0^t f(t) \, P_i(t_0 - t_i) \, dt$$

$$P_i = \frac{t_i}{i!} e^{-ct}$$

These models of the signal can be substituted in equation 2.7. This gives a weighted series of derivatives of the impulse distribution. The parameters are now independent of time and derivatives of the input/output signals. Equating like orders of derivatives of the impulse distributions results in linear algebraic equations in the unknown parameters.

Some of the previously mentioned methods utilise the least squares technique to arrive at a parameter estimate. While this has certain advantages, namely it has an averaging effect on the parameter estimates, it introduces a further complication, and increases the computational burden.

2-4-1 Unified Approach

Although the MFM as an approach to system identification seems to stand apart from other methods, it is possible to link it to most of the continuous methods which have been discussed here. This approach to parameter estimation can be classed as the weighting function

method (WFM). In the weighting function method equation (2.7) is multiplied by a weighting function and a scalar product formed with each term. The scalar product is defined as,

$$\int_0^t \emptyset(t)y(t)dt$$

This allows a set of N linear equations to be solved for the N unknown parameters. The weighting functions can be of several different forms, eg Shinbrot's and Takaya modulating functions. Mironovskii and Vudovich [70] have shown that Walsh functions are a type of weighting function.

Least squares and instrumental variables can be considered as a WFM's [70]. These methods are based on the minimisation of a cost function. A residual can be formed from equation (2.7).

$$e(t) = \overset{n}{y}(t) + a_{n-1}\overset{n-1}{y}(t) + \cdots + a_0 y(t) - b_{m-1}\overset{m-1}{u}(t) - \cdots - b_0 u(t) \qquad (2.15)$$

and the cost function is, $J(e) = \int_0^T e^2(t) \, dt$

and is a minimum when, $\dfrac{dJ}{d\hat{\theta}} = 0$ $\qquad (2.16)$

substituting equation (2.15) into equation (2.16) gives,

$$\int_0^T [\dot{y}(t) + a_n \dot{y}(t) + \cdots + a_1 y(t) - b_0 u(t)] f_i(t) \, dt = 0$$

$$f_i(t) = \begin{cases} y(t) & i = \overline{1,n} \\ -u(t) & i = \overline{n+1,N} \end{cases}$$

This is the form of the WFM with weighting functions $f_i(t)$. The least squares method will produce biased results by using the input-output signals as weighting functions if there exists a correlation between them and the noise, as discussed earlier. The instrumental variable method overcomes this by using a weighting function not correlated with the noise.

Thus it can be seen that most of the continuous methods of system identification belong to a set of methods called the Weighting Function Method, of which the Modulating Function Method is a sub-set.

2-5 Fault Detection

Fault detection has developed considerably, in the last twenty years or so, from the standard techniques of sensor output monitoring to the sophisticated parameter and state estimation techniques employed today. The standard techniques employed to detect faults in industrial plant involve checking the output level of various sensors around the system. These sensors usually measure parameters such as temperature, pressure

and velocity. When a parameter exceeds a certain level, an alarm is signalled and the system operators will take the appropriate action to locate and rectify the fault.

This standard approach has several disadvantages, it only gives an indication of the system's health at one moment in time, and does not give any indication that the system may fail in the near future. Also this kind of static test gives no indication if the system's dynamic response is not as it should be. It may not respond to disturbances as quickly as it should.

With the cost of maintenance increasing all the time and large amounts of capital being tied up in providing backup systems and a spares inventory, it makes sense to be able to plan the maintenance amd repair of a system, thus reducing the maintenance and capital costs. In addition there is an ever increasing emphasis on safety plus the need to eliminate unexpected catastrophic failures. The standard go/no go tests on measurements of system parameters do not provide the necessary information to accomplish this, and as a result complex dynamic tests on the system have been developed to try to achieve these aims.

Dynamic testing has several advantages over conventional methods of fault detection, the tests are

fast and the results obtained quickly - essential if further damage is to be avoided. The tests are standardised and under the control and supervision of a computer, therefore operator intervention and skills are usually not required. Usually, for dynamic tests, only two access points are required, the input and output, thus there is no need to provide extra access points for test measurements, which might reduce reliability. The number and type of dynamic tests available is large, figure (2.22) shows most of these. These tests can be split into two sections as in system identification, Non-Parametric and Parametric. Both will be considered here.

There appears to be two main influences in the development of fault detection methods. During the early sixties NASA's space flight programme became increasingly complicated and the number of equipment failures per mission increased. This led to a drive into the area of fault detection and isolation to overcome this problem. Fault detection using transfer functions is just one of the areas which were explored. The techniques employed were relatively straightforward and simple, and remained that way until the advent of the small low cost micro-computer.

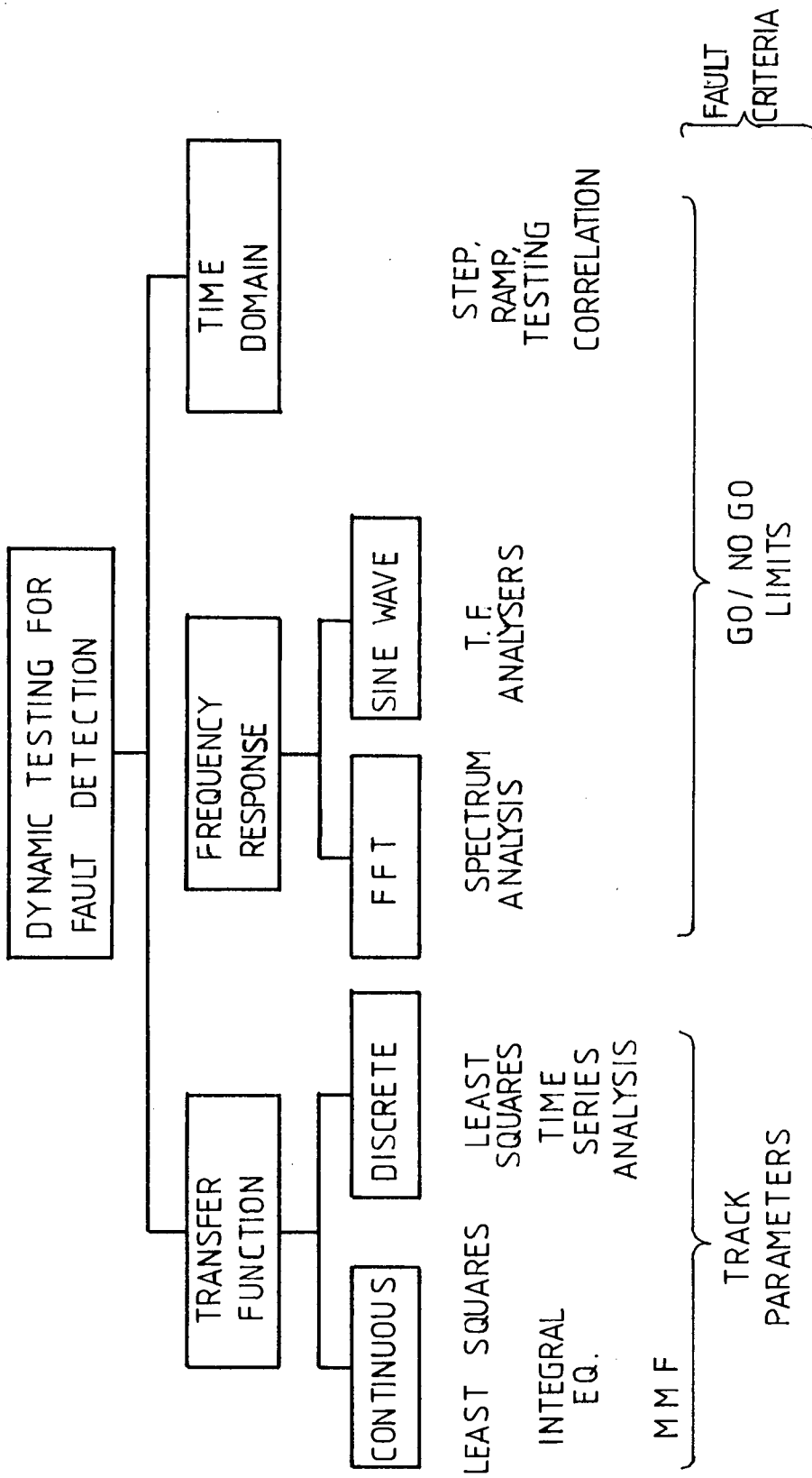The micro-computer has had a profound effect on fault

FIGURE 2-22

DYNAMIC TESTING FOR FAULT DETECTION

detection. It allowed complex algorithms with large data processing requirements to be implemented on-line. This increased capability has led to a reduction in the time between a fault occurring and it being detected

2-5-1 Fault Detection Based on Non-Parametric Methods

Non-Parametric fault detection like non-parametric identification utilises a graphical representation of the system. For fault detection this representation or signature of the system is compared against a standard signature and a decision made on whether a fault exists or not. The graphical representation can be either a time domain model such as step, ramp or impulse response, or a frequency domain representation, such as frequency response or power spectra.

On establishing that a fault in the system exists a diagnosis of that fault is sought. In non-parametric fault detection methods this is nearly always done using some kind of pattern recognition technique.

Dynamic time domain testing is usually associated with step, ramp and correlation testing. Correlation testing has been used extensively to obtain the system impulse response, and a few authors have extended this to fault detection.

94

Steps and ramps as test signals for fault detection have not received as much attention. This may, in part be due to the nature of the test signal. It is not easy to perform step inputs and not disturb the process, or send it into a non-linear region. If a small step is used many tests must be performed to average out the noise. This would require long test times, thus increasing the time to detect a fault condition.

Impulse, step and ramp testing is usually implemented by placing tolerance gates on the responses. For example a gate may be placed on the maximum overshoot or peak impulse response, fig (2.23). These tolerances or gates are usually chosen at the system design stage, using simulation knowledge [71]. Payne [72] has used this technique to test batches of servo-mechanisms, in conjunction with frequency response testing.

The impulse response, obtained by using correlation techniques is perhaps the most popular time domain test procedure. Towill [73] compares the measured impulse response with a standard impulse response, the measured response is subtracted from the standard response and the resulting function examined to see if it exceeds a threshold (fig. (2.24)). This threshold is chosen to allow some variation between the test and the standard responses. If the variation between the two
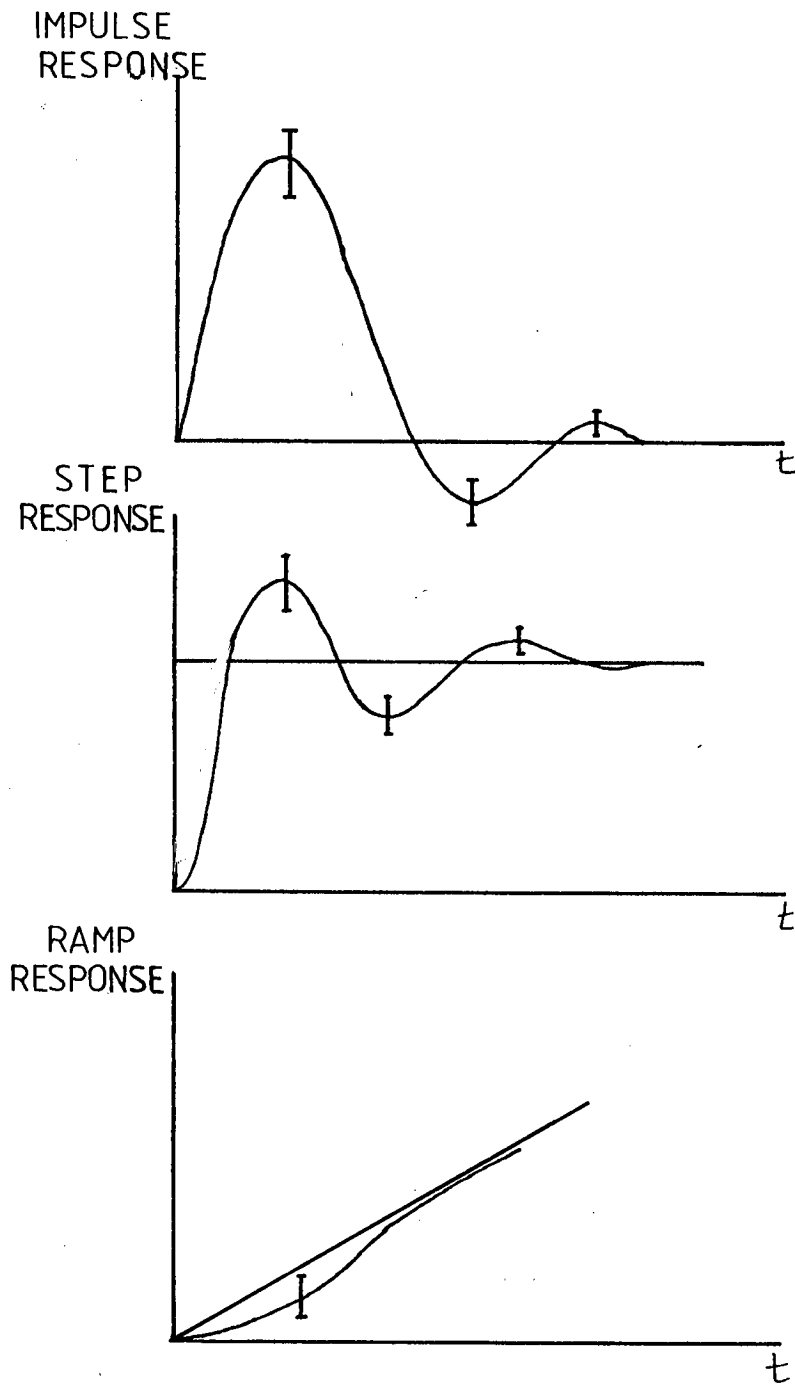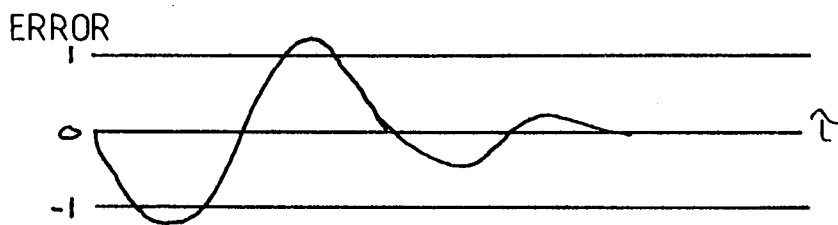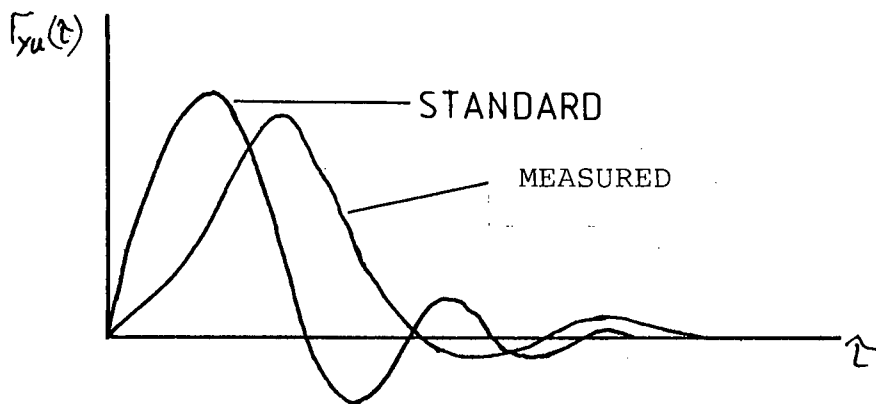
FIGURE 2-23
FAULT TOLERANCE GATES

96

FIGURE 2-24
ERROR BETWEEN STANDARD AND MEASURED
CORRELATION FUNCTIONS
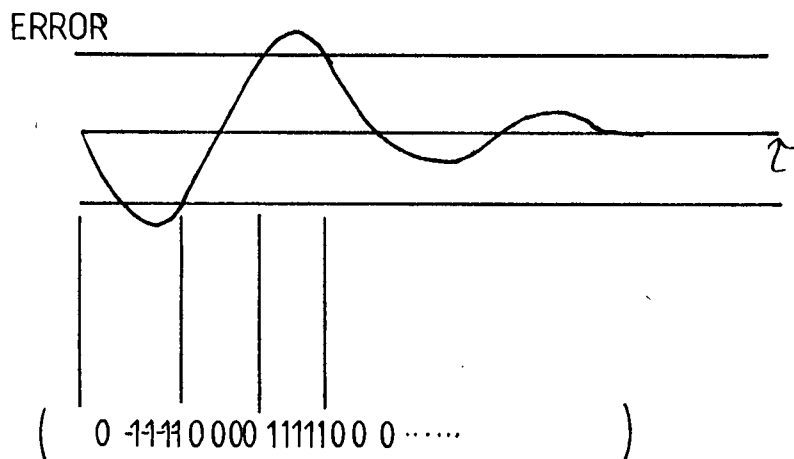

(0 -1-1-11 0 000 111110 0 ⋯⋯)

FIGURE 2-25
FAULT VECTOR

becomes large, the threshold is exceeded and a fault is declared. On establishing that a fault exists, pattern recognition techniques are then used to locate the fault. This is achieved by producing a fault vector (fig (2.25)) which describes, in binary form, whether the threshold has been exceeded at each particular instant on the impulse response. The fault is located in the system by comparing this fault vector with a table of fault vectors which are related to particular types of failure. Once the pattern in the test fault vector has been matched with that in the table, the location and type of fault will then be known.

Towill carried out tests on complex electro-hydraulic servo systems. To set up the table of fault vectors, simulation studies were carried out using a mathematical model of the system, a parameter in the model was altered and the effect on the impulse was noted. This was used to construct a fault vector for that particular fault. Alternatively the fault vector can be constructed by physically changing a component in the system to introduce a fault and again noting the effect on the system.

Garzia [74] in his comprehensive review of the fault detection methods available in the early 1970´s reports on a time domain method that is similar to Towill´s. He

determines the auto and cross spectra of a system and from this calculates the transfer function. Then the impulse response is calculated and compared with a standard impulse response. A measure of how well these two responses match each other is made using the chi-squared goodness of fit criteria,

$$\chi^2 = \sum_{i=1}^{n} \frac{\left[h_T(t_i) - h_s(t_i) \right]^2}{h_s(t_i)}$$

where $h_s(t_i)$ is the ith point on the standard. From this it is possible to determine if the system is acceptable. If not, the fault can be located using a similar approach to that used by Towill.

Frequency domain testing is the most popular of the non-parametric fault detection methods. Several techniques have been used, and most utilise some part or parts of the system's frequency response. This involves checking the response of a system at a particular frequency or set of frequencies, and if these do not prove satisfactory, a diagnosis of the fault is made.

Allen [75] proposed that the system gain and phase could be monitored on-line. This would be done by injecting test signals at frequencies above and below

the normal operating range of the system under test. This would not disturb the system's normal operation and hence the system could be continuously monitored, resulting in a fast fault detection method. The output of the system is bandpass filtered for each of the test frequencies and changes in the gain and phase monitored. On-line fault detection using frequency response is not usual. This method offers the possibility of detection of incipient failures [75], but fault diagnosis is not possible. The test frequencies do not give sufficient information about the system to allow this.

A more usual procedure is to inject several sine waves at critical frequencies into the system, and measure the gain and phase at these frequencies. From this a fault dictionary can be set up to diagnose any faults discovered.

Seshu [76] injects test frequencies at the break points on the frequency response curve for a system. Only the magnitude response is measured and if the gain exceeds a bound then a fault is declared. The bounds on the gain are calculated by using the worst case acceptable for the maximum and minimum gain. From the test data a gain signature is obtained, this is equivalent to the fault vector in the time domain's case, and as before, a fault table is used to determine

the cause of failure. A similar approach is taken by Hsieh [77], complex systems are split into modules, reducing the burden of generating a fault signature capable of pinpointing the fault, to that of producing a fault signature that will indicate which module is faulty.

Towill has extended his time domain fault detection methods to the frequency domain [78]. He proposes that three test frequencies are sufficient to test a dynamic system. These test frequencies are identified by generating a range of unacceptable systems and determining the frequencies at which these unacceptable systems are most easily recognised. A diagnosis of the fault is achieved by relating the change in the frequency response through the Bode approximation to that part of the system responsible for the fault.

Sriyananda has proposed a voting technique for the frequency response [79] methods. This is a similar approach to that used in the time domain [73]. Here both the gain and the phase are measured to create the fault vector.

Varghese [80] has proposed a method for determining the best test features, which will classify a system good or faulty with the minimum computation effort, ie.a

minimum number of test frequencies.

The basic approach to frequency response testing is as follows:-

1. Model the system and/or use test data to establish the frequencies at which the test signals have to be injected into the system. Also set the limits on the normal gain and phase of the system at these limits.

2. Set up the fault dictionary by simulating faults on the model or creating faults on the real system.

3. To locate faults in the system under test, match the fault vector with an element from the fault dictionary.

This is the basic approach used in the previous methods, it has also been used by Garzia [74], Polovko [81] and Himmelblau [82] in a similar manner.

The methods described above suffer from several drawbacks. It is difficult to set the limits on the gain and phase, and thus change may be difficult to detect, Himmelblau [82]. More complex characteristics such as slope may be required. Also the method fails if the fault vector is not in the fault dictionary,

allowing no diagnosis of the fault. This may be due to multiple faults. If there is a total failure no output is measured. In addition these methods do not monitor the system continuously, with the exception of the method proposed by Allen [75], and are therefore unlikely to detect incipient failure.

2-5-2 Fault detection based on Parametric methods

The non-parametric methods described in the previous section offer a technically simple method of detecting and locating faults in systems. A number of problems with this method have been noted, but perhaps the biggest restriction on the method is the inability to predict faults. Only those faults which have occurred, are detected and located. An on-line method of failure prediction is needed. One method which may achieve this is to monitor continuously the parameters of the system model for any change which may be due to an incipient failure of the system. The parameters of the model are obtained by one of the parametric system identification methods described earlier in this chapter. If a failure in the system is detected, it may be possible to relate deviations in the model parameters to failures in actual system components through the physical relations that formed the model. For instance, failure in a heat exchanger due to fouling could be detected by a change in the flow rate and temperature differentials which

would be reflected in a change in the relevant model coefficients. For this reason it is thought that continuous models are more appropriate for fault detection than discrete models [83], as the physical parameters such as length, mass, velocity are all combined in the continuous model parameters. Thus if identification is carried out on - line the fault detection procedure can be carried out simultaneously, so that the health of the system can be continuously monitored, and any incipient faults detected and located quickly, before any damage is done. Figure 2.26 shows a fault detection and isolation system. In addition to monitoring the parameters of the model it may be possible to monitor the character of the system noise, or look for change in the bias of the parameter estimates, also state estimation techniques may be used to monitor system performance.

In general parameter estimation is thought to be more sensitive to faults, and to give a better indication of the system performance than state estimation [83]. Also a trade off exists between noise and speed of detection. If a system is noisy, many averages must be performed on the parameter estimates if false alarms are to be avoided. This obviously reduces the speed of detection.
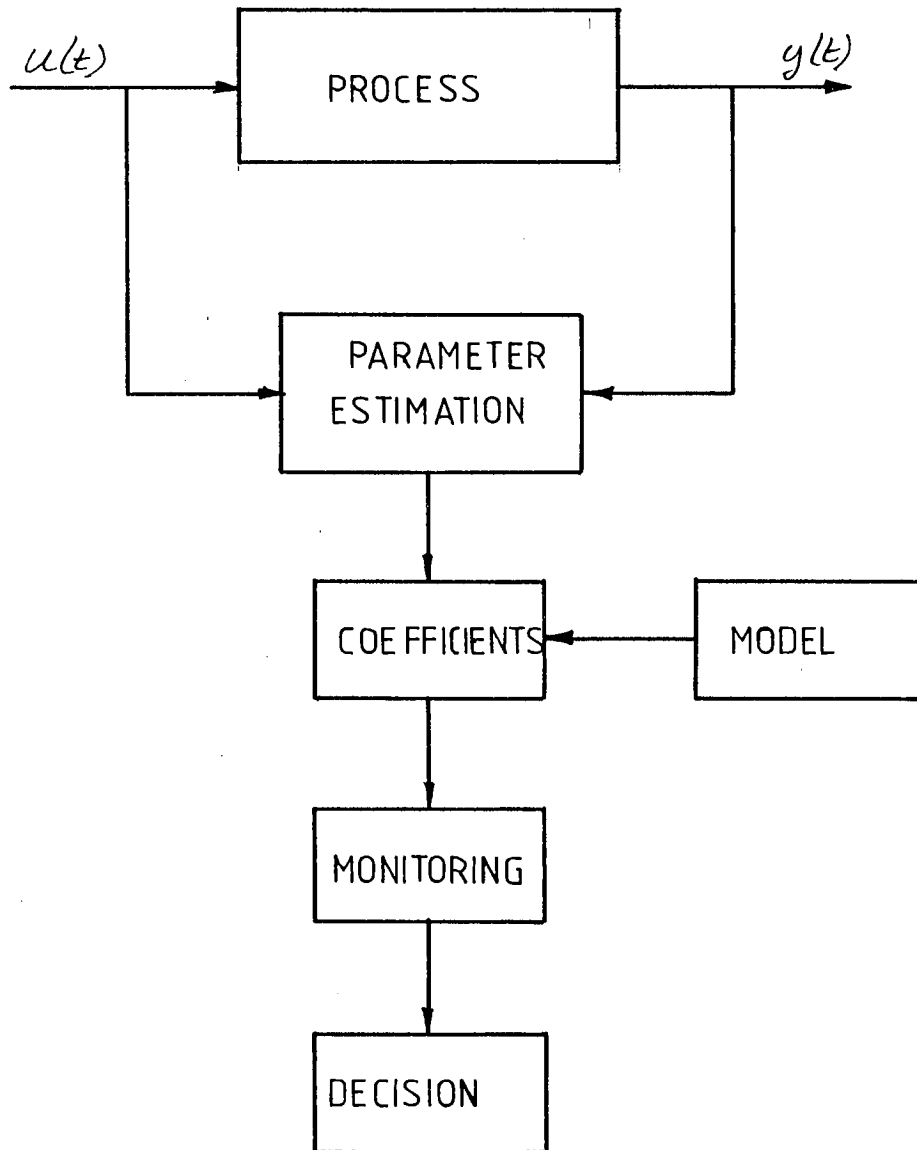
Valstar [84] proposed a method of checking aircraft

FIGURE   2 - 26
FAULT DETECTION PROCESS

dynamic systems during flight. This method tracked the parameters of the continuous transfer function. It was hoped that a gradual deterioration in the system could be detected using this method, and thus provide an early warning for the pilot. Valstar recognised the effect of modelling errors on the accuracy of the detection system, and suggested an optimisation method to reduce the error to a minimum.

Garzia [74] has reviewed many different methods of fault detection and fault isolation. A method similar to that of Valstar is proposed for parametric fault detection. Parametric fault detection did not develop much until the mid-seventies when low cost computers became available to perform the large computational requirements of most of the parameter estimation methods.

Although the continuous model of a system is thought to be more useful than the discrete model for fault detection, there is little literature published on this technique. Slightly more work has been carried out on discrete techniques. Pau [85] has used the discrete approach, as has Willsky [86] who has employed Kalman Filtering techniques, Mehra [87] used a time series forecasting method to try to predict a system going "out of limit". Park [88] used the Kalman Filter method in a

fault detection and isolation system applied to a chemical reaction tank. Various faults were created to determine the practicability of the method. It was found when the model of the system was accurate and there was no measurement noise, good fault detection was achieved. However if an accurate model cannot be achieved, and the measurement noise is significant, fault detection will be seriously impaired.

Duhamel [89] has reviewed many techniques for fault detection in analogue circuits. For parametric fault detection the least squares technique is used. Duhamel notes several problems with this approach. If several components are faulty, fault isolation is not possible. Variations of components within their tolerance band can obscure other faulty components, or cause false alarms. In addition the advantage of this method, ie two test points and only two signals to be measured, can be outweighed by the computational burden of parameter estimation, and modelling problems if the system is large.

In one of the few studies using continuous models Isermann [83] reviews the possible fault detection methods. Parametric fault detection is carried out using continuous least squares methods developed in Young [40]. Isermann notes that a need exists for a

parameter estimation method which will estimate the parameters of a continuous model for higher orders, more accurately than is possible at present. Two case studies are presented. In the first, an electrically driven centrifugal pump is monitored. A complex model was developed. Isermann suggested that to measure the transfer function accurately it should be split into sub-components so that low order elements are measured. In the second case study, detection of leaks in pipelines was examined. Isermann concluded that the use of dynamic models will enable smaller leaks to be detected.

Isermann suggests that fault detection is more suitable for abrupt faults than detecting slowly developing faults. In addition accurate models and parameter estimates are needed if fault detection is to be reliable and not generate false alarms.

Goedecke [90] has used a continuous technique in an experimental fault detection system. This was developed to detect faults in a tubular heat exhanger. The heat exchanger was modelled by a continuous differential equation, and parameter estimates were obtained using least squares methods outlined by Young [40]. Artificial faults were created, and the method detected a change in the model parameters.

CHAPTER THREE

MODIFIED MODULATING FUNCTION METHOD

3-1 Modified Modulating Function Method

In this chapter the theory of the modified modulating function method of linear system identification will be developed. The Modified Modulating Function Method (MMFM) is a development of the Modulating Function Method (MFM) reviewed in chapter 2.

When using parameter estimation techniques in a fault detection and identification environment it is better to work with continuous time models rather than sampled data models. Continuous models are closer to the physical laws governing the operation of the system, and hence it will be easier to connect parameter deviations in the model to a particular cause or fault in the system. The MMFM like the Modulating Function Method estimates the parameters of the small signal differential equation of the continuous model of the system, but has the advantage over other methods in that it is a two stage method. An initial correlation stage is used to check data quality before further signal processing is attempted. A visual indication of measured correlation function can be included to provide

operators with a diagnostic check on the performance of the measurement system.

Instead of using the system input/output signals, as the Modulating Function Method uses, the modified method uses the auto correlation of the system input signal in place of the input signal and the cross correlation function of the input and output signals in place of the output signal. This has the advantage in that it will be possible to eliminate bad data. If no cross correlation can be obtained, this may indicate a fault in the test equipment or a wrong assumption in the method of identification. Hence this makes this method suitable for on-line use in remote locations. Properly used the correlation function also gives clues as to what type of model to use.

The MMFM is similar to the approach taken in the correlation analysis with least squares parameter estimation method [31]. In that method the auto and cross correlations are formed and then least squares estimation applied to them.

The shape of the cross correlation function in relation to the input auto correlation function can give a strong indication as to the order of the system. First and second order systems are easy to distinguish

but higher order systems can present a problem. As the system order increases it is often possible to obtain similar responses from a second order model plus a time delay.

Time delay is a serious problem that badly affects attempts to identify the parameters of the system. In many identification algorithms that use the system input/output signals the detection of time delay is a considerable problem. Visual inspection of the measured cross correlation functions can be used to detect pure time delay terms, but as noted above, confusion with responses generated by multiple lag systems is a possibility.

The input and output signals of a single-input, single-output system represented by the differential equation,

$$a_n \hat{y}(t) + a_{n-1} \hat{y}(t) + \cdots + a_0 y(t) = b_m \hat{u}(t) + \cdots + u(t) \qquad (3.1)$$

are related through the convolution integral,

$$y(\tau) = \int_{-\infty}^{\infty} h(\tau - t) u(t) \, dt$$

or $\quad$ $y(\tau)=h(t)*u(t)$

where * indicates convolution.

or through the transfer function in the s-plane,

$$Y(s)=H(s)U(s)$$

where U(s) is the Laplace Transform of u(t).

A similar convolution and transfer function relation exists for the cross and auto correlation functions,

$$r_{yu}(\tau)=\int_{-\infty}^{\infty}h(\tau-t)r_{uu}(t)dt$$

or $\quad$ $r_{yu}(\tau)=h(t)*r_{uu}(t)$

and $\quad$ $R_{yu}(s)=H(s)R_{uu}(s)$

Where $r_{uu}(\tau)$ is the auto correlation of the input signal, and $r_{yu}(\tau)$ is the cross correlation of the input and output signals. A two-sided Laplace Transform is required in this case.

Thus the system differental equation (3.1) may be

re-formulated,

$$a_n \overset{n}{r}_{yu}(\tau) + a_{n-1} \overset{n-1}{r}_{yu}(\tau) + \cdots + a_o\, r_{yu}(\tau) =$$

$$b_M \overset{M}{r}_{uu}(\tau) + \cdots + r_{uu}(\tau) \qquad (3.2)$$

As in the modulating function method, equation (3.2) is multiplied by the modulating function Ø(t), and integrated by parts to form an equation where no reference is made to the differentials of the auto and cross correlation functions,

$$(-1)^n a_n \int_0^T \overset{n}{\phi}(\tau)\, r_{yu}(\tau)\, d\tau + \cdots + a_o \int_0^T \phi(\tau)\, r_{yu}(\tau)\, d\tau =$$

$$(-1)^M b_M \int_0^T \overset{M}{\phi}(\tau)\, r_{yu}(\tau)\, d\tau + \cdots + \int_0^T \phi(\tau)\, r_{uu}(\tau)\, d\tau \qquad (3.3)$$

Where $\overset{n}{\phi}(\tau)$ is the nth derivative of the modulating function. Figure 3.1 shows a diagramatic representation of the MMFM.

In the modulating function method κ ($\kappa = n + m + 1$) parameters are found by κ samples of the input/output signals. For the modified method, successive samples of the auto and cross correlation function are not

113
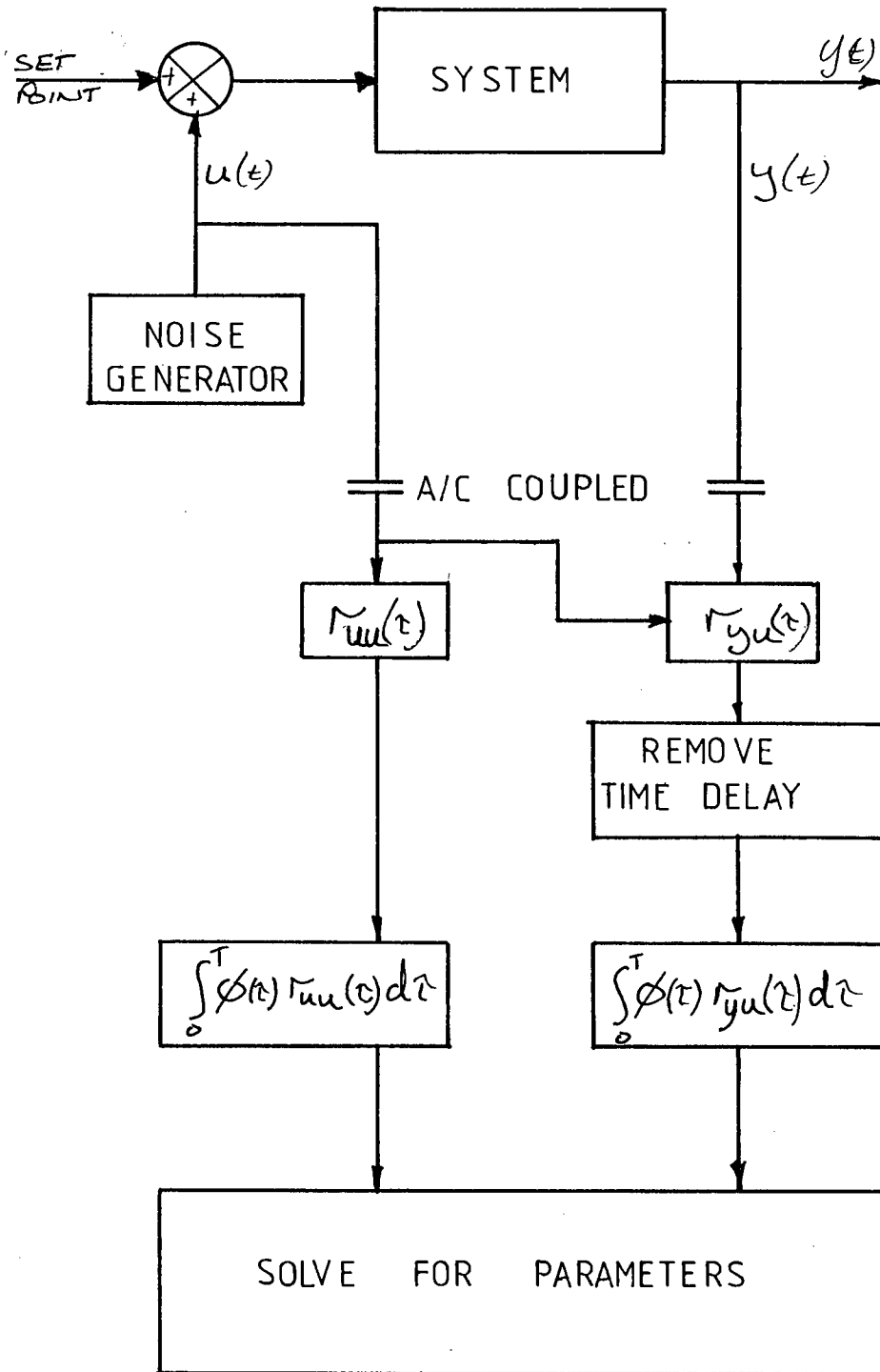
FIGURE 3-1

MODIFIED MODULATING FUNCTION METHOD

advisable, because the auto and cross correlation functions will not change significantly with successive samples. Any equations formed with them will be ill conditioned and therefore poor estimates of the parameters will be obtained. Differences between successive samples of the auto and cross correlations will be due to variance caused by short integration times in the computation of the functions and not a system effect. It is advisable then to form the set of simultaneous equations by using different modulating functions.

$$
\begin{bmatrix}
(-1)^n \int_0^T \overset{n}{\phi_1}(t)\, r_{yu}(t)\, d\tau & \cdots & -\int_0^T \phi_1(t)\, r_{uu}(t)\, dt \\
& | & | \\
& | & | \\
& | & | \\
& | & | \\
& | & | \\
(-1)^n \int_0^T \overset{n}{\phi_k}\, r_{yu}\, d\tau & \cdots & -\int_0^T \overset{n}{\phi_k}\, r_{uu}\, dt
\end{bmatrix}
\times
\begin{bmatrix}
a_n \\ | \\ \vdots \\ a_6 \\ b_M \\ \vdots \\ | \\ b_6
\end{bmatrix}
=
\begin{bmatrix}
\int_0^T \phi_1(t)\, r_{uu}(t)\, dt \\ | \\ | \\ | \\ | \\ | \\ \int_0^T \overset{n}{\phi_k}\, r_{uu}\, d\tau
\end{bmatrix}
$$

These modulating functions can be of a different shape or type so long as they satisfy the boundary condition,

$$
\overset{n}{\underset{k}{\phi}}[0,T]=0 \qquad n = 1, 2, \cdots N
$$
$$
k = 1, \cdots n+m+1
$$

In practice it was found convenient to use the higher order derivatives of the modulating functions to form each new simultaneous equation. This made the

generation of the modulating function easier to implement. Thus the solution to the parameter estimation problem using the Modified Modulating Function Method is,

$$
\begin{bmatrix}
(-1)^n \int_0^T \phi r_{yu} d\tau & ---- & -\int_0^T \phi r_{uu} d\tau \\
\vdots & & \vdots \\
(-1)^n \int_0^T \phi^{2n} r_{yu} d\tau & ---- & -\int_0^T \phi r_{uu} d\tau
\end{bmatrix}
\times
\begin{bmatrix}
a_n \\ \vdots \\ a_0 \\ b_M \\ \vdots \\ b_0
\end{bmatrix}
=
\begin{bmatrix}
\int_0^T \phi\, r_{uu} d\tau \\ \vdots \\ \vdots \\ \int_0^T \phi r_{uu} d\tau
\end{bmatrix}
\quad (3.4)
$$

This method also has an advantage over the modulating function method in that k samples of the input and output signals are not needed.

3-1-1 Noise

Noise has so far been omitted from the problem. Equation (3.3) can be re-stated with a noise term.

$$
(-1)^n a_n \int_0^T \phi(\tau) r_{yu}(\tau) + \cdots + a_0 \int_0^T \phi(\tau) r_{yu}(\tau)\, d\tau =
$$

$$
(-1)^m b_M \int_0^T \phi(\tau) r_{uu}(\tau) d\tau + \cdots + \int_0^T \phi(\tau) r_{uu}(\tau) d\tau +
$$

$$
\int_0^T \phi(\tau) r_{nu}(\tau) d\tau
$$

116

The MMFM has a considerable advantage over the MFM, in dealing with the problem of noise. The system is injected with a test signal free from corrupting noise and the auto correlation function is derived from this. The output signal is corrupted by noise, figure 2.5, but a cross correlation function formed between these two signals will be noise free so long as there is no correlation between the test signal and the corrupting noise, and the integration interval of the correlation function is sufficiently long. In other words the noise is rejected from the system.

This is a great advantage when identifying in noisy environments, where noise of a particular frequency or frequency band is automatically rejected without the need to use sophisticated filters or lose information about the system in a particular frequency range. The MMFM is also superior in this respect to most of the well known techniques of system identification of noisy data, such as Instrumental Variables or Maximum Likelihood which require prior knowledge of the noise.

The long integration times which are required to obtain noise free correlation functions, are not practical for identification of real systems. The system might drift during the experiment, also the data

storage requirement for batch processing would be prohibitive. In practice a limited integration interval is used, and the resultant correlation function will have a non-zero variance, but this can be reduced by taking successive averages of the correlation function.

3-1-2 Modulating Functions

The modulating function is central to this method of parameter estimation. Various modulating functions have been proposed. The most effective for parameter estimation are the modulating functions based on Hermite functions proposed by Takaya [61].

Modulating functions based on Hermite functions have been used in this work. The Hermite function is defined as,

$$H_n(r) = (-1)^n \exp\left(\frac{r^2}{2}\right) \frac{d^n \exp\left(-\frac{r^2}{2}\right)}{dt}$$

To form the modulating function, weighted Hermite functions are used (see chapter two).

$$\phi_n(r) = \frac{(-1)^n \exp\left(\frac{r^2}{2}\right) H_n(r)}{\sqrt{2\pi}} \qquad (3.5)$$

The Hermite function may be expressed in polynomial form,

$$H_n(r) = \sum_{p=0}^{[n/2]} (-1)^p (2p-1)!! \binom{n}{2p} r^{n-2p} \qquad (3.6)$$

where [n/2] is the largest integer that does not exceed n/2, and,

$$\binom{n}{2p} = \frac{n!}{(2p)!(n-r)!}$$

$$(2p-1)!! = (2p-1)(2p-3)\ldots\ldots 5.3.1$$

Mason [91] has shown that a recursive version of (3.6) can be developed,

$$H_{n+1}(r) = rH_n(r) - nH_{n-1}(r)$$
$$H_0(r) = 1$$
$$H_1(r) = r$$

This greatly improves the speed and ease of computation of the Hermite functions.

Hermite modulating functions, as described by equation (3.5), exist for a finite time window of approximately ten seconds, see figure 2.18 page 80. The modulating functions decay to zero at the limits of this window. If the dynamics of a fast system are to be examined,

119

then the auto and cross correlation functions may only exist for a time window considerably shorter than that of the modulating function window, figure 3.2. For a slow system the cross correlation function may not have decayed to zero at the end of the time window, and as a result much information about the system may be lost. Thus a time scale factor is needed for the modulating function.

For the diesel engine the modulating function window had to be shortened by approximately a factor of ten . The sample window of the auto and cross correlation functions being approximately two seconds.

Normalisation of the modulating function is necessary if the higher order derivatives of the modulating functions are to be used. This is because the modulating functions have the following property [61],

$$\lim_{\substack{n\to\infty \\ t}} \max \left| \overset{n}{\emptyset}(t-T/2) \right| = \infty$$

Therefore there may be a large difference in amplitude between the lower derivative modulating function and the higher order derivatives. Thus difference in range may exceed computational limits. Takaya proposed the use of the following normalisation factor to overcome this difficulty,
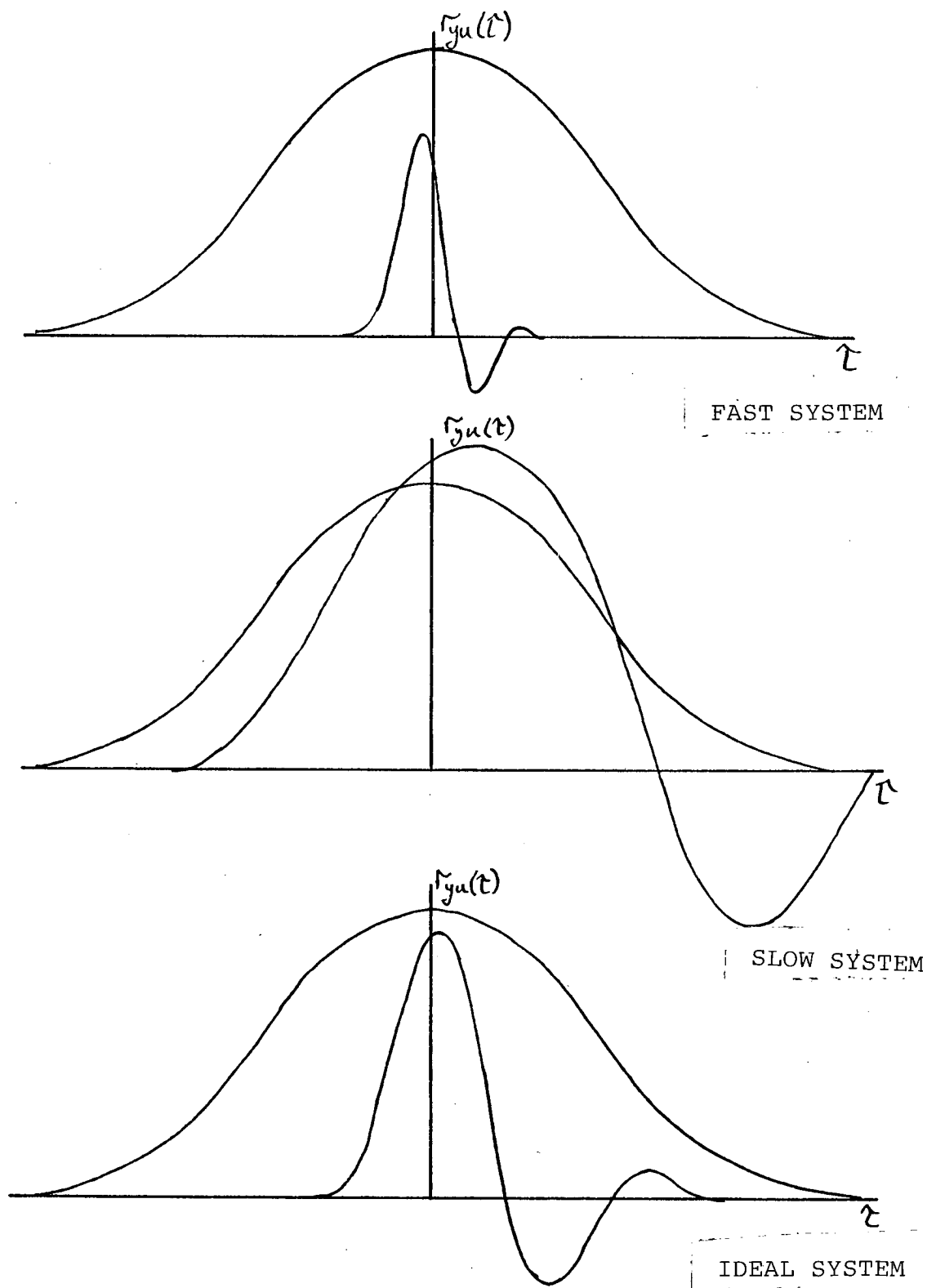
FIGURE 3·2
TIME SCALE FACTOR

$$c_n = \int_0^T \left| \overset{\wedge}{\phi}(t-T/2) \right| dt$$

The coefficient c_n is divided through the system differential equation.

In practice, use has not been made of this method of normalisation. Each modulating function has been normalised to a scale height so as to reduce rounding errors in the discrete integration of the correlation and modulating function products. The correlation functions have been normalised likewise, and equation (3.3) now becomes,

$$(-1)^n S^n A^n B \int_0^T \phi(t) \, \Gamma_{yu}(t) \, dt + \cdots = (-1)^m S^m A^m C \int_0^T \phi(t) \, \Gamma_{uu}(t) \, dt$$

where A^n is the scale factor for the nth derivative of the modulating function, B is the cross correlation scale factor, C is the auto correlation scale factor, S is the time scale factor.

3-2 Correlation

When using the MMFM the input auto correlation and cross correlation functions of the system output and input signals are computed before identification takes

122

place. It is therefore important that a fast and accurate correlation algorithm is used. The auto and cross correlations can be performed directly on the system input and output signals, but these rarely contain sufficient noise bandwidth to excite all the modes in the system. It is often better to inject a test signal into a system under test rather than rely on the naturally occurring noise signals. This ensures that a sufficiently wide bandwidth of noise signal is used to excite all the modes in the system.

It has been found that when testing a diesel engine used in a power station the injection of noise becomes essential. All natural noise in the signals is swamped by the 50 Hz interference present. This is a serious problem that can only be overcome if careful attention is given to filtering out the noise, without losing valuable information about the system at this frequency range. The MMFM eliminates this problem by cross correlating not between the actual input and output but between the injected noise signal and the output signal (see figure 3.3). Because the 50 Hz noise is not correlated with the injected noise it is eliminated from the correlation function. Although care has to be taken to ensure that the noise does not cause an overload of the ADC used in the measurement system.
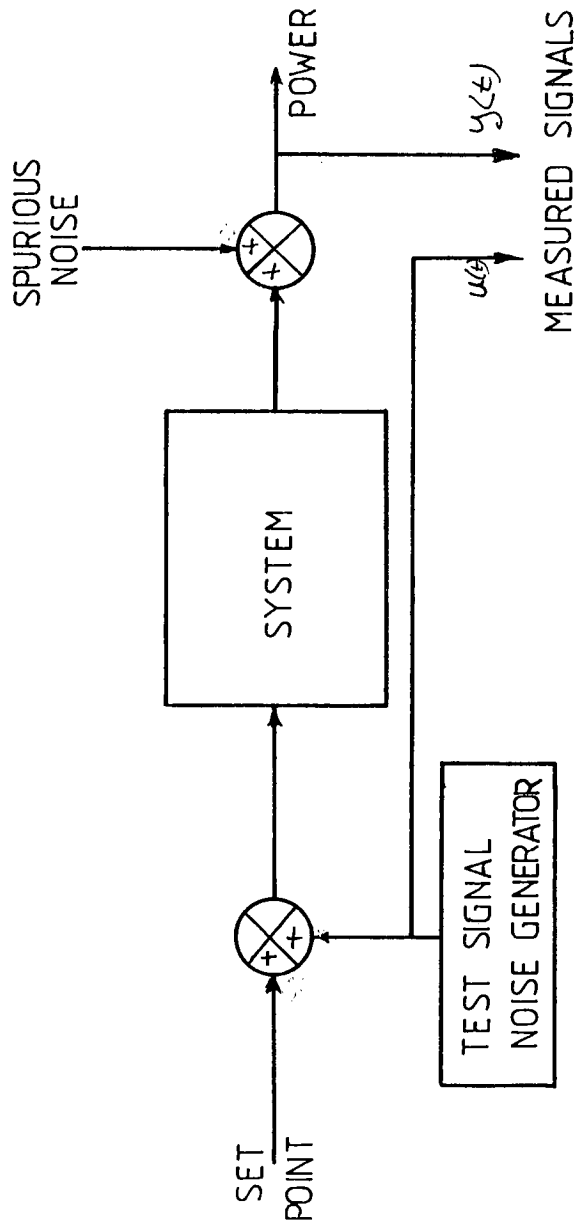
FIGURE 3·3

SIGNALS MEASURED USING THE MMFM

Pseudo Random Binary Sequences (PRBS) have been found to be very useful input signal [92]. The auto correlation of the PRBS signal has a triangular shape and the cross correlation between the PRBS and the output signal resembles the impulse response of the system under test. When testing a diesel engine PRBS noise was not used because it induced a large variance in the cross correlation function. Gaussian noise was used and gave a much smoother cross correlation function. The probable reason for this is that the Gaussian noise has lower high frequency harmonic content than the PRBS noise signal.

A brief description is given here of the various methods of performing correlation. Attention will be restricted to software methods. It should be noted that considerable progress has been made towards using silicon circuit technology to realise hardware correlators which will be of particular interest in applications involving fast data rates.

Correlation can be thought of as a matching process where a common pattern in two stochastic signals is to be detected. Given two stochastic signals, $y(t)$ and $u(t)$. The correlation between these two signals is dependent only on the time difference between the two signals, $(t-\tau)$, if the signals have stationary

characteristics. The correlation function is given by the expected value of y(t) and u(t-τ),

$$r_{yu}(\tau)=E[y(t)\ u(t-\tau)]$$

Where E is the expectation operator. Ideal cross-correlation is defined by.

$$r_{yu}(\tau)=\lim_{T\to\infty}\left\{(1/2T)\int_{-T}^{T}y(t)\ u(t-\tau)dt\right\} \qquad (3.7)$$

Similarly the auto correlation function, which matches a signal with itself, is given by,

$$r_{uu}(\tau)=\lim_{T\to\infty}\left\{(1/2T)\int_{-T}^{T}u(t)\ u(t-\tau)dt\right\} \qquad (3.8)$$

3-2-1 Discrete Correlation

If the correlation function is to be calculated on a micro computer, discrete versions of (3.7) and (3.8) are needed. The correlation operation consists of a time shift, τ , between the two signals, a multiplication between the signals and then a summation. This is shown schematically in figure (3.4). To ensure that no

information is lost the data should be sampled at twice
the Nyquist rate at least. The integration time is also
limited. Thus we arrive at an approximation for the
correlation function,

$$r_{uu}(\tau) = 1/T \int_0^T u(t-\tau)\ u(t)dt$$

If a limited length of data is available some loss of
data will occur as one signal is delayed, thus a further
modification will be necessary.

$$r_{uu}(\tau) = 1/(T-\tau) \int_0^{T-\tau} u(t)\ u(t-\tau)dt$$

This will restrict the size of τ to $\tau << T$ otherwise
$r_{uu}(\tau)$ will become increasingly inaccurate.

Assuming no restriction on the data available the
discrete correlation function becomes,

$$r_{uu}(k\tau_s) = (1/N)\sum_{r=1}^{N} u_r\ u_{r-k}$$

To implement this correlation requires a large computational effort with many multiplications and summations. A significant reduction in the complexity of the arithmetic operation can be obtained using some form of quantised data. Polarity and relay correlation are commonly used methods that use quantised data.

3-2-2 Polarity Correlation

When using polarity correlation the data signals are quantised into their sign form, figure 3.5. Correlation is then performed on this quantised data, and the correlation function becomes,

$$r_{uu}(\tau)=1/T\int_0^T sgn(u(t))sgn(u(t-\tau))dt$$

Performing the correlation in terms of the signal polarity will lead to a loss of information, but an approximate doubling of the integration time will compensate for this [93]. A very considerable time saving is achieved by eliminating the need for 8 bit digital multiplication. Using polarity correlation distorts the shape of the correlation function by sharpening the peak of the auto correlation function, for example see figure 3.6. A relationship between polar

128
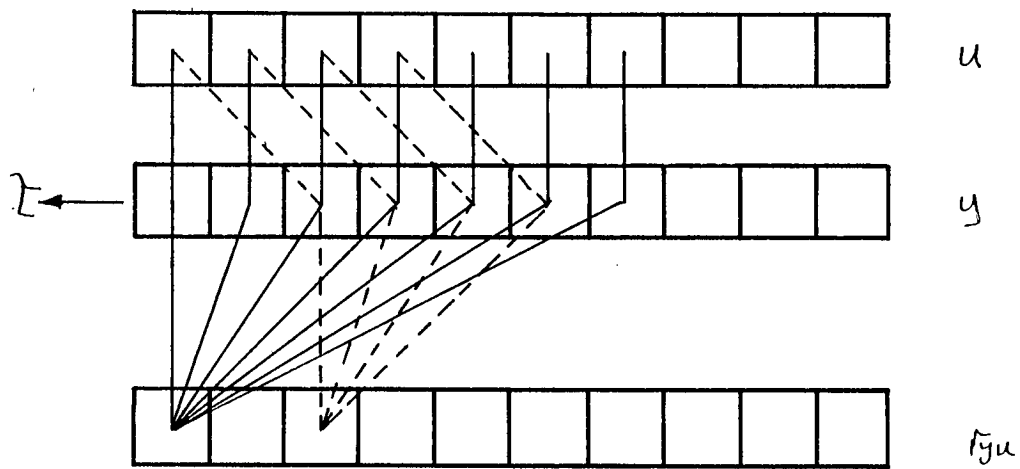
FIGURE 3-4

DISCRETE   CORRELATION
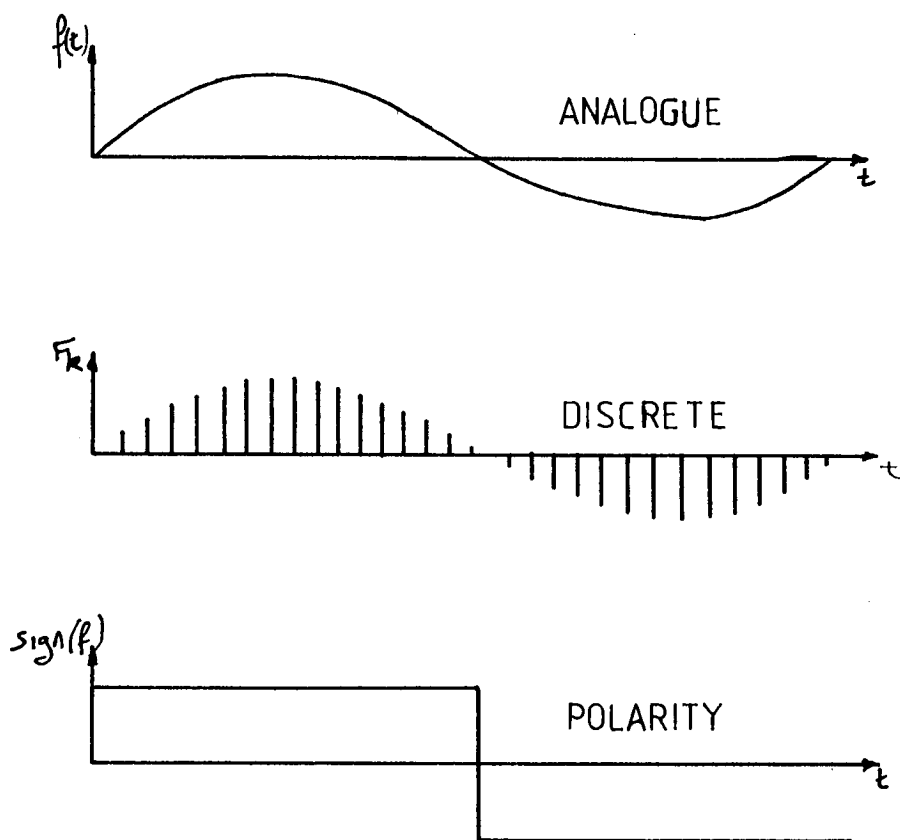
FIGURE   3-5
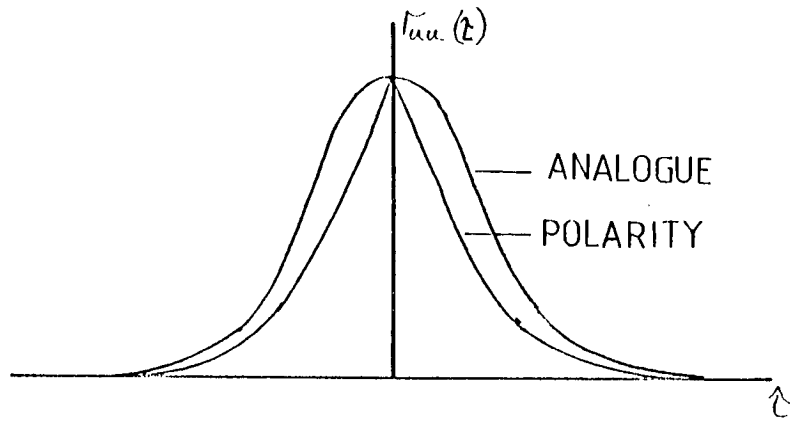
FIGURE 3-6

ANALOGUE AND POLARITY AUTO CORRELATION
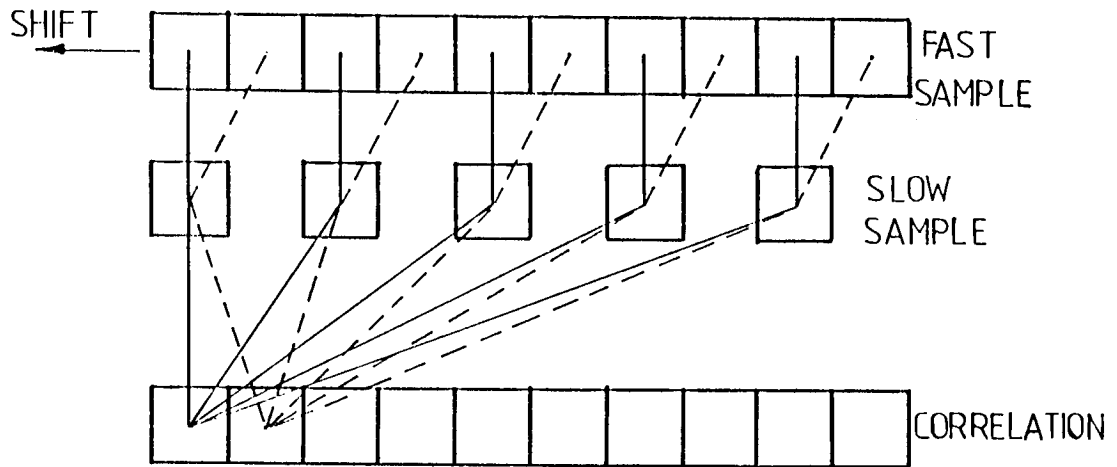
FIGURE 3-7

SKIP ALGORITHM

and normal correlation has been developed for the case when the signal to be correlated can be described by Gaussian statistics, [93]. It can be shown that,

$$r_p = (2/\pi)\sin^{-1}r_n$$

where r_p - polarity

and r_n - normal

3-2-3 Relay Correlation

Relay correlation can be considered a combination of polarity and normal correlation. It combines some of the speed advantage of polarity, with the shape benefits of normal correlation. With relay correlation one signal is quantised and the other analogue. The Relay correlation function is defined by,

$$r_{uu}(\tau) = 1/T \int_{O}^{T} \text{sgn}((u(t))u(t-\tau)dt$$

The computational time savings are not as large as polarity, but a significant improvement over normal correlation is achieved by reducing the multiplication of $u(t-\tau).u(t)$ to a sign test of $u(t)$ which will determine whether $u(t-\tau)$ will be added or subtracted to the accumulated $r_{uu}(\tau)$. A relationship between Relay

131

and analogue correlation exists for signals with
Gaussian statistics.

$$r_r = \sqrt{2/\pi} \; \frac{r_a}{r_{uu}(o)}$$

where r_r - Relay correlation

and r_a - Normalised analogue correlation

3-2-4 Skip Algorithm

The skip algorithm is a correlation algorithm designed
to speed up the calculation of the correlation function
still further. With normal correlation both signals are
sampled at the same rate. In the skip algorithm one
signal is sampled at a rate which ensures the necessary
resolution of the correlation function, and the other
signal is sampled more slowly, typically up to one
eighth the sample rate of the fast channel. This slow
sample rate should not fall below the Nyquist frequency,
otherwise information loss will occur. Figure 3.7 shows
a schematic illustration of the skip algorithm.

Fell [94] has shown that the fast sampling rate, used
to obtain resolution in the correlation signal is
usually much faster than the minimum sampling rate,
twice Nyquist, by typically, a factor of ten. It
is thus argued that if the cross correlation products

132

are formed at a slower sample rate while performing the shifting operation at the faster rate no loss of resolution in the correlation function will be suffered while a great reduction in redundant arithmetic will be achieved, typically a factor of ten. . Thus a large saving in the computational requirement needed to perform the correlation algorithm is achieved.

The skip algorithm can be used in conjunction with analogue, polarity or relay correlation algorithms, improving the computational speed of them. Although polarity correlation has the greatest speed advantage, relay correlation in conjunction with the skip algorithm has been used here. Although slower than polarity this enables us to retain the shape of the correlation function while achieving substantial computational savings over normal discrete correlation.

3-3 Diesel Modelling

The creation of a system model is a very complex process and a great deal of effort in the field of system identification has gone into this difficult area. Most of the effort has concentrated on producing discrete models for the discrete identification techniques developed over the past twenty years. For fault detection it is thought that continuous time

models are more relevant to the problem.

An accurate model of the system to be identified is essential for a good estimate of the parameters of the model to be obtained. The normal approach is to model the system using the physical laws governing that system. This often results in a model that is very complex, as in the case of distributed systems, or in a model that is an approximation because the system processes are not fully understood. Either way when it comes to identifying such a model a compromise must be arrived at in order to identify a model that is not over complex but will be able to respond to a disturbance in a similar manner as the system.

Diesel engines are very complex and some of the processes, for example combustion, are so complex with so many factors influencing them that it is not possible to model them accurately, so a very crude model is often used to model this process. Normally the diesel engine is so complex it is not modelled with linear transfer function techniques, but requires non-linear partial differential equations. Some work however has been done on producing transfer function models of the diesel, eg Thiruarooran et al,[95] but even these end up using a very high order model, principally due to the time delays in the system. A block diagram of the diesel

engine is shown in figure (3.8), and the corresponding transfer function model is shown in figure (3.9)

It can be seen that the combustion system has been modelled with a first order lag. The time delays in the system are due to the fuel system and the turbo charger characteristics. These were modelled using the Padé approximation and resulted in the closed loop transfer function,

$$G(s) = \frac{b_3 s^3 + b s^2 + b_1 s + b_0}{s^6 + a_5 s^5 + \cdots + a_0}$$

This model is too complex to use in most identification algorithms as the variance of the estimates of the higher order terms becomes large. Other authors [96] have produced simpler models, but these model the time delay with a first order lag and as such are probably too simplistic.

A simple linear model which takes into account time delays is needed for the identification algorithm, in which the variance of the estimates is not large and the computational effort does not prohibit on-line use. The problem of non-linearity can be overcome using piecewise linear models of the system.
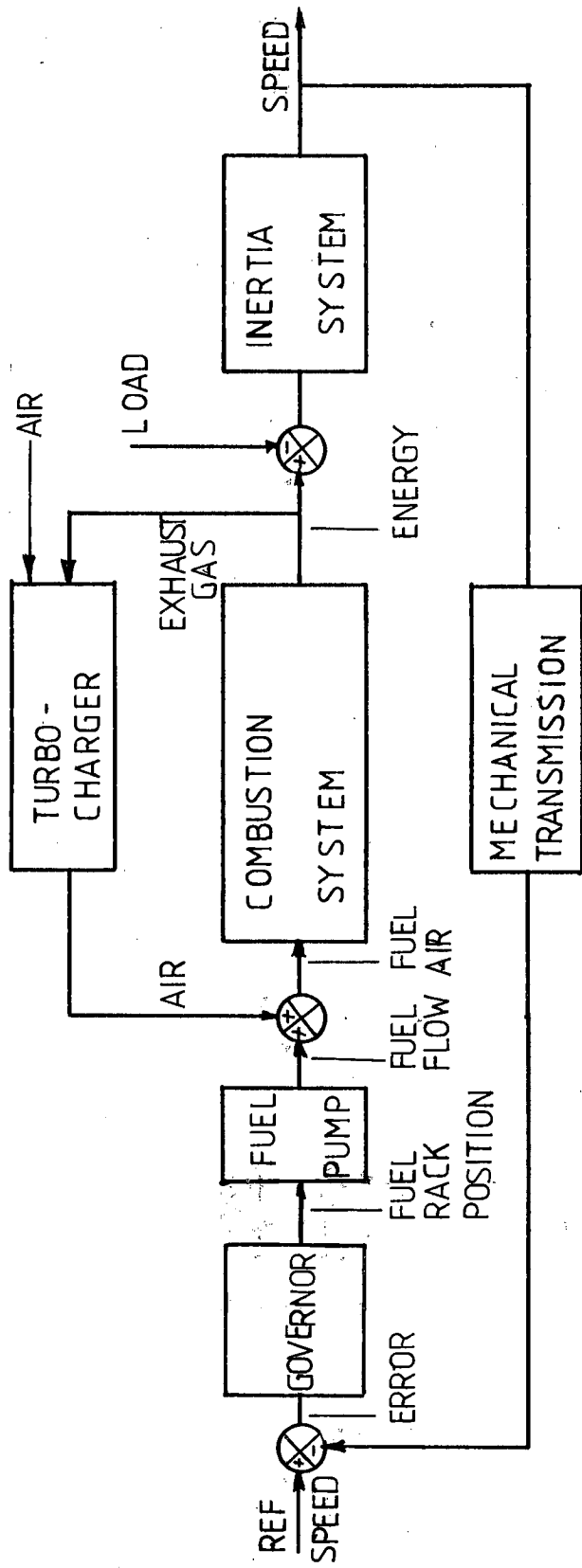
FIGURE 3-8

BLOCK DIAGRAM OF A DIESEL

136
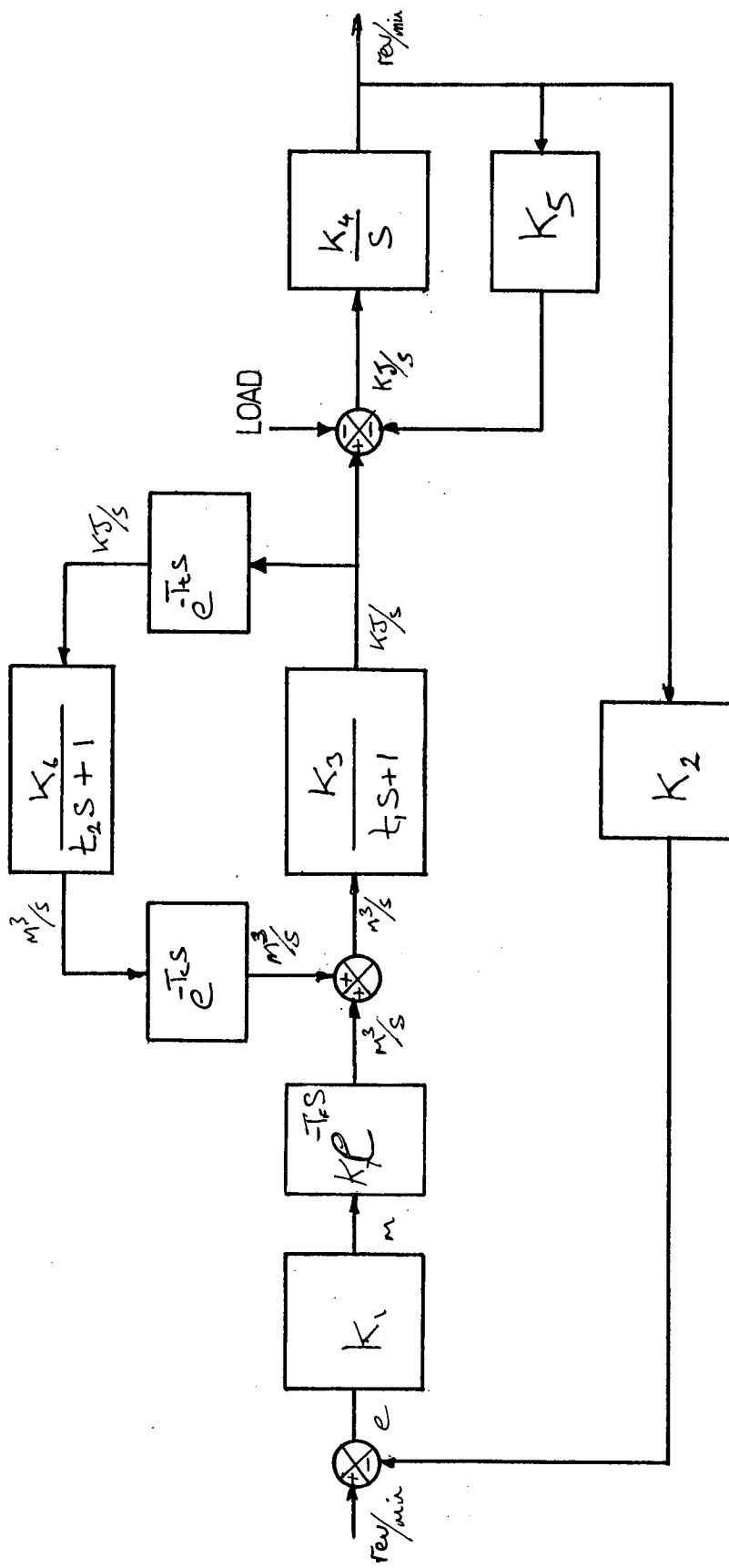
FIGURE 3-9

DIESEL TRANSFER FUNCTION

The choice of model is often difficult to make without prior knowledge of the system. The MMFM allows a study of the dynamics of the system before identifying the system, thus allowing a choice of model to be made. The MMFM uses the auto and cross correlation functions of the input and output signals. A study of the shape of the cross correlation will reveal clues about which order of model to use, and whether time delay is present. Higher order terms in the system will also give the appearance of time delay, thus it is thought possible to model most systems with a second order model plus a time delay. A study of the cross correlation of the diesel engine indicated that this type of model is appropriate, see chapter 5. This is a very simple model with only four parameters to identify, and as such will not present problems in identification. However this model is a long way removed from the physics of the actual system and as such will present problems for subsequent fault detection and isolation algorithms.

There is, unfortunately, a difference in requirements between identification and fault detection. For identification the model should be as simple as possible, but for fault detection the model should describe every detail of the original system. The closer the model is to the original system the more

138

likely it is that deviations in the parameters of the model can be associated with actual faults in the system. Unfortunately as we have shown it is not possible to model the diesel accurately, therefore a trade off between identification accuracy and fault detectability is required.

It is thought that if a fault occurs in the diesel it will affect the response of the diesel in some way, and hence the cross correlation function will change. This should affect the parameters of the model, which are being monitored for change. If the model was a comprehensive one it would be possible to relate these parameters back through the system and identify the faulty component in the engine. Because the model we are using is not comprehensive it may not be possible to isolate the fault, but the detection of a deterioration in the engine performance, due to a fault, may be possible.

3-3-1 Model Order

When modelling such a complex system as a diesel engine, with a simple model, modelling errors will be introduced as the higher order terms are truncated.

The auto and cross correlation functions are related

through the transfer function G(s).

$$R_{yu}(s) = G(s).R_{uu}(s) \tag{3.9}$$

where $G(s) = N(s)/D(s)$ (3.10)

and $D(s) = Q_1(s) + Q_h(s)$ (3.11)

and $N(s) = P_1(s) + P_h(s)$ (3.12)

The subscripts 1 and h represent the low order and high order parts of the polynomials.

If we then model the system with a low order model,

$$G'(s) = P_1(s)/Q_1(s)$$

An error will be introduced, this can be expressed in terms of a modified auto correlation function.

substituting (3.10) and (3.11) into (3.9) gives

$$Q_1(s).R_{yu}(s) = N(s).R_{uu}(s) - Q_h(s).R_{yu}(s) \tag{3.13}$$

from equation (3.10)

$$Q_h(s).R_{yu}(s) = Q_h(s).(N(s)/D(s)).R_{uu}(s) \tag{3.14}$$

substituting (3.14) into (3.13) gives

$$Q_1(s).R_{yu}(s) = N(s).R_{uu}(s).(1 - Q_h(s)/D(s))$$

$$Q_1(s).R_{yu}(s) = P_1(s).R'_{uu}(s)$$

$$R_{yu}(s) = G'(s).R'_{uu}(s)$$

where $R'_{uu}(s) = (1 + P_h(s)/P_1(s)).(1 - Q_h(s)/D(s)).R_{uu}(s)$

Therefore the effect of truncating the model of the
system can be expressed in terms of a distorted input
signal, ie $R'_{uu}(s)$ becomes the new input to the model if
the same output is to be obtained for the low order
model as for the system.

Thus the differential equation becomes,

$$a_n \frac{d^n r_{yu}(\tau)}{d\tau^n} + \cdots a_o r_{yu}(\tau) = b_m \frac{d^m r'_{uu}(\tau)}{d\tau^m} + \cdots + r'_{uu}(\tau)$$

$$r'_{uu}(\tau) = r_{uu}(\tau) + \triangle(\tau)$$

and $\triangle(\tau)$ is the effect of,

$$\left(1 + \frac{P_h(s)}{P_L(s)} \right) \left(1 - \frac{Q_h(s)}{D(s)} \right)$$

on $R_{uu}(s)$.

Thus the scalar products formed with the modulating
functions $\emptyset(\tau)$ and the auto correlation $r_{uu}(\tau)$ will
contain errors.

141

$$\int_0^T \phi(\tau) \cdot r'_{uu}(\tau) d\tau = \int_0^T (\phi(\tau) \cdot r_{uu}(\tau) + \phi(\tau) \cdot \Delta(\tau)) d\tau$$

Thus $\int_0^T \phi(\tau) \cdot \Delta(\tau) d\tau$ is the error induced by using a model order lower than that of the actual system. The time available for the programme of work reported here was not long enough to allow a detailed interpretation of the errors introduced by low order models. A continuation of this work has been reported by Jalali-Naini and Jordan [97].

3-3-2 Time Delay

Time delay in system identification is a serious problem. Discrete models tend to overcome this by increasing the order of the model [34], but there are difficulties if the time delay is not an integer value of the sample period. Continuous time models are not amenable to this solution, and there have recently been different techniques proposed to eliminate this problem [55]. The MMFM offers an alternative solution by using a manual method that requires no computational effort during identification, and only a setting up

procedure to implement. Care must be taken to avoid introducing errors caused by the incorrect removal of time delay.

The technique requires that the auto and cross correlation functions are studied for a time lag between them, this can then be removed in the software computation of the cross correlation function. The parameters of the model can then be identified in a straightforward manner. Thus, the time delay would be removed and the parameters of G´(s) identified. The final model of the system G(s) would be,

$$G(s) = G'(s) \cdot e^{-sT_d}$$

Alternatively a mechanised method [97] could be used. Assuming a second order model with time delay,

$$H(s) = \frac{e^{-sT_d}}{(a_2 s^2 + a_1 s + a_0)}$$

$$= \frac{1}{(a_2 s^2 + a_1 s + a_0) e^{+sT_d}}$$

Using a power expansion of e^{-sT_d}

$$H(s) \simeq \frac{1}{(a_2 s^2 + a_1 s + a_0)(1 + sT_d + s^2 \frac{T_d^2}{2} + \cdots)}$$

$$H(s) \simeq \frac{1}{\left(a_0 + (T_d a_0 + a_1)\right)s + \left(\frac{1}{2}T_d^2 a_0 + T_d a_1 + a_2\right)s^2 + \left(\frac{T_d^3}{3!}a_0 + \frac{T_d^2}{2}a_1 + T_d a_2\right)s^3 + \cdots}$$

143

Hence the measured coefficient of s^3 and higher order terms become zero when Td=0. Therefore an iterative delay removal procedure which checks the estimate of the s^2 parameter can be adopted.

3-3-3 Closed Loop Transfer Function

System identification is usualy best carried out with the feedback loop open. In many systems it is not possible to break the feedback loop for safety, or practical reasons. The diesel engine is such a case where the feedback loop cannot be broken for safety reasons. The open loop transfer function cannot be determined , thus we must be content with identification of the closed loop transfer function. Gustavsson et al [98] have produced a survey of closed loop identification, which reviews methods of discrete identification of closed loop systems. In their study they found that in some cicumstances closed loop identification is as good as open loop identification, for example when an additional test input is used.

CHAPTER FOUR

EXPERIMENTAL SYSTEM

The aim of this research work was to investigate the Modified Modulating Function Method of parameter estimation and use it in a fault detection system. An experimental system was developed and applied to a diesel engine used for the generation of electricity at Stornoway Power Station. It was necessary to have an experimental system which was both portable and flexible. The equipment was developed in the laboratory using R-C networks. and then the equipment was transported to Stornoway for on-site testing. This enabled the flexibility of on-line system identification to be retained, which was considered an advantage over the normal method of using recorded data. The system could then be adapted according to the results obtained.

4-1   TEST HARDWARE.

Given that the equipment needed to be portable, an HP85 computer was used to control the experiments and process the results. The data from the input/output signals was obtained from a Datalab Transient Recorder under the control of the HP85 via an IEEE 488 Bus. A noise generator provided the necessary stimulation for the system. This formed the basis of the test equipment

145

as shown on figure 4-1. A Wavetech Spectrum Analyser was added occasionally to verify the results.

The test equipment could be applied to any SISO process for identification of the transfer function parameters, provided the process had suitable access points. Therefore, once the system software was operational, it was a simple step to transfer the effort from laboratory study to field study. In the laboratory the method was first developed on software simulated systems, which will be discussed later, and then on simple analogue networks. Much of the work was concentrated on a second order network, figure 4-2, as the diesel was thought to be predominantly second order. This network was given a natural frequency approximating the diesel natural frequency, and stimulated using the same noise source as would be used on site. In this way a reasonable approximation to the diesel was obtained and any problems overcome before site testing.

Test equipment was installed at Stornoway Power Station to allow access points to the diesel so that a test signal could be injected into the engine and to enable the response of the diesel alternator set to be measured.

As a prerequisite for use of the MMFM, detailed access to the system under test is essential. The ability to
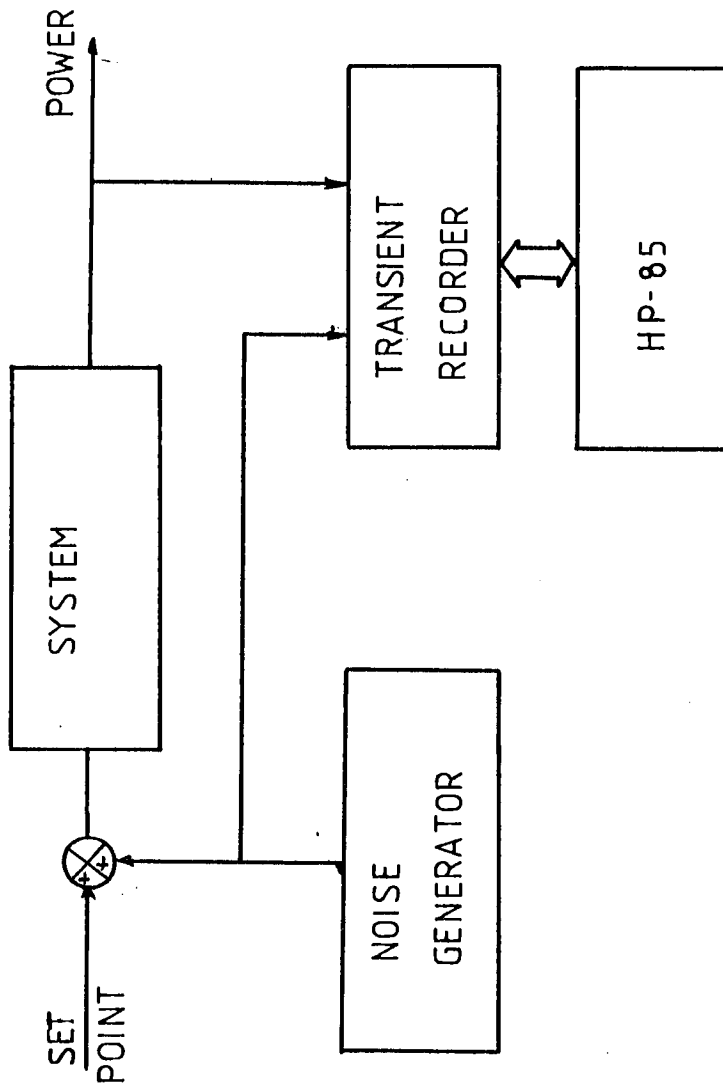
FIGURE 4-1
EXPERIMENTAL SYSTEM

147
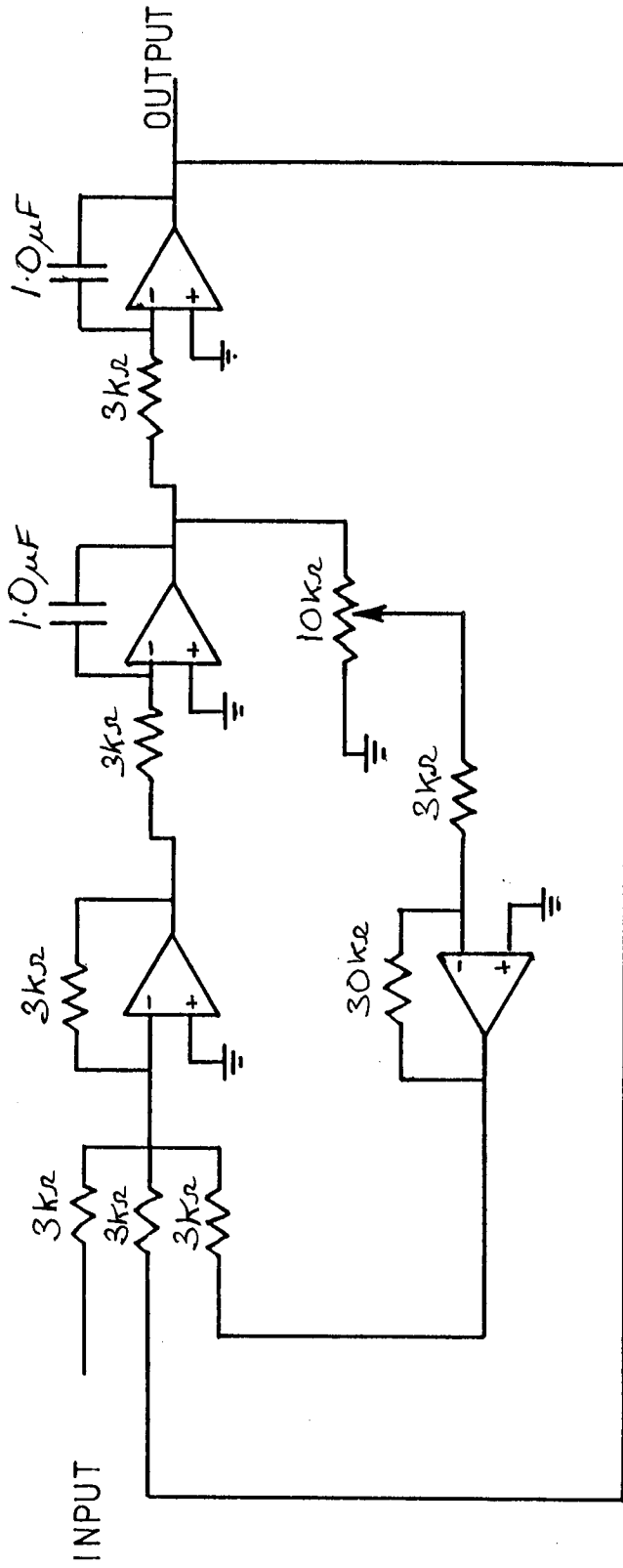
FIGURE 4-2

SIMULATED SECOND ORDER SYSTEM

148

stimulate the system and measure the effect is required.
On a conventionally governed diesel it is very difficult
to stimulate the engine in any useful manner, because it
is normal to fit a hydraulic governor which has limited
or no access to the summing junction. For this reason
the engine under test had an electronic governor fitted
which permitted access to the summing junction, and
enabled the injection of various test signals.
Measurement of the response of the engine to this
stimulation was achieved by the use of current and
voltage transducers, on the output of one phase of the
alternator, which enabled the power output to be
measured. All test equipment was isolated from the
power station control equipment by opto-isolation
amplifiers to avoid the problem of floating earths
causing damage or false alarms.

The diesel engine used in this work was a
Mirrlees-Blackstone KV12 Major ⊓. This is a twelve
cylinder engine configured in a ´V´ formation, the power
output is approximately 4.3 MW. The engine has twin
turbo-chargers, one for each half of the engine. A
block diagram of the engine and ancillaries is shown in
fig. 4-3.

4-2 TEST SOFTWARE

The software was written to perform two distinct
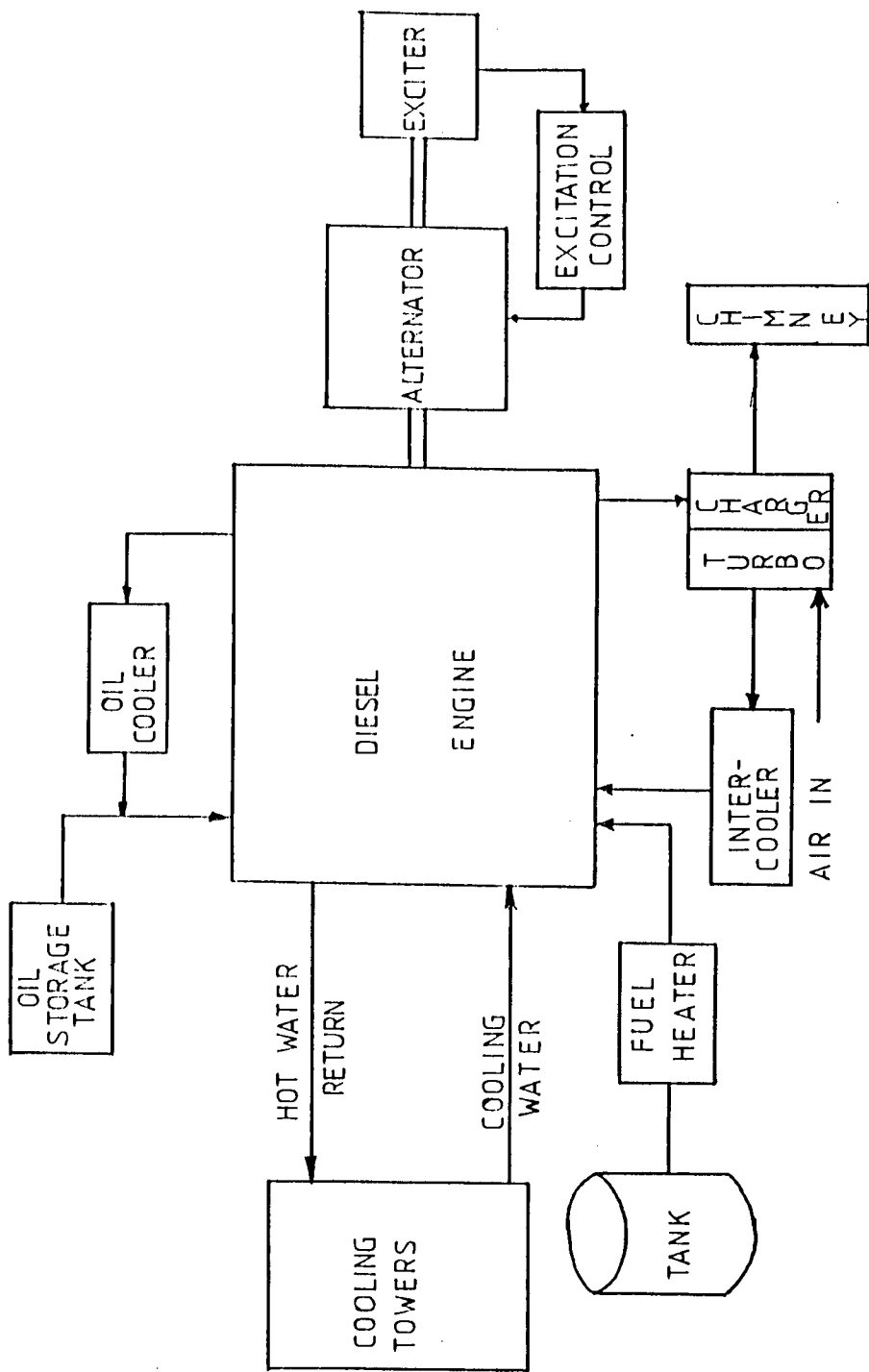
FIGURE 4·3
BLOCK DIAGRAM OF DIESEL / ALTERNATOR
AND ANCILLARY EQUIPMENT

tasks and consisted of two programs. The first program was written in BASIC and its function was to control the experiment. This included loading data from the Transient Recorder, controlling the assembler level calculations of the correlation and modulating functions, and storing the results. The second was an assembly language program designed to calculate the auto and cross correlation functions, and the integral of the product of a correlation function and a modulating function. This use of the assembler language program considerably improved the speed of calculation.

Two software designs were produced, one to control the experiments on hardware, and the other to perform simulation experiments. The flowchart for control of experimentation is shown on figure 4-4. This software was designed to enable a sequential processing of many samples of the input/output signals. Initially the Transient Recorder was triggered to sample the two signals. When this was complete, the data, 2048 bytes, was loaded into the HP85. The Transient Recorder was then triggered again so that data sampling could be performed while the current data set was processed. The data was then passed to the correlation routine where auto and cross correlation functions were calculated. The correlation functions were returned to the BASIC program for averaging.
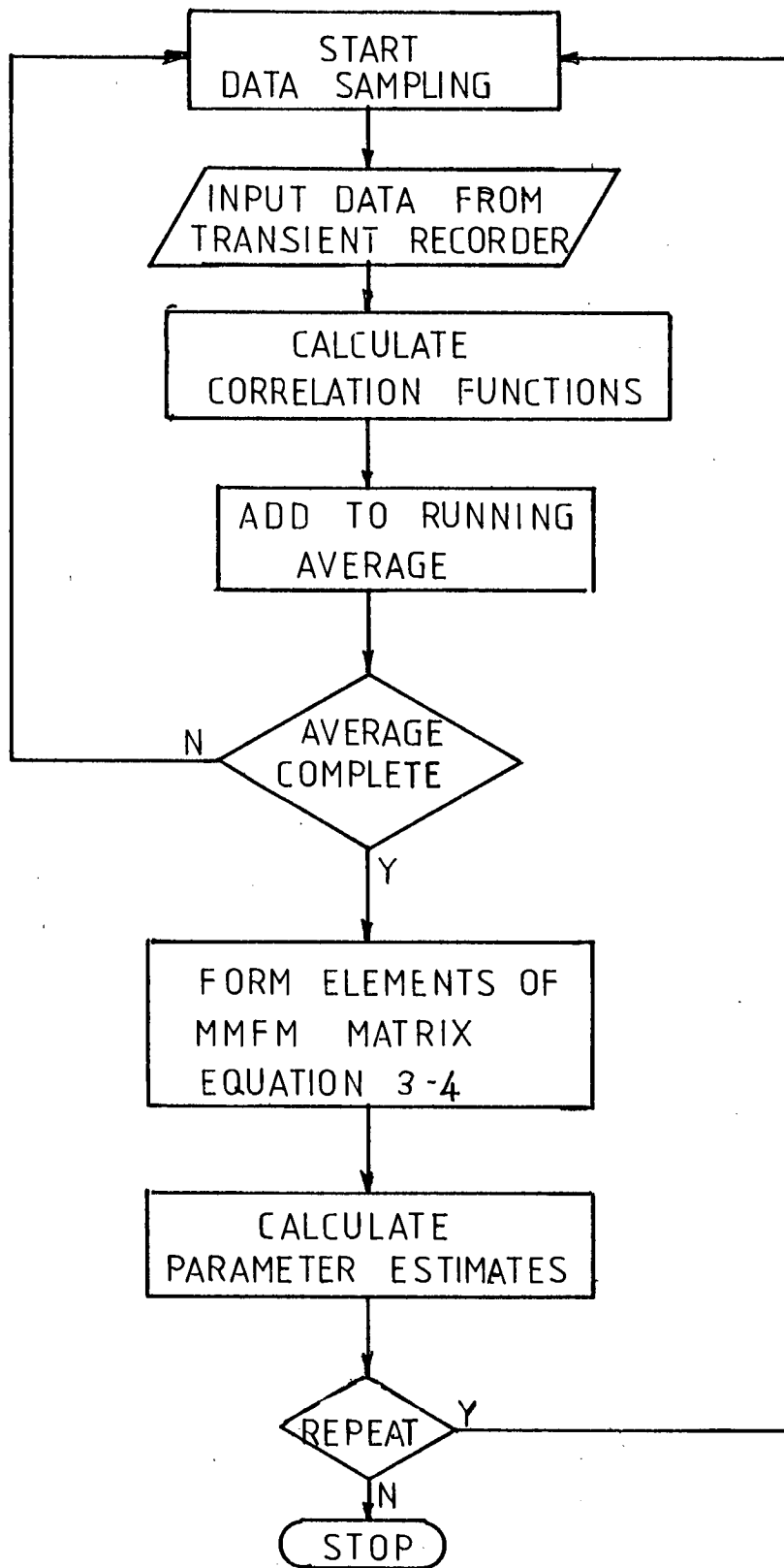
FIGURE 4-4
EXPERIMENT CONTROL FLOW CHART

The variance on the correlation functions causes a wide spread in the estimates of the parameters, it is therefore necessary to introduce a form of averaging. The program was designed for continuous monitoring of the system, so a moving average was used. This took the form of exponential averaging, which weighted the most recent correlation function in preference to the older functions. Figure 4-5 shows a flow chart of the averaging routine.

If the total number of correlations has not reached the running total another correlation is performed until the total is reached. When this happens the correlation functions are multiplied by the modulating functions and the results integrated. This operation is handled by an assembler program. Before the correlation function array is passed to the integration routine it must be converted into binary and amplitude scaled to ±127. The modulating functions are similarly scaled to minimise the integration error.

When the integrations have been performed, the matrix equation 3-4 can be formed. Matrix inversion is performed using a program stored in a ROM module fitted to the HP85. The parameter estimates calculated were then stored on tape for analysis.

The assembler programs were written in HP assembler
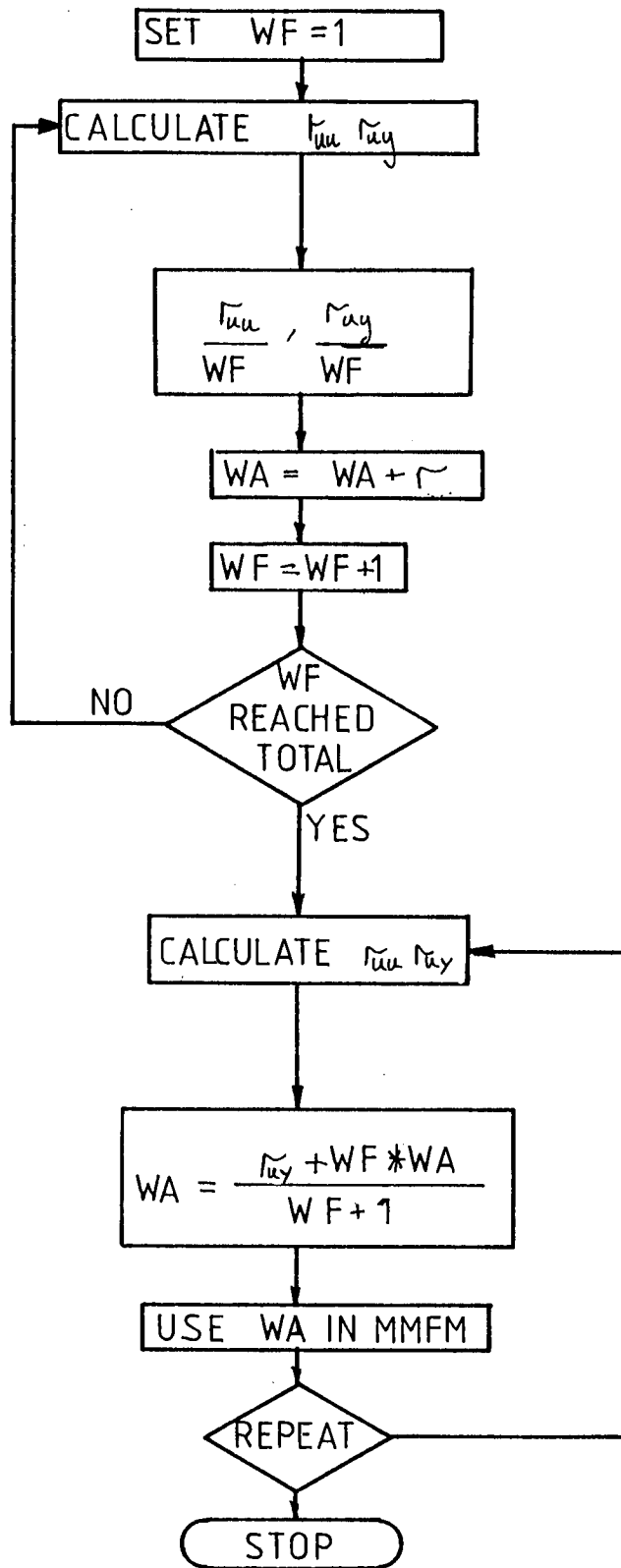
FIGURE 4-5

EXPONENTIAL   AVERAGING

154

lanquage. A significant improvement in processing speed over BASIC was achieved, thus making it possible to process large amounts of data. Only the two main assembler programs will be considered, these are the correlation routine and the evaluation of the correlation and modulating function integral. The other assembler routines handle simple communication tasks with the basic program, for example transfer and conversion of data.

The correlation program was designed as a Relay Correlator which also implemented Fell's [94] skip algorithm. The program was designed to be flexible with a variable shift and variable skip. This allowed experimentation into accuracy with decreasing data samples. A flow chart of the correlation routine is shown on figure 4-6. The raw data from the transient recorder was down loaded into the HP85 in a string of 8 bit bytes. Because Relay correlation is defined as,

$$\Gamma_{yu}(\tau) = \frac{1}{2T} \int_{-T}^{T} y(t) \, Sgn[u(t-\tau)] \, dt$$

the correlation can be simplified to a sign test on u and then either an addition or subtraction on the accumulator for the particular shift. The raw data signals u and y were accessed by indirect addressing

LOAD DATA
U   Y

SIGN TEST ON U
-VE   +VE

CHANGE SIGN OF Y

ADD Y TO ACCUMULATOR

INCREMENT INDIRECT CONTROL REGISTER BY SKIP N°

END OF DATA
N

INCREMENT SHIFT INDIRECT ADDRESS REGISTER

END OF SHIFT
N   Y

STOP
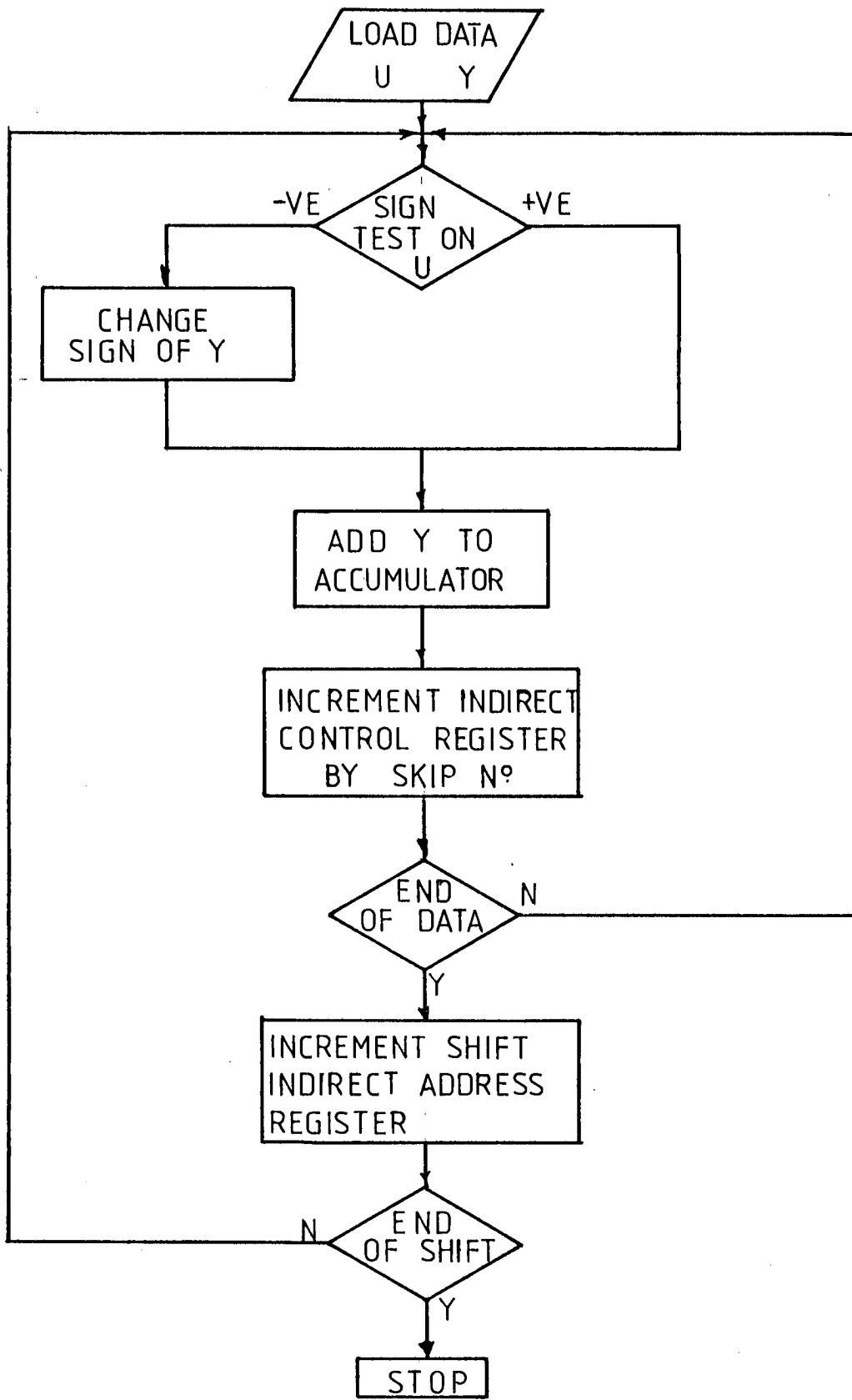
FIGURE 4-6
CORRELATION

156

which was incremented after each correlation cross product. After a complete pass on the data the delay shift register was incremented and another correlation cross product and summation was completed. For each pass the accumulator total was stored thus building up the correlation function. The number of shifts was limited to 256. This was considered sufficient and yielded a good resolution in the correlation function as the size of shift was set by the sample rate on the transient recorder. The data was sampled over 2048 bytes and the cross multiplication performed over 1600 bytes to allow for the overlap on maximum shift. Once the data had been returned to the BASIC program the correlation function was multiplied by the scale factors to create the true function. The number of shifts used could vary 2,4,8,....256.

The skip algorithm was implemented by skipping a number of cross products, i.e. a skip of one would have 1600 cross products. A skip of 2 would have 800 cross multiplications. The skip is variable under program control, the skip options available are, no skip (1), 2,4,8,--64.

The modulating functions routine calculates the integral of the correlation function multiplied by a modulating function. This task is performed for both the auto and cross correlation function with 2N

derivatives of the modulating function. A flow chart of the modulating function routine is shown on figure 4-7. As there was no multiplication operation in the assembler a short routine to perform the multiplication was written, this forms the bulk of the flow chart. The modulating function and its derivatives (up to the eighth) were stored in data arrays of 256 bytes per derivative.

The modulating function data was generated using a BASIC program, which calculated the functions using Hermite polynomials and magnitude scaled them to ± 127. The functions were converted into octal code which was used in the data arrays.

One other important BASIC program written was the software simulation program. This generated simulation of the auto and cross correlation functions of different systems, and was used to test and experiment with the MMFM. It had the great advantage that noise was not present, and accurate results could be achieved, enabling detailed analysis of the method to be performed. This program used the calculated impulse response of 1st, 2nd, 3rd and 4th order systems in a convolution with a triangular shape representing the auto correlation to give the cross correlation. The impulse response for first and second order systems is well known but the third and fourth order responses had
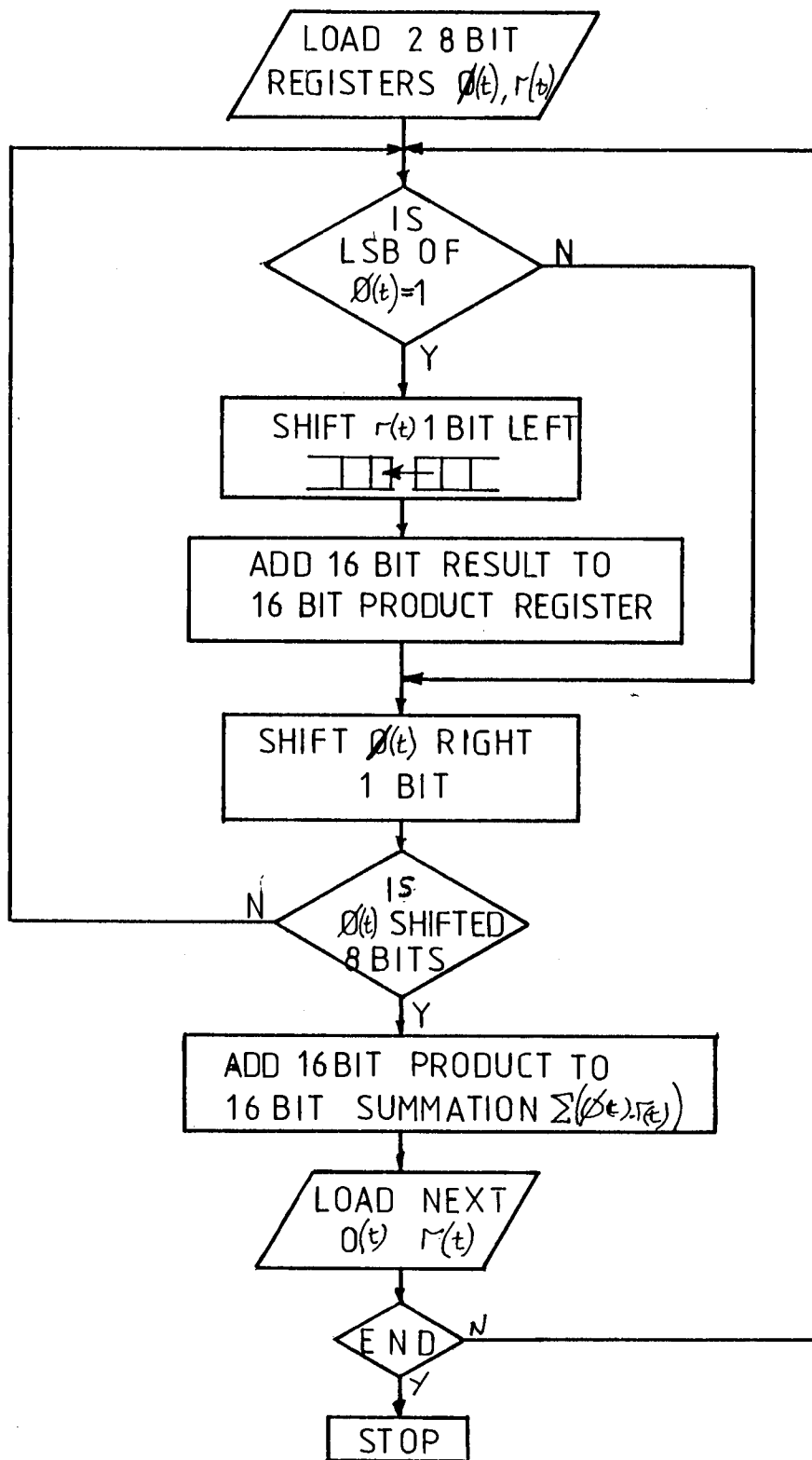
FIGURE 4-7
CALCULATION OF $\int \phi(t) r(t) dt$

to be calculated by convoluting a first order and a second order system to make a third order system. Similarly a fourth order system was made by convoluting two second order systems. This was achieved by multipling the two systems in the s - plane and then performing an inverse Laplace Transform.

4 - 3 TEST PROGRAM.

The testing carried out using the MMFM followed the classical technique of validating the method using simulated data, i.e. the simulation program previously discussed. Software experiments were carried out to determine the effect of time delay, and of measuring higher order systems with low order models on the parameter estimates. A hardware simulator was then used, this was a 2nd order analogue system which enabled the correlation routine to be verified. This simulator was also used to experiment with the number of shifts in the correlation routine and also the number of skips to be used. Once the best set up had been achieved experimentation at Stornoway Power Station was carried out. Long term tests were undertaken to determine the transfer function and an attempt made to detect failure on the engine. The results of these tests are detailed in chapter 5.

CHAPTER FIVE

EXPERIMENTAL RESULTS

In this chapter the application of the MMFM, and the results obtained, will be studied. As the aim of this research work was to produce a low cost, remote operation on-line monitoring device for a Diesel Power Station, the experimentation was biased towards this.

Experimentation on the MMFM followed the classical technique of software simulation experimentation to verify the method and then application of the method to a laboratory system before testing at Stornoway Power Station. Results obtained on the diesel were corroborated by results obtained from a commercial transfer function analyser.

5-1 Software Simulation

The software simulation experimentation program was carried out with the aim of proving the MMFM. It was also used to investigate the various problems created by using a simple model of the diesel engine, such as model order and time delays, which make accurate identification impossible. The use of computer simulated data had the advantage that there were no noise problems to confuse

the results. Simulated auto and cross correlations were generated by convolving the known impulse response of a system with a representation of the auto correlation function of a PRBS noise signal. A narrow triangle is a good approximation to the auto-correlation of PRBS noise. The bias of the auto-correlation can be considered negligible if we assume a long sequence length of the PRBS noise. The MMFM was applied to this simulated data. First order, second order, second order with time delay, third order and fourth order simulated systems were studied.

5-1-1 Verification of the MMFM

Verification of the MMFM is a straightforward procedure. Simulated system responses are generated by using the convolution routine described in chapter 4. The impulse response was convolved with a triangular shape to obtain the simulated auto and cross correlation functions, figure 5-1 shows a first order simulated system. The MMFM was then applied to these systems and the parameter estimates compared with the known system parameters. On performing this test the MMFM estimated the parameters of the transfer function accurately, however, it was noted that movement of the correlation functions within the modulating function window caused a variation in the accuracy of the parameter estimates.
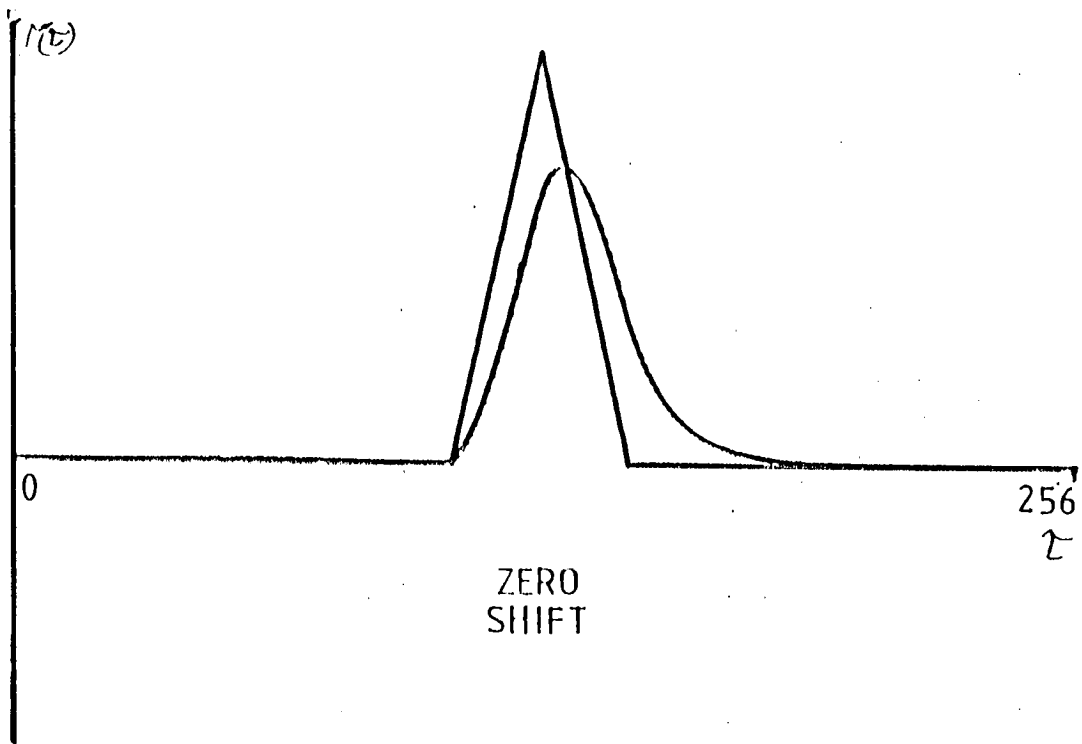
FIGURE 5-1
FIRST ORDER SIMULATED RESPONSE

$$G_{T(s)} = \frac{g}{(1 + s\beta)}$$
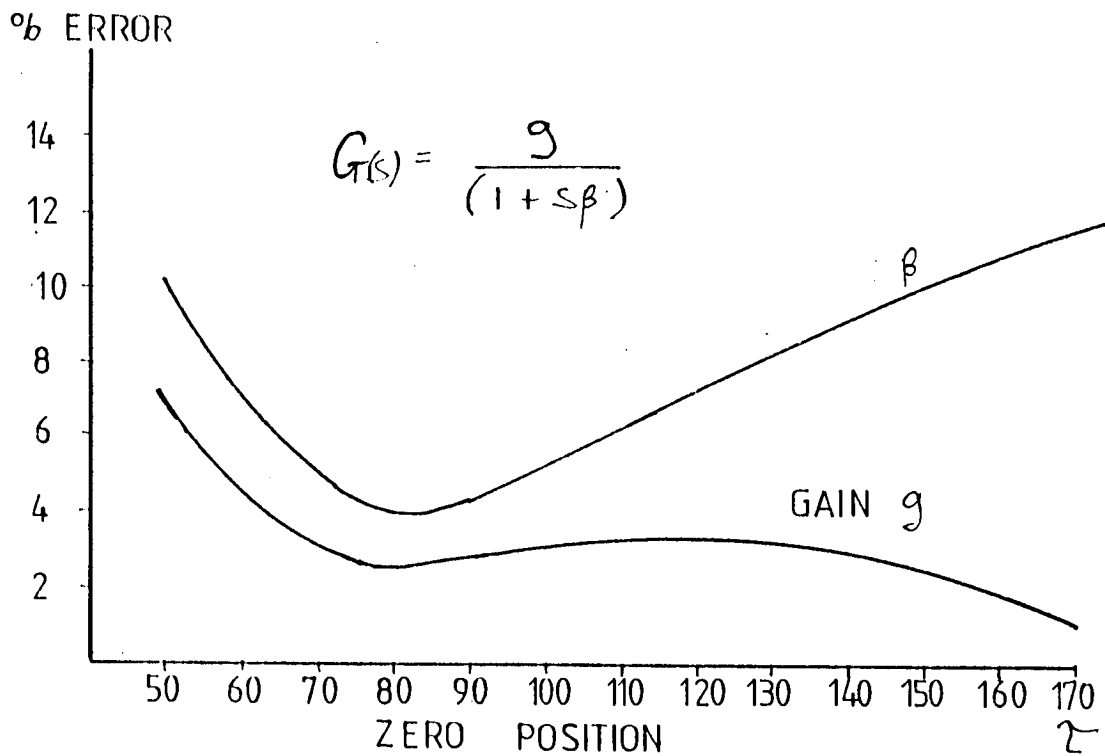
FIGURE 5-2
FIRST ORDER POSITION ERROR

163

For a first order system the zero shift point on the correlation functions was shifted within the modulating function window, from the extreme left, to the extreme right. The percentage error in the parameter estimates, the gain and time constants, is shown in figure 5.2. This reveals that the MMFM estimates of the parameters is most accurate for a first order system when the zero shift of the correlation functions is set between 70 and 100, of the sample length. The minimum error on the parameter estimate obtained, being in the region of two to four percent. A similar study was carried out with a second order system, figure 5.3. This shows the most accurate estimates to be in the region 100 to 150, fig. 5.4. The minimum percentage error achieved is lower for the second order system, between 0.1% and 1%.

The reason for a variation in accuracy between individual parameter estimates may be due to truncation error in the integrals formed with high order modulating functions, although the difference in the accurate region is not significant. The error caused by the variation of the zero within the window may be caused by the shape of the modulating function giving more weight to the centre of the window, thus reducing the truncating effects of the numerical integration at the centre, compared to the tails.
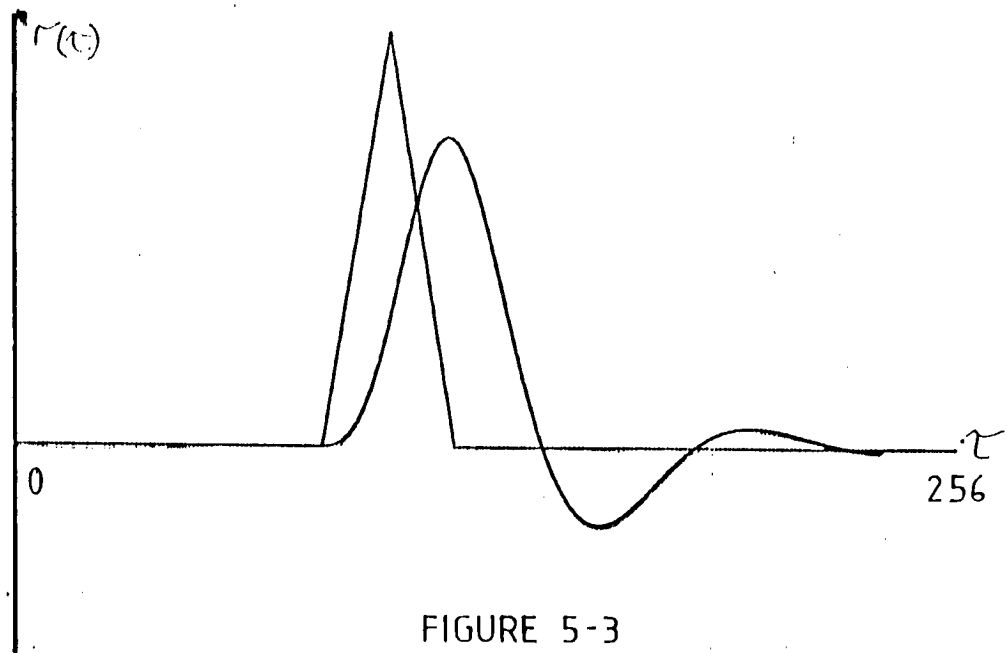
FIGURE 5-3
SECOND ORDER SIMULATED SYSTEM

$$G(s) = \frac{9}{\left(\frac{s}{\omega_n}\right)^2 + 2\delta\left(\frac{s}{\omega_n}\right) + 1}$$
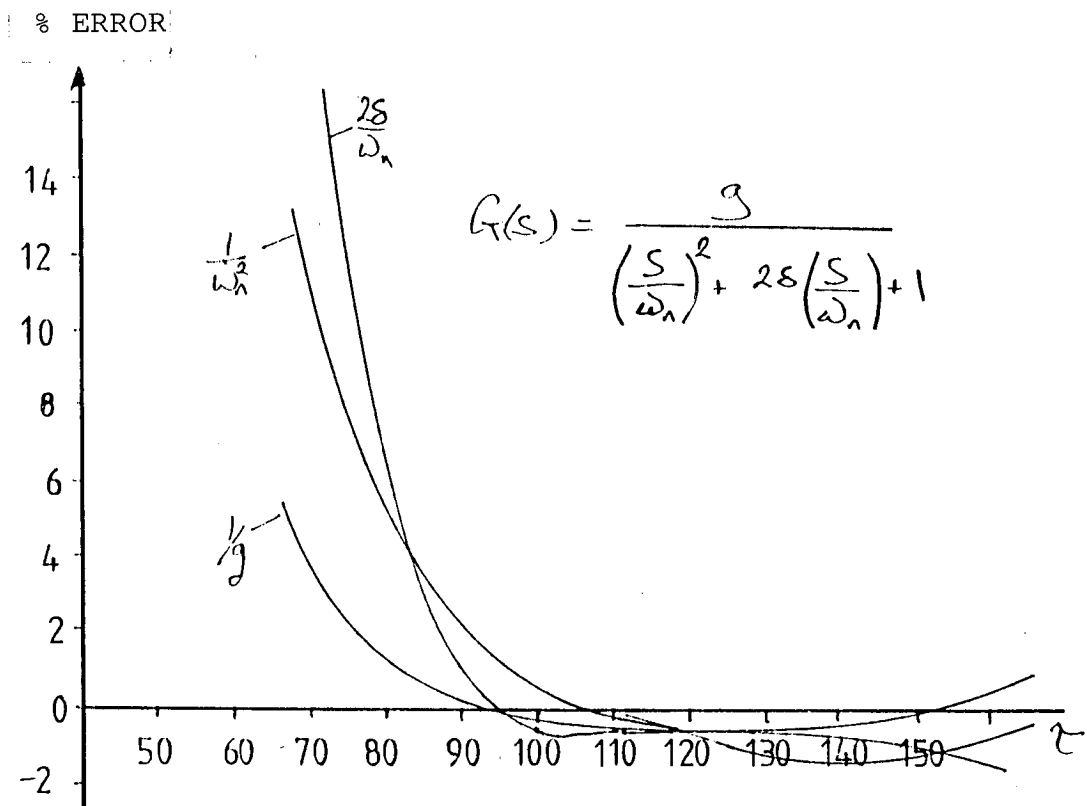
FIGURE 5 − 4
SECOND ORDER POSITION ERROR

165

Selection of the sample time is critical, the whole cross correlation function should be contained within the modulating function window, thus the sample rate should be selected to allow this. A slow sample rate should be avoided so that the correlation functions do not become narrow thus increasing the truncation error in the numerical integration. Errors due to this effect can be expected to be exaggerated due to noise present in real systems. Equally, fast sample rates should be avoided because, if the correlation function extends beyond the modulating function window, information about the system will be lost.

5-1-2 Time Delay

A serious problem was encountered during preliminary testing on the diesel engine. A time delay between the auto and cross correlation functions was detected. This was due to the fuel transport lag between the fuel pump and the fuel injectors. Any time delay would affect the estimates of the parameters, and thus make fault location, by way of relating parameter deviations to system failures, difficult. For this reason it was decided to investigate the effect on the parameter estimates of a simulated second order system when time delay and time advance were introduced.

166

This simulated system was modelled by a second order transfer function, and parameter estimates obtained. These were used to calculate the impulse response, which was convolved with the simulated auto correlation to give the cross correlation of the model. This was then compared with the cross correlation of the system.

Even small delays quite markedly affected the parameter estimate. Figure 5-5 shows the effect on the model when a time delay of 10% of the settling time is introduced between the auto and cross correlation functions. As can be seen, the response of the model is much more sluggish than that of the system. Figure 5.6 shows a similar and more exaggerated effect for a time delay of 20% of settling time.

A similar exercise can be carried out using time advance, although this situation would not normally be met in an engineering system. However in the winding out procedure adopted to remove time delay, there is a danger of over correcting and introducing time advance into the system. Time advance has the opposite effect on the model, the response becomes much more active. Figures 5.7 and 5.8 show the response for a time advance of 5% and 10% of the settling time.
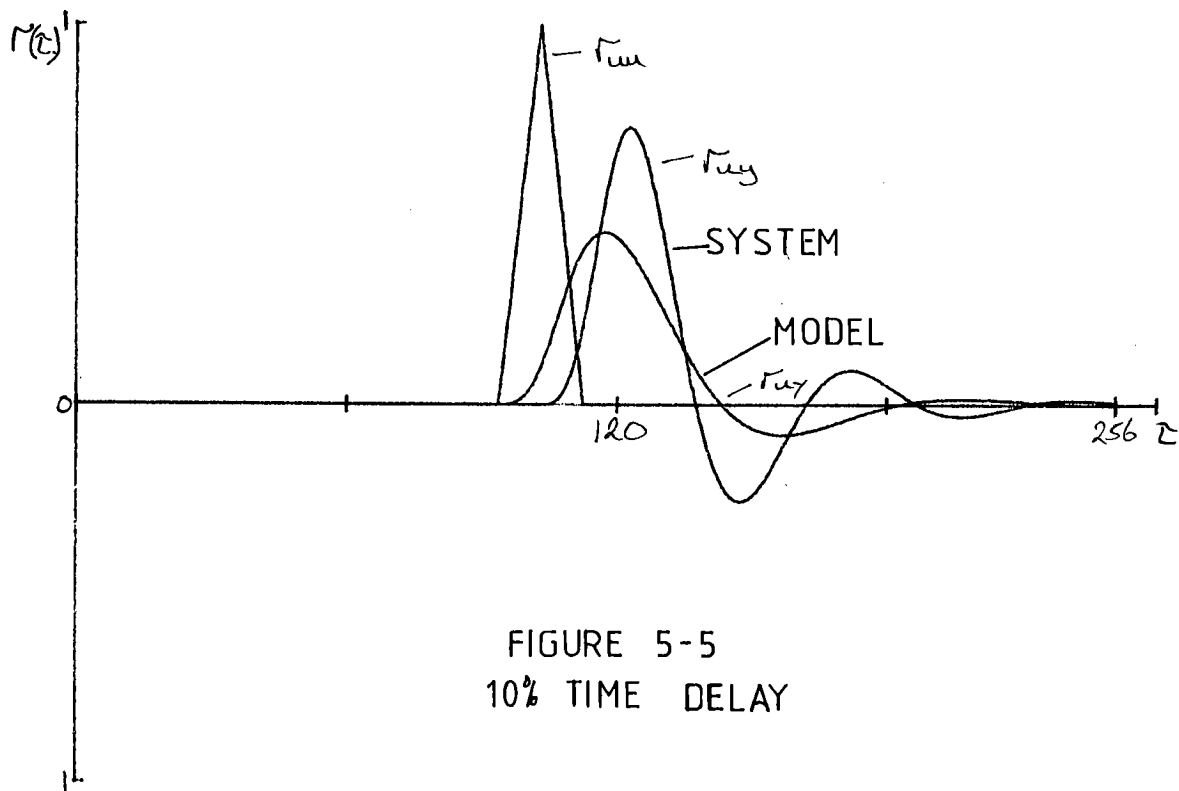
FIGURE 5-5
10% TIME DELAY
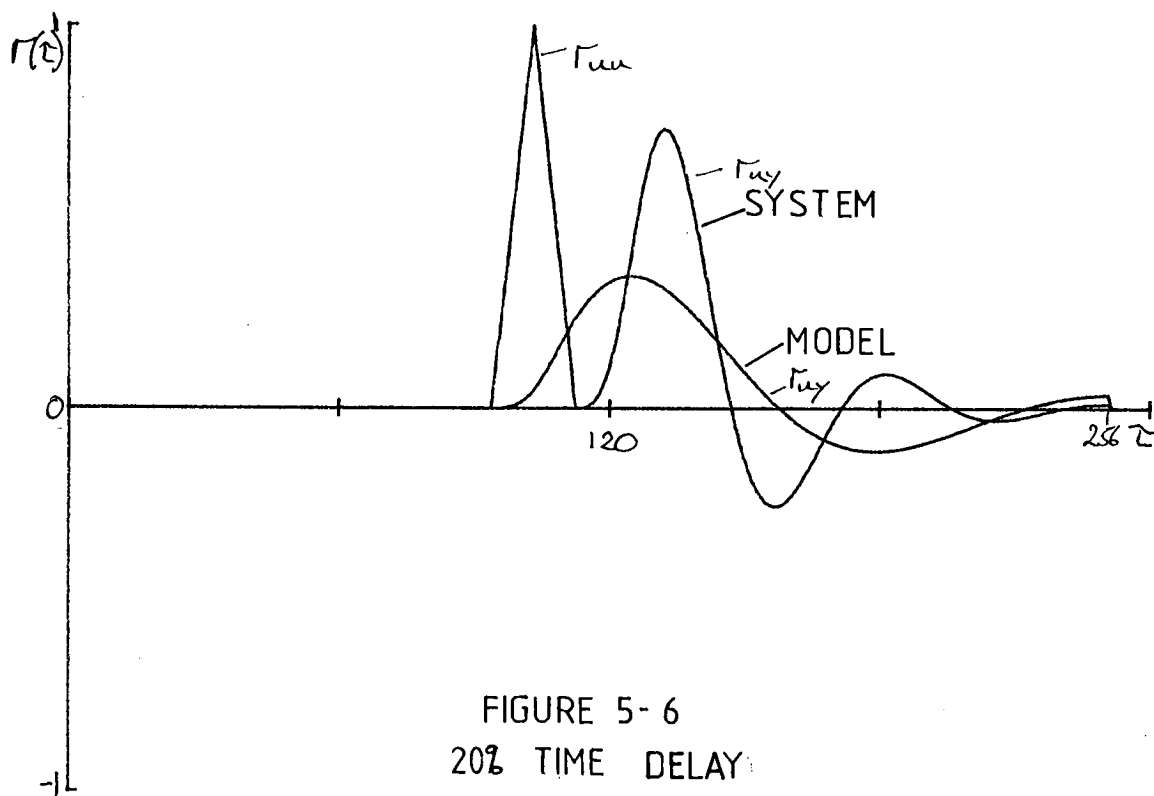
FIGURE 5-6
20% TIME DELAY

168
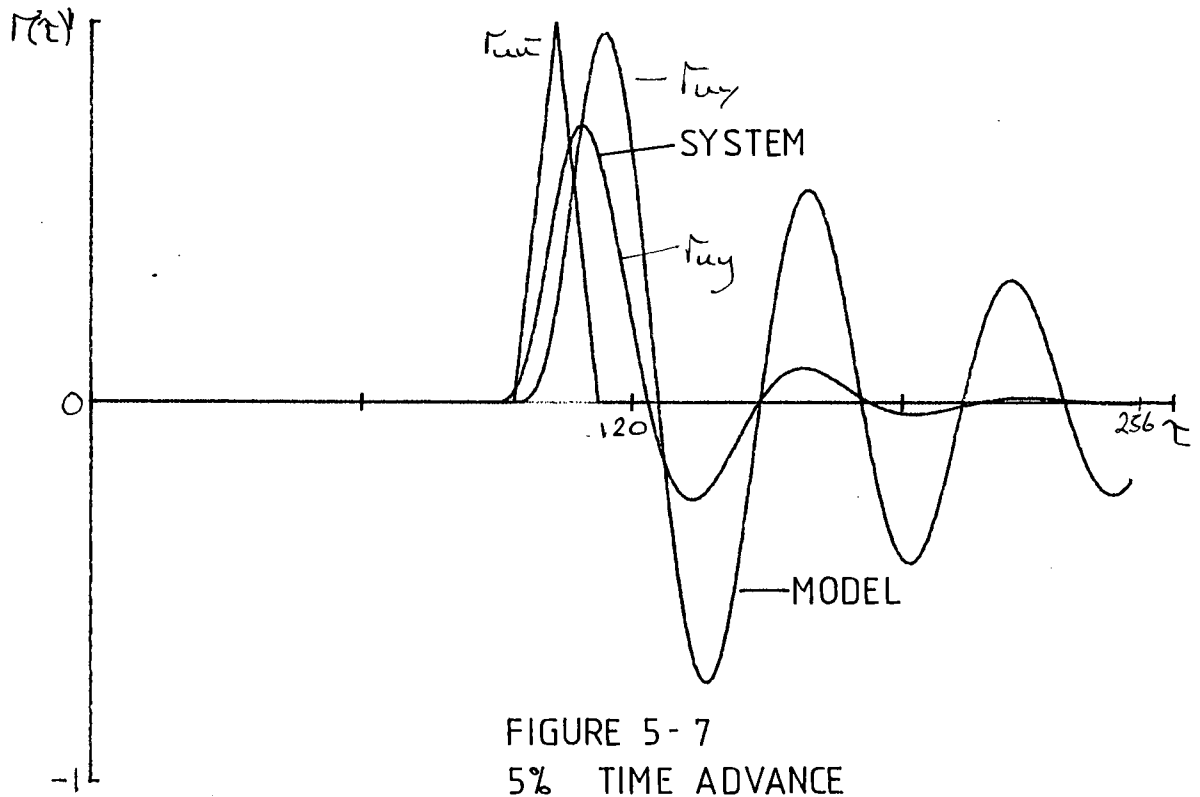
FIGURE 5-7
5% TIME ADVANCE
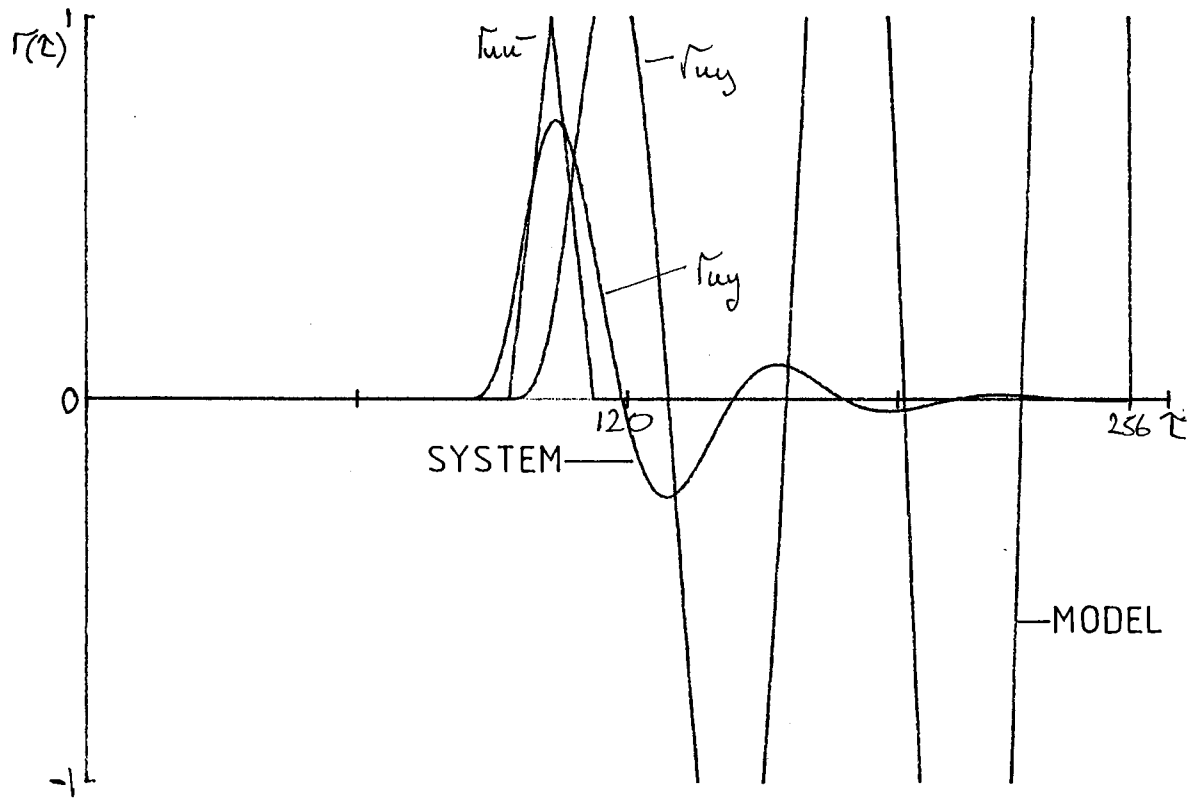
FIGURE 5-8
10% TIME ADVANCE

169

The effect on the parameter estimates of introducing a time advance or time delay into the system is shown on figure 5.9.

When time delay is introduced the coefficients, gain, damping factor and natural frequency all fall in value. This is reflected in the response of the model, which becomes increasingly sluggish. This is shown in figures 5-5 and 5-6. From figure 5-9 it can be seen that there is no characteristic, in the curves, which actually defines the zero time delay accurately. This meant that a manual winding out of the time delay had to be adopted, and the accuracy depended on judgment.

5-1-3 Model Order

A diesel engine is a complex system and has a complex transfer function. It was felt that any attempt to identify such a complicated transfer function with a large number of parameters would not be successful. Initial tests to obtain the cross correlation function of the diesel engine had shown that the cross correlation "impulse response" approximated to a second order response with a time delay. It was felt that greater accuracy of identification could be achieved by limiting the model to a second order transfer function.
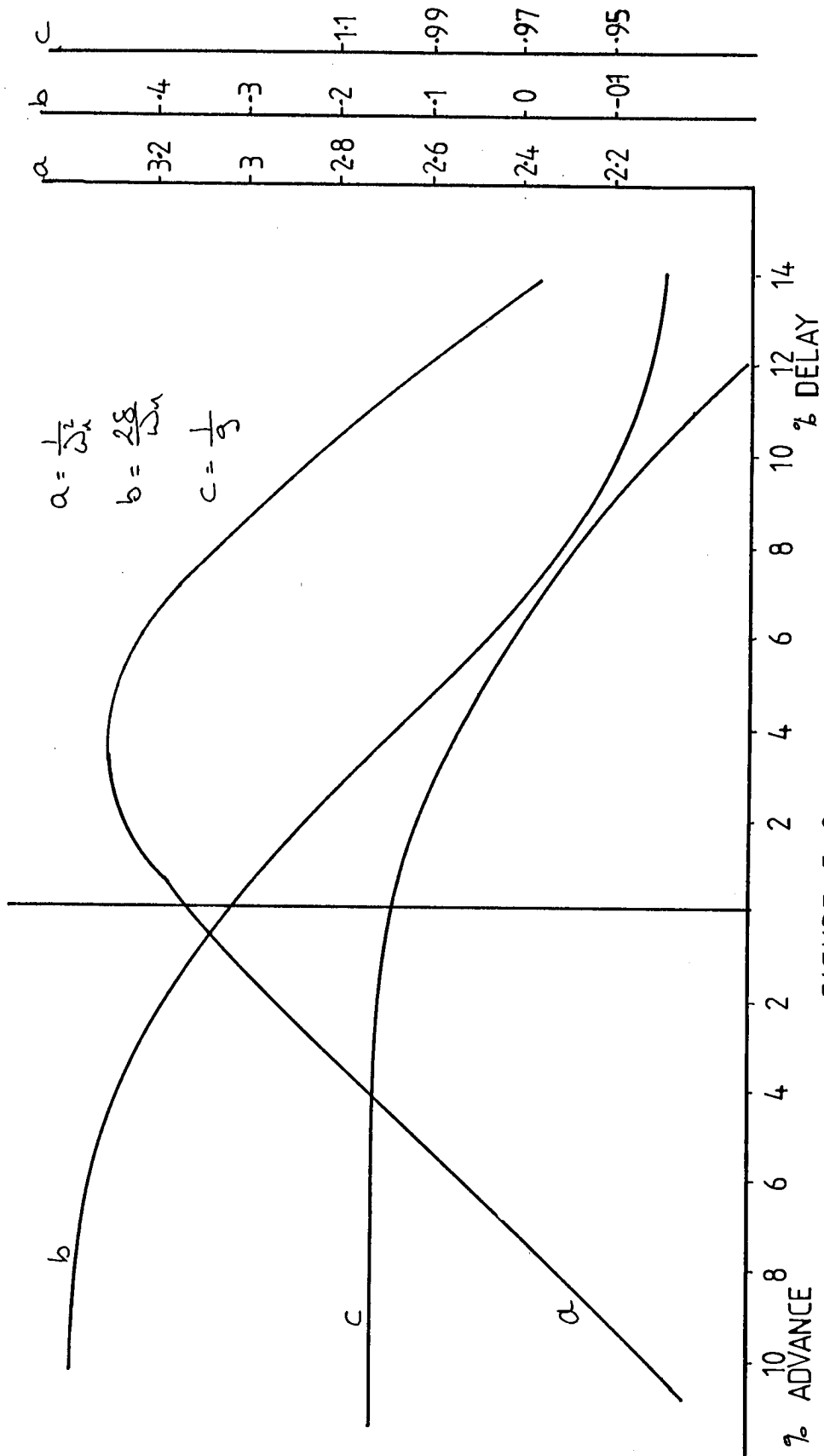
FIGURE 5-9

PARAMETER ESTIMATES IN THE PRESENCE OF TIME DELAY / ADVANCE

$$a = \frac{1}{\omega^2}$$

$$b = \frac{2\delta}{\omega_n}$$

$$c = \frac{1}{g}$$

171

In using a low order model, errors will be introduced. In section 3.3.1 it has been shown that these errors can be thought of as a distortion of the input signal. To study these errors, higher order simulated systems 3rd and 4th order, were modelled with a 2nd order model, and compared with the system responses. For the third order system a pole pair was located at $(-1,\pm3)$ on the pole zero map and a single pole located at various points on the real axis. The further out the single pole the less effect it has on the system and the model response should resemble that of the dominant pair. As the pole is brought closer, the greater the influence it will have, and the parameter estimates will change. Figure 5.10 shows this effect on the s^2 and s parameter estimates as the pole is brought in towards zero. Figure 5.11 shows the changes in the coefficients, natural frequency, damping factor and gain.

It is interesting to note that the gain remains flat and unaltered until the first order component begins to dominate, then falls away rapidly. As would be expected the damping factor increases as the first order term becomes prominent. The natural frequency tends to zero as the pole becomes the dominant feature. The series of figures 5.12 to 5.17 show the simulated system and model response to a triangular impulse. In figure 5.12 the pole is located at -20. At this location the poles'
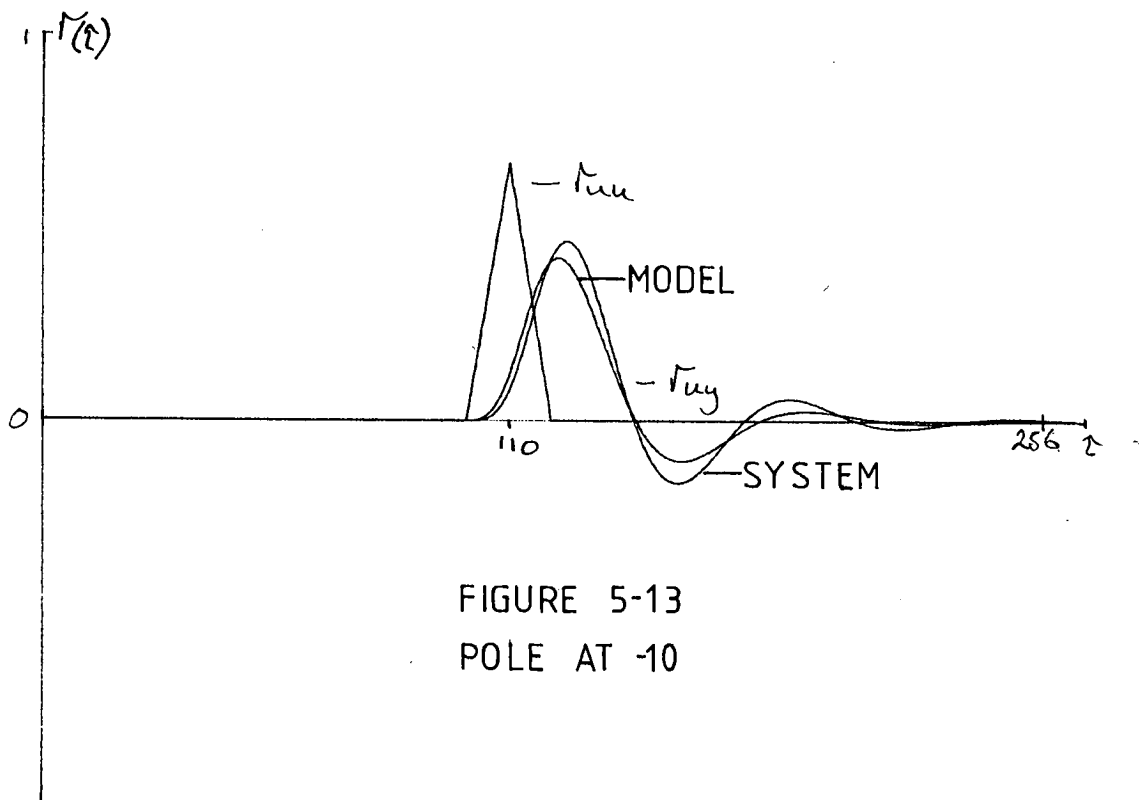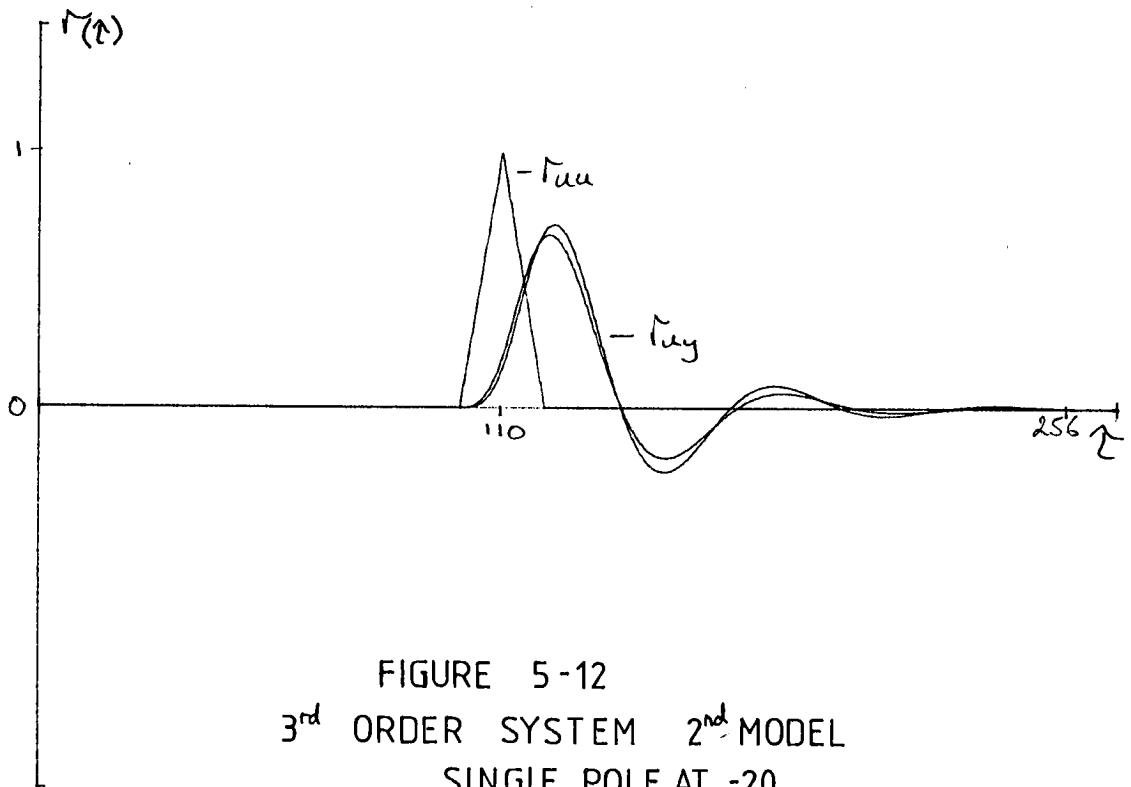
172

SYSTEM

MODEL

POLE POSITION

COEFFICIENT OF S^2 TERM

SYSTEM

MODEL

POLE POSITION

COEFFICIENT OF S TERM

FIGURE 5-10

173
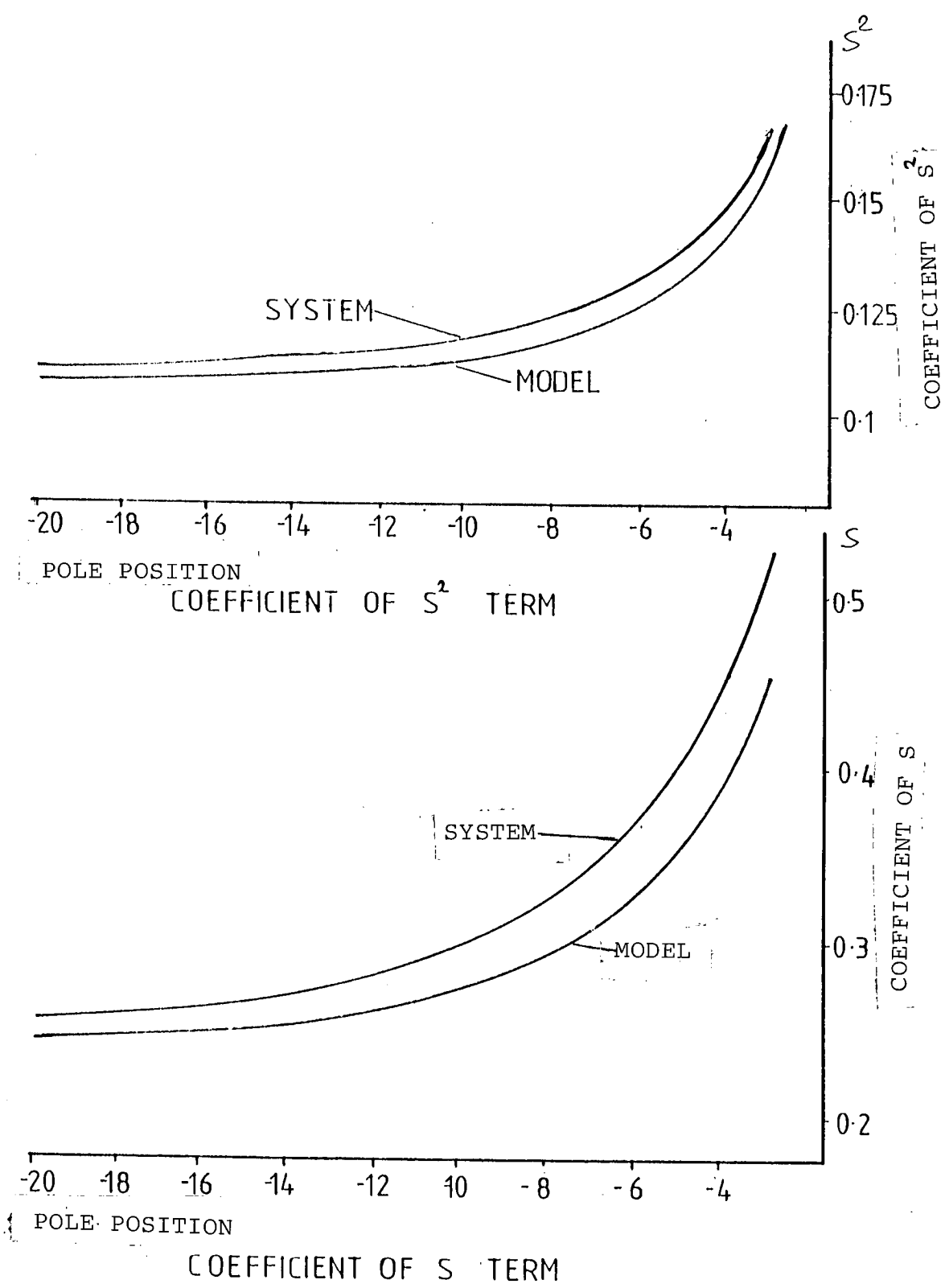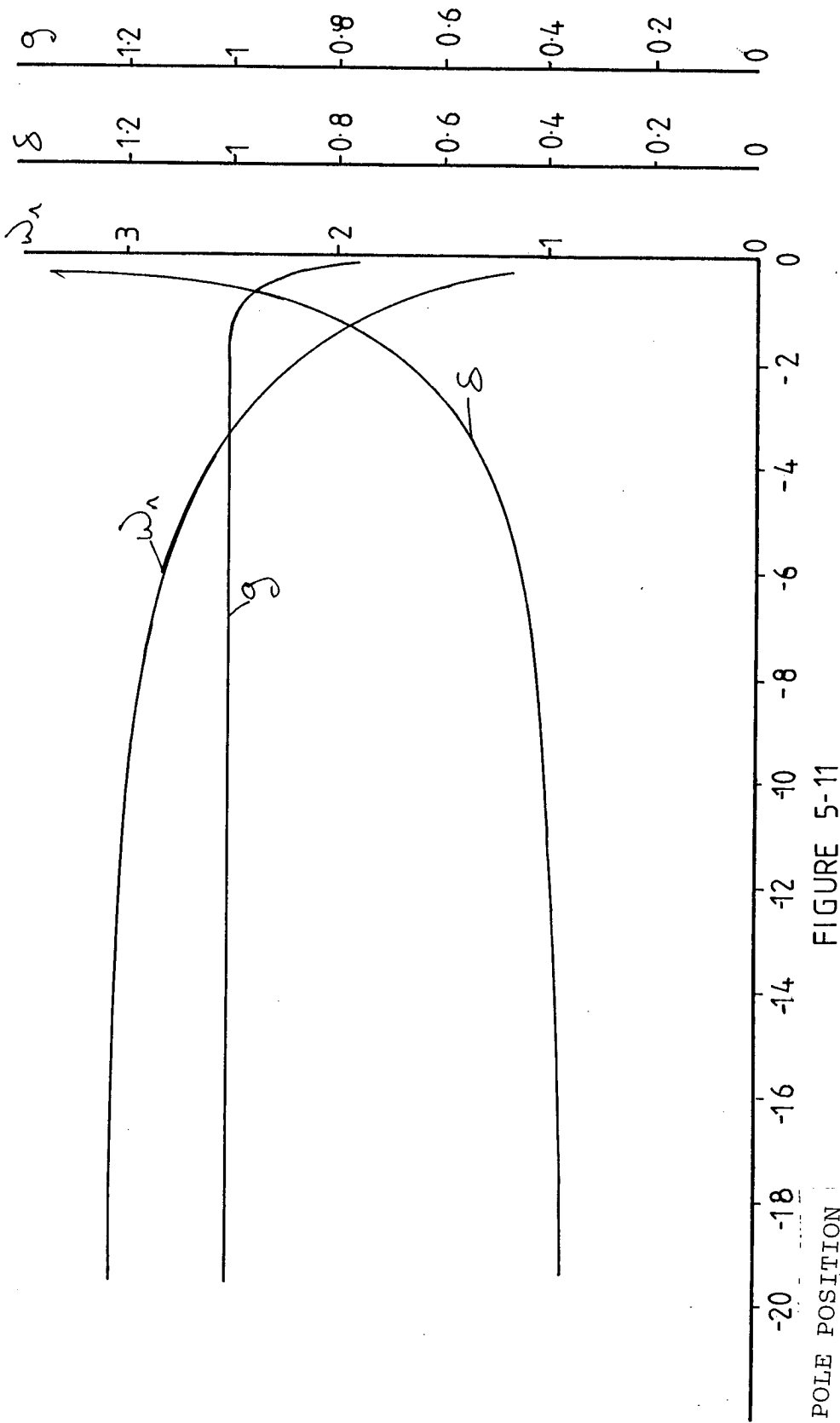
FIGURE 5-11

2nd ORDER MODEL PARAMETER ESTIMATES
OF A 3rd ORDER SYSTEM

POLE POSITION

174

FIGURE 5-12
3rd ORDER SYSTEM 2nd MODEL
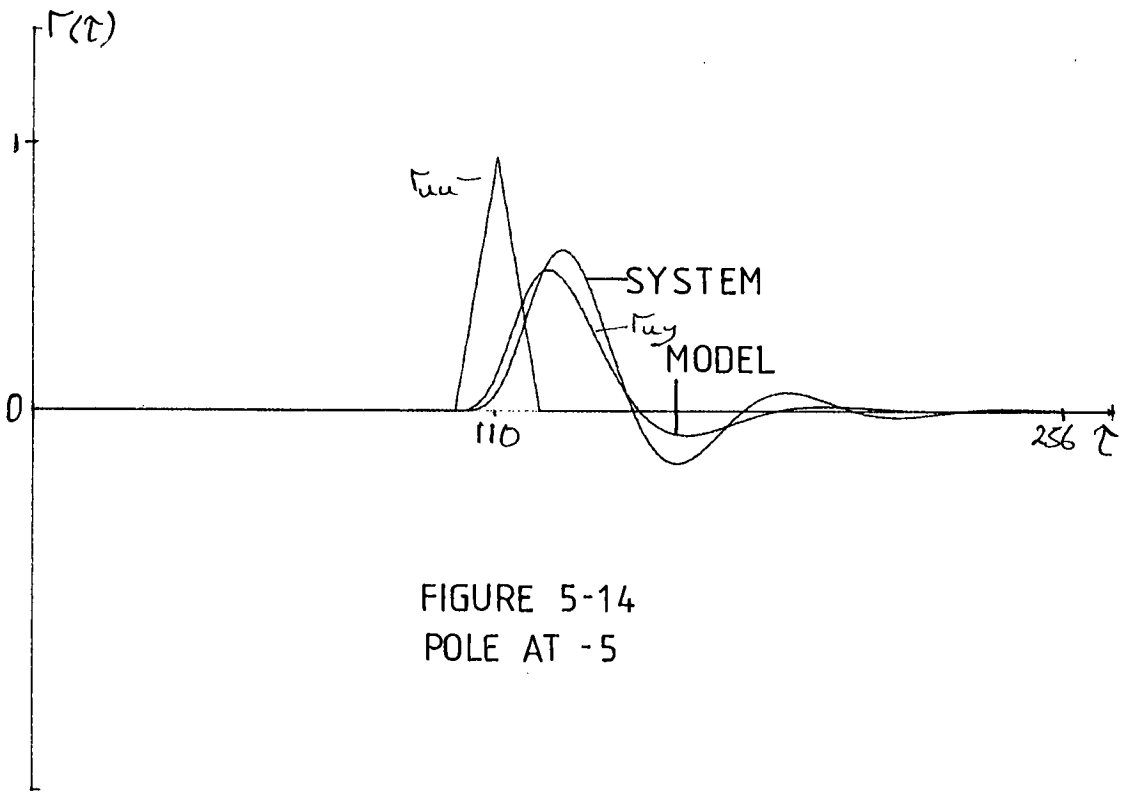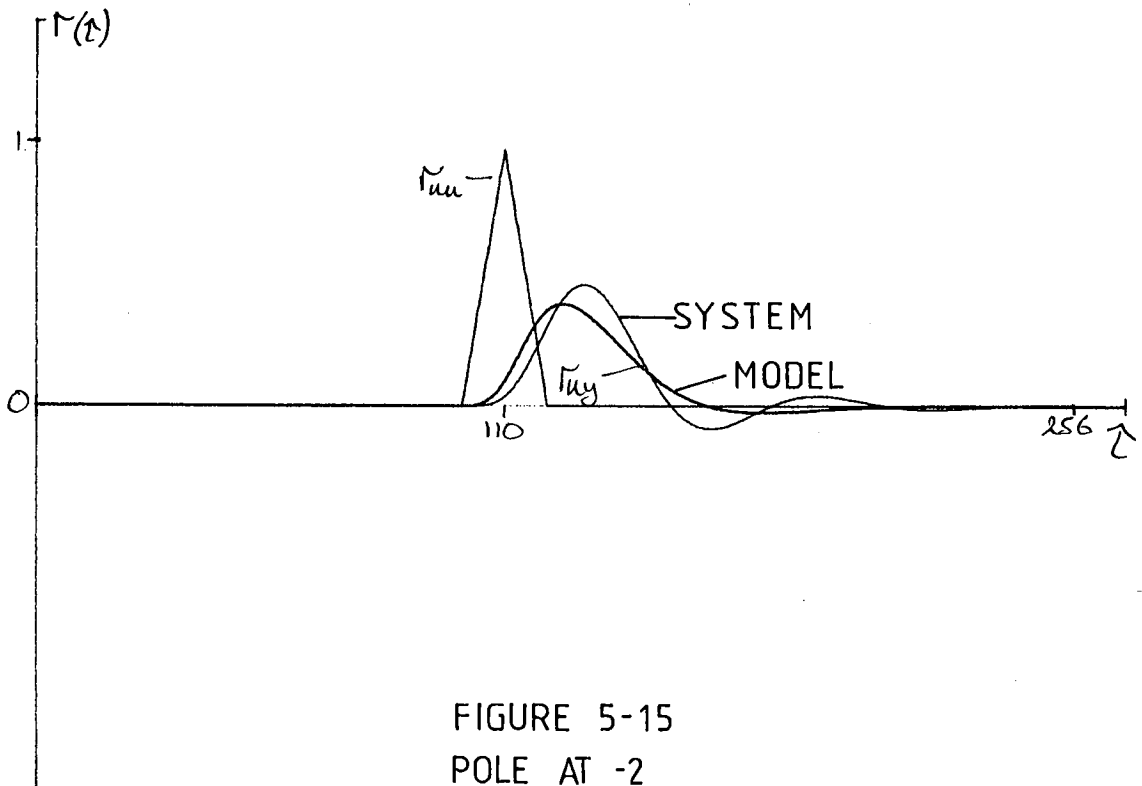SINGLE POLE AT -20

FIGURE 5-13
POLE AT -10

FIGURE 5-14
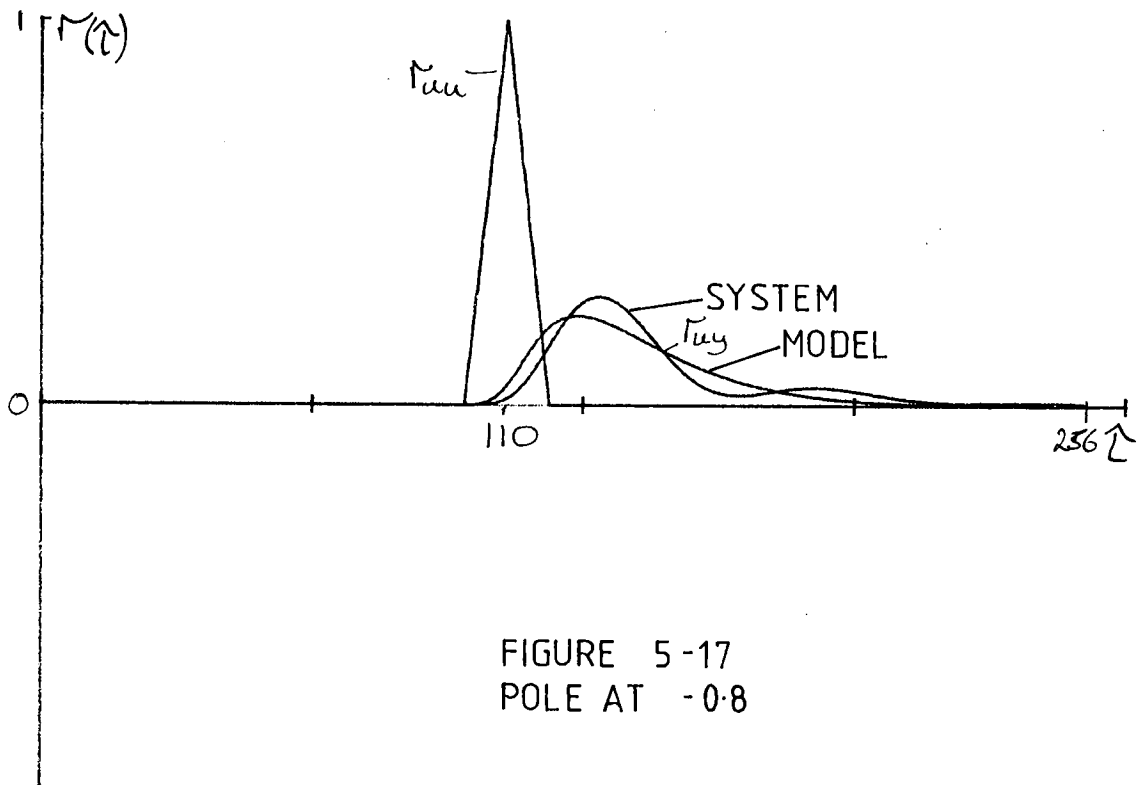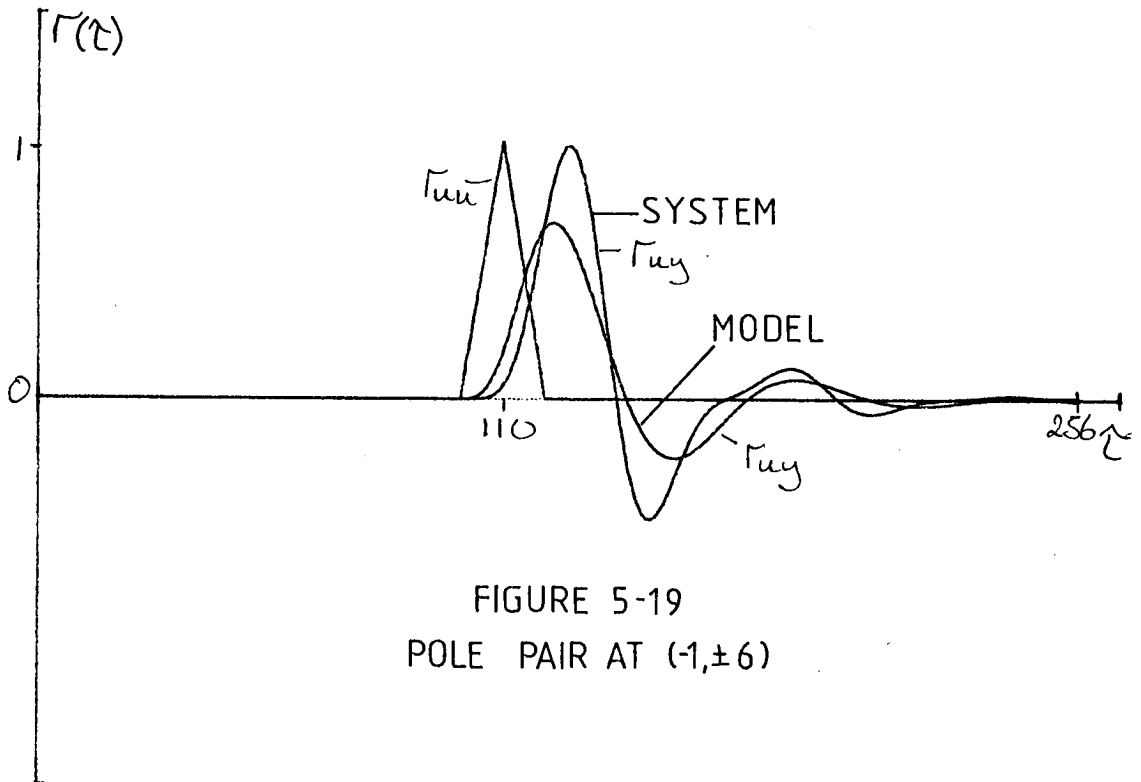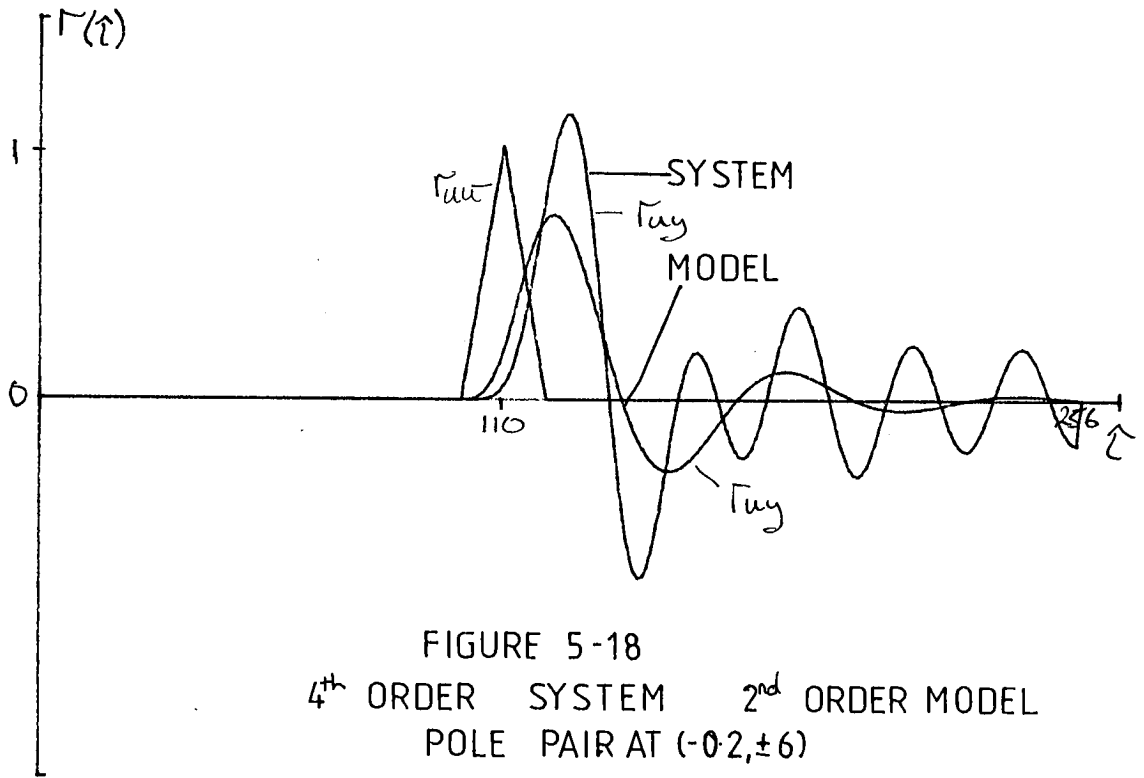POLE AT -5

FIGURE 5-15
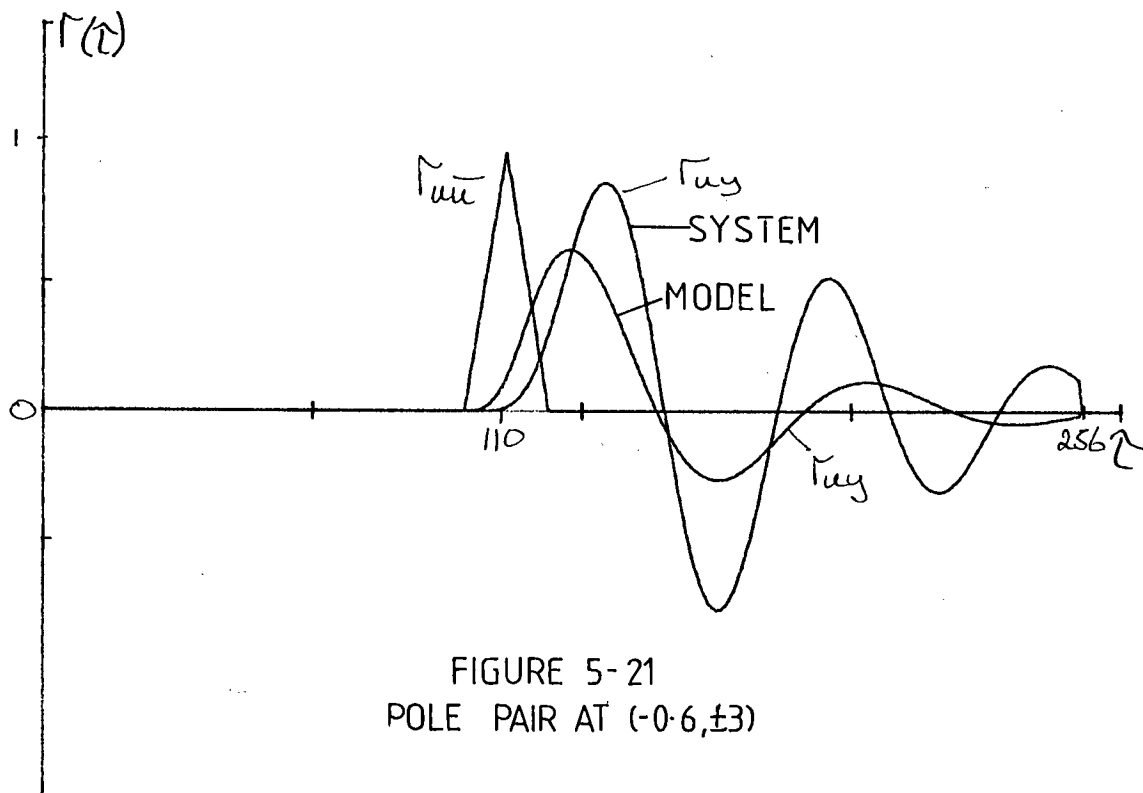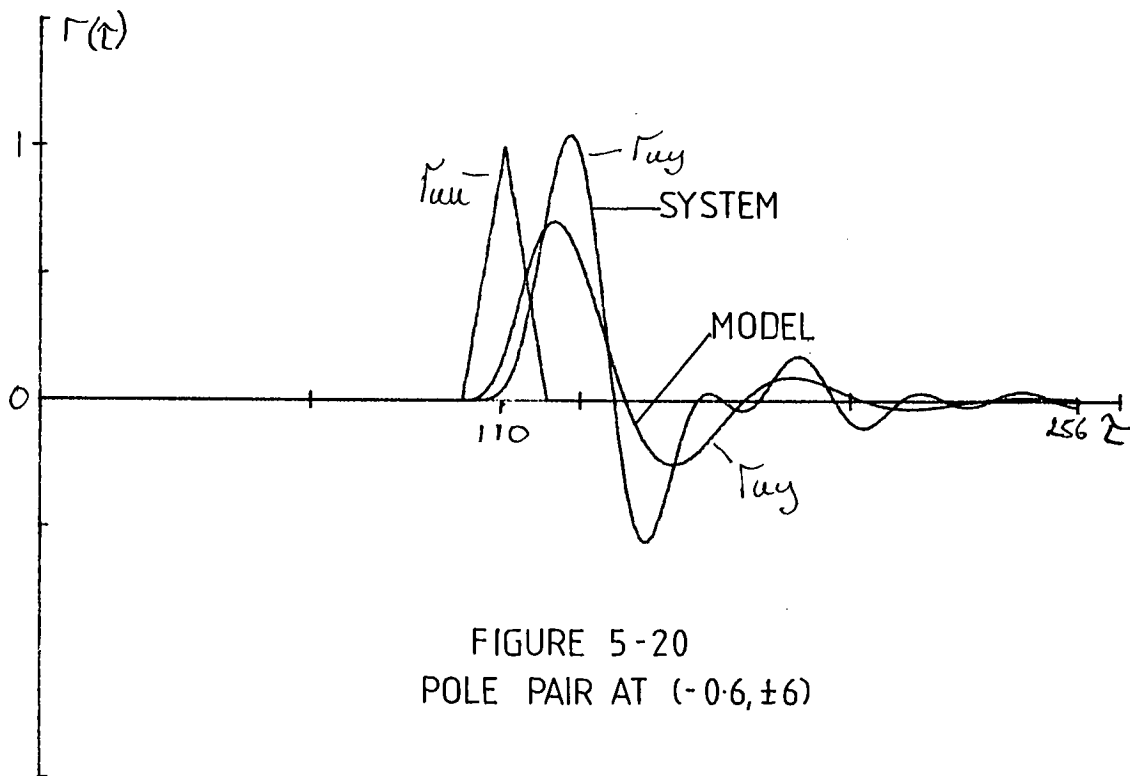POLE AT -2

176
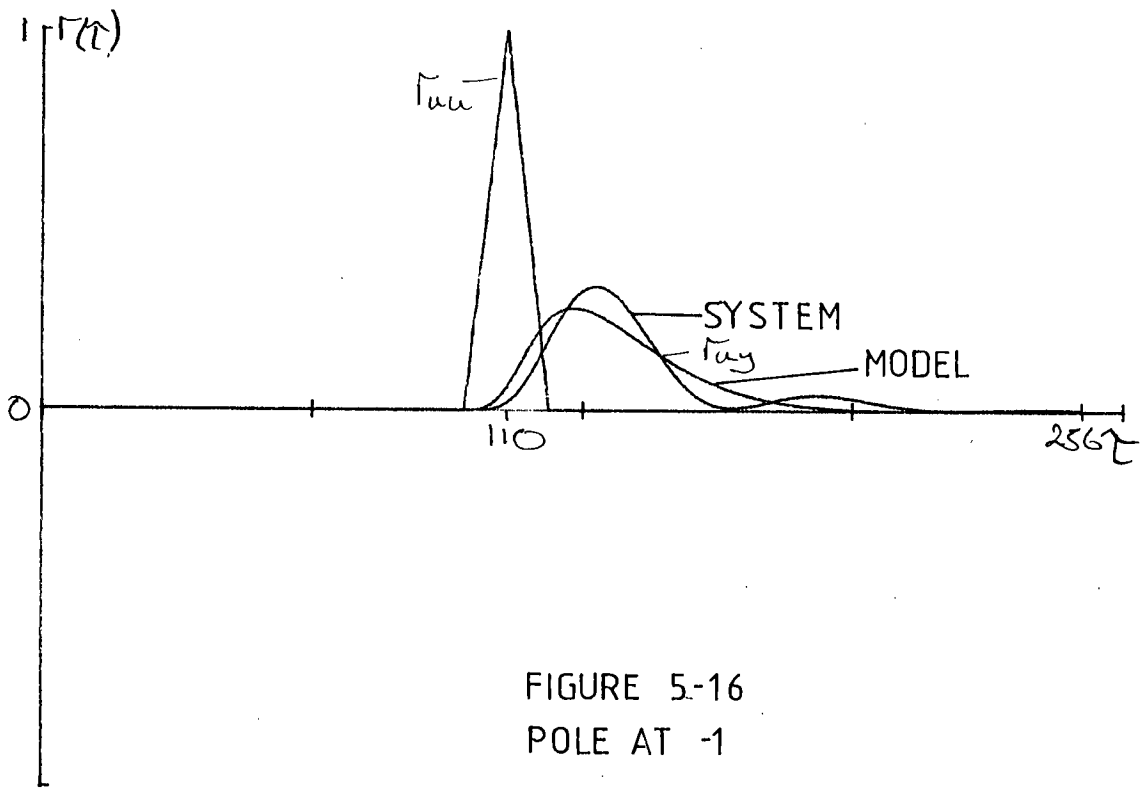
FIGURE 5-16
POLE AT -1

FIGURE 5-17
POLE AT -0·8

177

influence should be negligible, being 20 times further out from the imaginary axis than the pole pair. However there is still some influence exerted by the pole, indicating that the modulating function method is sensitive to small changes in the parameters. Figure 5.13 is a similar plot with the pole at -10, again the influence of the pole, which should be negligible, can be clearly seen. At -5, figure 5-14, the pole is clearly affecting the estimates, with the model becoming more sluggish than the system. This is to be expected as figure 5-11 shows. The parameter estimates have started to show large deviations from the true values, figure 5-10. As the pole reaches -2, figure 5.15, the effect of the pole is very significant with the model only roughly approximating the system. In figures 5.16 and 5.17 the model begins to look first order and only approximates the system by a kind of best fit.

Similarly a fourth order system can be modelled by a second order model. A second order pole pair is located at (-1,+3) on the pole zero map and a second pole pair is introduced at various positions around the pole zero map. The effect of this is shown in the series of figures 5-18 to 5-21. This series of figures show that the MMFM model adopts a best fit to the system, extracting what seems to be the dominant pole pair at (-1,+3). If each of the model responses in figures 5-18

FIGURE 5-18

4ᵗʰ ORDER   SYSTEM   2ⁿᵈ ORDER MODEL

POLE PAIR AT (-0·2, ±6)

FIGURE 5-19

POLE PAIR AT (-1, ±6)

FIGURE 5-20
POLE PAIR AT (-0.6, ±6)

FIGURE 5-21
POLE PAIR AT (-0.6, ±3)

180

to 5-21 are compared with the the response of a second
order system, located at (-1±3),figure 5-12, there is a
strong similarity between them.   However errors are
introduced as   the   theory   predicts,   and   the   model
responses are not exactly that of the second order
system.

It is interesting to compare a measured second order
model response with the response of the actual third
order system, truncated to second order.   For a third
order system as shown below.

$$G(s) = \frac{1}{a_3 s^3 + a_2 s^2 + a_1 s + a_0}$$

the response of the truncated system

$$G'(s) = \frac{1}{a_2 s^2 + a_1 s + a_0}$$

is compared with the estimated model G''(s)

$$G''(s) = \frac{1}{a_2' s^2 + a_1' s + a_0'}$$

181

This is shown in figures 5.22 and 5.23. It is interesting to note that G´(s) and G"(s) compare favourably until the error between the estimate and the system increases, then the model and truncated system do not approximate so well.

5-2 Analogue Simulations

The purpose of the hardware simulator was to assist in the development of a correlation routine that could be successfully used in the MMFM. A second order analogue computer, figure 4.2, was used as the simulation of the diesel engine. This had a natural frequency of approximately 2Hz, which corresponds to the diesel´s natural frequency. Although it was not possible to simulate the time delay with this method, the model still allowed development of the correlation algorithms which would be used, on-site, at Stornoway Power Station.

Initially correlation was accomplished using the normal Relay correlation algorithm with a 256 point shift. The auto and cross correlation of the second
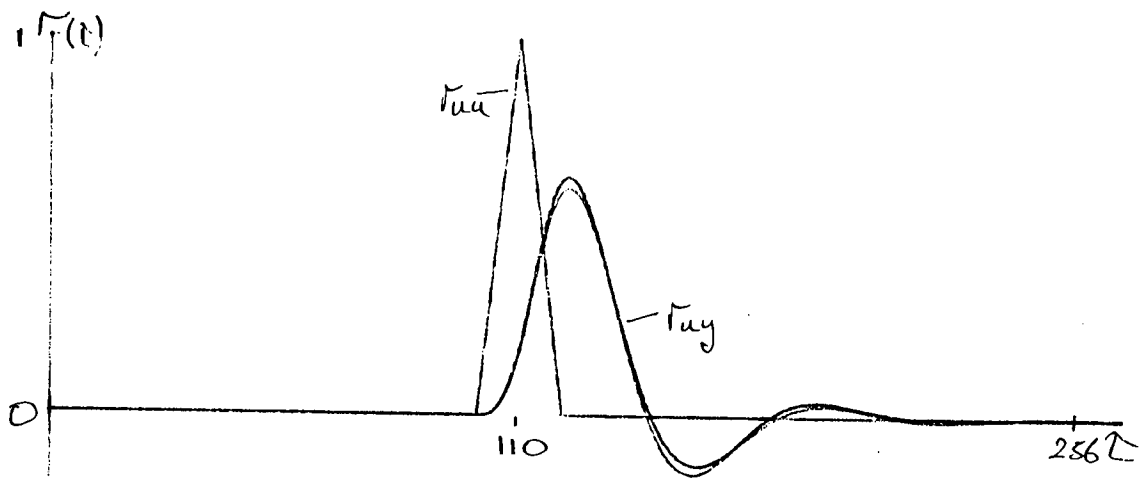
FIGURE 5-22
3rd ORDER SYSTEM
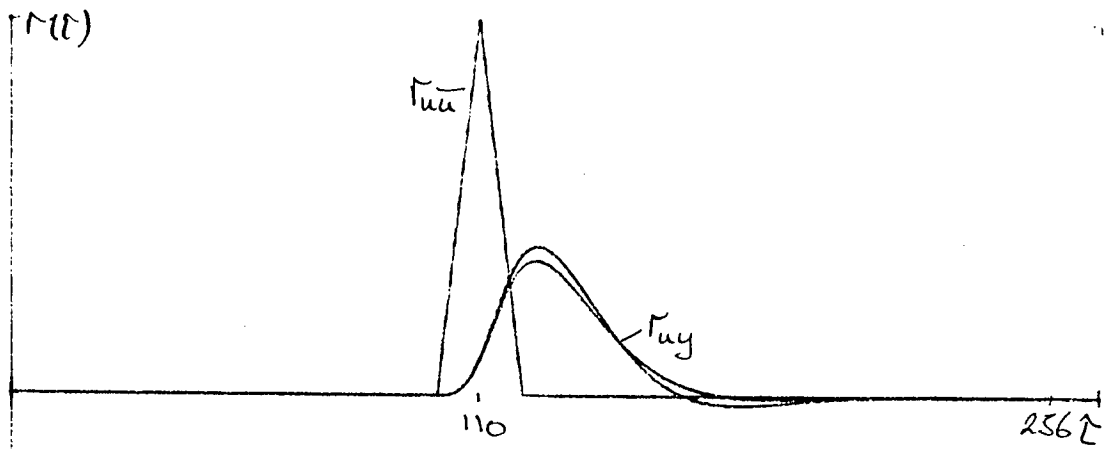TRUNCATED 2nd ORDER MODEL
POLE AT -10

FIGURE 5-23
POLE AT -2

183

order analogue system are shown in figure 5.24. In this
figure quite a degree of variance in the correlation
functions is apparent, especially noticeable in the
region before the auto correlation spike. This region
should be flat, and the imperfect correlation obtained
is due in part to the limited integration time used, and
to the loss of data when Relay correlation is used.
This variance in the correlation functions can result in
a large spread in the parameter estimates.

One method of overcoming this difficulty is to obtain
an average of the correlation functions over a number of
data samples. Figure 5.25 shows the auto and cross
correlation functions averaged five times, and it is
already apparent that the variance has been
significantly reduced. By ten averages, figure 5.26, it
has been all but eliminated. Performing this type of
correlation, then averaging, is slow, and can take up to
an hour to perform twenty averages. A faster correlation
algorithm is needed.

5-2-1 Skip Algorithm

The skip algorithm as described in chapter 3 provides
a faster method of obtaining the correlation function.
This correlation algorithm relies on speeding up the
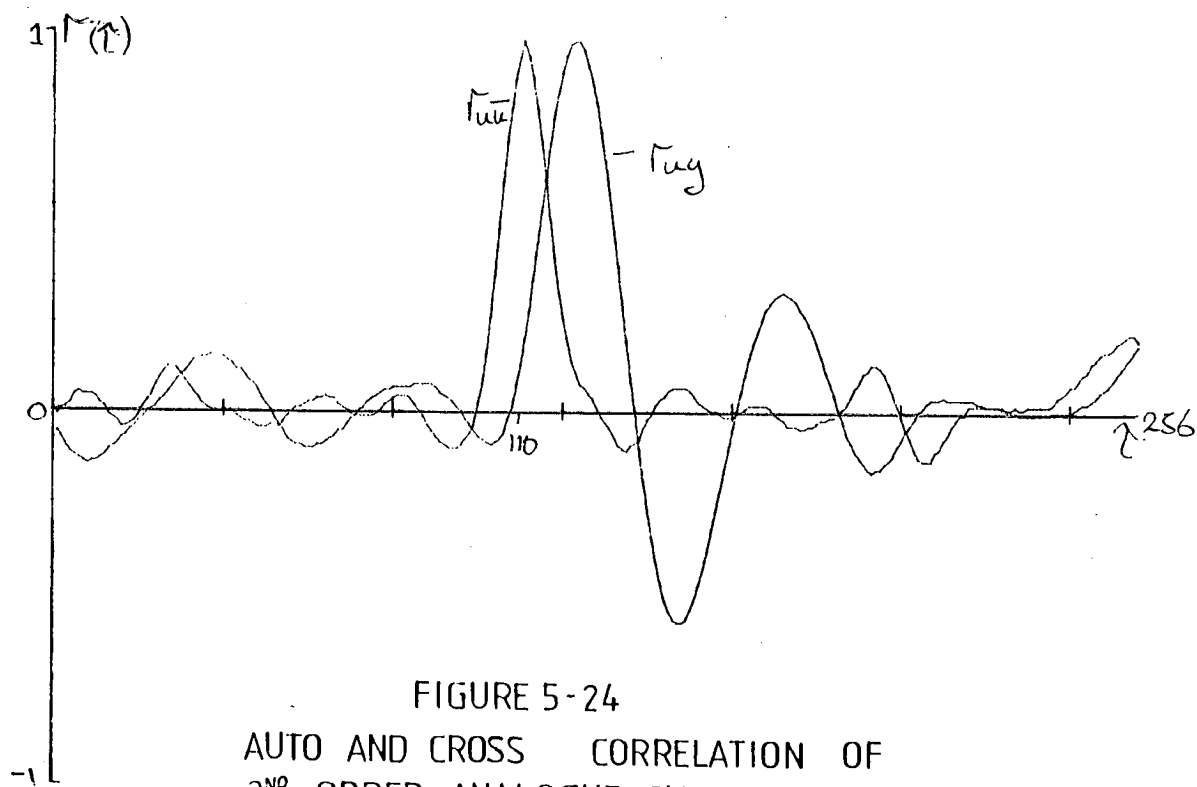computation of the correlation function by skipping

FIGURE 5-24

AUTO AND CROSS   CORRELATION   OF

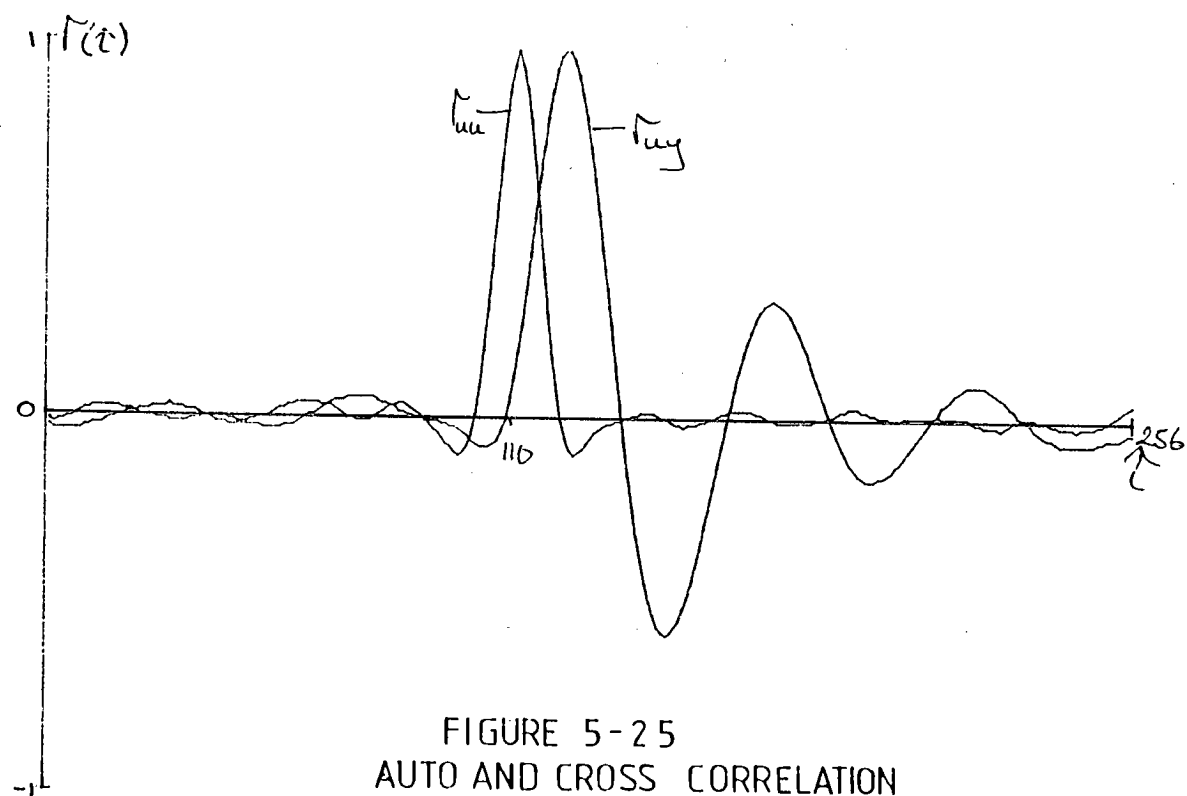2ND ORDER ANALOGUE SYSTEM   USING
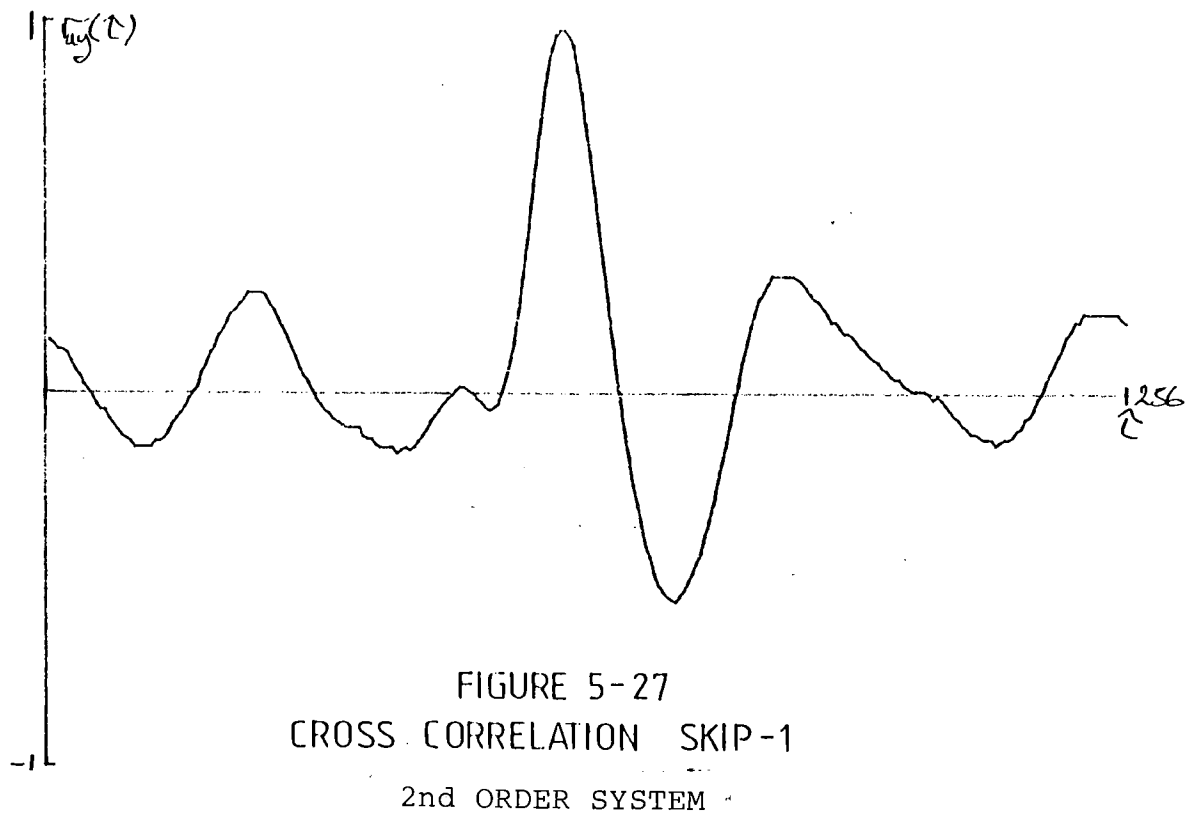
A   PRBS   INPUT   SIGNAL

FIGURE 5-25

AUTO AND CROSS CORRELATION

AVERAGED FIVE TIMES

185
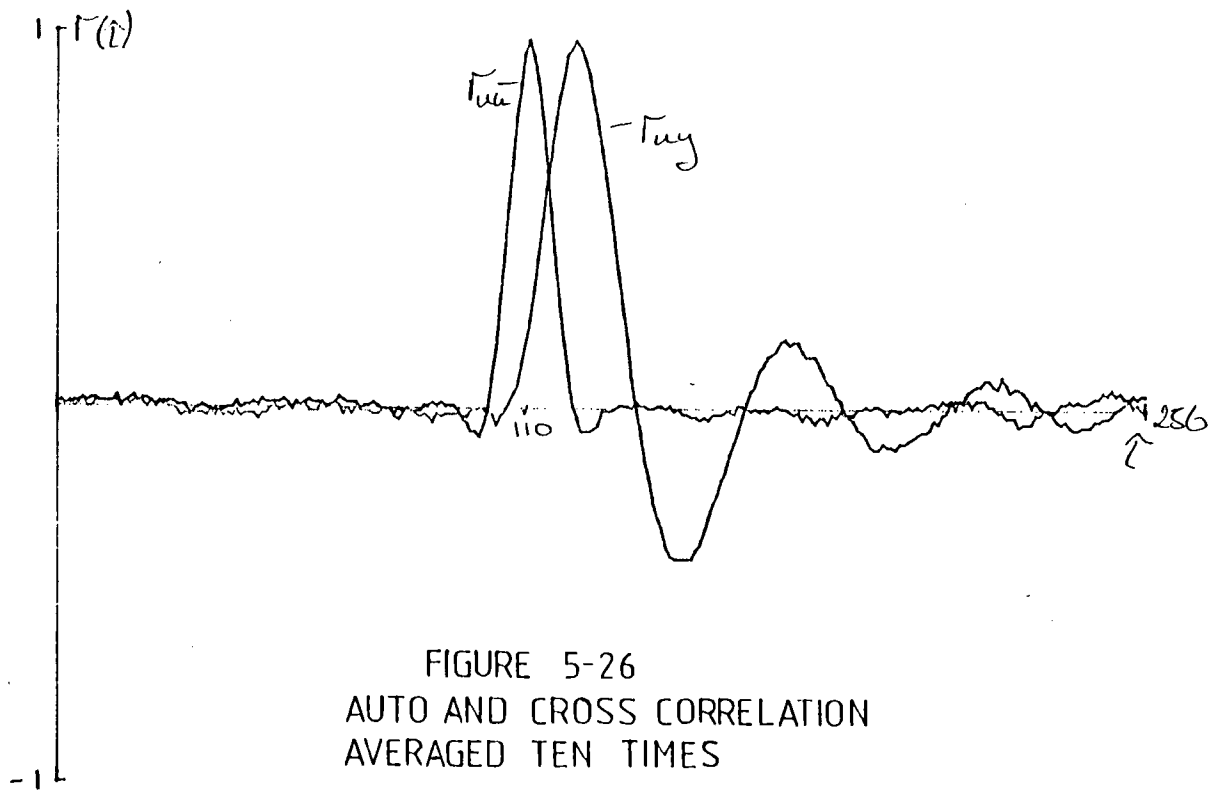
FIGURE 5-26
AUTO AND CROSS CORRELATION
AVERAGED TEN TIMES

FIGURE 5-27
CROSS CORRELATION SKIP-1

2nd ORDER SYSTEM

cross products. For example in a skip of two, only every second cross product would be formed, in a skip of eight only every eighth product would be formed and added to the accumulator, and so on.

This reduction in the data available to form the correlation at a given time shift, will increase the variance of the correlation function. The series of figures 5.27 to 5.34 show the cross correlation function formed using a skip of 2^n n=0,1,2,..,7. Only the cross correlation is shown for clarity, but both the auto and cross correlation functions would be formed using the skip algorithm.

It is apparent from these figures that as the number of skips is increased, the variance on the correlation function increases until it completely breaks down. However, it is interesting to note that the correlation function retains its basic shape up to a skip of sixteen, compare figure 5.27 with figure 5.31. In practice a skip factor of eight was found to give a good result. An average of twenty cross correlations with a skip of eight is shown on figure 5.35, this compares well with figure 5.26.

An analysis of the parameter estimates of a second order analogue system, using the skip algorithim, was
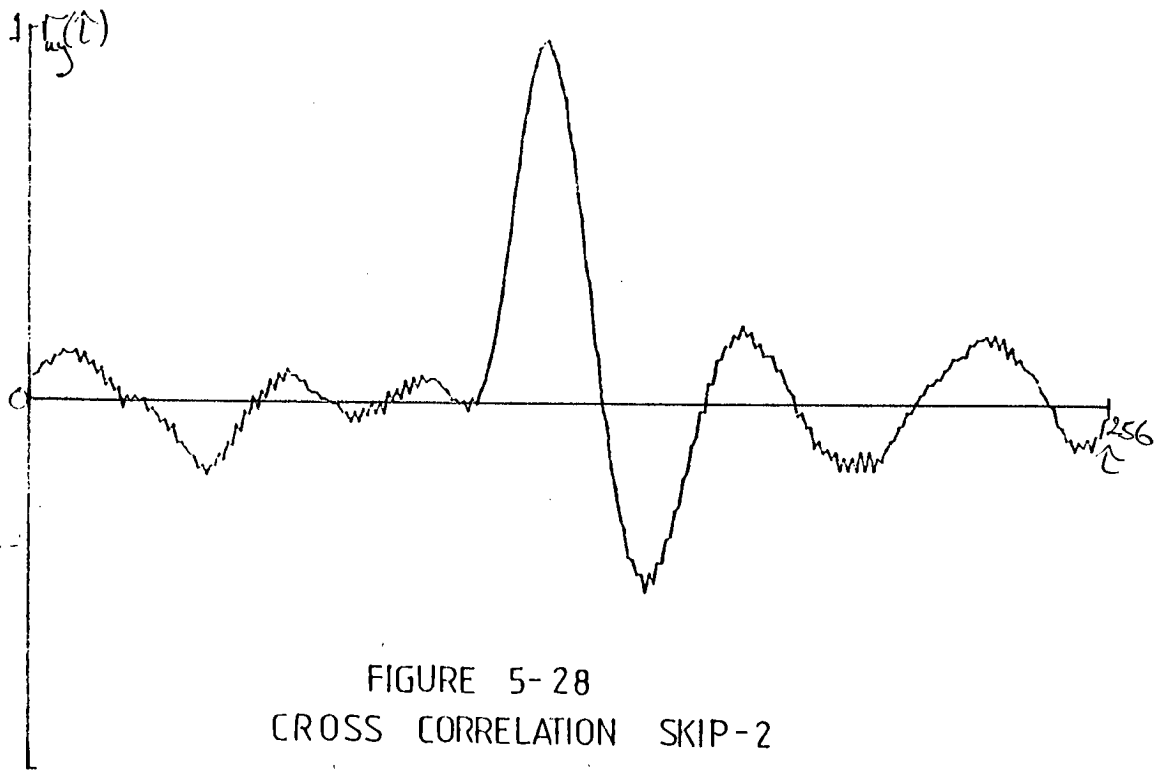
FIGURE 5-28
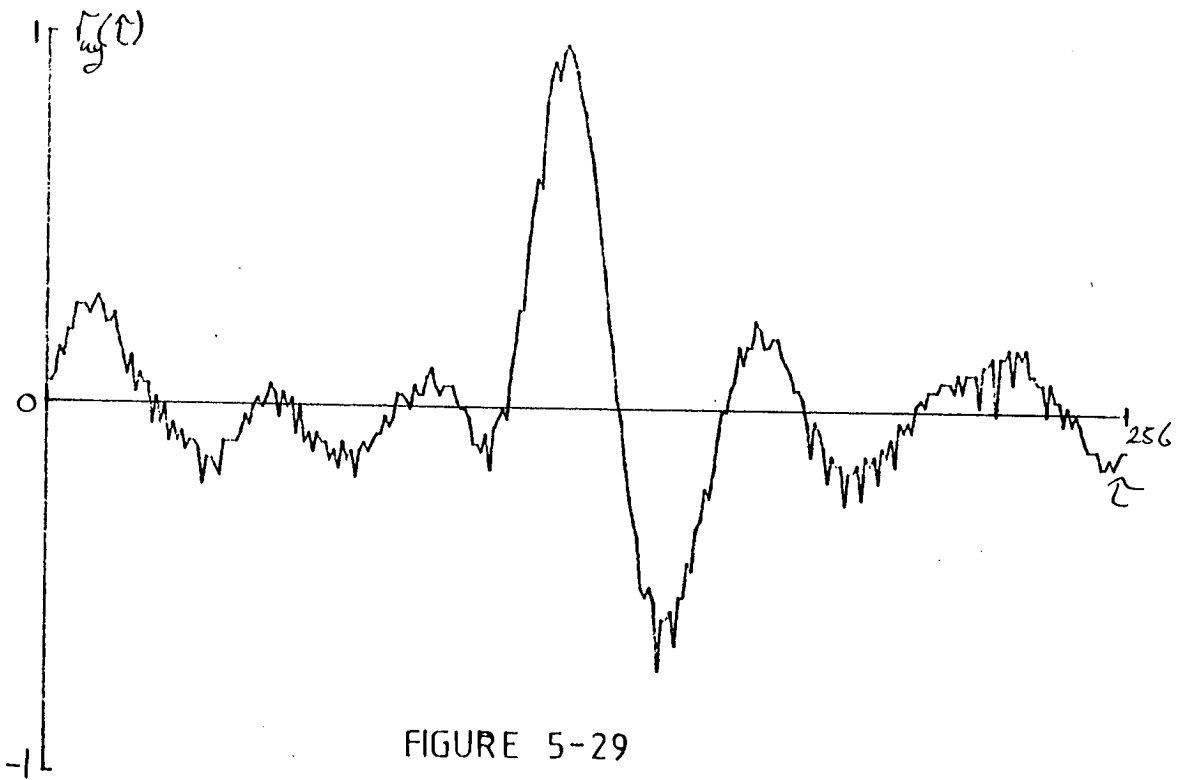CROSS CORRELATION SKIP-2
2nd ORDER SYSTEM

FIGURE 5-29
CROSS CORRELATION SKIP-4
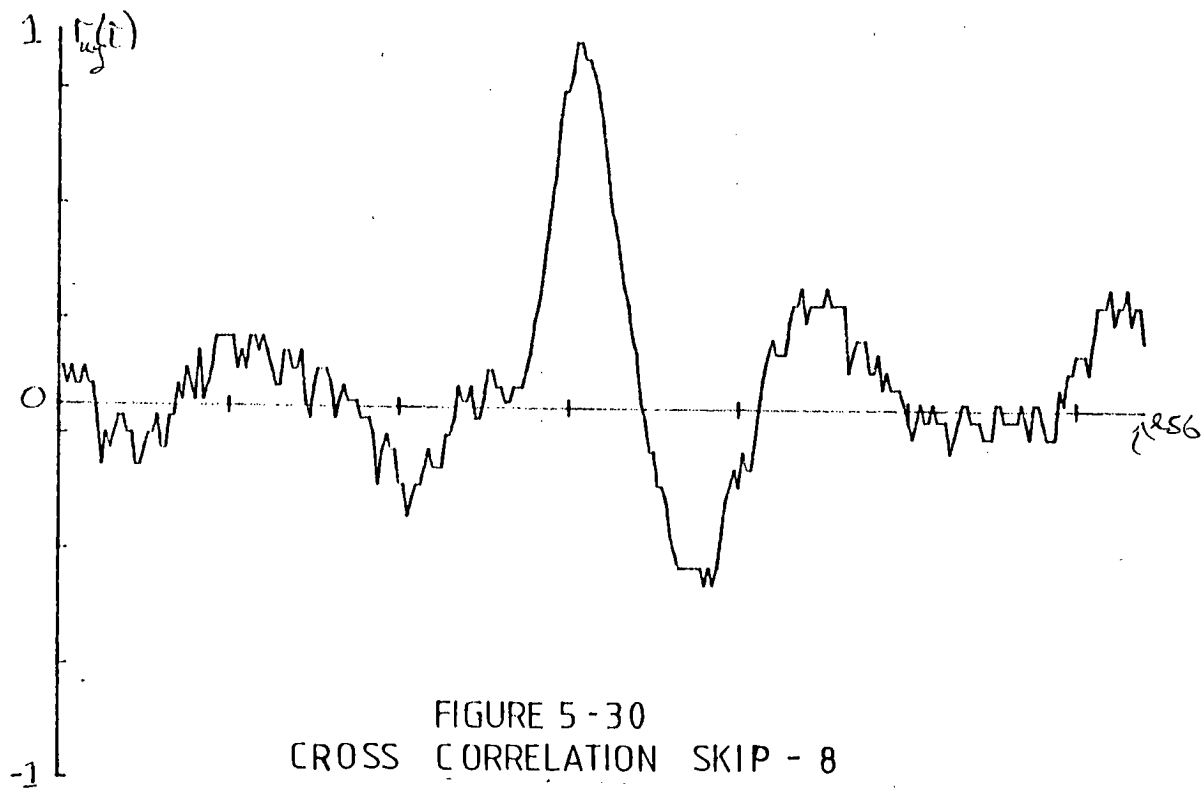2nd ORDER SYSTEM

188

FIGURE 5-30
CROSS CORRELATION SKIP - 8
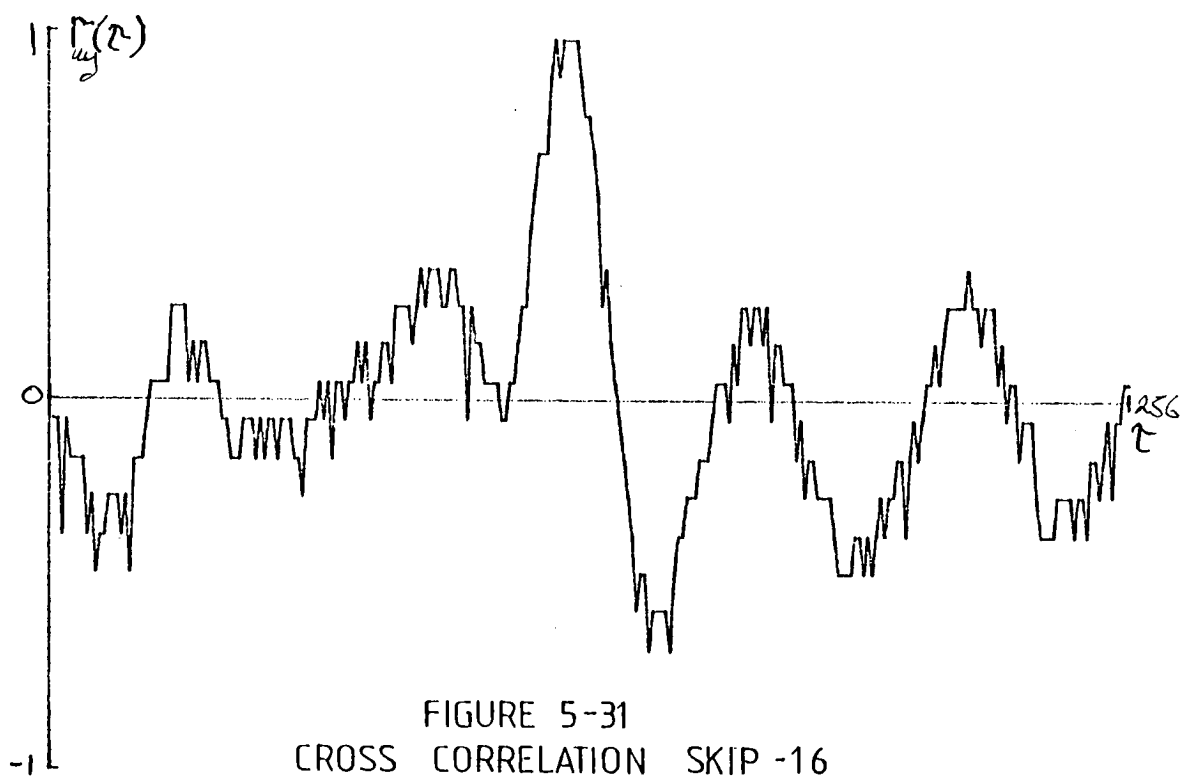2nd ORDER SYSTEM

FIGURE 5-31
CROSS CORRELATION SKIP -16
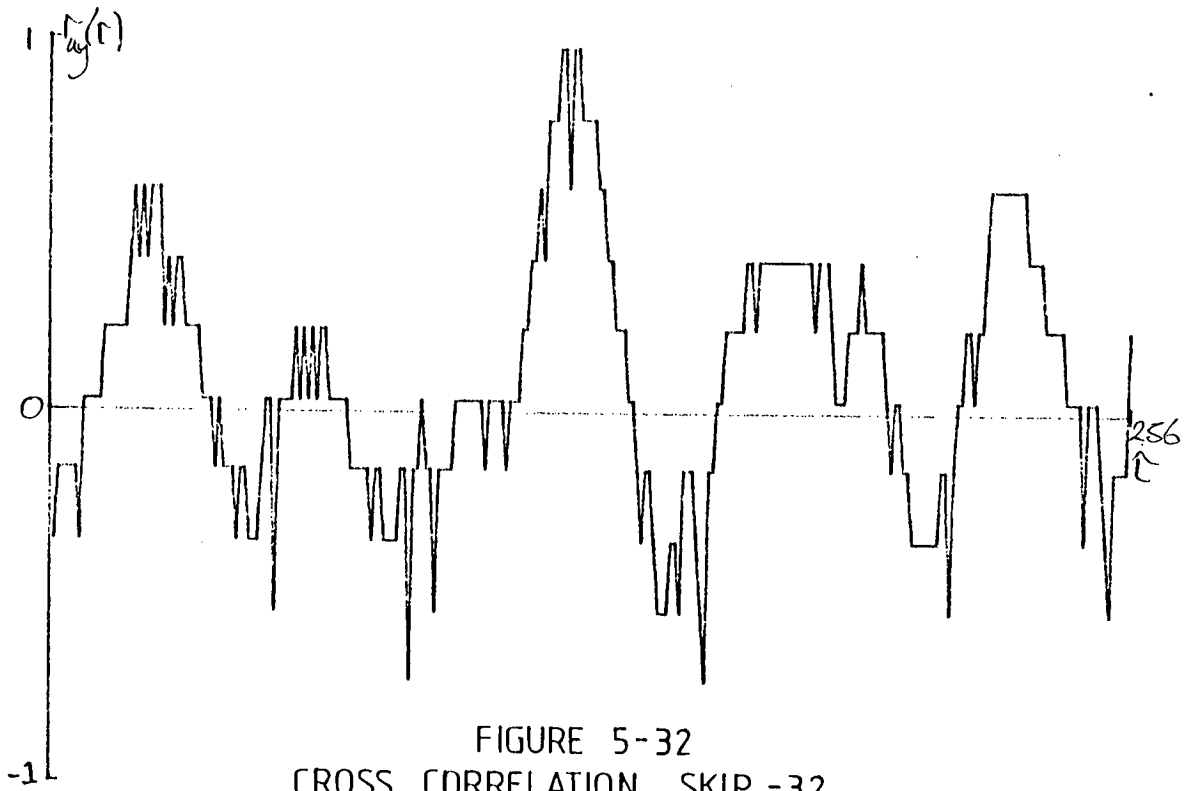2nd ORDER SYSTEM

189

FIGURE 5-32
CROSS CORRELATION SKIP -32
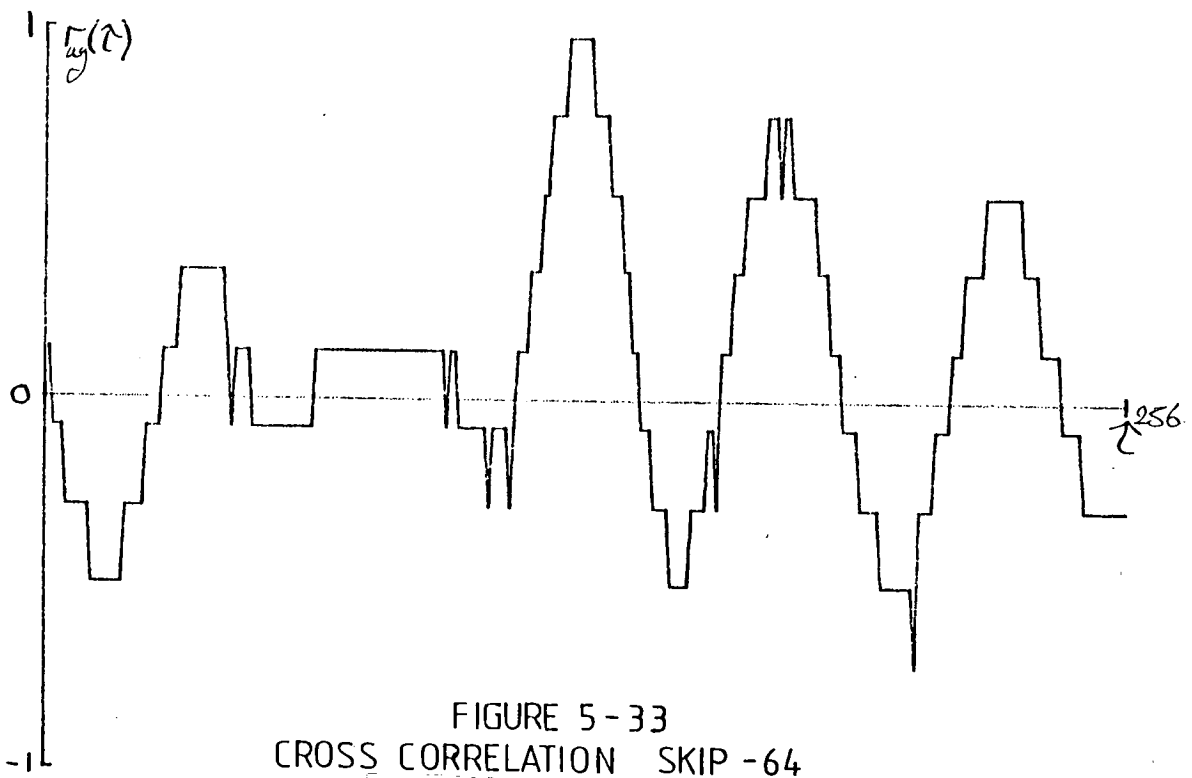
2nd ORDER SYSTEM

FIGURE 5-33
CROSS CORRELATION SKIP -64
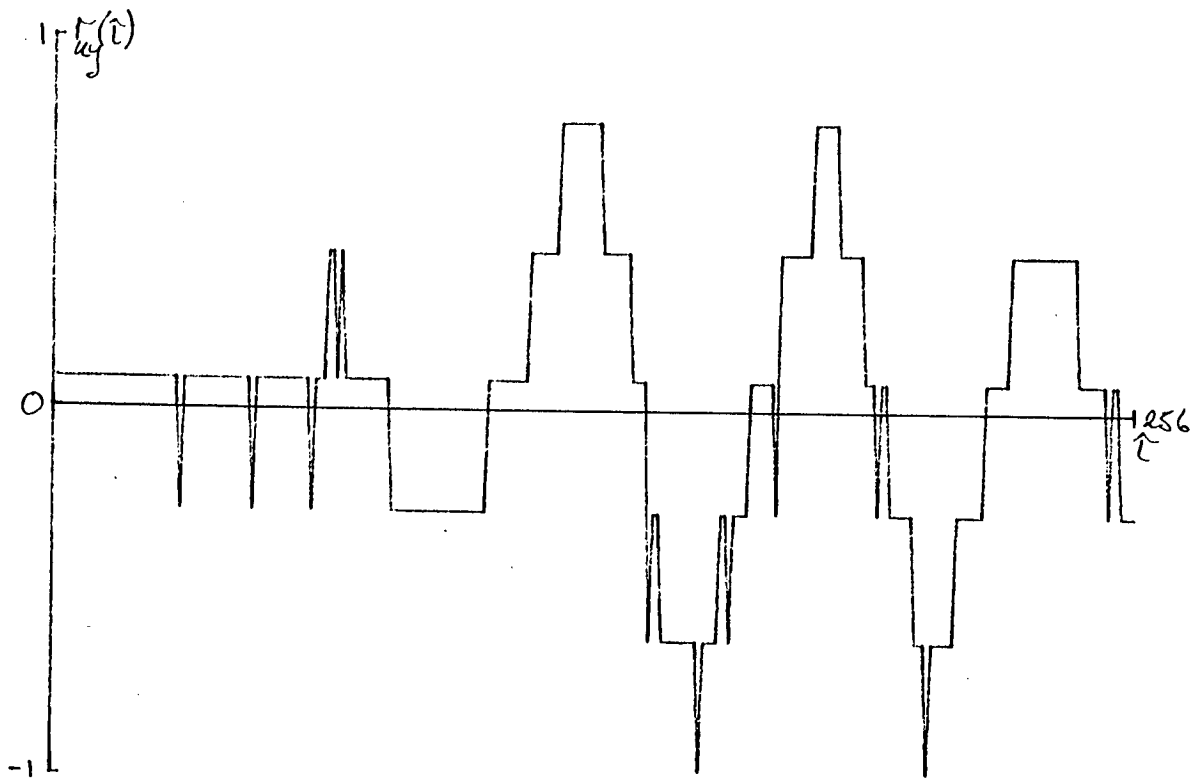
2nd ORDER SYSTEM

190

FIGURE 5-34
CROSS CORRELATION   SKIP-128
2nd ORDER SYSTEM
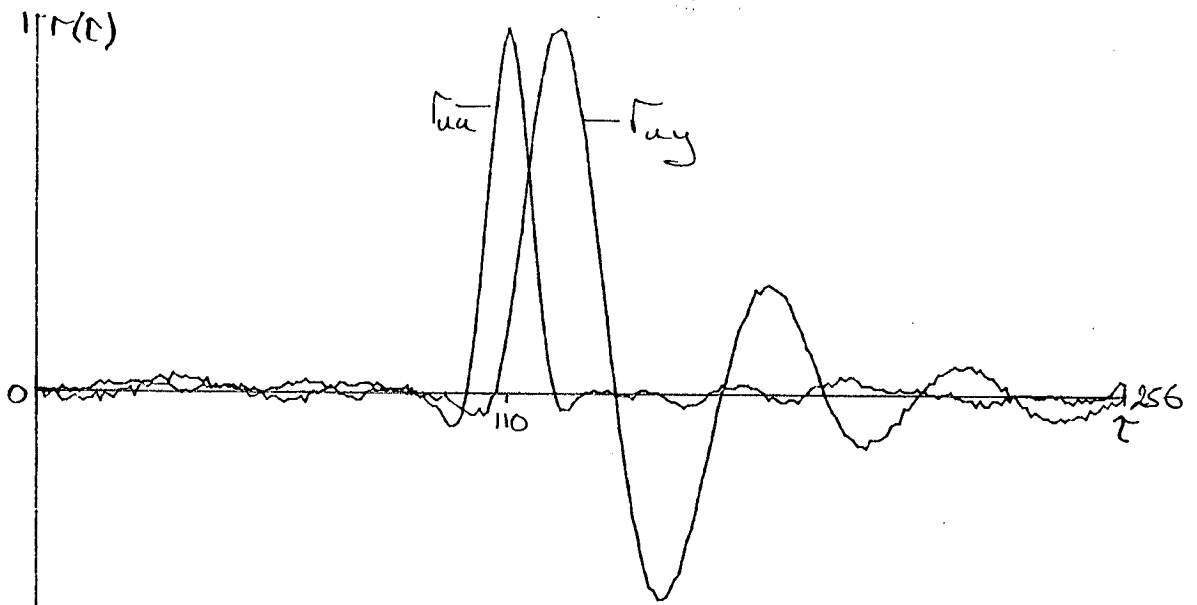
FIGURE 5-35
CROSS CORRELATION   SKIP-8
AVERAGED TWENTY TIMES

performed. The parameters of the system were estimated 100 times for each skip and a mean and standard deviation for each skip calculated. These results are presented in table 5-1.

It is interesting to note that the mean of the parameter estimates for different skips does not vary greatly. This indicates that increasing the skip does not introduce a bias.

As expected the standard deviation of the parameter estimates increases as the skip increases. However there is a marked change in the standard deviation at a skip of 32. This is interesting because in this case a skip of 16 is the maximum that still satisfies the criteria of the slow sampling rate, being twice the Nyquist rate.

5-2-2 Fault Simulation.

A test on the fault detection properties of the MMFM was performed. The second order analogue computer model was used, and a parameter estimate of the system was obtained using correlation functions that had been averaged using an exponential weighting factor of ten. The variable resister controlling the damping factor was adjusted while the experiment was running. The cross

	FREQUENCY		DAMPING		GAIN	
	$\omega_n \times 10^{-3}$		$\delta \times 10^{-2}$		g	
SKIP	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
1	7·6	4·7	3·3	0·37	1·1	0·17
2	7·7	5	3·3	0·43	1·1	0·17
4	7·5	5·5	3·3	0·41	1·1	0·19
8	7·5	5·9	3·4	0·4	1·1	0·2
16	8·5	8·2	3·2	1·0	1·1	0·28
32	8·8	13·8	3·9	3·7	1·1	0·7

TABLE 5-1

SKIP PARAMETER ESTIMATE
STATISTICAL RESULTS

correlation function began to change, and the parameter
estimates reflected this. Figure 5.36 shows the change
in the cross correlation function. As would be expected
with an increase in the damping factor the exponential
decay of the cross function becomes shorter, and the
parameter estimates reflect this. The only parameter to
change is the b term,

$$b = \frac{2\delta}{\omega_n}$$

figure 5.37 shows the variation in parameter estimates
as the experiment progressed. As can be seen with an
increase in the damping factor, δ , the parameter
estimate b will increase. A fault detection system
would be able to trace the fault to the element
controlling the damping by comparing the change in b
with the stationary estimates of a and c. Similarly a
change in the natural frequency could be detected by a
variation in the b and a estimates. A change in the
gain would only affect the c parameter estimate.

It is therefore possible to trace the source of a
fault through variation in the parameter estimates
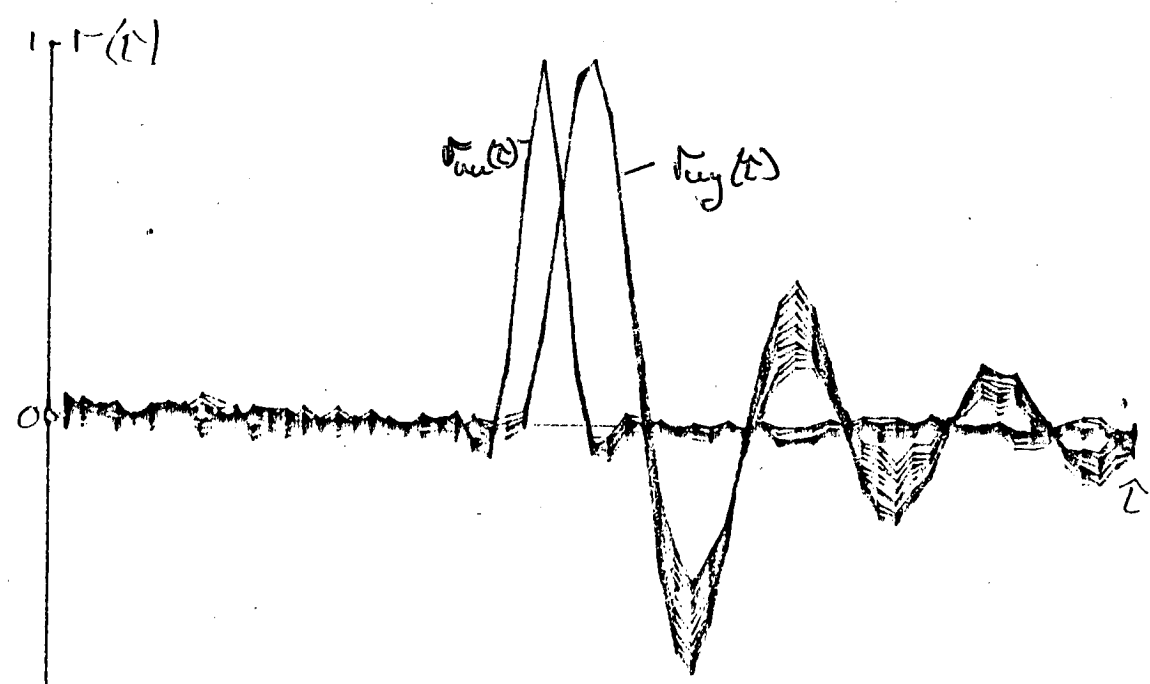obtained using the Modified Modulating Function Method.

FIGURE 5-36

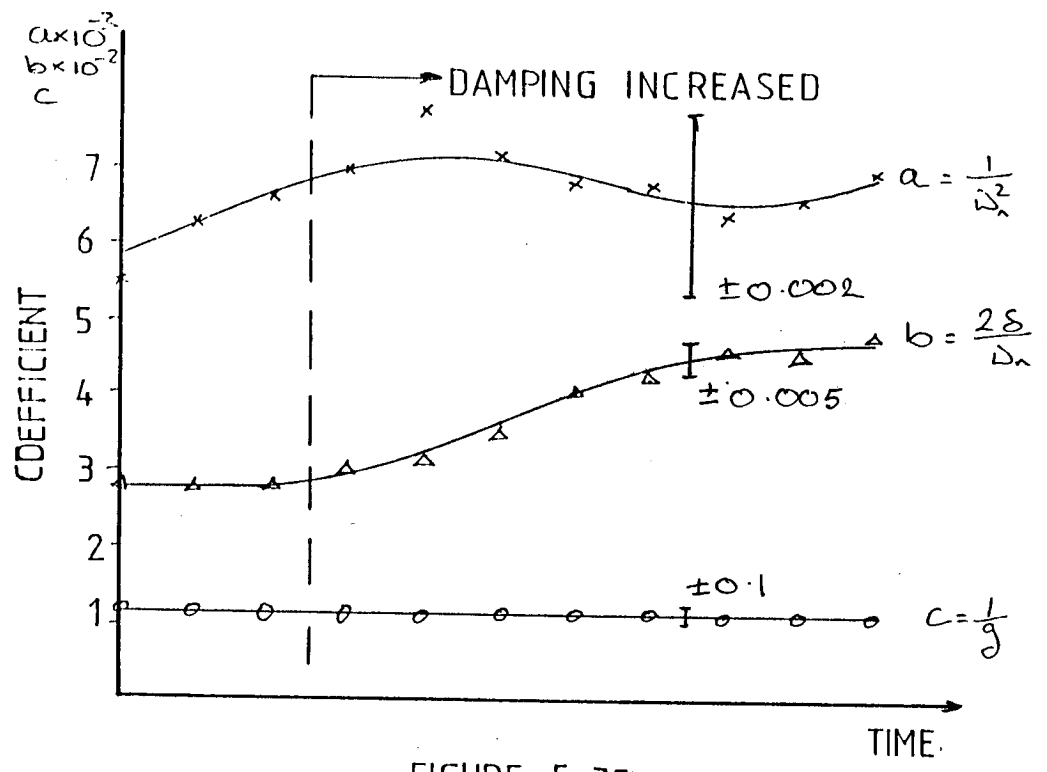CHANGE IN CROSS CORRELATION WITH
VARIATION IN DAMPING

FIGURE 5-37

SIMULATED FAULT PARAMETER ESTIMATES

5-3 Diesel Engine Results.

In this section the experimental results obtained from the diesel engine at Stornoway Power Station will be presented. The MMFM was used to obtain the transfer function of the diesel alternator set at different loads, also different amplitude test signals were used to check for non-linear condition. A spectrum analyser was also used to obtain the transfer function at these load conditions, the results were compared with the MMFM. An attempt to simulate a fault condition on the diesel was also made.

5-3-1   Diesel-Alternator   Set   Transfer   Function Measurement.

The transfer function of the Diesel Alternator set was measured using the MMFM. To eliminate the noise problem an exponential averaging weighting factor of fifty was used. With such a large averaging the computation of the correlation functions needed to be quick, for this reason a sixty four point correlation function was used. A skip of eight was used in the skip algorithm.

Previous measurements of the auto and cross correlation functions had shown that variance on the correlation measurement would prove to be a serious

problem, figure 5.38 shows such an auto and cross correlation measurement. A series of parameter estimates were performed using non-averaged correlations and the results obtained were not impressive. Table 5-2 shows the spread of parameter estimates for a second order model of the diesel at a power output of 3.5 MW.

In addition to this problem it was noted in the initial experiments that a time lag was present in the system. Before any parameter estimation was attempted, on the engine, this time lag was removed by winding out the time lag manually. This was achieved by re-defining the zero point of the cross correlation function. Since these experiments were performed a rigid method of measuring has been developed, thus accurately eliminating the time lag. This has been described in Chapter three.

A cross correlation function of the diesel-alternator set, obtained by correlating between the injected Gaussian noise and the power output of the alternator is shown in figure 5.39.

$$G(s) = \frac{e^{-sT_d}}{a_2 s^2 + a_1 s + a_0}$$

With the time delay removed and therefore a known fixed value, the parameters to be measured are,
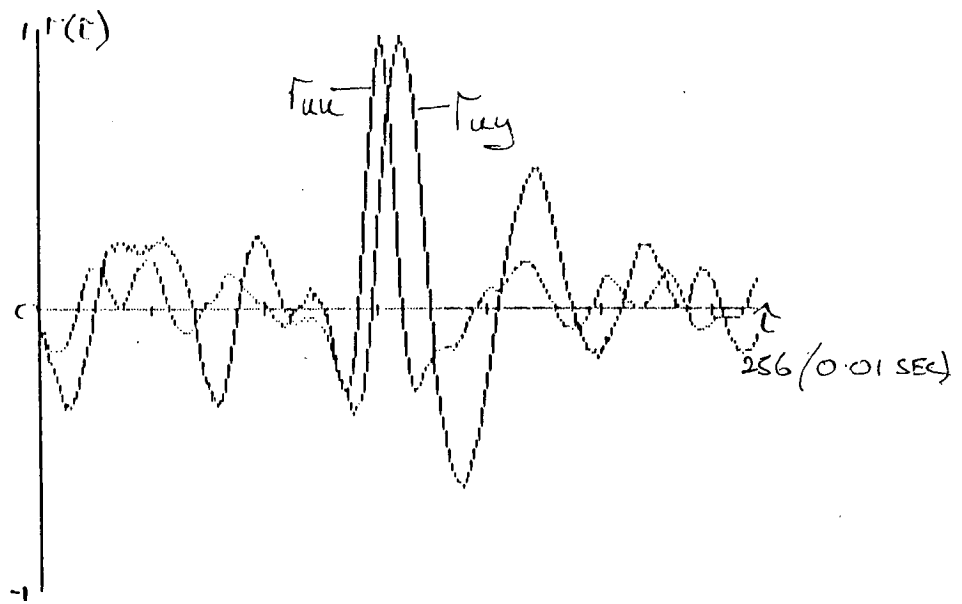
FIGURE 5-38
AUTO AND CROSS CORRELATION
FUNCTION OF THE DIESEL
ALTERNATOR SET

PARAMETER	σ	$\bar{X}$	HIGH LOW
a	0·147	0·11	0·48 −0·067
b	0·407	0·1	0·63 −0·71
C	7·6	14	33·7 6·7

TABLE 5-2
DIESEL PARAMETER ESTIMATES
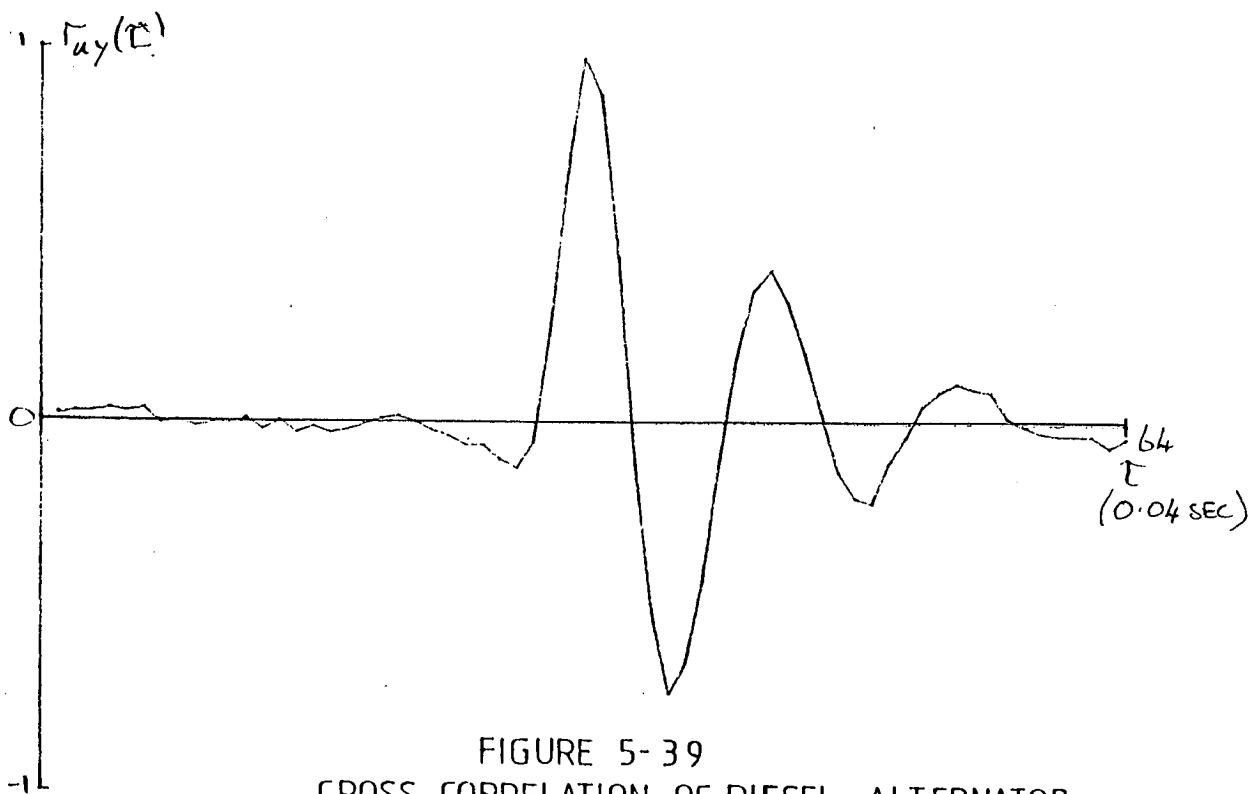
FIGURE 5-39
CROSS CORRELATION OF DIESEL ALTERNATOR
SET AT 3·5MW AVERAGED FIFTY TIMES
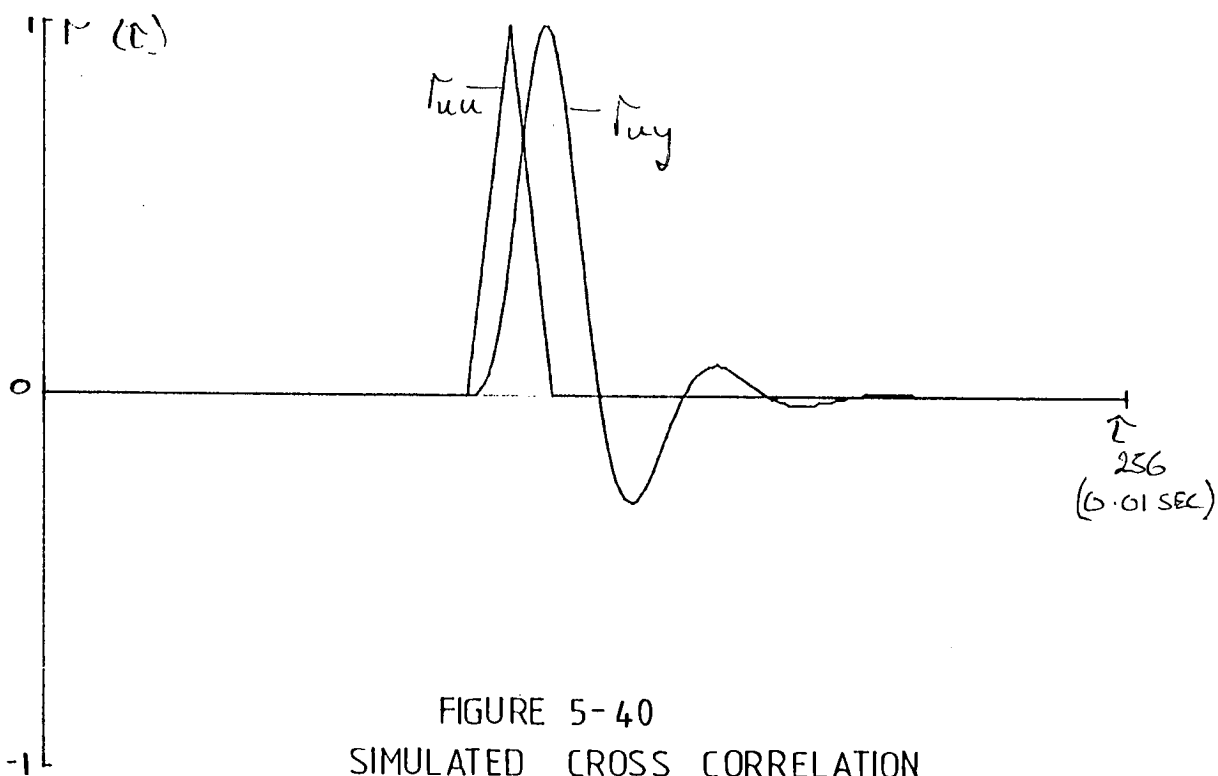
FIGURE 5-40
SIMULATED CROSS CORRELATION
AT 3·5 MW

$$a_2 = \frac{1}{g\omega_n} \quad , \quad a_1 = \frac{2\delta}{\omega_n} \quad , \quad a_0 = \frac{1}{g}$$

The parameter estimates for two different loads are shown in table 5.3. This indicates that there is little difference in the transfer function at these loads, suggesting that the diesel is linear in the test range. This is also confirmed by the frequency response tests shown later. In addition to this a stationary input/output curve was obtained, figure 5-41. This shows power output against voltage on the governor set point. This is also linear in the test range, only going non-linear at high loads.

These parameter estimates were used in the simulation program to produce responses which can be compared with the original, figure 5.40. This gave a response which is much more heavily damped than the system. A similar set of results for the engine at 2MW is shown in figures 5.42 and 5.43. Here the model closely approximates the system , indicating that the model is susceptible to small variations in the parameter estimates.

A fault detection experiment was performed on the diesel. Two fuel injectors were disconnected during an experiment, but a change in the transfer function parameters was not noticed. Because of the large
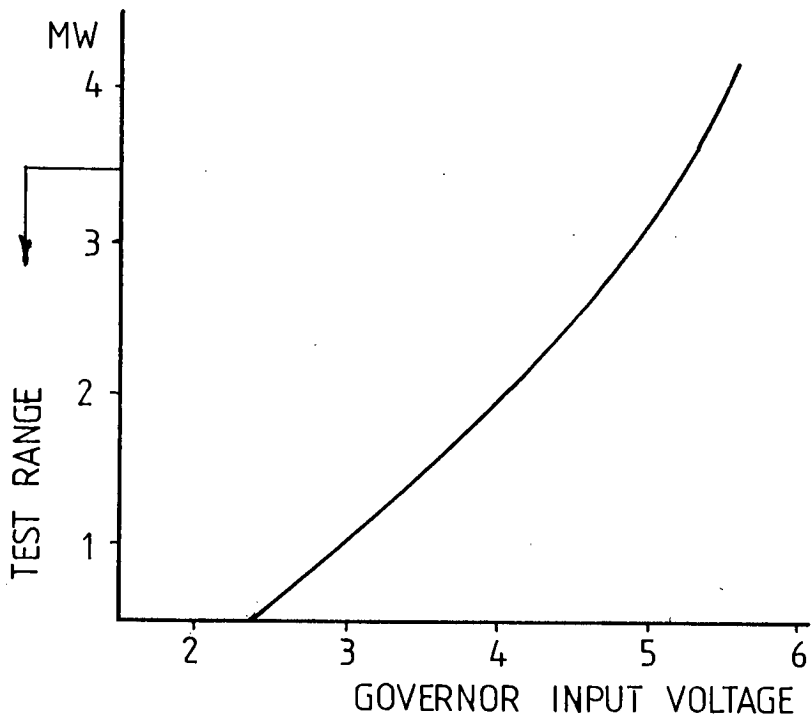
FIGURE 5-41
INPUT/OUTPUT RELATIONSHIP

PARAMETER	2MW		3MW	
	LOW	HIGH	LOW	HIGH
a	0·05	0·09	0·0075	0·054
b	0·22	0·54	0·37	0·44
C	10·5	12·5	9·5	11·6

TABLE 5-3
DIESEL PARAMETER ESTIMATES
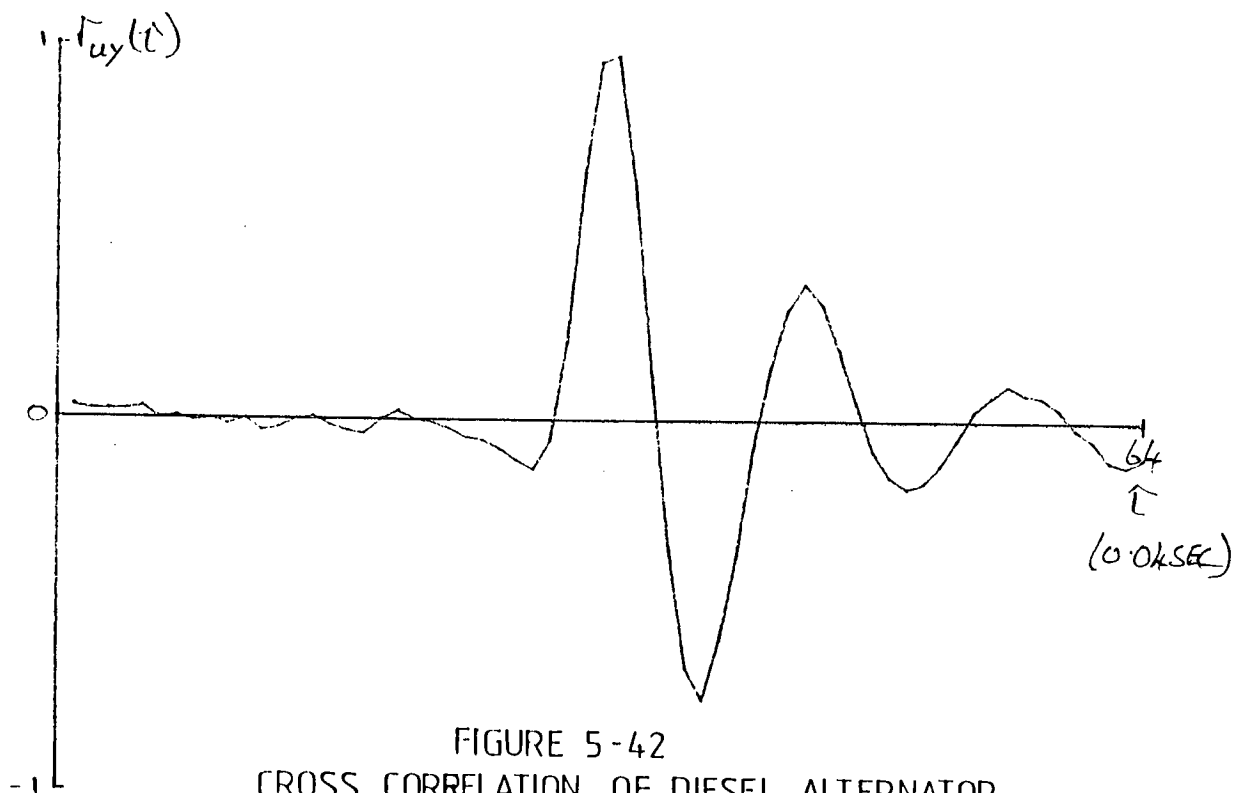
FIGURE 5-42
CROSS CORRELATION OF DIESEL ALTERNATOR
SET AT 2MW AVERAGED FIFTY TIMES
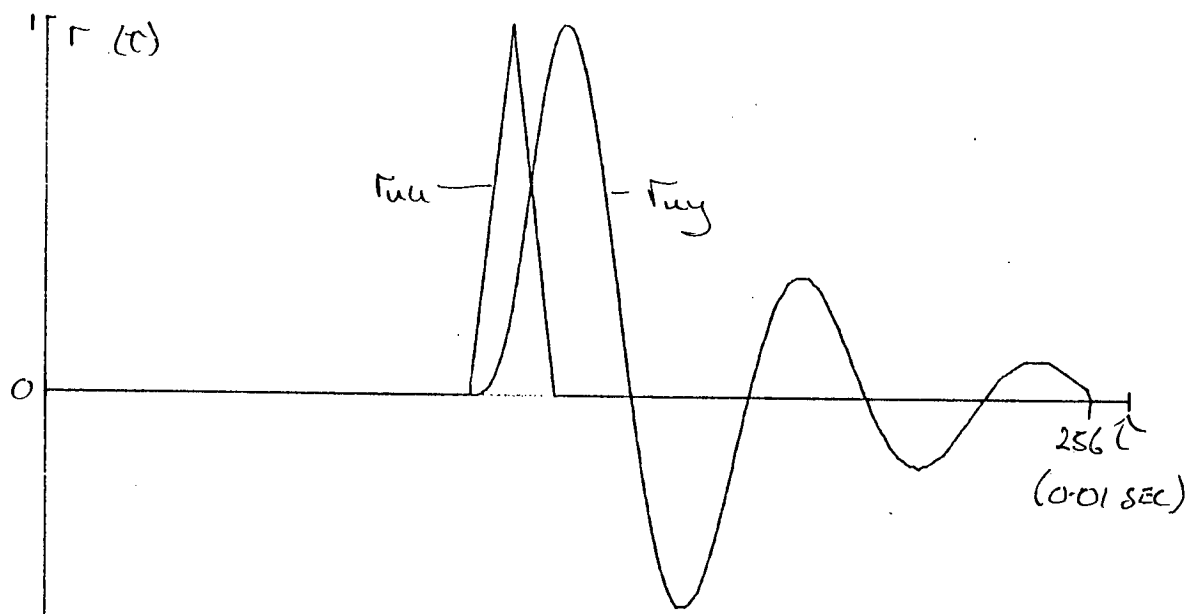
FIGURE 5-43
SIMULATED CROSS CORRELATION
AT 2MW

202

exponential averaging factor, 50, needed to overcome the variance in the parameter estimates, changes in the diesel engine response take some time to affect the parameter estimates. This experiment was not run for a sufficiently long period after the fuel injectors were disconnected, because an operating staff shift change at the power station required the termination of the experiment.

5-3-2 Diesel-Alternator Set Frequency Response.

As a check on the transfer function measured by the MMFM, a commercial spectrum analyser was used to measure the system frequency response. The frequency response was measured at various loads, and again there appears to be very little difference in the results obtained. This confirms the MMFM results and the static input/output curve previously shown. Different amplitude test signals were also used to obtain the frequency responses, again there was no variation in the results obtained. This all points to the diesel alternator set being linear in the test range. The frequency response for the engine at 2 MW load is shown in figures 5-44 to 5-46. This can be compared with the frequency response plot obtained from the engine at 3.5 MW shown in figures 5-47 to 5-49. As can be seen there is no significant difference in these frequency response

FIGURE 5-44
MAGNITUDE RESPONSE AT 2MW 0.6V RMS
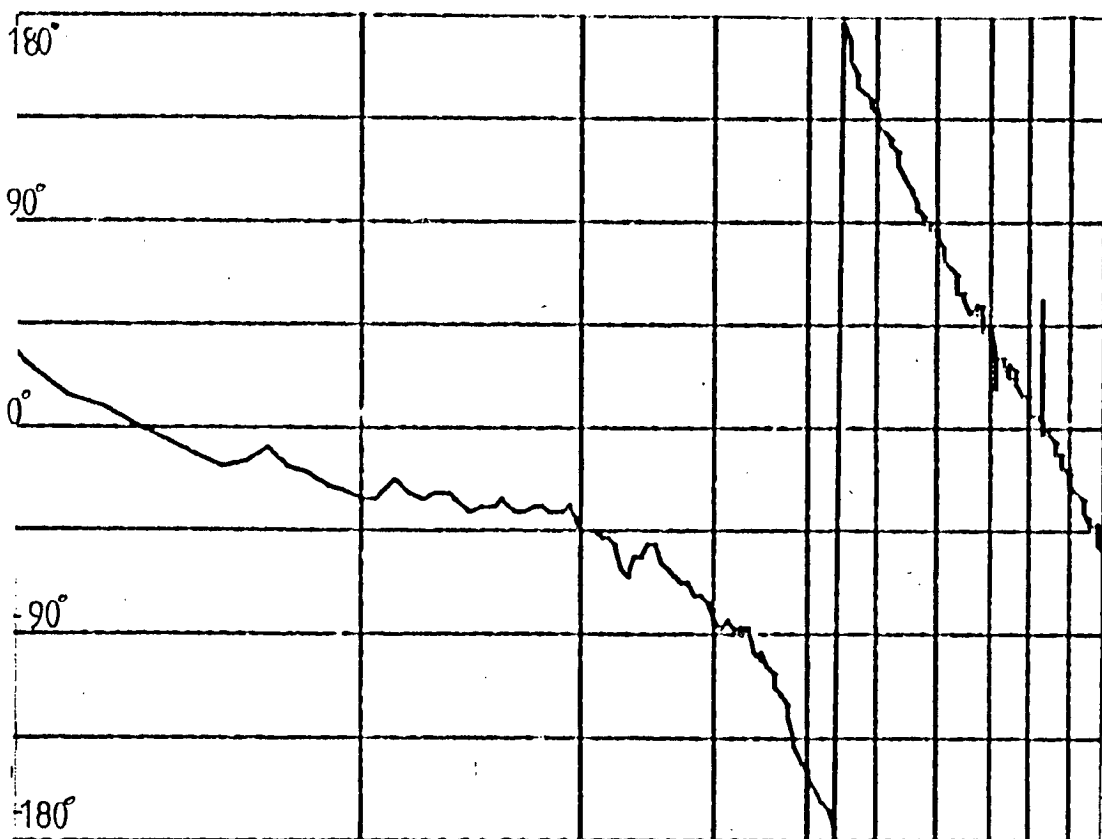
FIGURE 5-45
PHASE RESPONSE
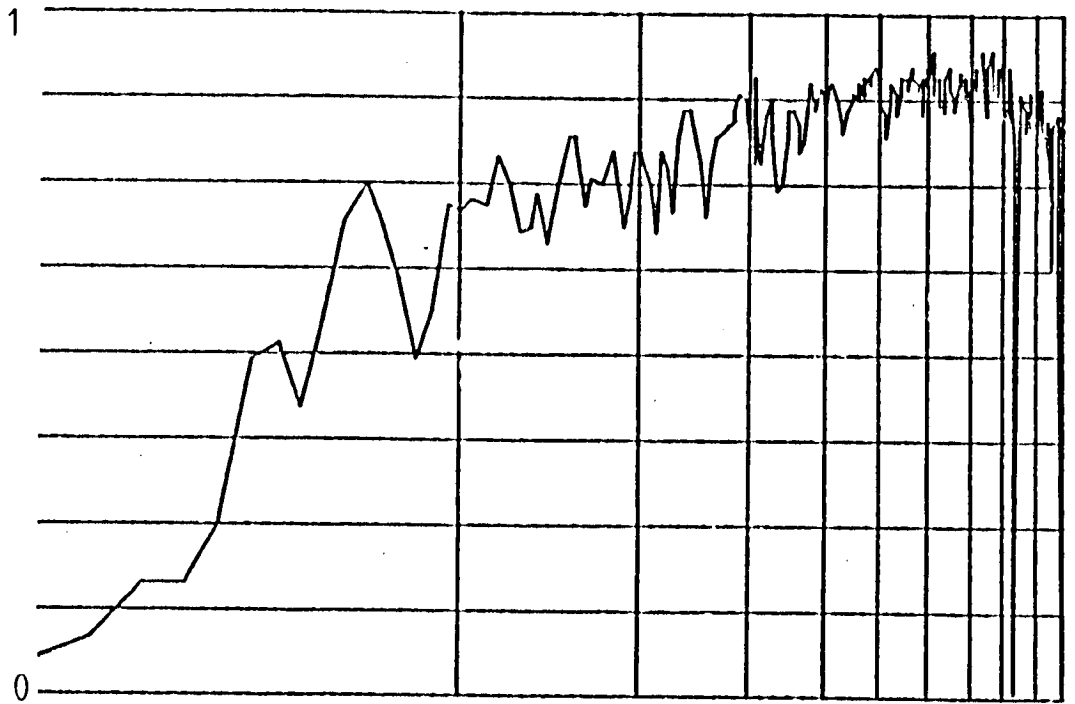
FIGURE 5-46
COHERENCE FUNCTION
AT 2MW 0·6V RMS
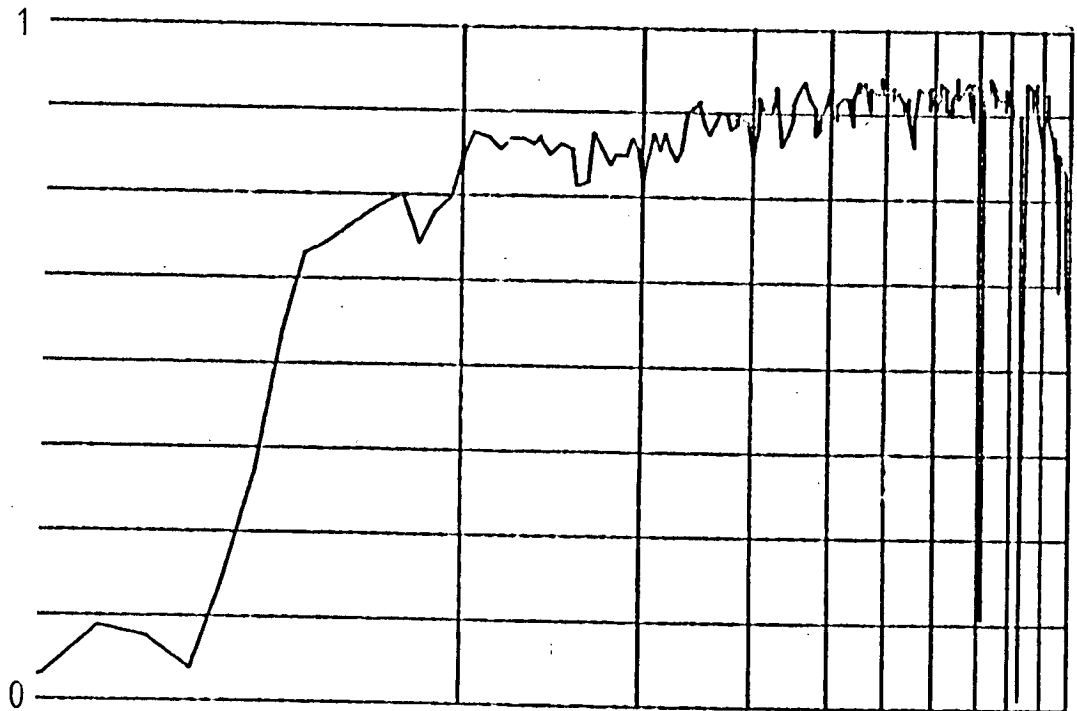
FIGURE 5-47
COHERENCE FUNCTION AT 3·5MW
0·6 RMS

205
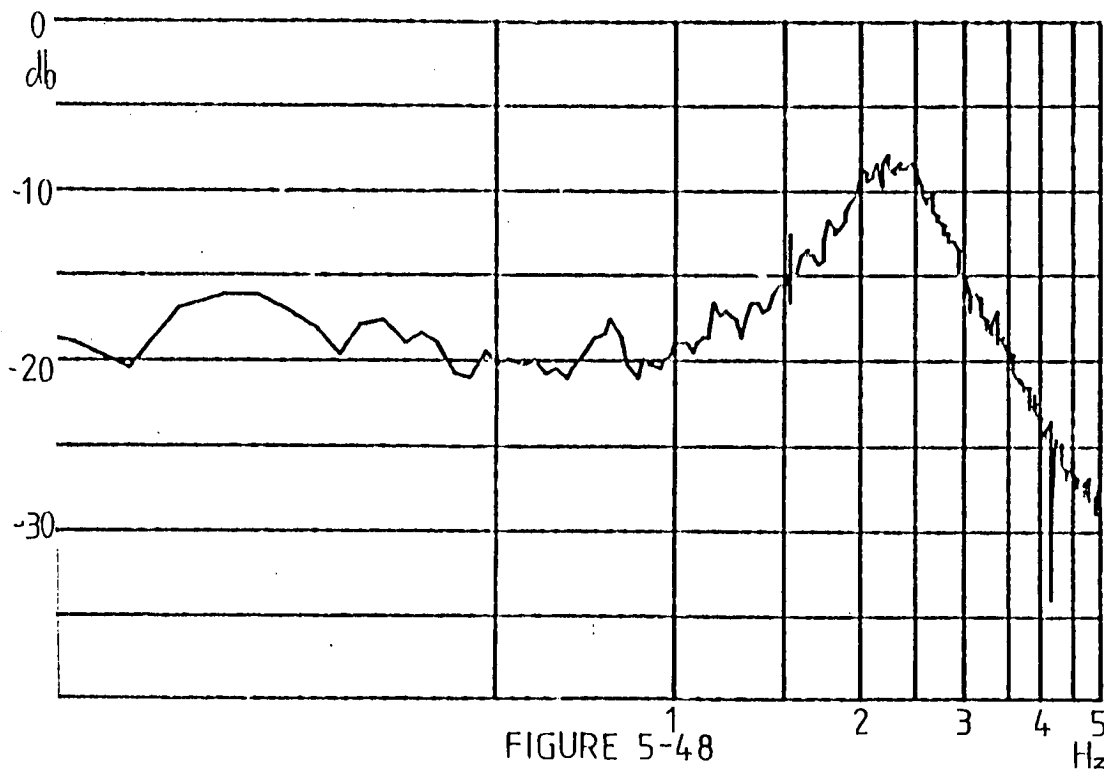
FIGURE 5-48
MAGNITUDE RESPONSE AT 3·5MW 0·6V RMS
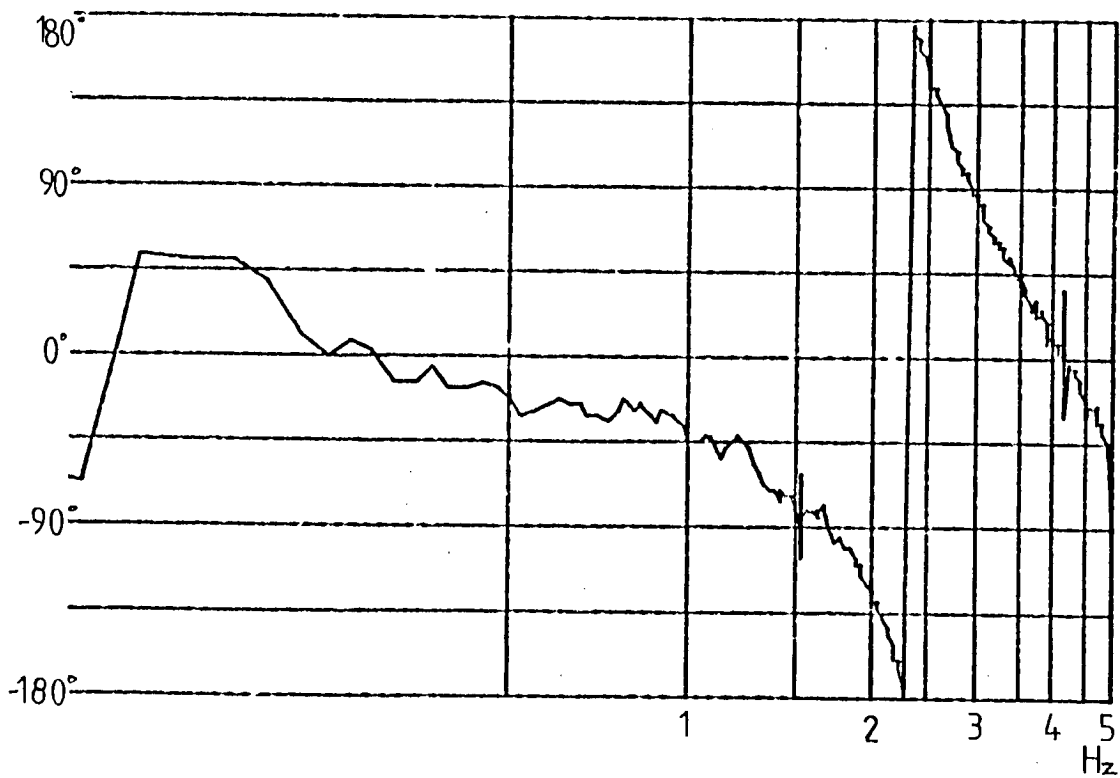
FIGURE 5-49
PHASE RESPONSE

206

plots. The coherence functions, figure 5-46 2 MW and figure 5.47 3.5MW, show that good coherence is obtained over the frequency range of interest, 0.5 H_z to 5H_z. This indicates that a high level of confidence can be placed in the results. These results were obtained by injecting a noise signal of 5H_z bandwidth and obtaining the frequency response from the spectrum analyser. The spectrum analyser also had the facility to inject pure sine waves and sweep through the desired frequency range. A similar response was obtained using this method as seen in figure 5.50 and figure 5.51.

Different amplitude noise signals were injected into the engine to determine the effect that any non-linearity present would have on the transfer function. Surprisingly there was very little difference in the reponses obtained. A test signal of 0.3v Rms which represented a disturbance of ±2% of the load at 3.5MW was injected and the responses shown in figures 5-52 to 5-54. Comparing this with a 0.6v Rms signal, giving a ±3% disturbance on the load at 3.5MW, figures 5-47 to 5-49, as can be seen there is little difference between the responses.
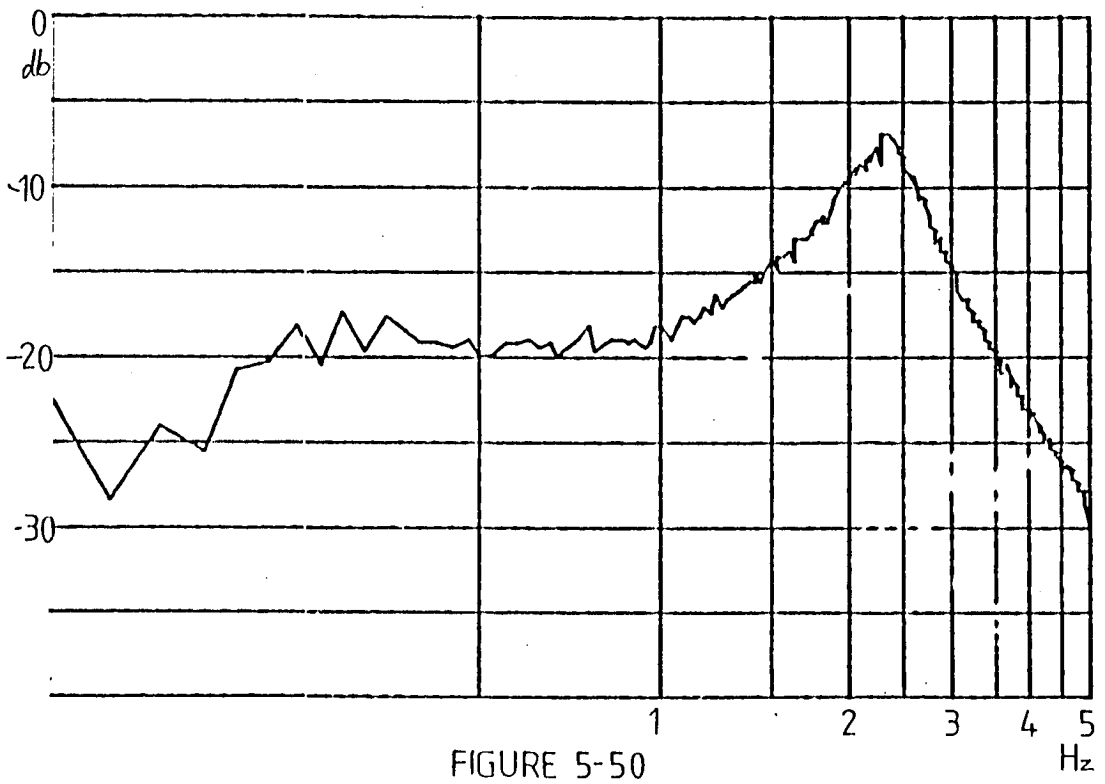
FIGURE 5-50
MAGNITUDE RESPONSE   3·5 MW SINE SWEEP 0·4V RMS

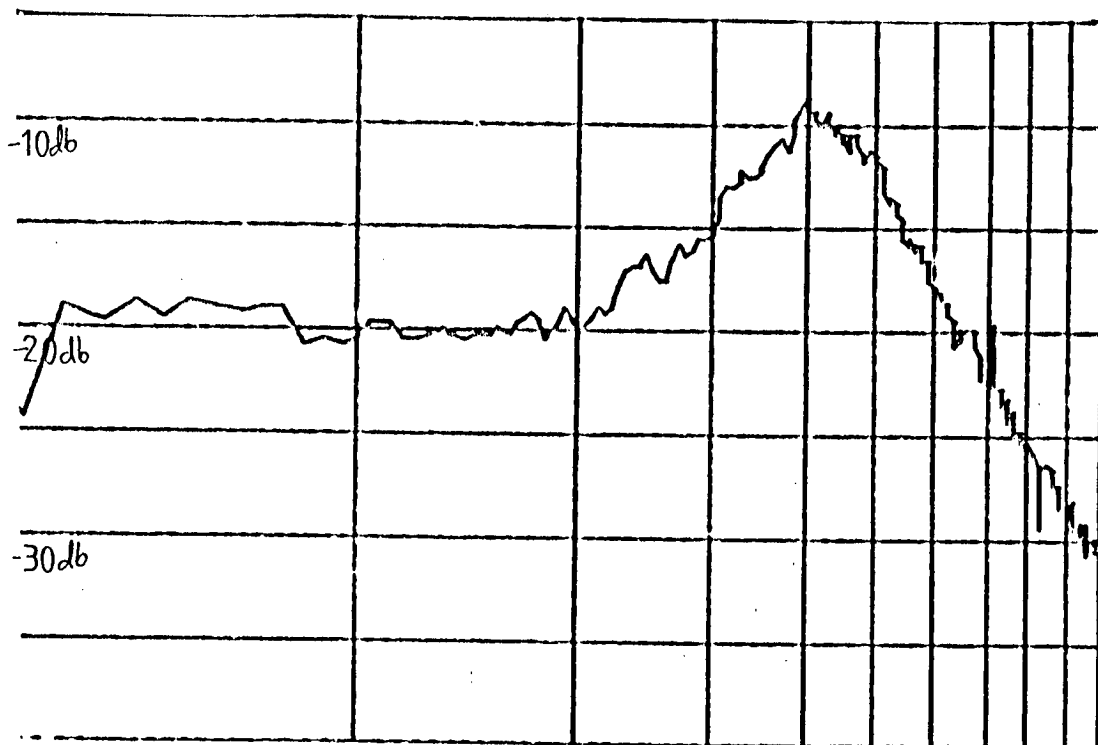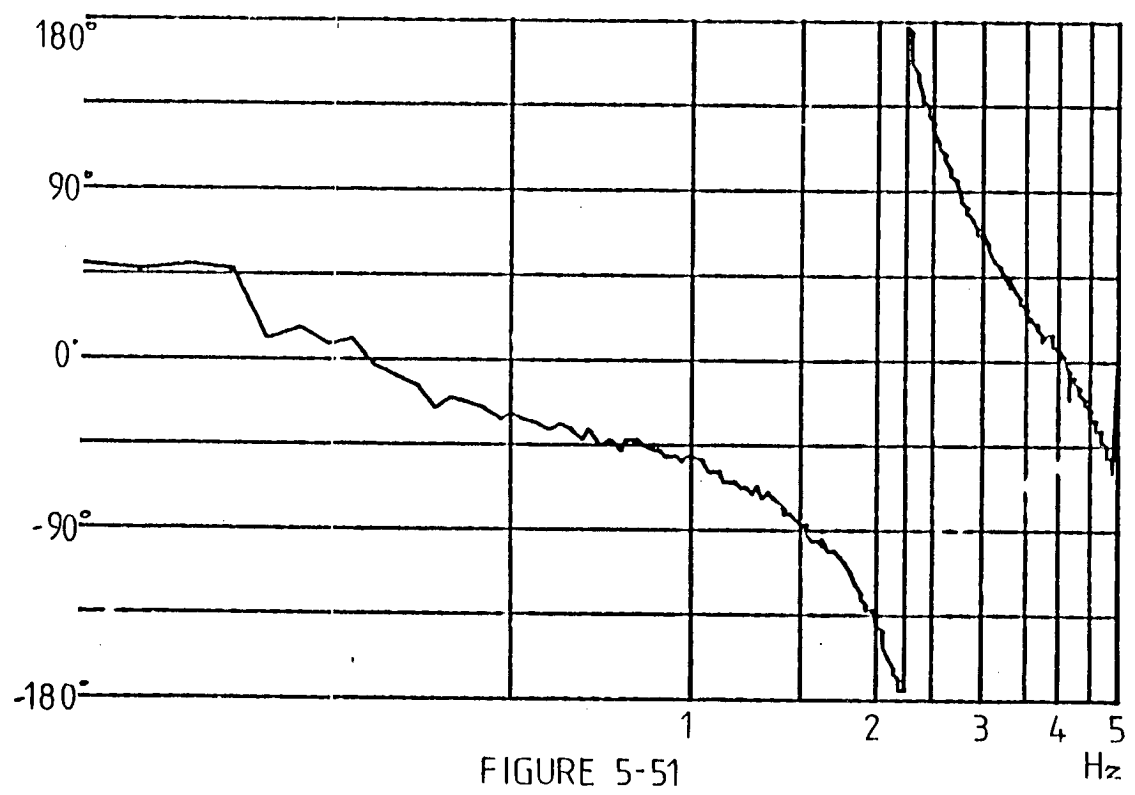FIGURE 5-51
PHASE RESPONSE   SINE SWEEP

208
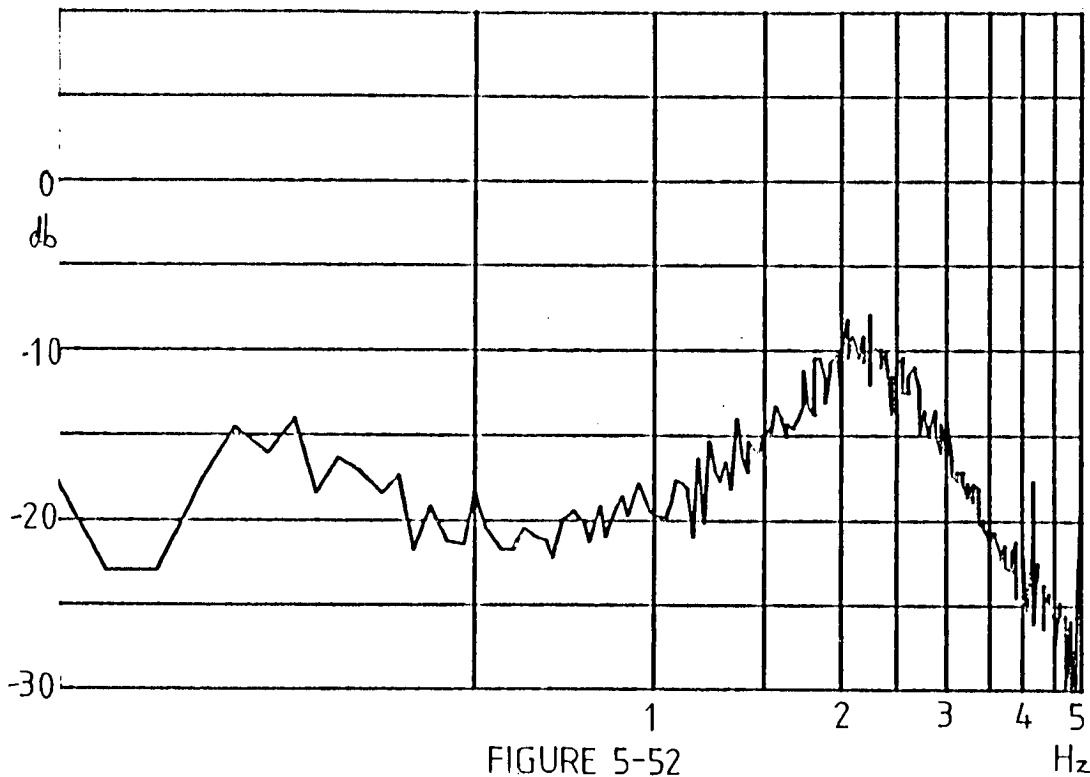
FIGURE 5-52
MAGNITUDE RESPONSE AT 35MW 0·3V RMS
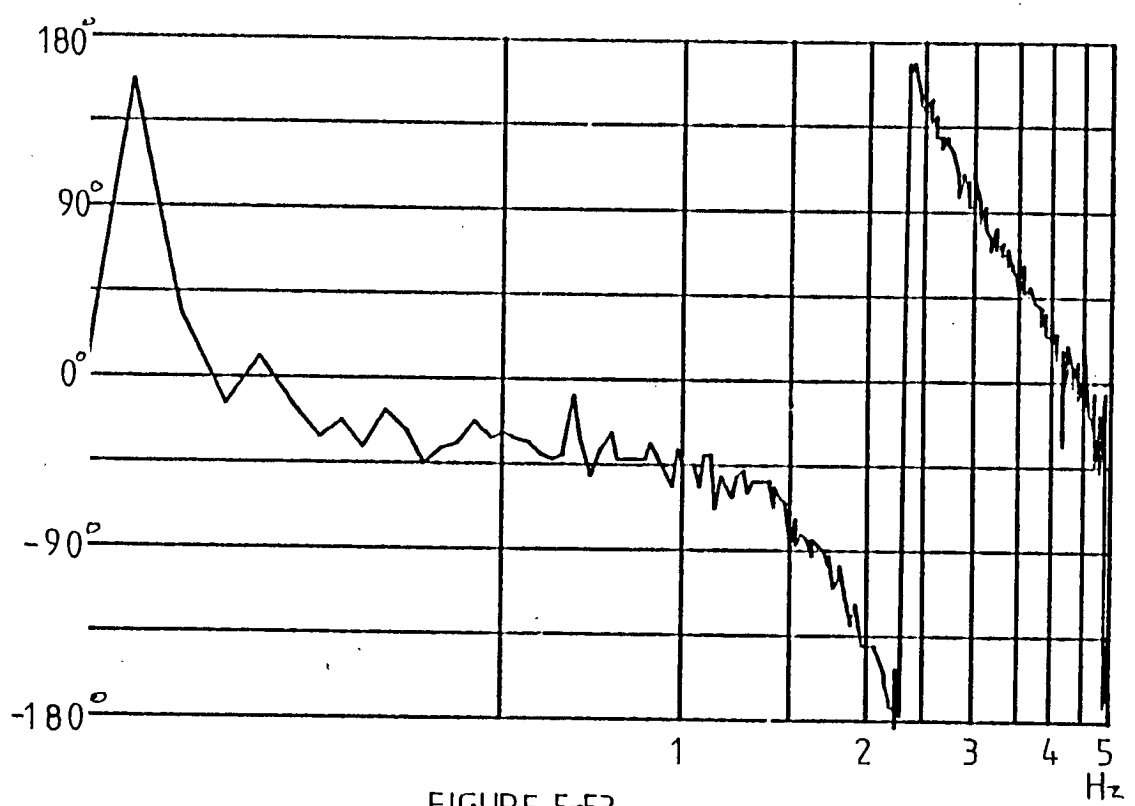
FIGURE 5-53
PHASE RESPONSE

209
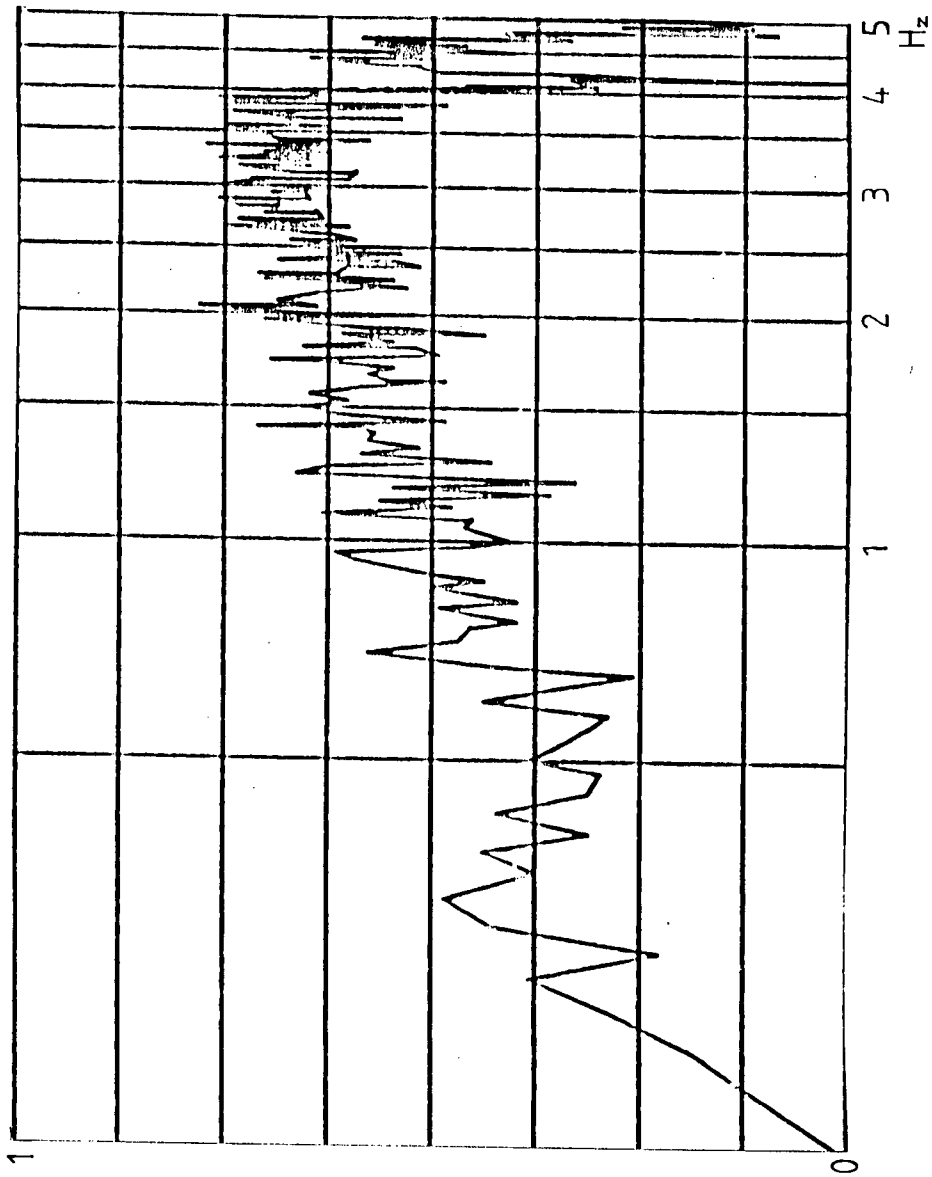
FIGURE 5-54

COHERENCE FUNCTION AT 3·5MW   0·3V RMS

5-3-3 Comparison of Frequency Response and MMFM Result.

As a check on the validity of the model obtained using the MMFM, a frequency response of the model was calculated. This was then compared with the frequency response obtained from the spectrum analyser.

A representative sample of the frequency response calculated from the MMFM models of the diesel-alternator set at 3.5MW are shown on figures 5.55 and 5.56. These compare favourably with the frequency response obtained using a spectrum analyser, figure 5.50 and 5.51. Additionally the MMFM transfer functions do not vary significantly with load or signal size, which is the same result obtained from the spectrum analyser. It is thought that this simple model of the diesel alternator set is a reasonably accurate model for control purposes, but is probably not sufficiently complex to allow fault isolation.
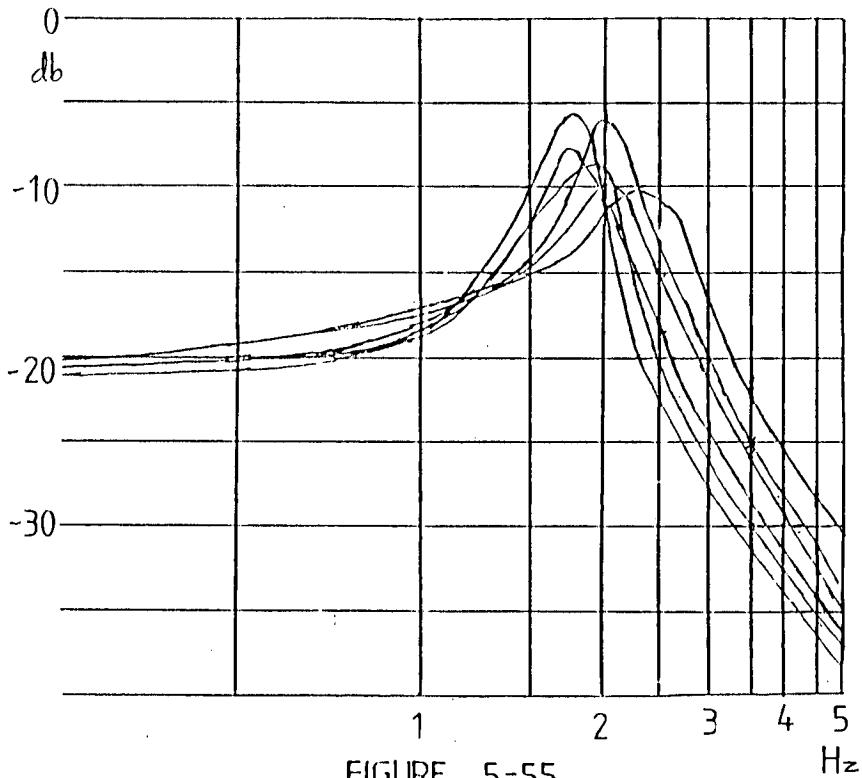
FIGURE 5-55

MAGNITUDE RESPONSE OF MODEL
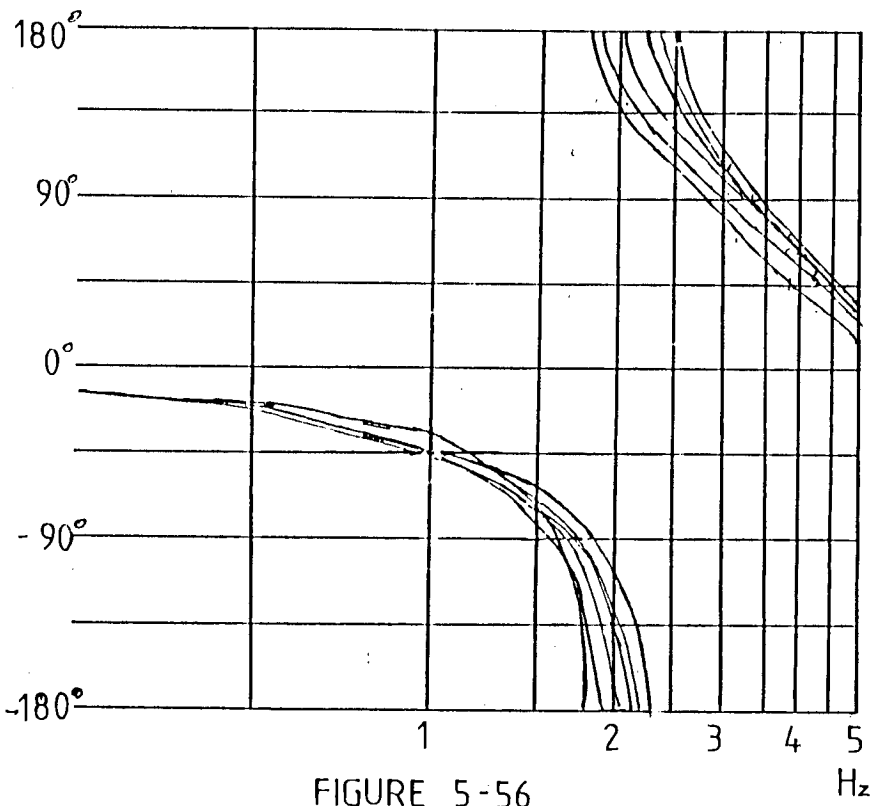
FIGURE 5-56

PHASE RESPONSE

212

CHAPTER SIX

CONCLUSIONS

A new method for identifying the parameters of a continuous transfer function has been presented here. This method is a development of Shinbrot's Modulating Function Method. The method was developed for use in a fault detection and fault isolation system. Other methods of continuous parameter estimation were examined, this showed that no standard method exists and that there are problems associated with all of them. The MMFM overcomes some of these problems, and was particularly suitable for the application chosen here.

The main advantage of this method lies in the pre-processing of the input and output system signals. It makes use of the input signal auto correlation function and the input/output cross correlation function. These functions are used in the identification algorithm in place of the input/output signals. This allows the problem of noise to be overcome. Also a pre-estimation check is automatically performed on the data which will detect bad data. Relay correlation was used in the correlation algorithm. This results in a faster processing time while not introducing any degradation in the correlation shape. A further improvement in processing speed was achieved by

implementing Fell's skip algorithm which resulted in little loss of accuracy up to a skip of eight.

The simulation tests have shown that the MMFM is a useful method of continuous system identification. Software simulations were carried out to study the effect on the parameter estimates of modelling a high order system with a low order model. In addition the effect of time delay on parameter estimation was studied and showed the need for accurate models of the system under test. This was a particular problem in this research due to the complex nature of the system under test at Stornoway.

A second order model with time delay was used to model the diesel/alternator set. The parameter estimates obtained were compared with independent results obtained using a commercial spectrum analyser. The results compared favourably. The second order model was chosen because the diesel auto and cross correlation functions closely resembled a second order response with time delay. Conventional modelling of diesels, results in high order transfer functions, which can be approximated by second order with time delay.

The problem of noise was of concern, especially at the power station where some noise is always present. The

MMFM overcomes this problem by using as the input signal an independent noise signal PRBS noise or Gausian noise, and performing the cross correlation with this and the output signal, thus eliminating the noise.

This method was developed for use in a fault detection and location system. Continuous methods are thought best for this purpose because it may be possible to relate changes in parameters back to faults in the system. Using this technique it was hoped to detect incipient failures, although several problems exist. Variance in the parameter estimate may mask any change and an accurate model is needed to detect and locate failure. Also some types of faults may not affect the input/output relationship significantly and therefore the method may be insensitive to these types of faults.

A fault was simulated on the diesel but was not detected. This may have been due to the heavy averaging needed to overcome the variance problem masking any change in the parameters.

To work effectively, this method of fault detection needs an accurate model and accurate parameter estimation if detection and isolation of faults is to be achieved. At present this is not possible and the

method is only of use as a fault detection scheme. The second order model used here is not sufficient to locate any fault. A better model of the diesel is needed.

Further work is required to determine how effective a second order model plus time delay representation of a complex system is, for detecting faults. The variance on high order parameter estimates needs to be studied, and compared with the variance of lower order estimates. In addition a range of practical examples should be studied now that the basis of the method has been established.

REFERENCES

[1]   Neale,M. ´A Guide to the Condition Monitoring of Machinery´. HMSO 1979

[2]   London Times 11-8-83

[3]   Fagerland,H. ´Condition Monitoring of Medium Speed Engines´. Norwegian Maritime Research, 1978, vol 1 (6) pp 24-29

[4]   Chambers,R. ´Experiences on Diesel Engines through Insurance´. British Engine Technical Report. 1981, vol 14, pp 24-36

[5]   Sair,K.F. ´Maintenance of Diesel Engines by Trend Analysis´. Marine Engineering Review. 1983, pp 5-7

[6]   Konstantinos,T.D. ´Maintenance of Diesel Engines´. MSc Thesis University College London 1981

[7]   Fagerland,H. ´Monitoring and Diagnosing Process Deviation in Marine Diesel Engines´. Norwegian Maritime Research. 1979 Vol 7 (2) pp 15-25

[8]   Langballe,M. ´Condition Monitoring of Diesel Engines´. Norwegian Maritime Research. 1975 Vol 3 (3) pp 2-16

[9]   Billings,S.A. ´Identification of Non-linear Systems-A Survey´. Proc. IEE pt.D 1980 Vol 127 (6) pp 272-285

[10] Kubrusly,C.S. ´Distributed Parameter System Identification A Survey´. Int. J. Control 1977 Vol 26 (4) pp 509-535

[11] Rake,H. ´Step Response and Frequency Response Methods´. Automatica 1980 Vol 16 pp 519-526

[12] Payne,P.A. ´An Improved Technique for Transfer Function Synthesis from Frequency Response Data´. IEEE Trans. on Auto. Control 1970 Vol AC-15 pp 480-483

[13] Kwiatkowski,A.W. ´Comparison of Time and Frequency Domain Approaches to Identification of Linear Systems´. Advances in Manufacturing Systems Research and Development. Editor J. Peklenik. Perganon Press 1971

[14] Eykhoff,P. ´System Identification´. Wiley 1974

[15] Godfrey,K.R. ´Correlation Methods´. Automatica 1980 Vol 16 pp 527-534

[16] Wellstead,P.E. ´Non-Parametric Methods of System Identification´. Automatica 1981 Vol 17 pp55-69

[17] Briggs,P.A.N. et.al. ´Correlation Analysis of Process Dynamics Using Pseudo-Random Binary Test Perturbations´. Proc. Instn. Mech. Engrs. 1964-65 Vol 179 (3H) pp37-51

[18] Hazelrigg,A.D.G. et.al. ´Application of Cross Correlating Equipment to Linear-System Identification´. PROC IEE. 1965 Vol 122 (12) pp 2385-2400

[19] Godfrey,K.R. ´The Theory of the Correlation Method of Dynamic Analysis and its Application to Industrial Processes and Nuclear Power Plant´. Measurement and Control, 1969 Vol 2 pp T65-T72

[20] Flower,J.O. and Windett,G.P. ´Dynamic Measurement of a large Diesel Engine Using PRBS Techniques. Part 1´ Int. J. Control 1976 Vol 24 (3) pp 379-392

[21] Flower,J.O. and Windett,G.P. ´Dynamic Measurement of a large Diesel Engine Using PRBS Techniques. Part 2´ Int. J. Control 1976 Vol 24 (3) pp 393-404

[22] Lang,R.D. ´Microprocessor-Based Identification System Applied to Synchronous Generators with Voltage Regulators´. IEE Proc. Pt.C 1983 Vol 130 (5) pp 257-265

[23] Cooley,J.W. and J.W. Tukey. ´An Algorithm for the Machine Calculation of Complex Fourier Series´. Math. Computing 1965 Vol 19 pp 297-301

[24] Ljung,L. ´Frequency Domain Versus Time Domain Methods in System Identification´. Automatica 1981 Vol 17 (1) pp 71-86

[25] Fasol,K.H. ´Principles of Model Building and Identification´. Automatica 1980 Vol 16 pp 505-518

[26] Strejc,V. ´Least Squares Parameter Estimation´. Automatica 1980 Vol 16 pp 535-550

[27] Goodwin G.C. and Payne R.L. ´Dynamic System Identification:Experimentation and Data Analysis.´ Academic Press 1977

218

[28] Young P. ´The Instrumental Variable Method:A Practical Approach To Identification and System Parameter Estimation´. IFAC Identification and System Parameter Estimation York UK 1985 pp1-15

[29] Franklin,G.F. and Powell,J.D. ´Digital Control of Dynamic Systems´. Addison Wesley 1978

[30] Astrom,K.J. ´Maximum Likelihood and Prediction Error Methods´. Automatica 1980 Vol 16 pp 551-574

[31] Isermann,R. ´Two-Step Process Identification with Correlation Analysis and Least Squares Parameter Estimation´. Trans. ASME J. Dynamic Systems, Measurement, and Control. 1974 pp 426-432

[32] Isermann,R. ´Comparison of Six On-Line Identification and Parameter Estimation Methods´. Automatica 1974 Vol 10 pp 81-103

[33] Astrom,K. and Eykhoff,P. ´System Identification-A Survey´. Automatica 1971 Vol 7 pp 123-162

[34] Isermann,R. ´Practical Aspects of Process Identification´. Automatica 1980 Vol 16 pp 575-587

[35] Strejc,V. ´Trends in Identification´. Automatica 1981 Vol 17 (1) pp 7-21

[36] Kallstrom,C.G and Astrom,K. ´Experiences of System Identification Applied to Ship Steering´. Automatica 1981 Vol 17 (1) pp 187-198

[37] Trankel,T.L. ´Identification of Ship Steering Dynamics Using Inertial Sensors´. IFAC Identification and System Parameter Estimation York UK 1985

[38] McDyer,F.J. ´Low Order Models for Large Generating Plant´. IFAC Identification and System Parameter Estimation York UK 1985

[39] Hogg,B.W. ´On-Line Identification of Power Generation Plant´. IFAC Identification and System Parameter Estimation York UK 1985

[40] Young,P. ´Parameter Estimation for Continuous-Time Models-A Survey´. Automatica 1981 Vol 17 (1) pp 23-39

[41] Diamessis,J.E. ´A New Method for Determining the Parameters of Physical Systems´. Proc. IEEE 1965 pp205-206

[42] Garnett,J.R. and Eisenberg,L. 'On the Determination of the Physical Systems'. Proc IEEE 1973 pp 490-491

[43] Smith,F.W. 'System Laplace-Transform Estimation from Sampled Data'. IEEE Transactions on Automatic Control 1968 Vol AC-13 (1) pp 37-44

[44] Sinha,N.K. 'Identification of Continuous-Time Systems From Sampled Data: A Comparison of 3 Direct methods'. IFAC Identification and System Parameter Estimation York UK 1985 pp1575-1578

[45] Sinha,N.K. 'System Identification Using Block-Pulse Functions'. Circuits and Systems,25th, Houghtion, Mich, Aug. 1982 pp325-328

[46] Sannuti,B.E. 'Analysis and Synthesis of Dynamic Systems via Block-Pulse Functions'. Proc. IEE Control and Science 1977 Vol 124 (6) pp569-571

[47] Prasad,T.and Sinha,N.K. 'Modelling of Continuous-Time Systems From Sampled Data Using Trapezoidal Pulse Functions'. Int.Conf Systems,Man.and Cybernetics Bombay 1983 pp427-429

[48] Chen,C.F. 'Time-Domain Synthesis via Walsh Functions'. Proc. IEE 1975 Vol 122 pp565-570

[49] Corrington,M.S. 'Solution of Differential and Integral Equations with Walsh Functions'. IEEE Transactions on Circuit Theory 1973 Vol CT-20 (5) pp470-476

[50] Rao,G.P. 'System Identification Via Walsh Functions'. Proc. IEE. 1975 Vol 122 (10) pp 1160-1161

[51] Rao,G.P. 'Improved Algorithms for Parameter Identification in Continuous Systems via Walsh Functions'. Proc. IEE 1983 Vol 130 Pt. D (1) pp9-16

[52] Bohn,E.V. 'Measurement of Continuous-Time Linear System Parameters via Walsh Functions'. IEEE Trans. on Industrial Electronics, 1982 Vol IE-29 (1) pp38-46

[53] Bohn,E.V. 'Estimation of Continuous-Time Linear System Parameters from Periodic Data'. Automatica 1982 Vol 18 (2) pp 27-36

[54] Rao,G.P 'Identification of Time-Lag Systems via Walsh Functions'. IEEE Trans. on Automatic Control. 1979 Vol AC-24 (5) pp806-808

[55] Bohn,E.V. ´Walsh Function Decoupled Parameter Estimation Equations for Dynamic Continuous-Time Models with Time Delay´. IFAC Identification and System Parameter Estimation York UK 1985

[56] Shinbrot,M. ´On the Analysis of Linear and Nonlinear System´. Trans. ASME 1957 pp547-551

[57] Shinbrot,M. ´On the Analysis of Linear and Non-linear Dynamical Systems´. NACA Tech Note TN-3288 1954

[58] Loab,J.M. and Cahen,G.M. ´More About Process Identification´. IEEE Trans. on Automatic Control 1965 Vol AC-10 pp 359-361

[59] Loeb,J.M.,Cahen,G.M. ´Extraction, Apartir des Enregistrements de Mesures, des Parametres Dynamiques d´un Systeme´. Automatisme 1963 Vol 8 (12) pp479-486

[60] Loeb,J.M. ´Les Erreurs systematique et aleatoires dans la determination Experimentale des Fonctions de Transfert´. Document 159/1 Congres IFAC Switzerland 1965 pp614-619

[61] Takaya,K. ´The Use of Hermite Functions for System Identification´. IEEE Trans. on Automatic Control 1968 Vol AC-13 pp446-447

[62] Matilensky,M. ´On-Line Parameter Estimation of Continuous Processes´. 6th IFAC Congress Boston 1975

[63] Matilensky,M. ´Identification of Continuous Dynamical Systems with "Spline-Type Modulating Functions Method"´. Identification and System Parameter Estimation. Iserman,R.[Ed] Pergamon 1979

[64] Belorusets,V.B. ´Method of Auxiliary Systems for the Identification of Dynamic Systems with Unknown Input´. Automation and Remote Control 1981 Vol42 pp1056-1061

[65] Pearson,A.E. ´Time Limited Identification of Continuous Systems Using Trigonometric Modulating Functions´. Proc. 3rd Yale Workshop on Applications of Adaptive Systems Theory, 1983 pp168-173

[66] Pearson,A.E. ´On the Identification of Polynomial Input-Output Differential Systems´. IEEE Trans. On Automatic Control 1985 Vol AC-30 (8) pp778-782 Vol

[67] Korobov,J. 'New Method for Analysis of Process Dynamics'.Control 1964 pp209-211

[68] Saha,D.C. and Rao,G.P. 'A General Algorithm for Parameter Identification in Lumped Parameter Continuous Systems'. IEEE Trans. On Automatic Control 1982 Vol AC-27 (1) pp223-225

[69] Fairman,F.W. 'Parameter Identification for Linear Time Varying Dynamic Processes'. Proc. IEE 1970 Vol 117 (10) pp2025-2029

[70] Mironovskii,L.A. and Yudovich,V.S 'Unified Approach to Identification of Linear Stationary Processes'. Automation and Remote Control. 1978 Vol 39 pp44-50

[71] Towill,D.R. 'Dynamic Testing of Control Systems'. The Radio and Electronic Engineer. 1977 Vol 47 pp505-521

[72] Payne,P.A., Towill,D.R. and Baker,K.J. 'Predicting Servomechanism Dynamic Perfomance Variation from Limited Production Test Data'. The Radio and Electronic Engineer 1970 Vol 40 (6) pp275-288

[73] Sriyananda,H. and Towill,D.R. 'Fault Diagnosis Using Time Domain Measurements'. The Radio and Electronic Engineer 1973 Vol 43 (9) pp5233-533

[74] Garzia,R.F. 'Fault Isolation Computer Methods'. NASA CR-1758 1971

[75] Allen,R.J. 'Failure Prediction Employing Continuous Monitoring Techniques'. IEEE Trans on Aerospace-Support Conference Procedures 1963 pp924-930

[76] Seshu,S. 'Fault Isolation in Conventional Linear Systems - A Feasibility Study'. IEEE Trans. on Reliability 1966 Vol R-15 (1) pp11-16

[77] Hsieh,C.C. and Chan,S.P. 'Some Practical Considerations in Circuit Design and Fault Diagnosis'. Proc. Asilomar Conf. on Circuits and Systems 1970 pp45-49

[78] Towill,D.R. and Payne,P.A. 'Frequency Domain Approach to Automatic Testing Of Control Systems'. The Radio and Electronic Engineer 1971 Vol 41 (2) pp51-60

[79] Sriyananda,H., Towill,D.R. and Williams,J.H. 'Voting Techniques for Fault Diagnosis from Frequency Domain Test Data'. IEEE Trans. on Reliability 1975 Vol R-24 (4) pp260-267

[80] Varghese,K.C 'Test Feature Selection for a Complex Electro-Hydraulic Servo System using Frequency Response Measurements'. Pattern Recognition 1981 Vol 14 (1-6) pp147-153

[81] Polovko,A.M. 'Fundamentals of Reliability Theory'. Academic Press 1968

[82] Himmelblau,D.M. 'Fault Detection and Diagnosis in Chemical and Petrochemical Processes'. Elsevier 1978

[83] Isermann,R. 'Process Fault Detection Based on Modelling and Estimation Methods - A Survey'. Automatica 1984 Vol 20 (4) pp387-404

[84] Valstar,J.E. 'In-Flight Dynamic Checkout'. IEEE Trans. Aerospace Support Conf. 1963 pp213-221

[85] Pau,L.F. 'Failure Diagnosis and Performance Monitoring'. Control and Systems Theory Vol 11 Marcel Dekker Inc. 1981

[86] Willsky,A.S. 'A Survey of Design Methods for Failure Detection in Dynamic Systems'. Automatica 1976 Vol 12 pp601-611

[87] Bohlin,T. 'System Identification' Ed. R.K. Mehra, Academic Press, 1976

[88] Park,S. and Himmelblau,D.M. 'Fault Detection and Diagnosis via Parameter Estimation in Lumped Dynamic Systems'. Ind. Eng. Chemical Process Des. Dev. 1983 Vol 22 pp482-487

[89] Duhamel,P. 'Automatic Test Generation Techniques for Analog Circuits and Systems: A Review'. IEEE Trans. on Circuits and Systems 1979 Vol CAS-26 (7) pp411-440

[90] Goedecke,W. 'Fault Detection in a Tubular Heat Exchanger Based on modelling and Parameter Estimation'. IFAC Identification and System Parameter Estimation York UK 1985

[91] Mason,C. 'Modulating Function method'. University of Edinburgh Report 1982

[92] Beck,M.S. 'Adaptive Control - Fundamental Aspects and their Applications'. Proc. First Annual Advanced Control Conference, Purdue University,IN, USA 1974 pp 1-25

[93] Henry,R.M. ´An Efficient Algorithm for Performing Polar Correlation on a Small On-Line Computer´. Open University 1976

[94] Fell,R. ´Microprocessor based Cross Correlators using the Skip Algorithm´. IEE Colloquium Digest on the Influence of Microelectrics on Measurement Instruments and Transducer Design 1982 pp 25-32

[95] Thiruarooran,C. ´The Assessment and Construction of Models for Transfer Function Evaluation of Turbocharged Diesel Engines´. 11th Israel Conf. on Mech. Eng. 1977

[96] Brodin,G. ´Power Oscillations in Parallel Connected Synchronous Alternators Driven by Diesel Engines´. 1962 Det Norske Veritas Pub No.31

[97] Jalali-Naini,S.A. and Jordan,J.R. ´System Identification with Hermite Modulating Functions´. To be published in Automatica

[98] Gustavsson,I. ´Identification of Processes in Closed Loop´. Automatica 1977 Vol 13 pp 59-75

APPENDICES

TRANSFER FUNCTION MEASUREMENT WITH A MODIFIED MODULATING FUNCTION METHOD

J.R. JORDAN, N.E. PATERSON and H.W. WHITTINGTON

TRANSFER FUNCTION MEASUREMENT WITH A MODIFIED MODULATING FUNCTION METHOD

J.R. Jordan, N.E. Paterson and H.W. Whittington

Department of Electrical Engineering, University of Edinburgh, King's Buildings
Mayfield Road, Edinburgh EH9 3JL, U.K.

Abstract. The transfer function identification method described in this paper is based on the modulating function technique. It has been modified to use measured cross-correlation and auto-correlation functions. This paper presents a brief introduction to the modified modulating function method and the results of a practical assessment of its use carried out at Stornoway Power Station.

INTRODUCTION

The transfer function analysis method described in this paper arose from a study of the monitoring requirements of diesel powered generating stations. Long term monitoring of the small signal parameters of the diesel/alternator system may provide useful diagnostic information about the health of the system. A low cost, microcomputer based, measurement technique is required.

This work is being carried out in collaboration with the North of Scotland Hydroelectric Board and field trials are being carried out at their more remote generating stations. Consequently a transfer function measurement method was required that could be left to operate, unattended for long periods. In practice, this suggests that bad data should be detected before it is processed by the transfer function algorithm and, that it should be easy to implement and check that small signals test conditions have actually been established. These considerations have led to the use of correlation preprocessing of the data before system identification is attempted.

The identification method for this work is based on the modulating function method introduced by Loeb (1965). This work can be related back to a report by Shinbrot published in 1954 (Eisenfeld 1979). Hermite modulating functions, proposed by Takaya (1968) have been used. Mironvuskii et al (1978) and Belorusets (1981) have related the modulating function method to other linear identification methods. It should be noted that the modulating function method directly measures the parameters of a systems, small-signal, differential equation.

The modulating function method was originally developed to operate on data directly derived from measurements made on the system under test. The method has been modified to use preprocessed data in the form of auto- and cross-correlation functions. The correlation functions provide a useful intermediate diagnostic check on the data. A number of efficient methods for microcomputer evaluation of correlation integrals have been produced. In this work use is being made of the Fell Skip Algorithm (Fell 1982) to implement the relay correlation function.

The measurement technique was initially validated by using a software simulation of the convolution integral to generate specimen auto- and cross-

correlation functions. Encouraging results were obtained from a more realistic test of the experimental system which was subsequently carried out by using a hardware simulation of a second-order system. In this paper the result of tests carried out on a large diesel/alternator set are presented.

Section 2 of this paper will introduce the modified modulating function method. Field trials of the experimental system were carried out at Stornoway Power Station in August 1984. The experimental system is described in Section 3 and the results obtained described in Section 4.

THE MODULATING FUNCTION METHOD

The output and input of a single-input linear system are related by the convolution integral. If x is the input and y the output then auto-correlation r_{xx} and cross-correlation r_{yx} may be defined and the convolution integral becomes

$$r_{yx}(a) = \int_{-\infty}^{\infty} h(a-t).r_{xx}(t)dt \tag{1}$$

or

$$R_{yx}(s) = H(s).R_{xx}(s) \tag{2}$$

where $R_{yx}(s)$, $R_{xx}(s)$ and $H(s)$ are the Laplace Transforms of the cross-correlation function, the auto-correlation function and the system impulse response, respectively.

It follows that the system differential equation may be reformulated in terms of correlation functions to give

$$q_n \cdot \frac{d^n r_{yx}(a)}{da^n} + q_{n-1} \cdot \frac{d^{n-1} r_{yx}(a)}{da^{n-1}} + \dots +$$

$$q_o \cdot r_{yx}(a) = r_{xx}(a) \tag{3}$$

This differential equation may be converted to an algebraic equation by multiplying by a suitably chosen modulating function, ϕ, and integrating by parts over the interval 0 to T.   If ϕ and all of its derivatives are zero outside of the interval 0 to T then it can be shown that

$$(-1)^n . q_n \int_0^T \phi^n(a) . r_{yx}(a) . da + \ldots$$

$$+ \ldots (-1)^{n-1} . q_{n-1} \int_0^T \phi^{n-1}(a) . r_{yx}(a) . da \ldots$$

$$+ \ldots + q_0 \int_0^T \phi(a) . r_{yx}(a) . da =$$

$$\int_0^T \phi(a) . r_{xx}(a) . da, \text{where } \phi^n(a) = \frac{d^n \phi(a)}{da^n} \qquad (4)$$

By choosing different sets of modulating functions a set of linear equations is obtained which can be represented by a matrix equation. The system parameters $q_n \ldots q_0$ are obtained by matrix inversion. In general, when phase advance terms are included, i.e. when the transfer function is of the form

$$H(s) = \frac{p_k s^k + \ldots 1}{q_n s^n + \ldots q_0} \qquad (5)$$

the matrix will be increased to include the parameters p_k to p_1. A correspondingly larger number of modulating functions will be required in this case.

Takaya (1968) introduced the use of Hermite functions as modulating functions. Hermite functions are defined by

$$H_n(t) = (-1)^n . \exp(t^2/2) . \frac{d^n \exp(-t^2/2)}{dt^n} \qquad (6)$$

The properties of Hermite functions guarantees that the modulating functions derived from them are continuous and differentiable. Takaya shows that the Hermite Modulating Functions and their derivatives approximately become zero at both upper and lower boundaries. Figure 1 shows the set of Hermite Modulating Functions used to assess the performance of this identification method.

Since the cross-correlation function is not required to be a good approximation to the system impulse response it is not necessary to design the input auto-correlation function to have a very small width compared with the decay time of the cross-correlation function. The experimental system used Gaussian filtered, pseudo random noise with a bandwidth approximately double the resonant frequency of the system under test. The input noise level was always adjusted to ensure that the developed power was never perturbed by more than ± 3% of its steady state value.

The power signal typically has a superimposed very low frequency drift term. To make best use of the dynamic range of the data logger and to avoid overload conditions this drift term must be eliminated by using an A.C. coupling network. If the power signal is A.C. coupled then, to balance the measurement system, the noise test signal must also be A.C. coupled into the data logger with an identical network. In this case the measured cross- and auto-correlation functions are defined by

$$R_{vw}(s) = H(s) . R_{ww}(s) \qquad (7)$$

since

$$V(s) = H(s) . H_c(s) . X(s)$$

and

$$X(s) = \frac{1}{H_c(s)} . W(s)$$

where $H(s)$ is the transfer function of the system under test and $H_c(s)$ is the transfer function of the coupling networks. The signals x, y, v and w are as defined in Figure 2. Algebraically the effect of the coupling networks has been completely cancelled but in practice care must be taken to use a coupling time constant which is large enough to ensure that sufficient of the low frequency content of the output power signal has been coupled into the data logger. A simple RC circuit with transfer function

$$H_c(s) = \frac{sRC}{1 + sRC} \qquad (8)$$

was used as the coupling network in the experimental system.

The measurement system, as currently developed, cannot automatically take account of system time delay terms. When a time delay term is observed it is necessary to interrupt the software and use key-board control to offset the cross-correlation data set until the delay has been eliminated. This is an area of this work that requires further attention.

EXPERIMENTAL SYSTEM

The engine used for these tests was a 4.6 MW Mirrlees Blackstone KV12 Major, used for base load generation at Stornoway Power Station. The engine was controlled by an electronic govenor and it was this circuit that enabled a Gaussian noise test signal to be injected. The power response of the system to this test signal was monitored by using current and voltage transducers on one phase of the alternator output. The noise signal and the power response were A.C. coupled to a data logger with a 0.1 Hz coupling time constant. Data was then down loaded via an IEE BUS to an HP85, desk top computer, where auto- and cross-correlations were generated using a relay correlation technique. In addition to this instrumentation a Spectrum Analyser was used to produce frequency response and coherence functions. A block diagram of the experimental system is shown in Figure 2.

The relay correlation implementation was based on the Skip Algorithm (Fell, 1982). This involved a batch processing of the data where two blocks of 2048 bytes were down loaded from the logger to the HP85 computer. The sample rate was set to define the required correlation function time delay intervals. Only every eighth cross-product was summed to form the correlation estimate. A negligible variance increase was observed compared with the results obtained by using the whole data set.

RESULTS

The engine was run at a constant load of 3.5 MW and the perturbing noise signal level adjusted to restrict the resulting changes in the power level to never exceed ± 3% of the steady state value. The correlation function estimates were

exponentially averaged and the modulating function method used to obtain the model parameters.

The measured cross-correlation function clearly indicates a response which includes pure time delay term. In the first instance a simple, resonant, low pass, second-order function was used to model the remainder of the response. The model transfer function was

$$H(s) = \frac{Kw_n^2}{s^2 + 2cw_n s + w_n^2} \cdot e^{-sT} \qquad (9)$$

where T = time delay
C = damping factor
w_n = undamped natural frequency
K = gain constant

The time delay term is removed before the modulating function method is used. A typical cross-correlation function after removal of this time delay is shown in Figure 3. The model transfer function becomes

$$H(s) = \frac{1}{q_2 s^2 + q_1 s + q_0} \qquad (10)$$

where q_2, q_1 and q_0 are the parameters to be measured.

Frequency response and coherence functions obtained from a Wavetek spectrum analyser monitoring the outputs from the A.C. coupling networks are shown in Figure 4. The coherence falls rapidly as the frequency decreases below 0.5 Hz. This reduction is caused by the A.C. coupling network.

The following transfer function parameters were obtained with the same power and noise levels used to obtain Figure 4 (i.e. 3.5 MW and 0.6 V r.m.s.).

q_2 ...... 0.04 to 0.1

q_1 ...... 0.14 to 0.23

q_0 ...... 9.6 to 12.5

Frequency response functions (including the effect of the measured system time delay) were calculated by substituting s = jw into the transfer functions obtained by using the measured parameters. Figure 5 shows the calculated magnitude and phase response. Good agreement is obtained between the measured response (Figure 4) and the calculated response.

CONCLUSIONS

Initial studies have shown that the modified modulating function method can be used to estimate the small-signal parameters of practical systems. Field trials at Stornoway Power Station of an experimental implementation of the method has shown that the method is easy to use and suitable for operation at remote sites.

More work is needed to (i) investigate the optimum choice of averaging procedure for correlation and parameter estimates, (ii) investigate the effect of the windowing action of the modulating functions on the cross-correlation function (iii) investigate the effect of choice of model order and (iv) investigate procedures for automatically off-setting system time delay.

REFERENCES

Belorusets, V.B. (1981), "Method of auxilary systems for the identification of dynamic systems with unknown input", Translated from Automatika i Telemekhanika, No.8, pp.76-82, Plenum Publishing Corporation, 1982.

Eisenfeld, J. (1979), "Remarks on the modulating functions method for impulse response identification", IEEE Trans. on Automatic Control, AC-24, 3, pp.498-499.

Fell, R. (1982), "Microprocessor-based cross-correlation using the skip algorithm", Proc. Conf. on the Influence of Micro-electronics on Measurement and Transducer Design, pp.25-32, Manchester.

Loeb, J.M. and G.M. Cahen (1965), "More about process identification", IEEE Trans. on Automatic Control, AC-10, pp.359-361.

Mironvouskii, L.A. and V.S. Yodovich (1978), "Unified approach to identification of linear stationary processes", Translated from Automatika i Telemekhanika, No.1, pp.66-67, Plenum Publishing Corporation, 1978.

Takaya, K. (1968), "The use of Hermite functions for system identification", IEE Trans. on Automatic Control, AC-13, pp.446-447.

Figure 1. Hermite modulating functions
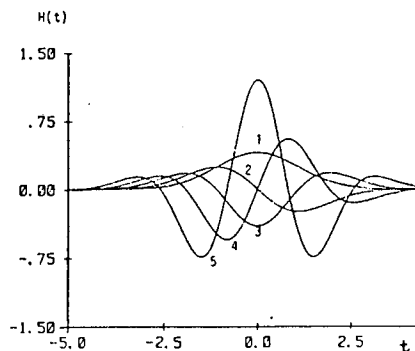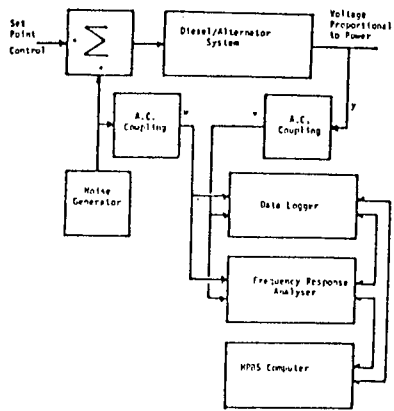
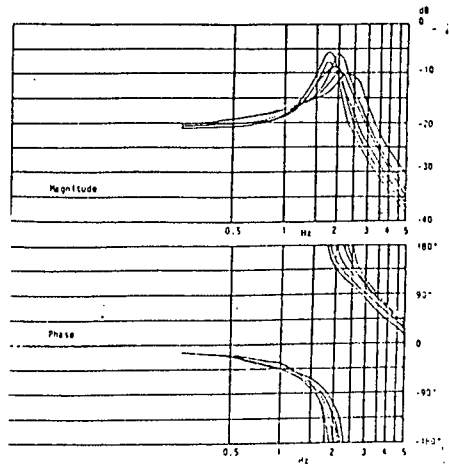Figure 2. Block diagram of the experimental system

Figure 5. Measured frequency response derived from model using measured parameters
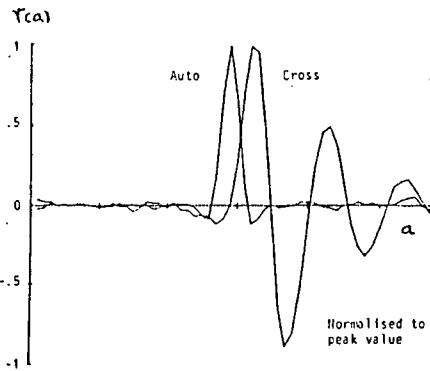

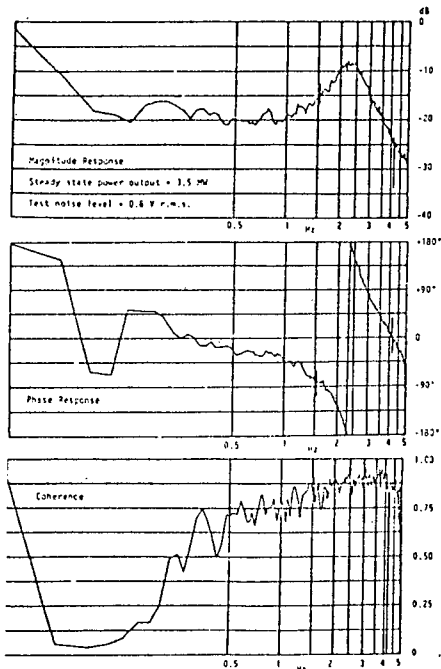Figure 3. Typical measured cross-correlation after removal of system time delay

Figure 4. Measured frequency response and coherence function

230

A MODULATING-FUNCTION METHOD FOR ON-LINE FAULT DETECTION

J.R. JORDAN and N.E. PATERSON

231

A modulating-function method for on-line fault detection

J R Jordan and N E Paterson
Department of Electrical Engineering, University of Edinburgh,
The King's Buildings, Edinburgh EH9 3JL, UK

Abstract. A modified modulating-function method for
identifying the parameters of the differential equations that
represent the small-signal performance of a system is described.
The method makes use of the input auto-correlation function
and the cross-correlation function relating the output to the
input of the system. The application of the method to the on-line
detection of faults is discussed.

1. Introduction

Systems for monitoring the health of machinery are receiving
more attention as new diagnostic algorithms are developed and
the cost of microelectronic implementation decreases. A study
of the monitoring requirements of diesel-powered generating
stations has indicated that it would be interesting to investigate
how the parameters of the small-signal transfer function vary as
the performance of the diesel control system changes throughout
its period of operation. It remains to be discovered whether the
small-signal parameters will change significantly as the control
system operates at different points on its typically non-linear
transfer characteristic and enable the state of tune of the engine
to be monitored.

This work is being carried out in collaboration with the
North of Scotland Hydroelectric Board and field trials are
being carried out at their more remote generating stations.
Consequently, a transfer-function measurement method was
required that could be left to operate, unattended for long
periods. In practice, this suggests that bad data should be
detected before they are processed by the transfer-function
algorithm and that it should be easy to implement and check
that small-signal test conditions have actually been established.
These considerations have led to the use of correlation
preprocessing of the data before system identification is
attempted.

One identification method chosen for this work is based
on the modulating-function method (Loeb and Cahen 1965).
This method can be related back to a report by Shinbrot
(Eisenfeld 1979) published in 1954. Hermite modulating func-
tions, proposed by Takaya (1968), have been used. Mironvuskii
and Yodovich (1978) and Belorusets (1981) have related the
modulating-function method to other linear identification
methods. It should be noted that the modulating-function
method directly measures the parameters of a system's small-
signal differential equation. The modulating-function method
was originally developed to operate on data directly derived
from measurements made on the system under test. In this paper
the method is modified to use preprocessed data in the form of
auto- and cross-correlation functions. The correlation functions
provide a useful intermediate diagnostic check on the data. A
number of efficient methods for microcomputer evaluation of
correlation integrals have been produced. In this work we use

the Fell skip algorithm (Fell 1982) to implement the relay correlation function.

A prototype transfer-function analyser (constructed from available computing and data-logging subunits) has been used to monitor simulated systems and the power control loop of a diesel powered alternator. It is expected that a wide range of applications will become feasible once a dedicated, low-cost, microcomputer-based, implementation has been developed.

Section 2 of this paper presents a brief review of fault detection using transfer functions. Section 3 introduces the modulating-function method and a discussion of the errors resulting from the use of models having a lower order than the actual system transfer function. The results of software simulation experiments demonstrating the use of the modulating-function method are described in § 4.

2. Fault detection using transfer functions

The method of detecting faults using changes in the small-signal, linearised, transfer function was investigated during the early 1960's. Interest in this area was particularly stimulated by NASA's space flight programme (Allen 1963). Recently this area has been given a considerable boost by the availability of low-cost, powerful, microcomputers. Typically, a signature characteristic of good performance is obtained by measuring frequency and time-domain transfer-function features when acceptable performance has been achieved and then predicting a fault condition when a change beyond some predefined level is observed. A major problem with this type of system is the setting of levels that minimise the number of false predictions of a fault condition and at the same time accurately indicate a damaging fault condition.

Many authors have proposed fault detection by monitoring gain and phase with various test frequencies, usually developing this further (i.e. by using pattern recognition techniques) to isolate the fault. Towill and Payne (1971), for example, used frequency response data and set limits to define an acceptable range of response values. Systems having responses that fall within these limits are passed as good systems and those that do not, fail the test. The limits were arrived at by creating a range of unacceptable systems (caused by a known fault condition) and identifying the frequencies at which these faulty systems are most easily recognised.

Test procedures have been based on cross- and auto-correlation of the input and output signals of the system under test. Naturally occurring noise signals do not usually have sufficient bandwidth to exercise all operating modes. Towill (1977) has stimulated the system under test with a pseudo-random noise signal and derived the impulse response of the system from the cross-correlation of the output signal with the input signal. Each measured impulse response was compared with a standard impulse response (representing good performance) and deviations used to indicate a fault condition. Voting techniques have been developed to optimise the process of fault location for frequency-domain (Sriyananda and Towill 1973) and time-domain test methods (Sriyandanda et al 1975).

Isermann (1984) has surveyed fault detection techniques based on the use of modelling and estimation methods to produce a parametric description of the system under test. For fault detection the system model should have a direct link with the physical description of this model. Hence a continuous-data model is preferable to a sampled-data model. Isermann has noted that 'there is still a need for robust parameter estimation methods with less computational effort which provide consistent and efficient estimates of the parameters of a continuous model under many noise conditions'. The modified modulating-function method described in this paper satisfies many aspects of this need.

3. The modulating-function method

The output and input of a single-input linear system are related by the convolution integral. If x is the input and y the output then auto-correlation r_{xx} and cross-correlation r_{yx} may be defined and the convolution integral becomes

$$r_{yx}(a) = \int_{-\infty}^{\infty} h(a-t)r_{xx}(t)\,dt \tag{1}$$

or

$$R_{yx}(s) = H(s)R_{xx}(s) \tag{2}$$

where $R_{yx}(s)$, $R_{xx}(s)$ and $H(s)$ are the Laplace transforms of the cross-correlation function, the auto-correlation function and the system impulse response, respectively.

It follows that the system differential equation may be reformulated in terms of correlation functions to give

$$\frac{q_n\,d^n r_{yx}(a)}{da^n} + \frac{q_{n-1}\,d^{n-1} r_{yx}(a)}{da^{n-1}} + \ldots + q_0 r_{yx}(a) = r_{xx}(a). \tag{3}$$

This differential equation may be converted to an algebraic equation by multiplying by a suitably chosen modulating function, φ, and integrating by parts over the interval 0 to T. If φ and all of its derivatives are zero outside of the interval 0 to T then it can be shown that

$$(-1)^n q_n \int_0^T \varphi^n(a)r_{yx}(a)\,da + (-1)^{n-1} q_{n-1} \int_0^T \varphi^{n-1}(a)r_{yx}(a)\,da$$

$$+ \ldots + q_0 \int_0^T \varphi(a)r_{yx}(a)\,da = \int_0^T \varphi(a)r_{xx}(a)\,da. \tag{4}$$

It should be noted that the modulating-function method is usually applied to the system differential equation relating input x to the output v.

By choosing different sets of modulating functions a set of linear equations is obtained which can be represented by the matrix equation shown below where $\varphi^n(a)$ has been used to represent the nth derivative of $\varphi(a)$

$$\begin{bmatrix} (-1)^n \int_0^T \varphi_1^n(a)r_{yx}(a)\,da & \ldots & \int_0^T \varphi_1(a)r_{yx}(a)\,da \\ \vdots & & \vdots \\ (-1)^n \int_0^T \varphi_m^n(a)r_{yx}(a)\,da & \ldots & \int_0^T \varphi_m(a)r_{yx}(a)\,da \end{bmatrix} \begin{bmatrix} q_n \\ \vdots \\ q_0 \end{bmatrix}$$

$$= \begin{bmatrix} \int_0^T \varphi_1(a)r_{xx}(a)\,da \\ \vdots \\ \int_0^T \varphi_m(a)r_{xx}(a)\,da \end{bmatrix}. \tag{5}$$

The system parameters $q_n \ldots q_0$ are obtained by matrix inversion. In general, when phase advance terms are included, i.e. when the transfer function is of the form

$$H(s) = \frac{(P_k s^k + \ldots + 1)}{(q_n s^n + \ldots + q_n)} \tag{6}$$

the matrix will be increased to include the parameters p_k to p_1. A correspondingly larger number of modulating functions will be required in this case.

Takaya (1968) introduced the use of Hermite functions as modulating functions. Hermite functions are defined by

$$H_n(t) = (-1)^n \exp(t^2/2)\frac{d^n \exp(-t^2/2)}{dt^n}. \tag{7}$$

The properties of Hermite functions guarantee that the

233

modulating functions derived from them are continuous and differentiable. Takaya shows that the Hermite modulating functions and their derivatives approximately become zero at both upper and lower boundaries. Figure 1 shows the set of Hermite modulating functions used to assess the performance of this identification method.
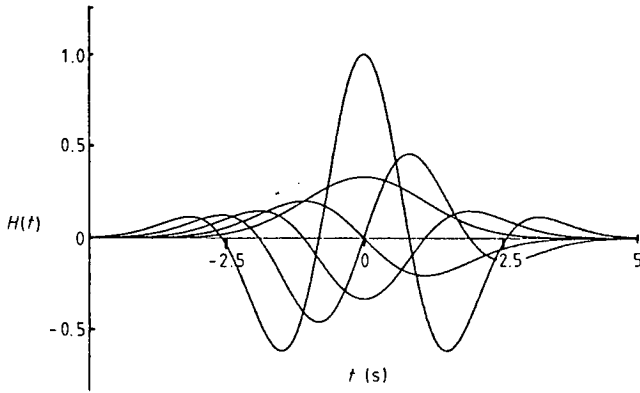

Figure 1. Modulating Hermite functions as used in the experimental system.

When the order of the model used is lower than the actual system order the modulating-function method will not measure the expected parameter values. If the system transfer function is written as

$$H(s) = \frac{N(s)}{D(s)} \qquad (8)$$

then the numerator and denominator functions can be partitioned into a high-order part and a low-order part i.e.

$$D(s) = Q_L(s) + Q_H(s) \qquad (9)$$

$$N(s) = P_L(s) + P_H(s) \qquad (10)$$

where subscript L indicates low order and subscript H indicates high order. Hence equation (2) becomes

$$Q_L(s)R_{yx}(s) = (P_L(s) + P_H(s))R_{xx}(s) - Q_H(s)R_{yx}(s).$$

Hence

$$Q_L(s)R_{yx}(s) = P_L(s)R_{xx}^1(s) \qquad (11)$$

where

$$R_{xx}^1(s) = \frac{H(s)}{H_L(s)} R_{xx}(s) \qquad (12)$$

and

$$H_L(s) = \frac{P_L(s)}{Q_L(s)}.$$

From (11) the time-domain differential equations for the model are

$$q_{nL} \frac{d^n r_{yx}(a)}{da^n} \ldots + q_{0L}r_{yx}(a) = P_{nL} \frac{d^n r_{xx}^1(a)}{da^n} + \ldots + r_{xx}^1(a) \qquad (13)$$

where $r_{xx}^1(a) = r_{xx}(a) + \Delta(a)$ where $r_{yx}(a)$ and $r_{xx}(a)$ are measured cross- and auto-correlation and $\Delta(a)$ represents the effect of $H(s)/H_L(s)$ acting on $R_{xx}(s)$. Hence the error introduced by a low-order model can be described by an error in the measured input auto-correlation functions.

Time-delay terms introduce high-order numerator terms which can introduce errors if they are ignored. Practical

experience has shown that a time delay observed in the measured cross-correlation can be backed-off before the modulating-function method is applied. The errors introduced by low-order models and partial cancellation of time delays can be investigated by convolving the impulse response of the model with the input auto-correlation and then comparing the result with the measured system cross-correlation. The results of an experimental investigation into the significance of the errors introduced by the use of low-order models will be presented in the next section.

4. Performance of the modulating-function method

The modulating-function method has been applied to a software simulated system having a dominant-transfer function in series with higher-order terms. The time constraints of the higher-order terms were adjusted and the effect on the parameters measured

Figure 2. Graph showing the variation of the coefficient of s^2 when system under test comprises a second-order low-pass filter in series with a first-order low-pass filter. Line 1 shows variation of coefficient of s^2 in overall third-order function as first-order pole position is changed. Line 2 shows variation of this coefficient in the measured second-order approximation.

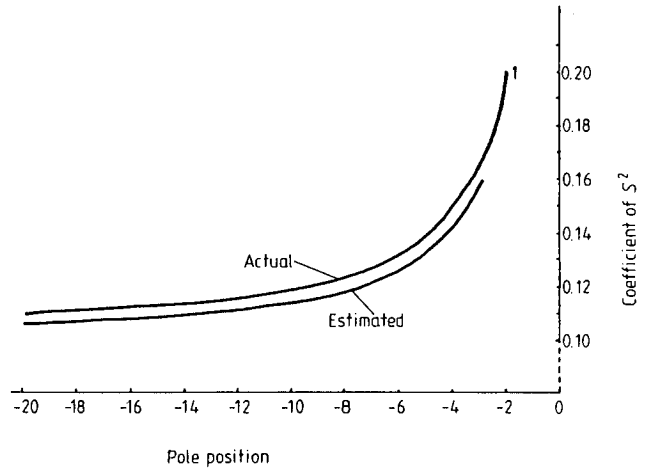
Figure 3. Graph showing the variation of the coefficient of s when the system under test comprises a second-order low-pass filter in series with a first-order low-pass filter. Line 1 shows variation of coefficient of s in overall third-order function as first-order pole position is changed. Line 2 shows variation of this coefficient in the measured second-order approximation.
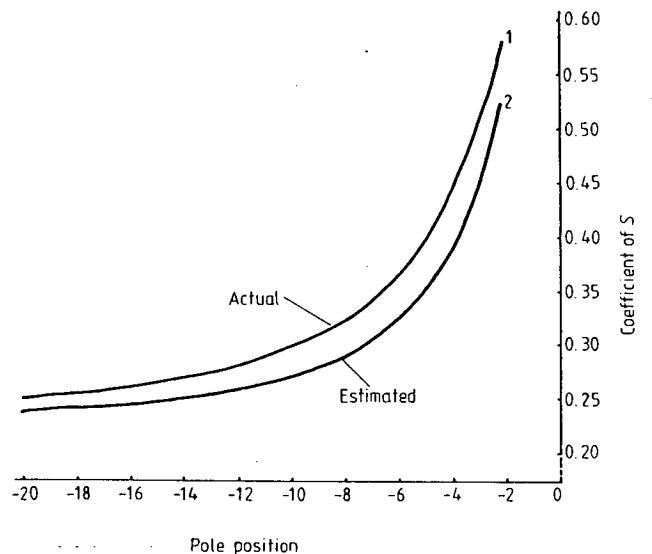
683

234

assuming the system to have a transfer function defined only by the dominant terms. Experiments involving a dominant second-order system will be described.

A second-order dominant-pole pair was positioned at −1 on the real axis of the pole zero map and a single pole moved towards it from −20 on the real axis. The third-order system so formed was modelled by a second-order function and the parameters of this model obtained by using the modulating function method. Figure 2 shows the error between the actual coefficient of s^2 and the model coefficient of s^2. It will be seen that the errors start to decrease significantly when the single pole is at least ten times further out in the plane than the second order poles. Figure 3 shows similar results for the coefficient of s. Figure 4 shows how the normalised parameters of the model (natural frequency, ω_n, damping factor c, and gain, g) vary as the single pole is moved towards the pole pair and the system becomes predominantly third order. The effect of partial elimination of time delay (i.e. giving a small residual time delay or time advance) on the measurement of the parameters of a second-order system is shown in figure 5. It will be seen that accurate results are obtained when the model order and system order are the same.
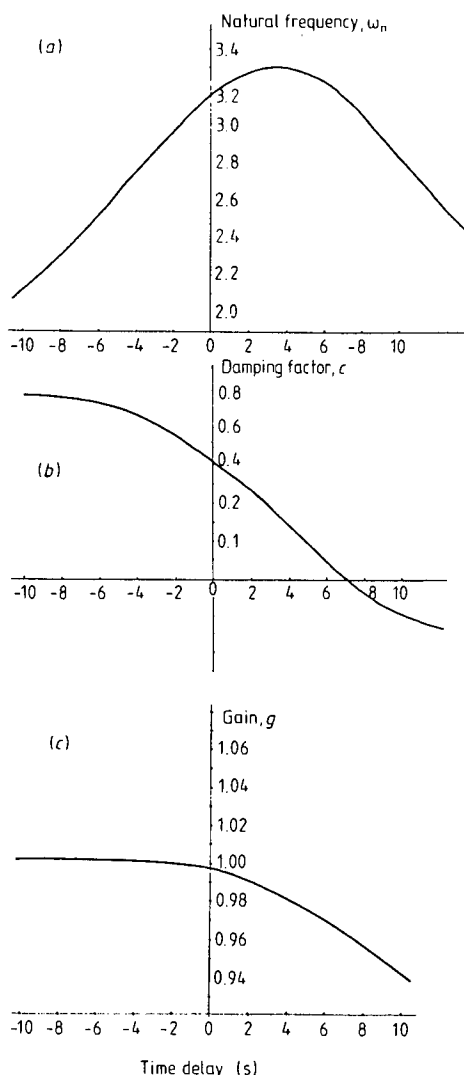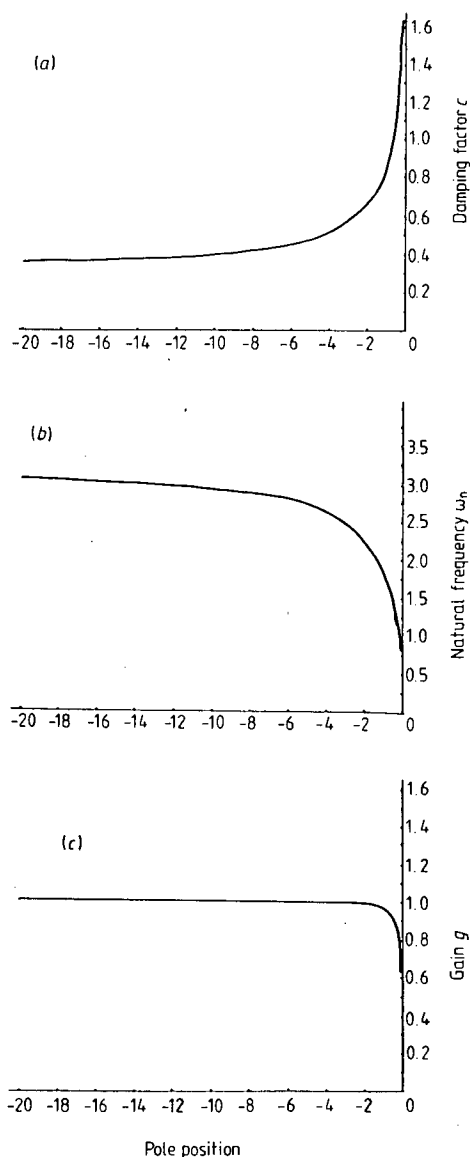
Figure 5. Graphs demonstrating the effect of time delay or advance on the measurement of standardised second order parameters.

5. Conclusions

Simulation studies and field trials of a prototype measurement system at Stornoway Power Station Centre (the power control loop of a diesel powered alternator was investigated) have shown that the modified modulating-function method can be used to reliably measure transfer-function parameters (Jordan *et al* 1985). Although an accurate model is not necessary for fault detection purposes it is clear that great care must be taken to ensure that observed changes are due to faults and not due to a variation in the neglected high-order terms. More work is required to relate changes of measured transfer-function parameters to particular fault conditions. A microcomputer-based implementation of the modified modulating-function method, which incorporates the transfer-function based fault diagnosis in a wider range of instruments and measurement systems, is being developed.

References
Allen R J 1963 Failure prediction employing continuous monitoring techniques
IEEE Trans Aerospace Support Conf. Proc. AS-1 924–30

Figure 4. Measured standardised second-order parameters as third-order function is changed by altering the position of its first-order pole.

Belorusets V B 1981 Method of auxiliary systems for the
identification of dynamic systems with unknown input
Translated from *Automatik i Telemekhanika* **8** pp 76–82
(New York: Plenum)

Eisenfeld J 1979 Remarks on the modulating functions method
for impulse response identification
IEEE Trans. Automatic Control **AC-24** 498–9

Fell R 1982 Microprocessor based cross correlation using the
skip algorithm
*Proc. Conf. on the Influence of Microelectronics on
Measurement and Transducer Design, Manchester* pp 25–32

Isermann R 1984 Process fault detection based on modelling
and estimation methods – a survey
Automatica **20** 387–404

Jordan J R, Paterson N E and Whittington H W 1985 Transfer
function measurement with a modified modulating function
method *Proc. 7th IFAC/IFORS Symp. on Identification and
System Parameter Estimation, York* pp 1085–8

Loeb J M and Cahen G M 1965 More about process
identification
IEEE Trans. on Automatic Control **AC-10** 359–61

Mironvouskii L A and Yodovich V S 1978 Unified approach to
identification of linear stationary processes
Translated from *Automatika i Telemekhanika* **8** 76–82
(New York: Plenum)

Sriyananda H and Towill D R 1973 Fault diagnosis using time
domain measurements
Radio and Electronic Engineer **43** 523–33

Sriyananda H, Towill D R and Williams J H 1975 Voting
techniques for fault diagnosis from frequency domain test data
IEEE Trans. Reliability **R-24** 260–7

Takaya K 1968 The use of Hermite function for system
identification
IEEE Trans. Automatic Control **AC-13** 446–7

Towill D R 1977 Dynamic testing of control systems
Radio and Electronic Engineer **47** 505–21

Towill D R and Payne P A 1971 Frequency domain approach
to automatic testing of control systems
Radio and Electronic Engineer **41** 51–60

PERFORMANCE MONITORING OF DIESEL ELECTRICITY GENERATION

H.W. WHITTINGTON, J.R. JORDAN, N.E. PATERSON

and P.M. JOHNSON

237

Performance monitoring of diesel electricity generation

H.W. Whittington, J.R. Jordan, N. Paterson, and P.M. Johnson

Abstract: The paper describes the characteristics of diesel generation plant and discusses the role of diesel generation in the context of bulk supply of electricity. The specific operational and maintenance aspects of diesel generation are presented and the need for the development of aids to routine maintenance are discussed. One technique for monitoring the performance of diesel plant is described and results are given from tests carried out in Stornoway power station on the island of Lewis.

Introduction

lthough the bulk supply of electricity is well established, he part to be played by small generating plant in electrifi- tion is still significant. For example, in developing coun- ies comparatively small plants will often suffice to serve he early needs of the larger towns, even of the capital city. ater, the growing demand will justify the installation of rger, more sophisticated and more efficient plant; erhaps even the interconnection of two or more town- ips to form the rudiments of a grid or power pool. When is happens it will often be possible to take up the earlier lant from their foundations, and to remove them to naller or more remote centres of population to perform seful pioneer work, further into the heart of the country.

Even in highly developed countries, small power plants rve an important function in remote places, either where ere is low population density or where consumers are nd- or sea-locked. These conditions are not conducive to e early establishment of grid systems and townships are, erefore, likely to remain electrically isolated from one nother for a long time.

Of the options available for the prime mover in such rcumstances, the diesel engine is one which is widely sed. With so many advantages to its credit, it is not sur- rising that, in the past, the diesel engine has found appli- tion in small power stations with total installed pacities of approximately 30 MW.

Larger stations are also becoming more common. This as certainly created a frontier of competition between the iesel engine and the steam turbine, as previously it had een assumed that above about 30 MW installed capacity, e steam turbine becomes economically attractive, and d been used in preference to the diesel engine, [1].

Because of their compactness, diesel stations can be ted at or as near to load centres as permitted by noise gulations. Although diesel engines tend to be rather oisy, tests and experience have shown that even without recautions to cut down noise, it is satisfactory to site rge diesel plants within 800 m of domestic dwellings [2].

Technical considerations of diesel generators

1   General features

he diesel-electric generating set is available in almost any ze over a wide range. Earlier size limitation was imposed y the need to use very low speeds. This meant very bulky

and heavy machines but modern designs have enabled higher speeds to be utilised, with the result that more powerful plant can be built without their physical size being unmanageable. Large diesel engines have long been used for marine propulsion, but today are finding an ever increasing role in land-based operation. For example, as standby generating plant in both large industrial and com- mercial premises or as completely stand-alone plant in iso- lated power systems such as on islands.

2.2   Security of supply

In isolated systems, it is necessary to provide some margin of spare plant so that supply may be maintained when part of the plant is out of service for routine maintenance or because of breakdown. For complete security, it could be argued that in any station two spare sets should be installed, so that when one set is out of service for routine maintenance, there is still a spare set standing by in case of the breakdown of a running set. Although this level of spare capacity may be justified on occasion, the provision of two spare sets is often an expensive luxury. For small, newly established electricity supply undertakings, it is usually of greater importance to produce relatively cheap electricity with the risk of very occasional interruptions, than to have almost perfect continuity of supply. Indeed, even with established systems, the supply authority will accept that there is a finite probability of failing to meet demand at any given time, even working with spare ca- pacity. This risk is of course dependent not only on the reliability of the plant, which in turn depends to a great extent upon the care with which it is maintained, but also on the proportion of time (in days per year) when the plant is out of service for routine maintenance.

Where a large number of units is installed, the risk of a 'double outage' i.e. one set out for maintenance and another set subject to breakdown, will be greater. No defi- nite ruling can be given as to the amount of spare plant which should be provided; much will depend on the size and nature of the load. The following margins are there- fore suggested tentatively, though it is recognised that many individual variations will be necessary:

(*a*) assuming the largest installed unit to be out of service for routine maintenance, and where the total number of sets is five or less, assume there is no coincident breakdown, i.e. one spare set

(*b*) where the total number of installed sets is six or more, assume the largest unit to be out of service for routine maintenance and any other set to be out due to breakdown, i.e. two spare sets.

The average time any machine is out due to a breakdown is usually about 5% of its total operational time in any year. This, combined with a typical planned outage time of

. Whittington, Dr. Jordan and Mr. Paterson are with the Department of Electri- Engineering. University of Edinburgh. The King's Buildings, Edinburgh 9 EJL. United Kingdom. Mr. Johnson is with the North of Scotland Hydroelec- : Board. 16 Rothesay Terrace, Edinburgh EH3 7SE, United Kingdom

238

9.7% per annum would result in a 0.035% chance of a double outage occurring during an average operating year of 4605 hours.

2.3 Life expectancy of plant

When planning diesel-powered electricity generation, it is usual to allow a life of 15 years for diesel plant. For very-high-speed diesels, or engines using fuel of low quality, a shorter period may sometimes be more realistic (10 or 12 years). For slow-speed diesels, or for diesel plant that is operated intermittently without overloading, rather longer periods (up to 20 years) may be taken as according with practice. However, when planning ahead, a 15-year period of utilisation in a given situation would appear to be prudent because diesels tend to become superseded through growth of demand rather than through wear and tear.

3 Operation of plant

Well maintained diesel engines can run for a long period between overhauls and have low wear rates. The average running hours between overhauls for most North of Scotland Hydroelectric Board (NSHEB) machines is greater than 10000 hours, which represents about 18 months continuous operation. Where skilled labour is scarce, station operation can often be adapted for automatic operation.

Maintenance may be simplified by provision of easily replaceable assemblies of parts, thus enabling reconditioning to be undertaken away from the generating plant. Periods between order and delivery are short and a diesel power station may therefore be extended rapidly by adding suitably sized units to meet growing demands. Compact, lightweight, high-speed diesels are obtainable for sites which are remote, cramped or difficult of access. Mobile diesel power units mounted on skids or trailers may be used for temporary or emergency purposes.

Despite the provision of installed redundancy and portable plant, unplanned outages do occur and can be expensive both in terms of lost revenue or increased fuel costs by meeting demand with less efficient plant. Additionally, the inconvenience to the consumers of loss of supply must be considered as of prime importance.

It follows that a system which can indicate the onset of fault conditions and thus allow time for action to be taken will be of considerable value to the supply authority. Maintenance could be scheduled on the basis of equipment performance rather than merely at regular intervals resulting in savings, both recurrent and capital. Recurrent costs will be reduced because of the reduced maintenance requirements and capital savings would be possible because of the reduced requirements for standby plant to cover for the occasions when unplanned outages occur.

Sea locked, small-scale electricity supply systems are generally diesel powered. This leads to unit costs for electricity which are significantly higher than those for a large grid system, mainly because of higher fuel costs. It follows that the incentive to reduce operating costs is high and the potential savings considerable if fuel costs can be reduced.

Machine health monitoring

Several studies have demonstrated that 'secondary effects' such as sound, vibration, temperature, pressure, and other physical phenomena exhibit detectable changes long before catastrophic mechanical failures occur and, therefore, can be used to predict and diagnose mechanical malfunctions. Sensors can usually be positioned to detect the secondary

effects of one component or mechanical subsystem, independent of similar emissions from other components or subsystems. Frequency- and time-domain signal processing is used to obtain 'signatures' representing normal and faulty performance.

It should be noted that identification of a fault condition from such measurements is often difficult and, in the case of complex systems involving many measurements and a large body of background knowledge (e.g. initial design information), it will be found that decision support software will be required to assist the engineer [3]. Operator confidence will diminish rapidly if the health monitoring system continually indicates a fault condition for no good reason, so time must be spent in developing reliable decision making software.

5 Present monitoring of diesel generating sets

Electronic instrumentation for machinery health monitoring based on both vibration data and on transfer function models has been investigated by the authors.

5.1 Vibration monitoring in Kirkwall power station
Initial experience with vibration monitoring equipment and hardware at Kirkwall power station on Orkney, Fig. 1, had shown that this equipment tended to be expensive if
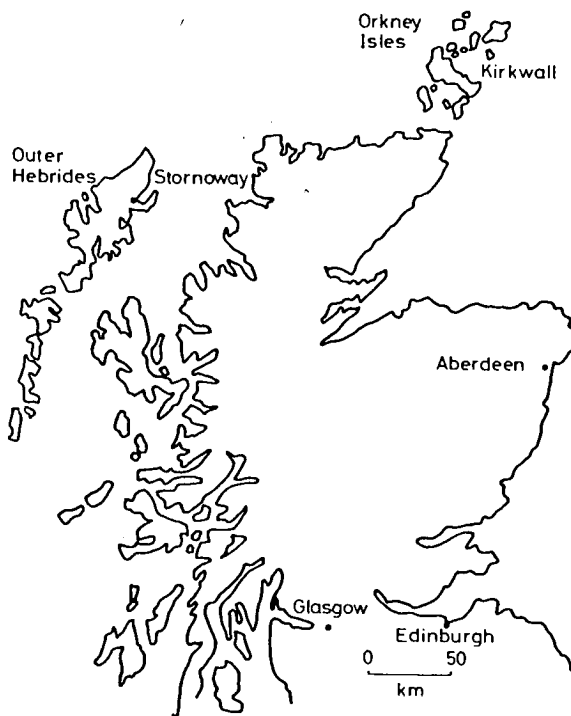
Fig. 1 Map of North of Scotland Hydroelectric Board area

a comprehensive coverage of sensitive monitoring points is attempted but could be designed to give information about clearly identified points in the machine, i.e. bearings and structural support.

The study was directed towards the development of monitoring equipment of reduced cost and complexity. In the knowledge that signature based fault detection requires only a change to be detected from a known good (i.e. healthy) condition, when a vibration spectrum is used, it was realised that it is not necessary for the spectrum to be mathematically accurate. It has been found that an approximate spectrum can be easily implemented if the conventional sine and cosine operations in the discrete Fourier transform (DFT) are replaced by a square wave-

form. Using this technique a 0–10 kHz spectrum analyser with 10, 100 Hz ZOOM bands has been implemented by a single-board computer with a single-board interface circuit. This work is reported fully elsewhere [4].

5.2 Transfer-function based monitoring at Stornoway power station

This paper concentrates on a description of a method of machine monitoring based on a knowledge of the transfer function, for, although the measured small signal parameters of a system transfer function cannot usually be easily related to identifiable parts of the system, such measurements can be used to give a global indication of the state of a system.

The identification method for this work is based on the modulating function method introduced by Loeb [5]. This work can be related back to a report by Shinbrot published in 1954 [6]. Hermite modulating functions proposed by Takaya have been used. Others [8, 9] have related the modulating function method to other linear identification methods. It should be noted that the modulating function method directly measures the parameters of a system small-signal differential equation.

The modulating function method was originally developed to operate on data directly derived from measurements made on the system under test. The method has been modified to use preprocessed data in the form of auto- and cross-correlation functions. The correlation functions provide a useful intermediate diagnostic check on the data. A number of efficient methods for microcomputer evaluation of correlation integrals have been produced. In this work, use is being made of the Fell skip algorithm [5] to implement the relay correlation function.

Normally, naturally occurring system noise does not have a sufficient bandwidth to excite all the modes in a system and psuedorandom noise is injected to overcome this problem. The modulating function method is then applied to measured correlation functions and if n parameters are to be measured, n equations must be solved simultaneously to estimate the transfer function parameters. Extensive simulation studies have shown this to be a reliable method and, as such, should be suitable for remote operations.

Modulating function method

The output and input of a signal-input linear system are related by the convolution integral. If x is the input and y the output then autocorrelation r_{xx} and cross-correlation r_{yx} may be defined and the convolution integral becomes

$$r_{yx}(a) = \int_{-\infty}^{\infty} h(a - t) r_{xx}(t) \, dt \qquad (1)$$

or $R_{yx}(s) = H(s) R_{xx}(s)$ (2)

where $R_{yx}(s)$, $R_{xx}(s)$ and $H(s)$ are the Laplace transforms of the cross-correlation function, the autocorrelation function and the system impulse response, respectively.

It follows that the system differential equation may be reformulated in terms of correlation functions to give

$$q_n \frac{d^n r_{yx}(a)}{da^n} + q_{n-1} \frac{d^{n-1} r_{yx}(a)}{da^{n-1}}$$
$$+ \cdots + q_0 r_{yx}(a) = r_{xx}(a) \qquad (3)$$

This differential equation may be converted to an algebraic equation by multiplying by a suitably chosen modulating function ϕ and integrating by parts over the interval 0 to

T. If ϕ and all of its derivatives are zero outside of the interval 0 to T, it can be shown that

$$(-1)^n q_n \int_0^T \phi^n(a) r_{yx}(a) \, da$$
$$+ \cdots + (-1)^{n-1} q_{n-1} \int_0^T \phi^{n-1}(a) r_{yx}(a) \, da$$
$$+ \cdots + q_0 \int_0^T \phi(a) r_{yx}(a) \, da$$
$$= \int_0^T \phi(a) r_{xx}(a) \, da = \frac{d^n \phi(a)}{da^n} \qquad (4)$$

By choosing different sets of modulating functions, a set of linear equations is obtained which can be represented by a matrix equation. The system parameters $q_n \ldots q_0$ are obtained by matrix inversion. In general, when phase advance terms are included, i.e. when the transfer function is of the form

$$H(s) = \frac{p_k s^k + \cdots + 1}{q_n s^n + \cdots + q_0} \qquad (5)$$

the matrix will be increased to include the parameters p_k to p_1. A correspondingly larger number of modulating functions will be required in this case.

Takaya [7] introduced the use of Hermite functions as modulating functions. The properties of Hermite functions guarantee that the modulating functions derived from them are continuous and differentiable. Takaya shows that the Hermite modulating functions and their derivatives approximately become zero at both upper and lower boundaries.

As the cross-correlation function is not required to be a good approximation to the system impulse response, it is not necessary to design the input autocorrelation function to have a very small width compared with the decay time of the cross-correlation function. The experimental system used Gaussian filtered, pseudorandom noise with a bandwidth approximately double the resonant frequency of the system under test. The input noise level was always adjusted to ensure that the developed power was never perturbed by more than $\pm 3\%$ of its steady-state value. The power signal typically has a superimposed very-low-frequency drift term. To make best use of the dynamic range of the data logger and to avoid overload conditions, this drift term must be eliminated by using an AC coupling network. The power signal is AC coupled into the data logger with an identical network. In this case, the measured cross- and autocorrelation functions are defined by

$$R_{vw}(s) = H(s) R_{ww}(s) \qquad (7)$$

because

$$V(s) = H(s) H_c(s) X(s) \qquad (8)$$

and

$$X(s) = \frac{1}{H_c(s)} W(s) \qquad (9)$$

where $H(s)$ is the transfer function of the system under test and $H_c(s)$ is the transfer function of the coupling networks. The signals x, y, v and w are defined in Figure 2. Algebraically, the effect of the coupling networks has been completely cancelled but in practice, care must be taken to use a coupling time constant which is large enough to ensure that sufficient of the low frequency content of the output power signal has been coupled into the data logger.

A simple RC circuit with transfer function

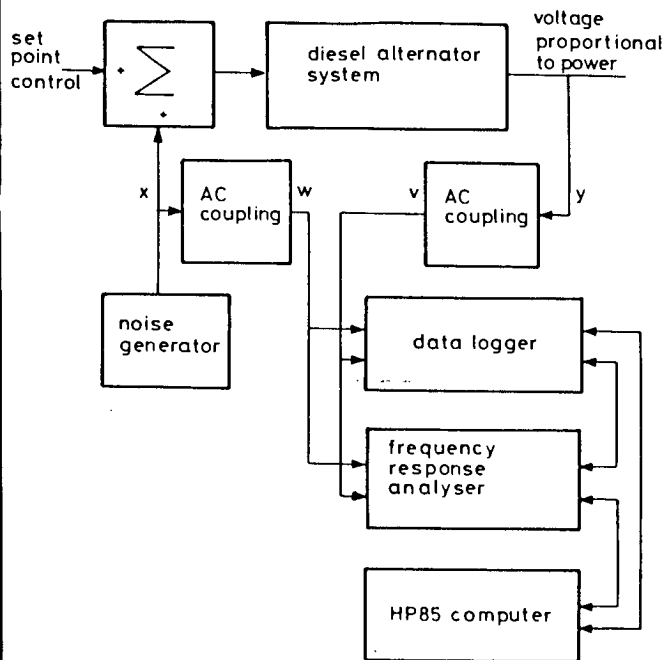$$H_c(s) = \frac{sRC}{1 + sRC} \qquad (10)$$

Fig. 2   Block diagram of experimental system

was used as the coupling network in the experimental system.

7   Experimental system

The measurement system, as currently developed, cannot automatically take account of system time-delay terms. When a time-delay term is observed, it is necessary to interrupt the software and use keyboard control to offset the cross-correlation data set until the delay has been eliminated.

The engine used for these tests was a 4.6 MW Mirrlees Blackstone KV12 Major, used for base-load generation at Stornoway power station, Fig. 1. The engine was controlled by an electronic governor enabling Gaussian noise test signal to be injected. The power response of the system to this test signal was monitored by using current and voltage transducers on one phase of the alternator output. The noise signal and the power response were AC coupled to a data logger with a 0.1 Hz coupling time constant. Data was then downloaded via an IEEE BUS to a desk top computer, where auto- and cross-correlations were generated using a relay correlation technique. Hermite functions were used to implement the modified modulating function identification method. In addition to this instrumentation, a spectrum analyser was used to produce frequency response and coherence functions. A block diagram of the experimental system is shown in Fig. 2.

The relay correlation implementation was based on the skip algorithm [10]. This involved batch processing of the data where two blocks of 2048 bytes were down loaded from the logger to the computer. The sample rate was set to define the required correlation-function time-delay intervals. Only every eighth cross product was summed to form the correlation estimate. A negligible variance increase was observed compared with the results obtained by using the whole data set. Fig. 3 shows the results of an

experiment demonstrating the benefits of skipping cross products.
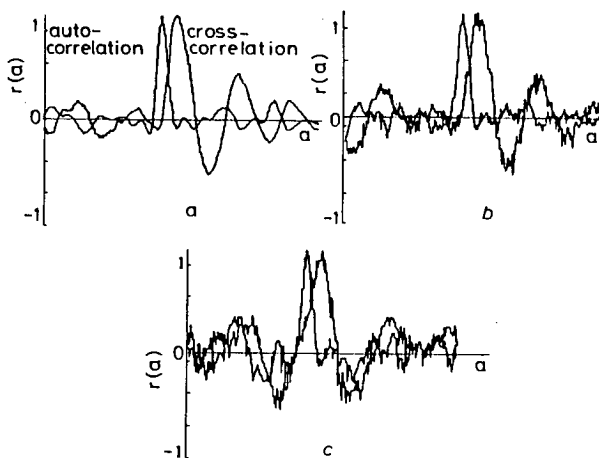
Fig. 3   Typical measured cross-correlations after removal of system time delay demonstrating effect of skip algorithm

a Skip = 1
b Skip = 8
c Skip = 16
Peak values are normalised

8   Results

The engine was run at a constant load of 3.5 MW and the perturbing noise signal level adjusted to restrict the resulting changes in the power level to never exceed ±3% of the steady-state value. The correlation function esti-

Fig. 4   Measured frequency response and coherence function

a Magnitude response
Steady-state power output = 3.5 MW
Test noise level = 0.6 V RMS
b Phase response
c Coherence

mates were exponentially averaged and the modulating function method used to obtain the model parameters.

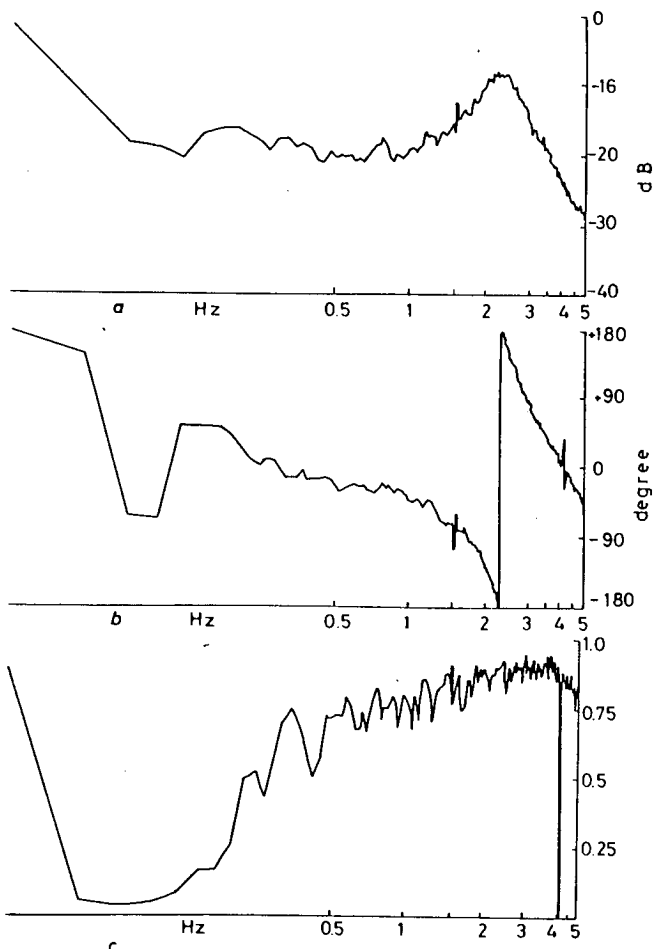The measured cross-correlation function indicated a response which included a pure time delay term. In the first instance, a simple, resonant, lowpass, 2nd-order function was used to model the remainder of the response. The time-delay term is removed before the modulating function method is used. A typical cross-correlation function after removal of this time delay is shown in Fig. 4. The model transfer function is

$$H(s) = \frac{1}{q_2 s^2 + q_1 s + q_0} \qquad (11)$$

where q_2, q_1 and q_0 are the parameters to be measured.

Frequency response and coherence functions obtained from a commercial spectrum analyser monitoring the outputs from the AC coupling networks are shown in Fig. 4. The coherence falls rapidly as the frequency decreases below 0.5 Hz. This reduction is caused by the AC coupling network.

The following average transfer-function parameters were obtained with the same power and noise levels used to obtain Fig. 4 (i.e. 3.5 MW and 0.6 V RMS).

$q_2 \ldots 0.07$

$q_1 \ldots 0.19$

$q_0 \ldots 11.05$

Frequency response functions (including the effect of the measured system time delay) were calculated by substituting $s = j\omega$ into the transfer functions obtained by using the measured parameters. Fig. 5 shows the calculated magnitude and phase response. Good agreement is obtained
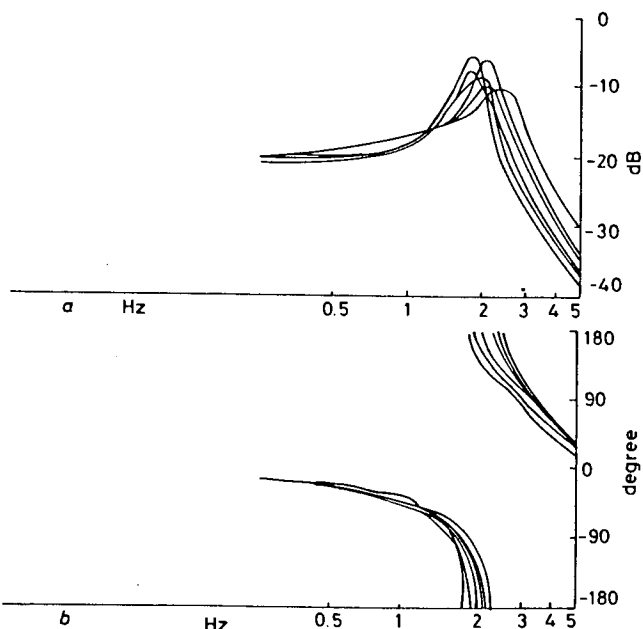

Fig. 5 *Measured frequency response derived from model using measured parameters*

a Magnitude
b Phase

between the measured response (Fig. 4) and the calculated response.

9 Conclusions

These studies have demonstrated that the modified modulating function method can be used to estimate the small-

signal parameters of a practical generating system. As it has been found that the parameters of the small-signal model change as a fault condition develops, it follows that the method may be used for machine health monitoring. Unfortunately, the small signal parameters will also, in general, change as the system static operating point changes and as a result the small signal parameter space must be combined with the static operating point space before a fault decision is made.

Although it can be difficult to relate small signal parameter changes to a specific fault condition, they do offer the considerable convenience of making measurements while the system is operating normally. The use of continuous modelling techniques (rather than sampled data models) has made it easier to relate to the actual structure of the practical problem but more work is necessary to establish the practical value of a continuous model in fault-detection algorithms.

However, these initial studies have resulted in a statistically reliable procedure involving a two-stage identification procedure: data are first analysed by correlation techniques and then a small-signal transfer function is developed from the correlation function. The correlation function provides a very convenient diagnostic check on the data collected and it can be assumed that the transfer-function analyser always operates on reliable data.

It is important to emphasise that although assessment of the different techniques used was possible by simulation in the laboratory, it was essential to validate the equipment and software in the field. The access to diesel sets at both Kirkwall and at Stornoway power stations proved to be an ideal test-bed for this study. Indeed, these tests have confirmed that the method is easy to use and suitable for operation at remote sites.

Furthermore, if, as has happened in Orkney, the islands become interconnected by submarine cable to the mainland grid, the role played by diesel stations will change to standby; here it becomes even more important to have confidence in the condition of the plant so that it is available when required, often in an emergency.

At present it is not possible to quantify the economic returns from performance monitoring, but they should be manifest as lower operational costs because of reduced fuel costs, lower maintenance costs because of the better knowledge of the specific maintenance requirements of plant, and lower capital costs because of the reduction in the level of installed plant to cover for unplanned outages.

However, with all health monitoring schemes, it is essential that the correct interpretation of information from such tests is made, and the authors envisage a decision support software package to be a natural complement to the transfer-function based monitor.

11 References

1 . . . .On: H.W.: 'Diesel generation in power system development'. Proceedings of 15th University Power Engineering Conference, Leicester, UK, 1980, Paper 7A.4
2 GAWKES, F.K.: 'Economics of reliability of supply in tropics'. *IEE Conf. Pub. 34*, 1967, pp. 53–57

3 STEWART, R.M.: 'A systematic approach towards automating the management of machinery complexes' in BUTCHER, D.W. (Ed.): 'Online monitoring of continuous process plants' (Ellis Horwood, Chichester, UK, 1983) pp. 293–312

4 MOULTON, G.A.: Ph.D. Thesis, University of Edinburgh, 1986, to be published

5 LOEB, J.M., and CAHEN, G.M.: 'More about process identification', IEEE Trans., 1965, AC-10, pp. 359–361

6 EISENFELD, J.: 'Remarks on the modulating functions method for impulse response identification', ibid., 1979, AC-24, (3), pp. 498–499

7 TAKAYA, K.: 'The use of Hermite functions for system identification', ibid., 1968, AC-13, pp. 446–447

8 MIRONVOUSKI, L.A., and YODOVICH, V.S.: 'Unified approach to identification of linear stationary processes', ibid., 1978, AC-13, pp. 446–447

9 BELORUSETS, V.B.: 'Method of auxiliary systems for the identification of dynamic systems with unknown input' (Plenum Publishing Corporation, 1982, translated from Automatika Telemek Hanika, 1981, (8), pp. 76–82)

10 FELL, R.: 'Microprocessor-based cross-correlation using the skip algorithm'. Proceedings of Conference on the Influence of microelectronics on measurement and transducer design, Manchester, UK, 1982, pp. 25–32

MACHINE HEALTH MONITORING OF ISLAND-BASED   DIESEL
ELECTRICITY GENERATION

H.W. WHITTINGTON, J.R. JORDAN and N.E. PATERSON

MACHINE HEALTH MONITORING OF ISLAND-BASED
DIESEL ELECTRICITY GENERATION.
HW Whittington , JR Jordan and N Paterson.
Department of Electrical Engineering,
University of Edinburgh.
PM Johnson, North of Scotland Hydroelectric Board.

1. INTRODUCTION

This paper describes work carried out to assess the condition of diesel generating plant. The eventual aim of the study is to develop a monitoring system which will assist maintenance staff to operate more effective repair and maintenance policies, based on plant condition or health rather than or a routine basis.

2. BACKGROUND.

Although the bulk supply of electricity is well established ,the part to be played by small generating plant in electrification is still significant. For example, in developing countries comparatively small plants will often suffice to serve the early needs of the larger towns, even of the capital city. Later, the growing demand will justify the installation of larger, more sophisticated and more efficient plant; perhaps even the interconnection of two or more townships to form the rudiments of a grid or power pool. When this happens it will often be possible to take up the earlier plant from their foundations, and to remove them to smaller or more remote centres of population to perform useful pioneer work, further into the heart of the country.

Even in highly developed countries small power plants serve an important function in remote places, either where there is low population density or where consumers are land- or sea-locked. These conditions are not conducive to the early establishment of grid systems and townships are, therefore, likely to remain electrically isolated from one another for a long time.

Of the options available for prime mover in such circumstances, the diesel engine is one which is widely used. With so many advantages to its credit, it is not surprising that, in the past, the diesel engine has found application in small power stations with total installed capacities of approximately 30 MW.

Larger stations are also becoming more common. This has certainly created a frontier of competition between the diesel engine and the steam turbine, since previously it had been assumed that above about 30MW installed capacity, the steam turbine becomes economically attractive, and had been used in preference to the diesel engine.

Because of their compactness, diesel stations can be sited at or as near to load centres as permitted by noise regulations. Although diesel engines tend to be rather noisy, tests and experience have shown that even without precautions to cut down noise, it is satisfactory to site large diesel plants within 800 m of domestic dwellings (Fawkes, 1967). Since the troublesome noise is in general in the higher frequency range, sound barriers are found completely effective in reducing disturbances to an acceptable level.

3. TECHNICAL CONSIDERATIONS OF DIESEL GENERATORS

3.1 General Features

Diesel engines are extremely versatile prime movers for the generation of electricity. They are obtainable from a large number of manufactures based in several countries and the market is therefore competitive.

The diesel electric generating set is available in almost any size over a wide range. Earlier size limitation was imposed by the need to use very low speeds. This meant very bulky and heavy machines but modern designs have enabled higher speeds to be utilised, with the result that more powerful plant can be built without their physical size being unmanageable. Very large, low-speed, diesel engines have long been used for marine propulsion, but today are finding an ever increasing role in land-based operation.

3.2 Security of supply

In isolated stations it is necessary to provide some margin of spare plant so that supply may be maintained when part of the plant is out of service for maintenance or because of breakdown. For proper security it could be argued that in any station two spare sets should be installed, so that when one set is out of service for maintenance purposes there is still another spare set standing by in case of the breakdown of a running set. Although this level of spare capacity may be justified on occasion, the provision of two spare sets is often an expensive luxury. For small, newly established electricity supply undertakings, it is usually of greater importance to produce relatively cheap electricity with the risk of very occasional interruptions, than to have almost perfect continuity of supply. Indeed, even with established systems, the Supply Authority will accept that there is a finite probability of failing to meet demand at any given time, even working with spare capacity . This risk is of course dependent not only on the reliability of the plant, which in turn depends to a great extent upon the care with which it is maintained, but also on the proportion of time (in days per year) when the plant is out of service for routine maintenance.

Where a large number of plant units is installed, the risk of a "double outage" i.e. one set out for maintenance and another set subject to breakdown, will be greater. No definite ruling can be given as to the amount of spare plant which should be provided; much will depend on the size and nature of the load. The following margins are therefore suggested tentatively though it is recognised that many individual variations will be necessary:

(a) assuming the largest installed unit to be out of service for routine maintenance, and where the total number of sets is five or less, assume there is no coincident breakdown, ie one spare set.

(b) where the total number of installed sets is six or more, assume the largest unit to be out of service for routine maintenance and any other set to be out owing to breakdown, ie two spare sets.

It is of interest to note that, in the United States, the term "firm capacity" is used; this is the capacity of the station with the largest set out. The average time any machine is out owing to a breakdown is usually about 5% of its total operational time in any year. This, combined with a typical planned outage time of 9.7% per annum (NSHEB, Private communication), would result in a 0.035% chance of a double outage occurring during an average operating year of 4605 hours.

3.3 Life expectancy of plant.

When planning diesel-powered electricity generation it is usual to allow a life of fifteen years for diesel plant. For very high speed diesels, or engines using fuel of low quality, a shorter period may sometimes be more realistic (ten or twelve years). For slow speed diesels, or for diesel plant that is operated intermittently

without overloading, rather longer periods (up to twenty years) may be taken as according with practice. However, when planning ahead, a fifteen year period of utilization in a given situation would appear to be prudent since diesels tend to become superceded through growth of demand rather than through wear and tear. However, a second lease of life is often found for old plant at other sites.

4. OPERATION OF PLANT.

If well maintained, diesel engines can run for a long period between overhauls and have low wear rates. The average running hours between overhauls for most NSHEB machines is greater than 10,000 hours which represents about 18 months continuous operation (NSHEB, private communication): the NSHEB is the world's largest user of land-based diesel engines, and as such has much experience in their operation and maintenance.

Where skilled labour is scarce, station operation can often be adapted for automatic operation; that is to say, a stand-by diesel set may be started up by remote control or by mains failure. Full automation may be provided for starting, synchronising and shutting down diesels in response with local demand.

Maintenance, may be simplified by provision of easily replaceable assemblies of parts, thus enabling reconditioning to be undertaken away from the generating plant. Periods between order and delivery are short and a diesel power station may therefore be extended rapidly by adding suitably sized units to meet growing demands. Compact, light-weight, high-speed diesels are obtainable for sites which are remote, cramped or difficult of access. Mobile diesel power units mounted on skids or trailers may be used for temporary or emergency purposes.

The NSHEB has several such power units which are kept at strategically central positions in readiness for quick transportation. Transportable diesel sets are also used for supplying construction power for large civil engineering works and for supplementing electricity supply systems which are temporarily short of power.

Despite the provision of installed redundancy and portable plant, unplanned outages do occur and can be expensive both in terms of lost revenue or increased fuel costs by meeting demand with less efficient plant. Additionally, the inconvenience to the consummers of loss of supply must be considered as of prime importance.

It follows that a system which can indicate the onset of fault conditions and thus allow time for action to be taken will be attractive to the supply authority. Indeed, if such a system could be arranged to monitor plant continuously and to give data on its health, maintenance could be scheduled on the basis of equipment performance rather than merely at regular intervals. The advantage would be that servicing would be carried out only on plant which required it and potentially considerable financial savings will result, both recurrent and capital. Recurrent costs will be reduced because of the reduced maintenance requirements and capital savings would be possible because of the reduced requirements for standby plant to cover for the occasions when unplanned outages occur. This approach to plant operation is usually termed machine health monitoring.

5. MACHINE HEALTH MONITORING.

Machine health monitoring and its associated instrumentation may be considered to involve designers, construction engineers, commissioning engineers and most importantly operational staff. It is in this ultimate stage where performance monitoring systems aids the operator by continually assessing the health of the equipment and plant

during its planned life. However, when considering the usefulness of such a system, it is in the early phases of system design that a great deal of attention must be given to the type of plant to be installed, and the expected variation of parameters during healthy operation together with the tolerance on these parameters that must be allowed prior to a probable fault condition being reported or acted upon. In much the same way data obtained at the construction and plant installation stage has an important bearing on the initial settings that are used. These two pre-operational stages are often neglected. Future construction programmes however offer the possibility of implementing effective systems.

Several studies have demonstrated that "secondary effects" such as sound, vibration, temperature, pressure, and other physical phenomena exhibit detectable changes long before catastrophic mechanical failures occur and, therefore, can be used to predict and diagnose mechanical malfunctions. Sensors can usually be positioned to detect the secondary effects of one component or mechanical subsystem, independent of similar emissions from other components or subsystems. Frequency- and time- domain signal processing is used to obtain "signatures" representing normal and faulty performance.

Manufacturers of monitoring systems claim savings of up to 40% on planned maintenance and up to 60% on breakdown repair. This has to be seen in the light of a maintenance expenditure which may be a significant proportion of the historic capital investment of the plant per annum, and a typical installation cost of a comprehensive monitoring system of approximately 15% of the capital investment in plant (Smith, JR , private communication).

Sea-locked, small-scale electricity supply systems are generally diesel-powered. This leads to unit costs for electricity which are significantly higher than those for a large grid system, mainly because of higher fuel costs. It follows that the incentive to reduce operating costs is high and the potential savings considerable if fuel costs can be reduced. Field trials of prototype monitoring instrumentation have been carried out at both Kirkwall Power Station on Orkney and at Stornoway Power Station on the Outer Hebrides, Figure 1.

6. MACHINE HEALTH MONITORING EXPERIENCE.

Electronic instrumentation for machinery health monitoring based on vibration data and on transfer function models has been investigated by the authors. Initial experience with vibration monitoring equipment and hardware at Orkney Power Station had shown that this equipment tended to be expensive if a comprehensive coverage of sensitive monitoring points is attempted but can be designed to give information about clearly identified points in the machine i.e. bearings and structural support.

In contrast, the measured small signal parameters of a system transfer function cannot usually be easily related to identifiable parts of the system and therefore such measurements can only be used to give a global indication of the state of a system.

In both cases "signatures" of normal performance are obtained and deviations from normal performance are monitored. Identification of a fault condition from these changes is often difficult and in the case of complex systems involving many measurements and a large body of background knowledge (e.g. initial design in formation), it will be found that "Expert System" software will be required to assist the decision making process [Stewart, 1992]. Operator confidence will diminish rapidly if the health monitoring system continually indicates a fault condition for no good

reason, so time must be spent in developing reliable decision-making software.

Signature-based fault detection only requires a change to be detected from a known good (i.e. healthy) condition. Subsequently when a vibration spectrum is used it is not necessary for the spectrum to be mathematically accurate. It has been found that an approximate spectrum can be easily implemented if the conventional sine and cosine operations in the DFT are replaced by a square waveform. Using this technique a 0 to 10 kHz spectrum analyser with 10, 100 Hz ZOOM bands has been implemented by a single board computer (using the 6809 processor) with a single board interface circuit [Moulton, 1985]. This hardware requirement will be reduced to a single card system when single chip microcomputer techniques are used.

Normally naturally occurring system noise does not have sufficient bandwidth to excite all the modes in a system and psuedo random noise is injected to overcome this problem. The modulating function method is then applied to measured correlation functions and if n parameters are to be measured then n equations must be solved simultaneously to estimate the transfer function parameters. Extensive simulation studies have shown this to be a reliable method that is suitable for remote operations. Full details of this technique are given elsewhere by Jordan et al (1985).

7. STORNOWAY EXPERIMENTAL SYSTEM.

A practical investigation of the transfer function monitoring method has been carried out on the main speed control loop of one of the diesel-generator sets installed in Stornoway Power Station. A 4.6 MW Mirlees Blackstone KV12 Major diesel engine was used and to allow injection of pseudo random noise signals part of the existing hydraulic governor was replaced by a commercial electronic unit.

The power response of the system to this test signal was monitored by using current and voltage transducers on one phase of the alternator output. The noise signal and the power response were A.C. coupled to a data logger with a 0.1 Hz coupling time constant. Data was then down loaded via an IEEE BUS to a microcomputer, where auto and cross-correlations were generated using a relay correlation technique.

The relay correlation implementation was based on the Skip Algorithm described by Fell(1982). This involved a batch processing of the data where two blocks of 2048 bytes were down loaded from the logger to the computer. The sample rate was set to define the correlation function time delay intervals. Only every eigth cross-product was summed to form the correlation estimate. A negligible variance increase was observed compared with the results obtained by using the whole data set.

8. RESULTS

The engine was run at a constant load of 3.5 MW and the perturbing noise signal level adjusted to restrict the resulting changes in the power level so as never to exceed ±3% of the steady state value. The correlation function estimates were exponentially averaged and the modulating function method used to obtain the model parameters. Typical measured auto and cross correlation functions are shown in Figure 2.

Figure 3 shows the system frequency response obtained by substituting s = jω in the transfer function derived from the measured small signal parameters. Good agreement has been obtained with frequency response measurement made directly with a Spectrum Analyser.

9. CONCLUSIONS

To be effective, machinery health monitoring equipment must have sufficient data processing and decision making complexity to ensure that spurious fault indications will be minimised. It must also be manufactured at a low enough cost to ensure that a financial return will be possible from the capital required to install the equipment.

Progress is being made with the development of lower cost machinery health monitoring equipment. Initial tests on realistic scale system (eg the diesel powered generating sets) have shown the approximate spectrum analyser and the transfer function analyser to be viable, low cost, machinery health monitoring units.

Although assessment of the different techniques used was possible by simulation in the laboratory, it is esential to validate the equipment and software in the field. The access to diesel sets at both Kirkwall and at Stornoway Power Stations proved to be a ideal test-bed for this study.

At present, indications are that the technique described could produce significant financial savings for any Authority responsible for operating of such generating equipment and that eventually operations and maintenance scheduling may be based around such a concept.

However, it is essential that the correct interpretation of information from this unit is made, and the authors envisage the type of decision support software described above to be a natural complement to the transfer-function based monitor.

Furthermore, as the islands become interconnected by submarine cable to the mainland grid, the role played by diesel stations will change to standby; here it becomes even more important to have confidence in the condition of the plant so that it is available when required, often in an emergency.

10. ACKNOWLEDGEMENTS.

11. REFERENCES.

Fawkes, FK (1967) 'Economics of Reliability of Supply in Tropics' IEE (London) Conf Pub. 34, pp53-57.

Fell,R. (1982) "Microprocessor-based cross-correlation using the skip algorithm" Proc Conf on the influence of Microelectronics on Measurement and transducer Design, pp25-32. Manchester..

Jordan,JR, Paterson,N and Whittington,HW (1985) "Transfer Function Measurement with a Modified Modulating Function Method" Accepted for publication, Proc IFAT, University of York,

Moulton,GA. (1985) PhD. Thesis, University of Edinburgh. To be published.

Stewart,RM (1983), A systematic approach towards automating the management of machinery complexes. On-line monitoring of continuous process plants. ed DW Butcher (Chichester: Ellis Horwood) pp 293-312. systems' Proc Symp Sci and Technol. of Condition Monitoring, Southampton Univ, pp 1.3-1.37 .

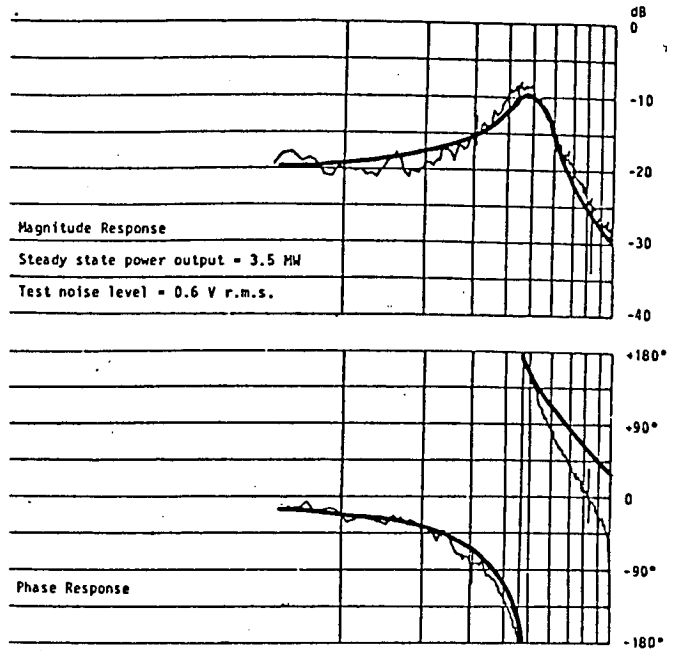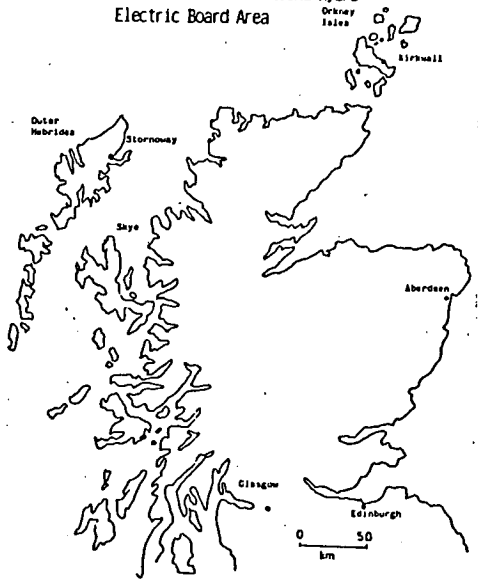Figure 1. Map of North of Scotland Hydro Electric Board Area

Magnitude Response

Steady state power output = 3.5 MW

Test noise level = 0.6 V r.m.s.

Phase Response

Figure 3 - Frequency response derived from second order parameters.
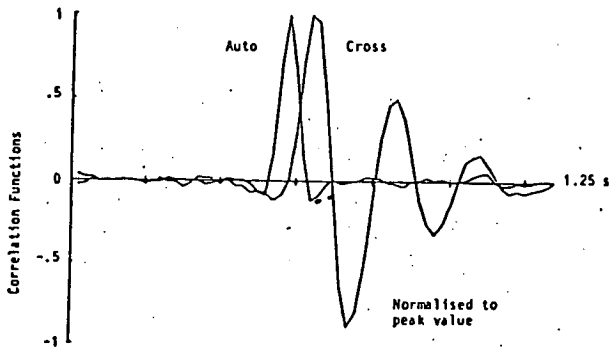
━━ Model results

∿∿ Measured frequency response

Figure 2 - Typical Measured Auto and Cross-Correlation Function