

AGARD-AG-14-VOL-14

ADA111412

AGARD-AG-14-VOL-14

AGARDograph No. 160

**AGARD Flight Test Instrumentation Series
Volume 14**

on

The Analysis of Random Data

by

D. A. Williams

This document has been approved
for public release and sale; its
distribution is unlimited.

DTIC
ELECTE
FEB 26 1982

S

A

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

02 26 096

(14) AGARD-AG-160
Volume 14

NORTH ATLANTIC TREATY ORGANIZATION
ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT
(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.160 Vol.14
THE ANALYSIS OF RANDOM DATA

by

D.A.Williams

Volume 14

of the

AGARD FLIGHT TEST INSTRUMENTATION SERIES

Edited by

A.Pool and K.C.Sanderson



Accession For	
DTIC GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD.

400073 all

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published November 1981

Copyright © AGARD 1981
All Rights Reserved

ISBN 92-835-1405-X



*Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St, London, W1P 1HD*

PREFACE

Soon after its founding in 1952, the Advisory Group for Aerospace Research and Development recognized the need for a comprehensive publication on flight test techniques and the associated instrumentation. Under the direction of the AGARD Flight Test Panel (now the Flight Mechanics Panel), a Flight Test Manual was published in the years 1954 to 1956. The Manual was divided into four volumes: I. Performance, II. Stability and Control, III. Instrumentation Catalog, and IV. Instrumentation Systems.

Since then flight test instrumentation has developed rapidly in a broad field of sophisticated techniques. In view of this development the Flight Test Instrumentation Group of the Flight Mechanics Panel was asked in 1968 to update Volumes III and IV of the Flight Test Manual. Upon the advice of the Group, the Panel decided that Volume III would not be continued and that Volume IV would be replaced by a series of separately published monographs on selected subjects of flight test instrumentation: The AGARD Flight Test Instrumentation Series. The first volume of the Series gives a general introduction to the basic principles of flight test instrumentation engineering and is composed from contributions by several specialized authors. Each of the other volumes provides a more detailed treatise by a specialist on a selected instrumentation subject. Mr W.D.Mace and Mr A.Pool were willing to accept the responsibility of editing the Series, and Prof.D.Bosman assisted them in editing the introductory volume. In 1975 Mr K.C.Sanderson succeeded Mr Mace as an editor. AGARD was fortunate in finding competent editors and authors willing to contribute their knowledge and to spend considerable time in the preparation of this Series.

It is hoped that this Series will satisfy the existing need for specialized documentation in the field of flight test instrumentation and as such may promote a better understanding between the flight test engineer and the instrumentation and data processing specialists. Such understanding is essential for the efficient design and execution of flight test programs.

The efforts of the Flight Test Instrumentation Group members (J.Moreau CEV/FR, H.Bothe DFVLR/GE, J.T.M. van Doorn and A.Pool NLR/NE, E.J.Norris A&AEE/UK, K.C.Sanderson NASA/US) and the assistance of the Flight Mechanics Panel in the preparation of this Series are greatly appreciated.

F.N.STOLIKER
Member, Flight Mechanics Panel
Chairman, Flight Test
Instrumentation Group

SUMMARY

At first sight, the presence of a Volume concerned with the analysis of random data in a series devoted to flight test instrumentation may appear to be extraneous. However, the availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment (the former is normally carried aboard the test vehicle, whilst the latter may be incorporated in the ground equipment). Further, the techniques described in this Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. It therefore follows that the "software" tools which are represented by such techniques are just as important to the modern flight test engineer as the hardware used to detect and collect the necessary measurements.

It is not intended that this Volume should be a reference document for the specialist analyst who is required to generate the software for analysing random data; a number of excellent publications exist which satisfy this requirement admirably (for example (reference 0.1, 0.2 and 0.3)). Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. For this reason the author has attempted to present mainly heuristic explanations of the techniques in the main body of the text, and has confined the more detailed mathematical treatments to appendices. A second, and equally important, objective of the Volume is to emphasise the strong interaction between the various elements and facts of a measurement programme.

In essence, the contents of this volume are intended as a "guide" to the analysis techniques which are available to the flight test engineer for analysing random data, and to the requirements and constraints which may be imposed upon the selection of hardware and the design of a trial if meaningful information is to be gained from the measurements gathered during a trial.

ACKNOWLEDGEMENT

A large number of people contributed to the preparation of this Volume. The author is particularly indebted to Professor E.J.Durbin (Princeton University) who contributed much of Section 2.4.4. and Appendices D and E; to Messrs. J.Brownlow (NASA Dryden), T.R.Twisdale and R.Lenz (AFFTC), K.D.Puch and P.Skudridakis (DFVLR), J.Buhrmann (NLR), J.R.Schiess (NASA Langley) and Dr R.Bach (NASA Ames) for their constructive advice and criticism; to the members of the AGARD Flight Test Instrumentation Group for their encouragement and patience, especially to the late Mr Owen Matthews (CIT); to Miss Carol Potheary who so ably prepared and corrected the manuscript.

CONTENTS

	Page
PREFACE	iii
SUMMARY AND ACKNOWLEDGEMENT	iv
LIST OF PRINCIPAL SYMBOLS	vi
Chapter 1. PRE-PROCESSING	
1.1 Introduction	1
1.2 Recording and Recovery of Random Signals	2
1.3 Analogue Pre-processing	8
1.4 Sampling	13
1.5 Digital Pre-processing	29
Chapter 2. PROCESSING TECHNIQUES	
2.1 Introduction	44
2.2 Fundamental Concepts	44
2.3 Time Domain Analysis	47
2.4 Frequency Domain Analysis	59
Chapter 3. PRACTICAL ASPECTS	
3.1 Measurement System Calibrations	79
3.2 Data Qualification	80
3.3 Data Handling	81
3.4 Data Presentation	82
Chapter 4. EXAMPLES OF RANDOM DATA ANALYSIS	
4.1 Examples of Spectral Analysis Computations	84
4.2 Measurement of Atmospheric Turbulence	98
4.3 Analysis of Records from Flight Flutter Trials	101
4.4 Applications in Evaluating Aircraft Handling Qualities	106
4.5 The Reduction of Random Measurements to a form suitable for Ground Certification Trials	108
5. REFERENCES	113
APPENDICES	115

LIST OF PRINCIPAL SYMBOLS

B	Frequency bandwidth of a (possibly filtered) data sequence
$D(y_0)$	The probability of a function y having a value less than y_0
$E[y]$	The expected value of a function y
f	Frequency (Hz)
f_0	Natural frequency (of a filter)
f_c	Folding, or Nyquist, frequency
δf	Interval between successive frequency estimates
$\hat{f}(t)$	An input function of time
$F_x(if), X(if)$	The Fourier Transform of a function $x(t)$
$F_y(if), Y(if)$	The Fourier Transform of a function $y(t)$
$G_{yy}(f)$	The Power Spectral Density function of y
$G_{xy}(if)$	The Cross Spectral Density function of x with respect to y
h	The time interval between successive samples
$H(if)$	The Transfer (or Frequency Response) function
i	The square root of -1
$K_{yy}(t)$	Auto-covariance function of y
$K_{xy}(t)$	Cross-covariance function of x with respect to y
M_n	The n^{th} moment of a function (Appendix D)
CM_n	The n^{th} central moment of a function (Appendix D)
N	The number of samples in a sequence
N_0	The expected number of zero crossings per unit of time
N_y	The expected number of level crossings per unit of time
N_m	The expected number of maxima per unit of time
$p(y)$	Probability density of a function y
$R_{yy}(t)$	The Auto-correlation function of y
$R_{xy}(t)$	The Cross-correlation function of x with respect to y
t	Time
T	Sample sequence time ($= Nh$)
W_n	See equation A.2.12
$x(t), y(t)$	Continuous functions of time
$x(i), y(i)$	Sampled sequences of the function $x(t), y(t)$, respectively
Y^*	The complex conjugate of Y
$\mu_y, \bar{y}$	The mean value of the function y
ψ_y^2	The mean square value of y
σ_y^2	The variance of the function y
ϵ_a	The standard error of an estimate a
ϵ_v	The variance error of an estimate
ϵ_b	The bias error of an estimate
$\gamma^2(f)$	The coherence function

THE ANALYSIS OF RANDOM DATA
BY
D.A. WILLIAMS

Cranfield Institute of Technology,
Cranfield,
Bedfordshire,
U.K.

CHAPTER 1 PRE-PROCESSING

1.1 Introduction

The validity of an interpretation of the results obtained from any flight trial is critically dependent upon the fidelity of the measurements. This in turn depends upon:

- the selection of appropriate transducers, signal conditioning and recording equipment
- the determination of total measurement system static and dynamic characteristics for each parameter
- selecting suitable sensitivities for all elements in the measurement chain for each measurement
- the correct siting and mounting of the transducers
- the correct functioning of the complete measurement system

It is the author's experience that of the items listed above, the first often receives least attention. This is, perhaps, because instrumentation is considered to be a "black art" and best left to a specialist. Or, perhaps, because a large and expensive measurement system already exists, it is easier to risk using this than it is to demand one which is more appropriate to the task in hand. In any event, it is unlikely that a flight trial will prove successful unless appropriate instrumentation is used; and nobody is in a better position to specify the requirements for this than the engineer who is going to interpret the results.

What, then, are the requirements for an instrumentation system which is suitable for gathering random data? The answer to this question will depend upon the application. Random signals may occur as a result of, for example, navigation errors, ILS approach path errors, structural loads resulting from take off and landing or flight in atmospheric turbulence, excitation by turbulent boundary layers or power plant exhausts, through to radio frequency interference. Each may be of interest to at least one discipline involved in the design of a flight vehicle, and each may make different demands of an instrumentation system. In each case, however, the sort of questions which need answering before an adequate instrumentation system can be defined are:

- is it possible to measure all the parameters of interest?
- if not, is it possible to derive them from other measurements?
- should this derivation be effected in real time, or on the ground during analysis?
- what type of analysis is to be performed on the results?
- what frequency bandwidth is required?
- how many channels are required?
- is it possible to condition the experiment so as to reduce the number of measurement parameters?
- how long do the records need to be?
- what is the likely maximum signal level to be encountered?
- is the presence of the transducers and/or the instrumentation system likely to affect materially the parameters of interest?
- are variations in the parameters of interest likely to affect the output of the transducers in an indirect way (e.g. pressure transducers are sensitive to acceleration, jet effluxes generate accelerations; great care is needed to measure the pressure fluctuations of jet effluxes correctly)?

- are steady state levels important?
- is it necessary to filter the signals before sampling?
- are inter-channel phase relationships important?
- what sort of environment will the measurement equipment be subjected to?
- are the transducers outputs likely to be affected by uncontrolled environmental factors?
- does this matter?
- how are the measurement units to be transformed into physical units?

It is not within the scope of this Volume to discuss in detail all, or even many, of the questions posed above in relation to the measurement of random data. However, the analysis techniques described in Chapter 2 are performed upon the end product of what is often a very long chain of decisions and events, all of which can affect the validity of the results obtained. The present section is devoted to matters which may be looked upon as forming the interface between the raw measurements and the analysis proper.

1.2 Recording and Recovery of Random Signals

Data gathered during flight trials may be handled in a very large number of ways. Some of those which are available from commercial sources are shown in figure 1.1. The range is so wide, and each technique is so expensive to implement, that the luckless engineer faced with the task of choosing a system can find it very difficult to arrive at a decision. His problems are exacerbated by the fact that each type of system has characteristics which are beneficial in some applications, but which can create difficulties in others.

It has to be stated that an author encounters a similar kind of dilemma when faced with the task of describing techniques which are available for handling random signals. His difficulties arise from the fact that random signals are of interest to all engineering disciplines ranging from radio technology with signals of many millions of Hertz to fatigue investigations in which "cycles" can last for several hours. Further, the capabilities of some techniques are increasing very rapidly: the packing density potential of commercially available magnetic tape recorders, for example, has recently been increased by a factor of ten, and ideas currently under development threaten to increase this by at least another order of magnitude.

In the event, rather than examining particular applications areas, this author has chosen to classify, in fairly broad terms, the techniques which are available for handling experimental information, to discuss the characteristics of each technique as it may affect the handling of random signals, and to indicate possible applications areas in the general field of flight testing.

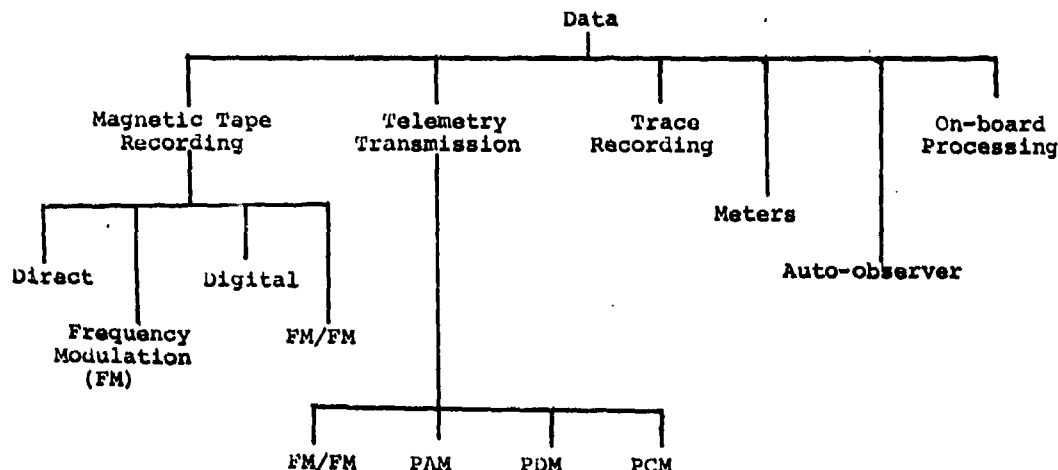


Figure 1.1 Methods of Recording and/or
Transmitting Flight Test Data

1.2.1. Analogue Magnetic Tape Recording

Magnetic tape provides the means for recording a large quantity of information in a very small volume onto a relatively permanent medium. Moreover, it is possible to recover information stored in this way directly in the form of electrical signals which can be fed to electronic analysis equipment without a great deal of difficulty. These are the features which make magnetic tape recorders so attractive to the flight test engineer when he is required to record flight trials information. In all other respects, magnetic tape is an unlikely medium for flight trials work. For example, recorders are complex and expensive devices which do not like operating in less than ideal conditions - they are sensitive to vibration, dirt and extremes of temperature, pressure and humidity. Further, they are incapable of recording directly information having frequencies less than several tens of Hertz. Finally, it is not possible to visualise recordings without recourse to additional expensive equipment. Despite these drawbacks magnetic tape recorders are now used almost universally for recording flight test information. It is therefore essential that engineers are familiar with the characteristics associated with magnetic tape and with the various techniques which can be used to encode information onto the medium. This Section is devoted to those encoding techniques which are generally classified as "analogue".

1.2.1.1. Direct Recording

Direct recording is, as the name implies, the process of recording analogue voltages directly onto magnetic tape, with the addition only of a very high frequency bias current to linearise the properties of the medium. It is the technique which is used in the majority of domestic audio recording equipment, largely because it affords the greatest frequency bandwidth for a given tape speed. However the technique has three major drawbacks which make it highly unpopular for measurement purposes. Firstly the apparent amplitude of a recorded signal is affected by time, tape-head contact, and by the lateral positioning of the tape relative to the replay head. Secondly it is necessary to "equalise" (filter to correct inherent variations in the frequency response characteristic) recovered signals - in the measurement context a very inexact process. Finally, as noted in Section 1.2.1., the medium is incapable of recording slowly varying signals.

The technique is of rather poor fidelity with accuracies quoted in decibels (normally a sure sign that all is not well). Therefore, despite its use in "Hi Fi" audio systems, it is not to be recommended for measurement purposes, unless the frequency bandwidth of interest is such that no other recording technique can be used. At the present time this implies an interest in frequencies in excess of 500 KHz.

1.2.1.2. Frequency Modulation Recording

The requirement to record signals with a reasonable degree of fidelity down to steady state levels can be satisfied by frequency modulation (FM) recording. The technique uses a voltage-controlled oscillator to convert the "instantaneous" value of a signal to an "instantaneous" frequency prior to recording on magnetic tape. It follows that, provided only that the frequency of the recorded signal can be detected, then the original undistorted signal can theoretically be recovered, irrespective of the distortions introduced by the recording medium.

Separate channels of data may be recorded on separate tracks of a multi-track tape recorder (up to eighty have been recorded on one inch wide tape). In this application modulations in frequency of up to forty percent of centre frequency are used to describe the amplitude variations of a signal. An increased number of recording channels may be accommodated by frequency multiplexing (FM/FM) up to eight channels onto each recorder track. In this case it is necessary to reduce the maximum allowable frequency deviation and to assign a different frequency band to each channel. Since the maximum bandwidth of a data channel is directly proportional to centre frequency, this has the effect of varying the data bandwidth between one channel and the next. For this reason the FM/FM technique is no longer widely used for recording random data.

The penalties of using FM as a recording (rather than transmission) technique are that the fidelity of the recording becomes critically dependent upon the stability of tape speed as it passes the record (and/or replay) heads, and that the "packing density" (maximum number of effective data cycles per unit length of tape) is only a tenth of that which can be achieved by direct recording.

The importance of recorder speed stability can be demonstrated by analogy with a chart recorder. If the pen of the recorder is made to oscillate at a constant frequency, then a sinusoid will be traced on the paper produced by the recorder when the paper travels at a constant speed. The distance, measured along the paper, between one peak and the next will be directly proportional to paper speed. If the paper speed varies without the knowledge of the user, then the frequency of the sinusoid will appear to vary. If the apparent frequency is to be interpreted as being proportional to the amplitude of a signal, then that amplitude will appear to vary. This is exactly analogous to the situation where signals are recorded on magnetic tape using frequency modulation.

Very precise tape speeds are achieved in instrumentation standard recorders either by using large flywheels (the Ampex AR 1200 used a flywheel weighing over 20kg) or by using tight servo control of the speed of a capstan (and consuming a lot of power). In applications where size and weight are critical, and where the recorder must be powered by a battery, neither of these stabilising techniques is appropriate. In such applications the recorder must be set to operate as best it can, and the hard work of recovering the recorded signal must be performed in the laboratory after flight.

A very good technique for recovering signals from frequency modulated information recorded under conditions where the tape speed is liable to vary uses the principle that a directly recorded signal of known frequency can be used to monitor the tape speed. If such a signal is recorded at the same time as the data, then it can be demodulated in the same way and the resulting time history used to remove speed induced amplitude changes in the data signals. For the system to be effective over the whole frequency band, the following rules should be followed:

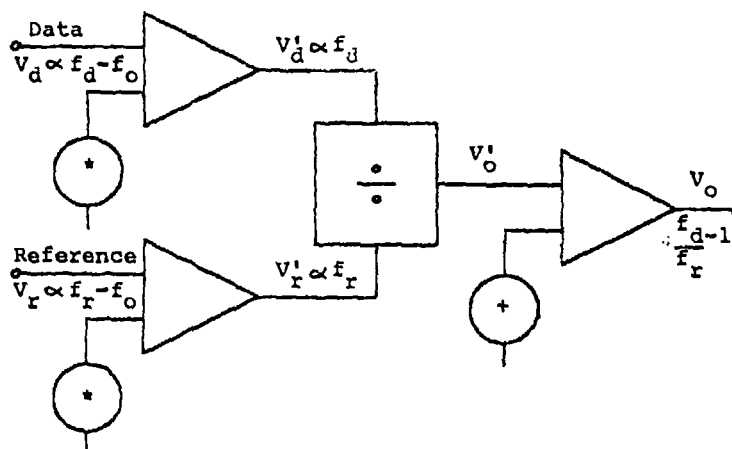
- the reference frequency should be of a similar value to the centre frequencies of the data modulators
- the output filters of all demodulators should be closely matched to limit phase errors between channels
- a reference track must be recorded on each stack of a multi-head tape recorder
- tape motion past the heads must have a low dynamic skew for accurate compensation
- best results are obtained when the azimuth of the replay head stack is matched to that of the recorder

A diagram of a "true" speed compensator is shown in figure 1.2. An illustration of the performance which can be achieved by the tape recorder is shown in figure 1.3. It may be seen that peak-to-peak errors of over forty percent (more than half the nominal range of the recorder) can be reduced to barely perceptible levels using this type of compensation. The particular recorder used to obtain the results shown in figure 1.3. was a lightweight recorder consuming approximately 2 amps at 20V DC. Its performance when operating in a severe environment compares favourably (after compensation) with laboratory standard instrumentation recorders.

A simplified form of speed compensation is offered by several manufacturers of instrumentation recorders in which it is assumed that the amplitude error caused by speed deviation is additive.

It is easy to demonstrate that this assumption is invalid, except at one data modulation frequency. The method does achieve a significant reduction in speed induced errors, but the author has found that the performance of systems using this method is generally inferior to that which can be achieved by "true" compensation.

In summary, frequency modulation allows magnetic tape recorders to be used as an accurate data recording medium, but the penalty is a decreased recording density and a greater complexity in tape transports and/or signal recovery equipment. FM tape recorders are however used almost universally for recording random signals in the frequency range from zero to 500KHz.



* Set to give zero output at zero frequency.

+ Set to give zero output at centre frequency.

If $V_d = K_1 \cdot \frac{S_t}{S_n} \cdot f_d - K_2$, where S_t is the actual tape speed and S_n is the nominal tape speed,

and $V_r = K_3 \cdot \frac{S_t}{S_n} \cdot f_r - K_2$

then $V'_d = K_1 \cdot \frac{S_t}{S_n} \cdot f_d$, and $V'_r = K_3 \cdot \frac{S_t}{S_n} \cdot f_r$

Hence $V'_o = K \cdot \frac{V'_d}{V'_r} = K \cdot \frac{K_1 \cdot f_d}{K_3 \cdot f_r}$

Finally $V_o = V'_o - K_4 = K \cdot \frac{K_1 \cdot f_d}{K_3 \cdot f_r} - K_4$

Figure 1.2. A "True" Speed Compensation Circuit.

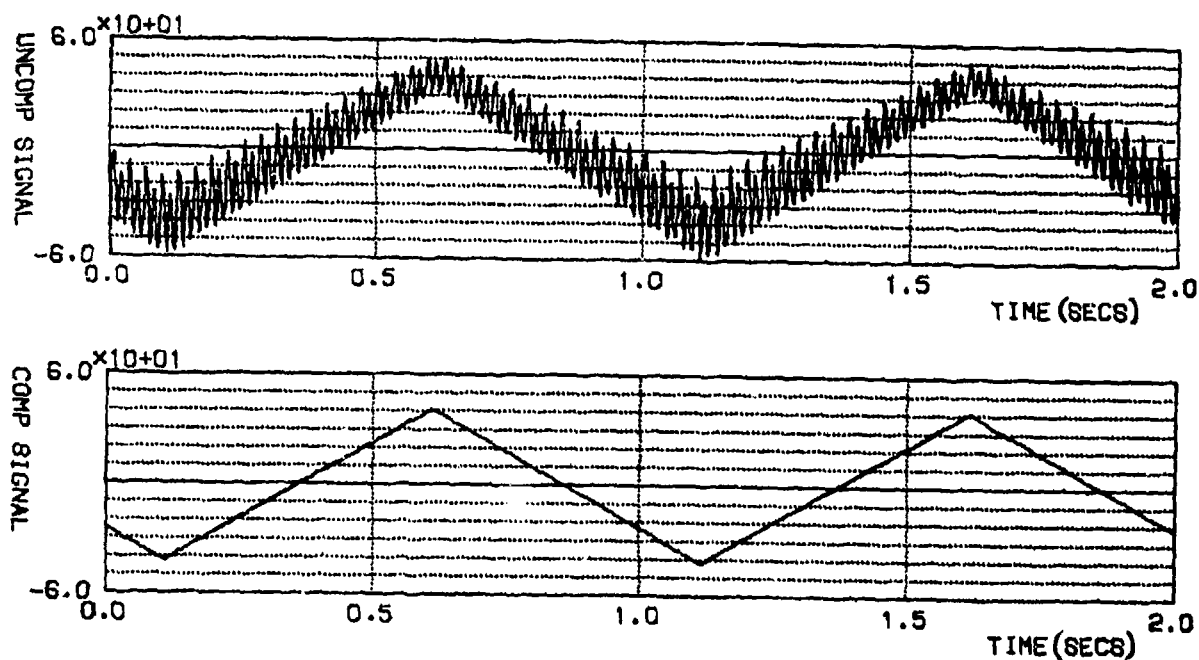


Figure 1.3. Performance Potential of the "True" Speed Compensator.

1.2.1.3. Digital Recording.

When data gathered in flight is to be analysed digitally, it is natural to consider the possibility of making the tape recorder completely transparent. This may be achieved by sampling and digitising the analogue signals at source and recording the digital information on magnetic tape in a format which can be read by a digital computer either directly (computer compatible) or indirectly, via an interface.

At the present time standard computer compatible recording formats are rarely used, for a number of reasons:

- data must be recorded on tape in discrete blocks. This can only be achieved by using a pair of interleaving storage buffers.
- computer tape recorders normally include hardware which prevents a block being read if it contains an error.
- computer standard packing densities are normally much lower than can be achieved technically.

A great many recording formats which are not, in general, computer compatible are, however, available commercially. Discussion of these is beyond the scope of the present manual; the interested reader is referred to References 1.1. and 1.2., for example.

All digital recording techniques employ the principle of sampling a number of analogue signals sequentially, digitising each sample, and encoding the resulting digital value prior to recording the information onto magnetic tape.

Provided that the record quality is sufficient to enable each recorded "bit" to be recovered, then the recording and replay processes are completely transparent (i.e. there is no difference between the recorded and recovered data.) This feature is unique to digital recording and makes the technique very attractive to flight test engineers. It is inevitable that the technique has disadvantages. These include:

- additional electronic equipment must be carried in the aircraft.
- the packing density of information is much lower than can be achieved with FM recorders.
- it is sometimes necessary to have a prior knowledge of the signal levels if the best use is to be made of limited resolution digitizers.
- a tape recorder having good tape speed stability is normally required.
- data sampling can cause difficulties (see Section 1.3.)
- complex and expensive equipment is normally required in order to recover flight records.

As a general rule, digital recording techniques are used for signals which vary comparatively slowly with time and which have known amplitude bounds. In the context of random phenomena, they can be used to good effect for monitoring navigation errors, rigid body aircraft motion and, perhaps, flexible aircraft responses up to frequencies of several tens of Hertz. They are not at present suitable for recording acoustic signals, boundary layer pressure fluctuations, or engine induced vibrations.

1.2.1.4. Telemetry.

Telemetry has been used for many years for transmitting measurements from flight vehicles which must be controlled remotely and/or which are unlikely to be recovered. More recently, the high cost of flight testing has made the technique viable for normal aircraft flight trials, when the potential for "on-line" analysis and display can enable the flight envelope to be explored more rapidly. However the technique remains very costly to implement and may restrict unduly the flexibility of flight trials since the aircraft must remain within range of a ground receiving station during the time measurements are required. The interested reader is referred to References 1.3. and 1.4. for reviews of telemetry techniques.

Telemetry transmission gives the engineer the capability of processing random information "on-line" to obtain statements about the quality of data being received. Statements of the peak, mean and RMS (root mean square) values of each channel are normally sufficient to satisfy this requirement. It is unlikely that the computing power necessary to make more detailed on-line statements about the data could be justified except in particular situations where the aircraft may be at risk, for example during flight flutter trials.

1.2.1.5. On-board Processing.

The arrival of the ubiquitous microprocessor, or microcomputer, has greatly reduced the size and cost of providing on-board processing capabilities, and it is certain that such systems will be used increasingly in multi-seat aircraft where telemetry is impractical. However the environment on board an aircraft is not conducive to decision making, so that the major application of on-board processing is to provide for relatively simple assessments in order to ensure that information of suitable quality has been obtained. Where random data is of interest, then useful on-board processing is likely to be restricted to estimates of peak, mean and RMS values for each data channel. The task of processing the measurements in detail is likely to remain firmly ground-based.

1.3. Analogue Pre-processing

Signals which have been transmitted from an aircraft or recorded in flight in an analogue format are normally recovered prior to analysis as continuous signals whose voltages are proportional to the physical parameters under investigation with the addition, perhaps, of measurement and reconstruction noise. They may then be processed using analogue, or pseudo-analogue, equipment. Alternatively, and this is the usual case, they may be sampled and converted to a digital sequence in preparation for analysis using a digital computer.

Signals which are recovered in this way have the characteristics that they are "soft", that is, no two attempts at recovering the signals produce identical results. Such a characteristic is caused by noise injection and/or fluctuations in sensitivity and offset of the recovery equipment. The obvious disadvantage of this characteristic is compensated for by a number of advantages which are afforded by analogue recording. These include:

- the recovery equipment can be optimised so as to recover successfully signals that would otherwise be lost (due to over modulation, for example)
- signals can be filtered and/or amplified to make best possible use of the available processing equipment
- the acquisition parameters can be varied to suit the requirements of a particular recording

The subject of this Section is the pre-processing techniques which can be used with analogue signals to maintain that the best possible analysis fidelity.

1.3.1. System Calibration

The "soft" nature of analogue recording systems has been the stimulus for considerable ingenuity in generating in-flight system calibration signals. At worst such signals allow the record/replay/acquisition system to be calibrated prior to analysis. In their most sophisticated form they allow the performance of the complete measurement system to be monitored automatically during flight.

System calibration signals recorded on the data tape can take a variety of forms, including:

- multiple steady-state levels
- sine waves of constant amplitude
- square waves of constant amplitude

Examples of typical calibration signals are shown in figure 1.4.

The type of signal chosen for any particular application depends upon tradition and/or upon the nature of the transducers employed and their associated signal conditioning equipment. It is certain that the most satisfactory calibration signal uses multiple steady state levels of known amplitude expressed in physical units. In this case the zero offset, sensitivity, and sign of each signal channel can be estimated with no ambiguity.

The interpretation of sine wave calibration signals presents some difficulty. The perceived level of a sine wave may be estimated by:

- computing the RMS amplitude of the fundamental frequency component
- computing the RMS amplitude of the fundamental and harmonics of the sine wave (both of these can be obtained from the power spectral density)
- computing the average of successive peak-to-peak excursions (the peaks must be well defined)
- computing the amplitude probability density of the signal

The four methods outlined above will, in general, yield four different estimates for the RMS value of the signal. The most appropriate therefore, is the method which most closely resembles that used to define the level of the signal prior to recording, provided that all of the intervening measurement and recording elements are linear. In the absence of other information the most satisfactory method is the first, provided that supporting information regarding the amplitudes of the harmonics and overall noise levels is also presented.

Square wave calibration signals can be processed in a manner similar to sine waves. However, filters are used to recover the modulated analogue signals. These will suppress higher harmonics of the square waves and may give rise to overshoot of the signal after each transition. Therefore the only satisfactory method of estimating the peak levels of a square wave calibration signal is to use its amplitude probability density.

It is, in general, possible to estimate neither the sign of the scale factor nor the value of the zero offset from sine or square wave calibration signals.

Multiple steady state calibration levels can be interpreted by associating the average value of each level with the physical value assigned to that level. Both the zero offset and sensitivity of the channel can then be computed by regression analysis of the results. Examples, both good and bad, of the results of this type of analysis are shown in figures 1.5. and 1.6.

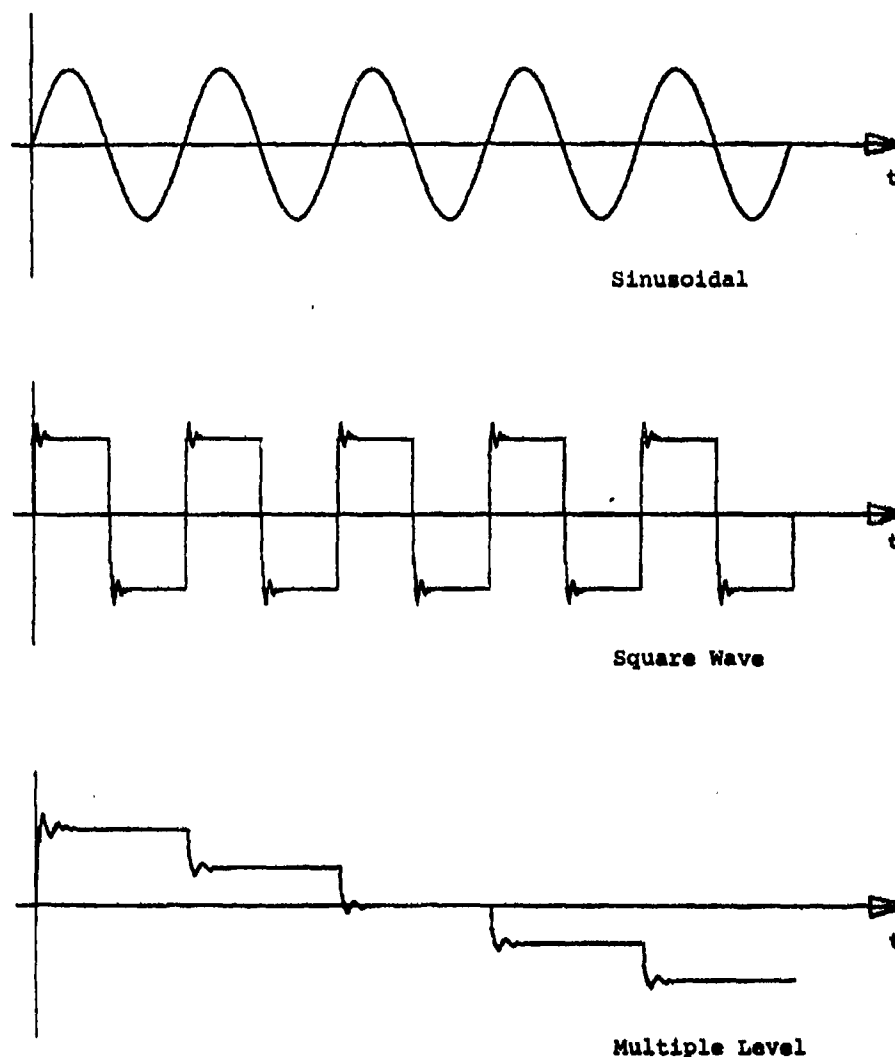


Figure 1.4. Examples of Recorded System Calibration Signals

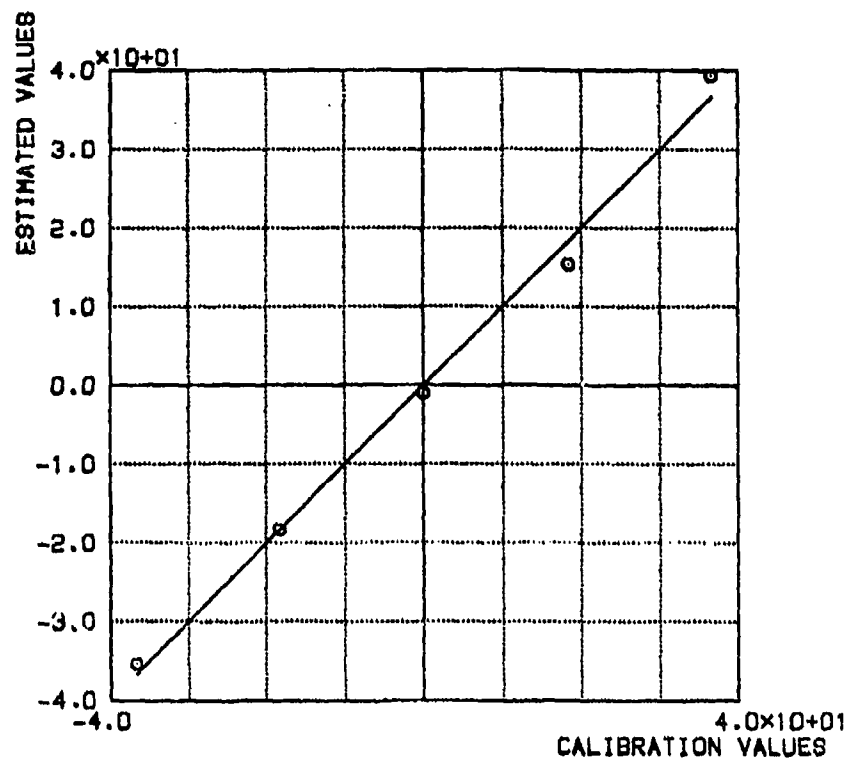


Figure 1.5. Example of Fitted Calibration Signals
(Poor Quality.)

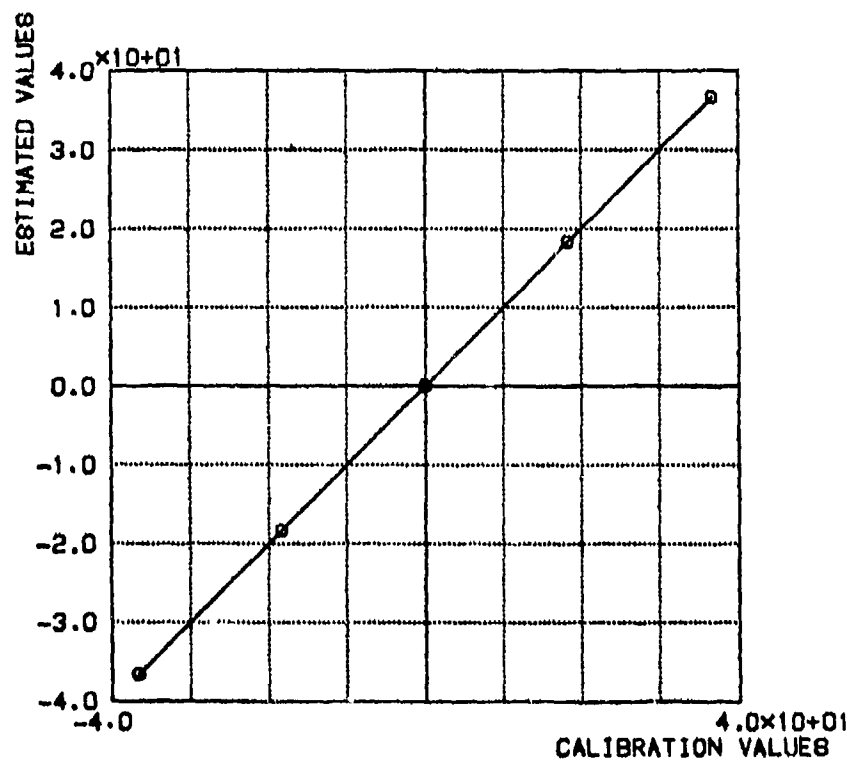


Figure 1.6. Example of Fitted Calibration Signals
(Good Quality.)

1.3.2. Analogue Filtering

Analogue filters have been used for conditioning random signals prior to analysis for a very long time. In particular, early analogue spectral analysers possessed a very limited dynamic range and it was common practise to "pre-whiten" random signals (q.v. reference 1.5.) to obtain satisfactory results. The objective of this process was to modify the characteristics of a signal to obtain, so far as was possible, a band-limited white noise spectrum. Provided that the modification was effected by filters having precisely known characteristics, then the spectrum of the original signal could be reconstructed by weighting the computed spectrum.

Modern analysis equipment has brought with it a dramatic improvement in dynamic range, and this has led to a widespread belief that conditioning filters of this sort are no longer necessary. However it should be remembered that modern analysis equipment has also permitted a greatly increased frequency resolution, more than a hundred times the resolution of early analysers, and applications still exist in which significant improvements can be effected by the use of simple pre-whitening filters.

In fact, all commercially available spectral analysers permit the user to condition signals in a simple way by high pass filtering. The filters are provided mainly to inhibit the acquisition of large signal offsets, but they have the additional characteristic of suppressing low frequency signals which may be of no interest to the analyst. Such high pass filters are in essence, simple pre-whitening filters, and computed spectra can be corrected to remove their effects.

1.3.3. Gain Adjustment

It frequently happens that the levels which may be attained by measured parameters during a flight trial are unknown prior to the trial. When this is the case, then it is common practice to set the channel sensitivities to a conservative value to ensure that over-ranging does not occur. As a result the analyst is sometimes called upon to extract information from signals having a maximum value which is only a small proportion of the available recorder range. In this event it is normal to amplify the signal level before sampling or analysis so as to make best possible use of the analysis equipment. The analysis results must, of course, be adjusted to take account of the resulting changes in sensitivity and, if appropriate, signal offset.

Measurement system noise (400Hz breakthrough, recorder resonances, etc.) is often predominantly narrow band. Hence when the processing involves spectral analysis, very small signals can be analysed in a meaningful way provided that the inherent system noise characteristics are known (by analysing a record which contains no signal, for instance). The author has, on occasion, amplified recorded signals destined for spectral analysis by more than 40dB (100 times) with no apparent detriment to the results.

Amplifiers used for sensitivity adjustment should have, ideally, switched gain settings which are known accurately. They should possess an adequate frequency bandwidth and, particularly when inter-channel time delays are important, should have matched phase characteristics. All of these parameters must be measured; experience has shown that it is not sufficient to rely upon manufacturers' specifications.

1.3.4. Tape Speed Compensation

Analogue tape recorders are not perfect instruments. The speed at which tape passes the record or replay head may be affected by vibration, temperature, supply voltage, and even the amount of tape on the take-up reel. Tape speed variations can cause sampling time "jitter" and, if frequency modulation is used, signal amplitude fluctuations. A method by which speed dependent amplitude fluctuations can be reduced has been described in Section 1.2. It remains to consider the problem of time jitter.

Time jitter is important when a record is being examined in detail to establish precise frequency information or to estimate accurate inter-channel phase characteristics. In such situations the resolution of the record can be limited by frequency smearing caused by jitter. More seriously, frequency and phase information may be biased if the tape speed is consistently different from nominal.

When a record is to be analysed digitally, then the effects of time jitter may be reduced significantly by using a reference track (it can be the same track as that required for amplitude compensation) to produce sampling clock pulses. This causes data samples to be obtained at points related to record, rather than replay, time. However, drop outs in the reference track can cause a degradation of the analysed results, rather than the hoped for improvement. When this happens, the insertion of a phase-locked oscillator (PLO) between the sampling clock and the pulse shaping circuit can restore matters to a degree. The phase-locked oscillator will "flywheel" over dropout areas, thereby ensuring that no data is lost. It is to be noted that a PLO should be used only when reference track drop-outs are a problem, because of its finite response time to tape speed changes and because the phase of its output relative to the reference frequency changes with the value of that frequency. A typical circuit is shown in figure 1.7.

Certain tape recorders offer a "tape servo" facility which allows the replay speed to be locked to a reference track. This facility can be used to limit the more extreme tape speed fluctuations, but the control loop frequency bandwidth is generally insufficient to remove the effects of time jitter completely.

When analogue equipment is to be used to analyse a record, then tape speed induced time jitter can be reduced by using a FIFO (first in, first out) buffer. Signals output by the tape recorder are sampled, digitised and input to the buffer under the control of a recorded reference frequency. Samples are then output from the buffer and passed to a digital-analogue converter under the control of a local oscillator. The arrangement is shown in figure 1.8.

The scheme works provided that the buffer is sufficiently long and provided that the average clock rates are identical.

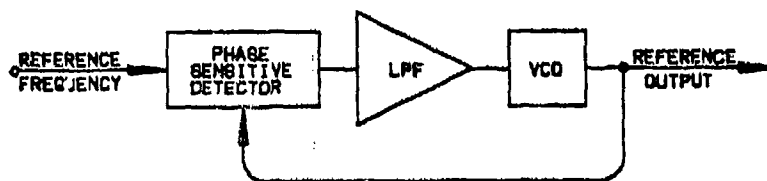


Figure 1.7. Schematic of a Phase-Locked Oscillator.

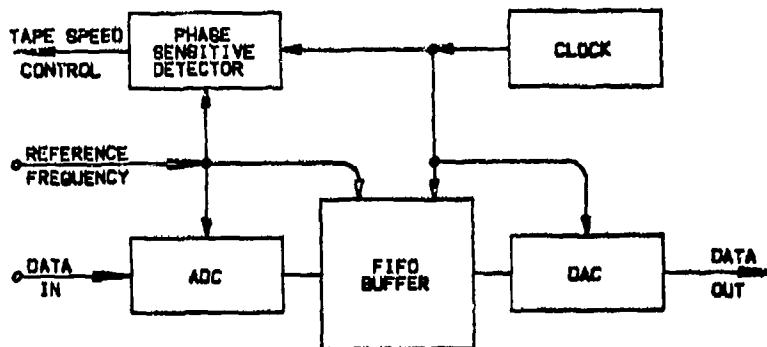


Figure 1.8. A Circuit to Remove Time Jitter From Analogue Signals.

1.4. Sampling

The majority of parameters of interest to the flight test engineer occur naturally as continuous functions of time. If the parameters are to be processed by a digital computer, and this is now normally the case, then the parameters must, at some stage, be converted into a set of digital numbers, each number representing the value of a particular parameter at a particular instant of time. The resulting sequence of numbers is referred to as a sampled time history or data sequence. Usually, although not always, a parameter is sampled such that the time interval between successive samples is constant. When this is not the case it is normally convenient for processing purposes to interpolate between samples so as to generate a sequence which does have a constant interval between successive samples.

A signal may be sampled at any stage in the measurement process, depending upon the nature of the signal and, often just as important, upon the measurement system available to the Flight Test Engineer.

It is current practice to transmit and/or record low frequency information (e.g. steady state and aircraft "rigid body" motions) in a digital format, but to use analogue methods for recording higher frequency information (e.g. flutter measurements, engine vibrations, etc.). The reasons for this are simply that the cost per unit volume or weight for recording each parameter is lower for an analogue system than it is for a digital system when the required frequency bandwidth is greater than a few tens of Hertz.

It is important to realise that the rules for sampling a continuous function of time successfully, and the characteristics of the resulting sampled sequence are the same regardless of the point at which the sampling is executed: the output of a shaft encoder has the same sampling characteristics as the output of an equivalent analogue measurement system after it has been sampled at the input to a digital computer.

The remainder of this Section is concerned with the characteristics of a sampled sequence, and the rules for sampling which result from those characteristics. Also included are sections devoted to the characteristics of the elements which may be used for effecting signal sampling.

1.4.1. Principles of sampling

Every engineer has, at some time, plotted a graph of measured extension against applied load and fitted a curve through the points to determine the stiffness of a spring. A number of decisions are made during this process. The most important decision relates to the choice of curve from the infinity of possible curves which pass through the measured points (see, for example figure 1.9.).

The process of fitting a curve to a set of measured data points is, in some respects, the inverse of the sampling process. In one case the task is to represent a set of sampled data points by a suitable continuous function; in the other the task is to represent a continuous function by a set of sampled data points.

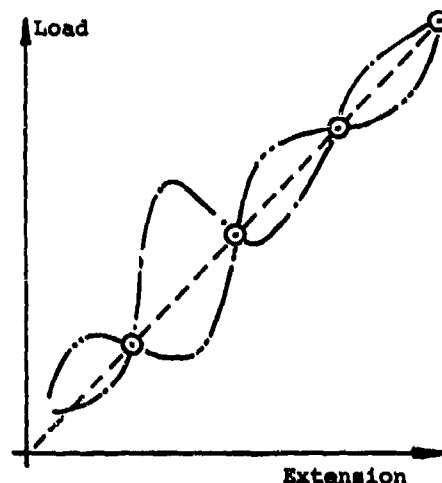


Figure 1.9. Some possible curves passing through one set of data.

It is not normally difficult to choose a suitable function relating, for example, the extension of a spring to the applied load since the choice is limited by the physics of the process. However, when there is no reason to restrict the choice of functional relationship, then an infinite number of possible solutions exist. The latter is usually the case for a sampled time history if no precautions have been taken to limit the choice. It clearly follows that for a sampled time history to have any meaning, some mechanism must be incorporated to limit possible excursions of the signal between successive samples.

If this limiting can be achieved then it becomes possible to interpolate between samples without incurring an excessive error, so that the sampled sequence can be said to represent fairly the original time history. It remains to define reasonable limits for excursions between samples and to decide upon the characteristics of the limiting mechanism.

1.4.2. The Sampling Theorem

The Sampling Theorem states that a sequence resulting from sampling a continuous signal at a constant rate of $2f$ samples per second completely describes that signal (in the sense that the original signal can be reconstructed) provided that the signal contains no frequency components greater than or equal to f Hz. The frequency f is known as the Nyquist or folding frequency. The Sampling Theorem is discussed further in Appendix A.

The underlined part of the above statement is often omitted: its relevance is clearly illustrated in figure 1.10., which shows a sinusoid of frequency f Hz. sampled at $2f$ samples per second!

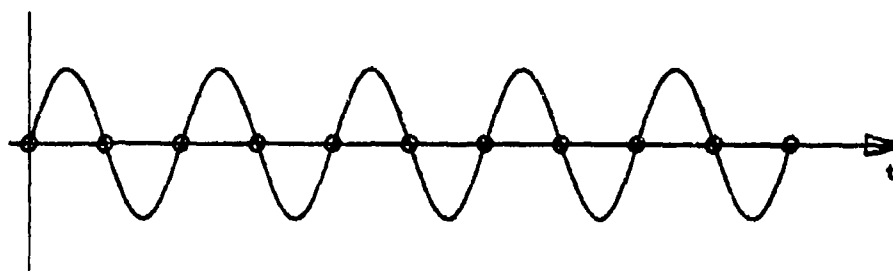


Figure 1.10. A Sinusoid Sampled Exactly Twice per Period.

The Sampling Theorem describes, in frequency domain terms, the excursions which are allowable between samples. It states that a signal can be sampled successfully, provided that it has been filtered prior to sampling so as to eliminate effectively all signal components greater than or equal to the folding frequency. It will be noted that the resulting sequence represents the filtered rather than the original analogue signal. Filters which are used in this way are called "anti-aliasing" filters. The design of suitable filters is discussed in Section 1.4.5.

The selection of an anti-aliasing filter for any particular application depends upon the nature of the signals under investigation and upon the proportion of the available frequency bandwidth which is of interest. For general applications a rather conservative rule is adopted. This demands that the amplitude of a white noise signal shall be reduced to one percent of its original value at the folding frequency. The available bandwidth is then dictated by the number of poles used in the filter and the permissible attenuation of the signal, bearing in mind that the effects of this can, under some circumstances, be removed after sampling. Figure 1.11. represents a filter order selection guide using the above rule (curve A). Also included is a guide based upon the rather less conservative rule of less than one percent aliasing at the highest frequency of interest (curve B). The latter rule implies that aliasing is acceptable provided that its effects are not significant in the frequency range of interest.

In general, a filter will be acceptable if its characteristics lie between curves A and B. It is interesting to note that a six pole filter only allows a frequency range of between 46 and 63 percent of that theoretically available.

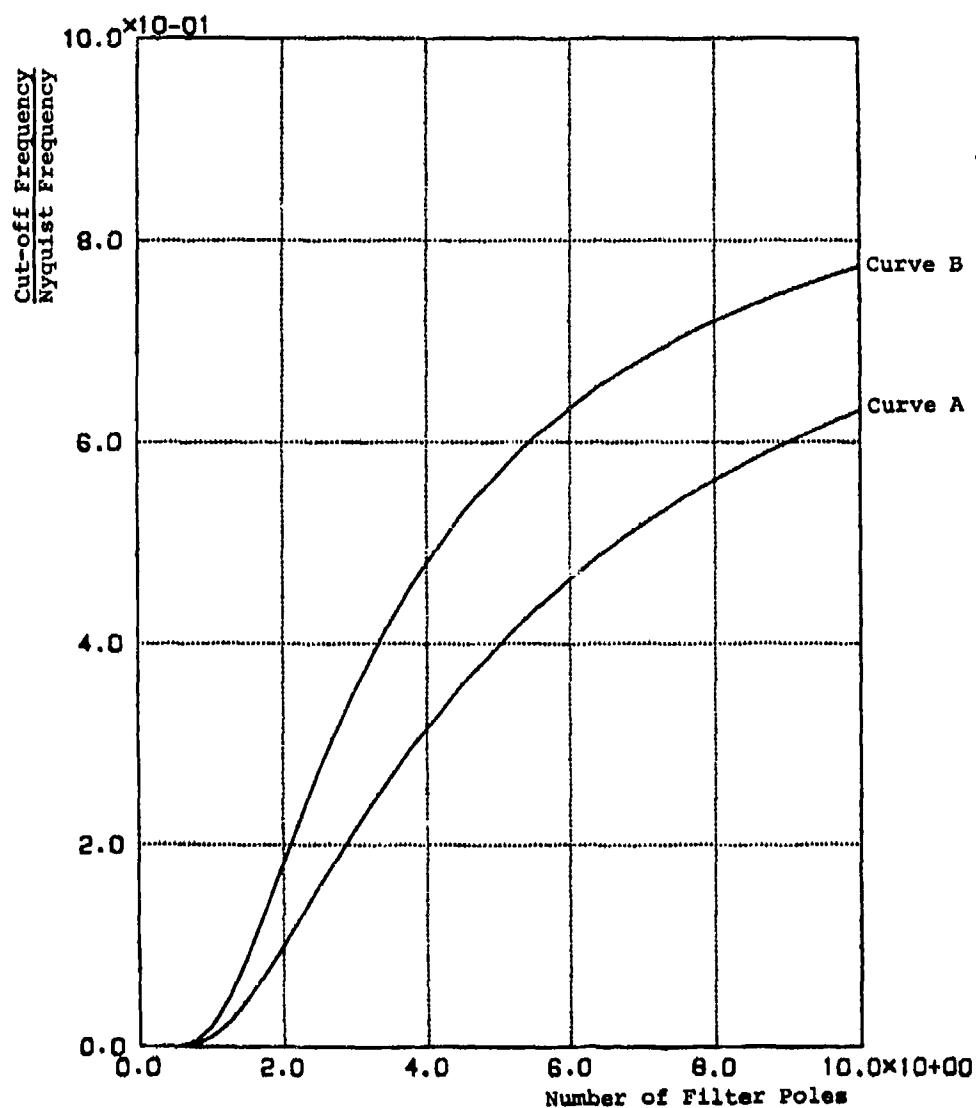


Figure 1.11. Attainable Bandwidth as a Function of Filter Order (No. of Poles)

1.4.3. The Effect of Aliasing

The rules for filtering an analogue signal prior to sampling have been discussed in the previous section. It is relevant to consider the effects of omitting such filtering, partly because some airborne digital systems do not incorporate filtering specifically to prevent aliasing, and partly because the characteristics of an aliased signal can be utilised in digital processing.

It is assumed that the signal under investigation is continuous and has characteristics which do not vary with time (i.e. is stationary). Suppose that a portion of the signal is sampled at constant interval such that the folding frequency of the signal is f_c . Then a periodic component of the signal of frequency f may be described as

$$\cos 2\pi ft \quad 1.4.1.$$

if a suitable time reference is chosen. The k th sampled value is then equal to

$$\cos 2\pi fkh = \cos \frac{2\pi f.k}{2f_c} \quad 1.4.2.$$

where the interval between successive samples is $h = 1/2f_c$.

An arbitrary frequency f can be re-defined as

$$f = 2n.f_c \pm f' \quad 1.4.3.$$

where n is a suitably chosen integer and f' lies between zero and f_c .

Equation 1.4.3. may be substituted into 1.4.2. to give

$$\begin{aligned} \cos \frac{2\pi f k}{2f_c} &= \cos \frac{2\pi k.(2n.f_c \pm f')}{2f_c} \\ &= \cos 2\pi k.(n \pm \frac{f'}{2f_c}) \\ &= \cos \frac{2\pi k f'}{2f_c} \end{aligned} \quad 1.4.4.$$

because $\cos 2\pi kn$ is unity for integer k and n .

Equation 1.4.4. is identical with 1.4.2., but with f replaced by f' . Note that no restriction was placed upon the value of f , but that f' was assumed to lie between zero and f_c . Thus equations 1.4.3. and 1.4.4. imply that any periodic component of a sampled signal will appear to have a frequency within the range zero and f_c . Equation 1.4.4. implies that the amplitude of the component will be unchanged.

Equation 1.4.3. states that the perceived frequency of a periodic component is related to the actual frequency as shown in figure 1.12.

Thus a frequency perceived at f may in fact be at any one of the following frequencies:

$$f ; (2f_c - f) ; (2f_c + f) ; (4f_c - f) ; (4f_c + f) ; \text{etc.}$$

Consideration of the effects of aliasing can be extended to aperiodic signals within the assumption of stationarity by a similar analysis if it is assumed that a portion of the time history can be represented by a Fourier Series. The conclusions reached are then identical to those above.

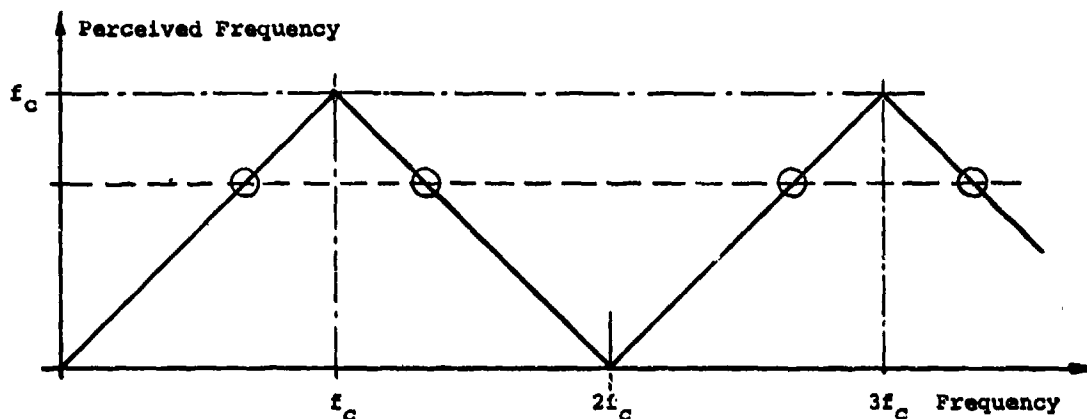


Figure 1.12. Graphical Interpretation
of Equation 1.4.3.

To summarise, a frequency component which is aliased by sampling will be perceived as a component having the same amplitude (RMS value for a random component) as the original, but having a predictable, incorrect frequency.

The consequences of aliasing can be catastrophic, particularly if the original signal cannot be re-examined, since it is impossible to differentiate between "true" and aliased components of a signal.

However, the predictable nature of aliasing can be used to enhance the analysis of a signal, as described in Section 2.4.

1.4.4. Typical Sampling Systems

Sampling systems are constructed using five basic elements:

- Pre-sampling filter
- Multiplexor
- Sample/hold module
- Analogue - digital converter (ADC)
- Controller

The elements can be arranged in a variety of configurations; three typical arrangements are shown in figure 1.13.

The purpose of the filters, one of which should be dedicated to each analogue signal, is to prevent aliasing of the digital output. (see section 1.4.3.). Sample/Hold modules are used to maintain a constant sampled analogue value for the period required by the converter to quantise that level. A multiplexor is commonly used at some point to limit the number of connections required and/or to reduce the number of modules required in a system.

It is relevant to consider in more detail the configuration shown in figure 1.13.

The first arrangement (a) employs a filter, a Sample/Hold module, and an Analogue - digital converter for each analogue signal. The digital outputs from the ADCs are multiplexed to obtain a single digital data stream. The arrangement is capable of very high sampling rates since the conversion process is executed in parallel. Moreover samples can be obtained with no time delay between one output and another. However the item of major cost in a sampling system is the analogue - digital converter (ADC), so that the arrangement is prohibitively expensive for all but a few specialised applications.

Arrangement (b) employs an analogue multiplexor to enable multiple analogue inputs to be sampled by one ADC. Sample/Hold modules are used to obtain apparent parallel sampling of the analogue signals. The arrangement has a lower acquisition rate potential than (a) but is much less expensive.

The configuration most commonly used in airborne acquisition systems is shown as arrangement (c). It is similar to (b) but requires only a single Sample/Hold module, thereby reducing the cost of the system. The arrangement introduces time delays between data channels which, if embarrassing, must be removed prior to processing.

Of the three, arrangement (b) has much to commend it provided that the increased size and weight in relation to arrangement (c) can be tolerated. The higher initial cost of the arrangement compared with (c) is offset by a reduction in processing time for every record in which inter-channel time delays must be removed.

It remains to consider in some detail the individual characteristics of the four elements used in sampling systems.

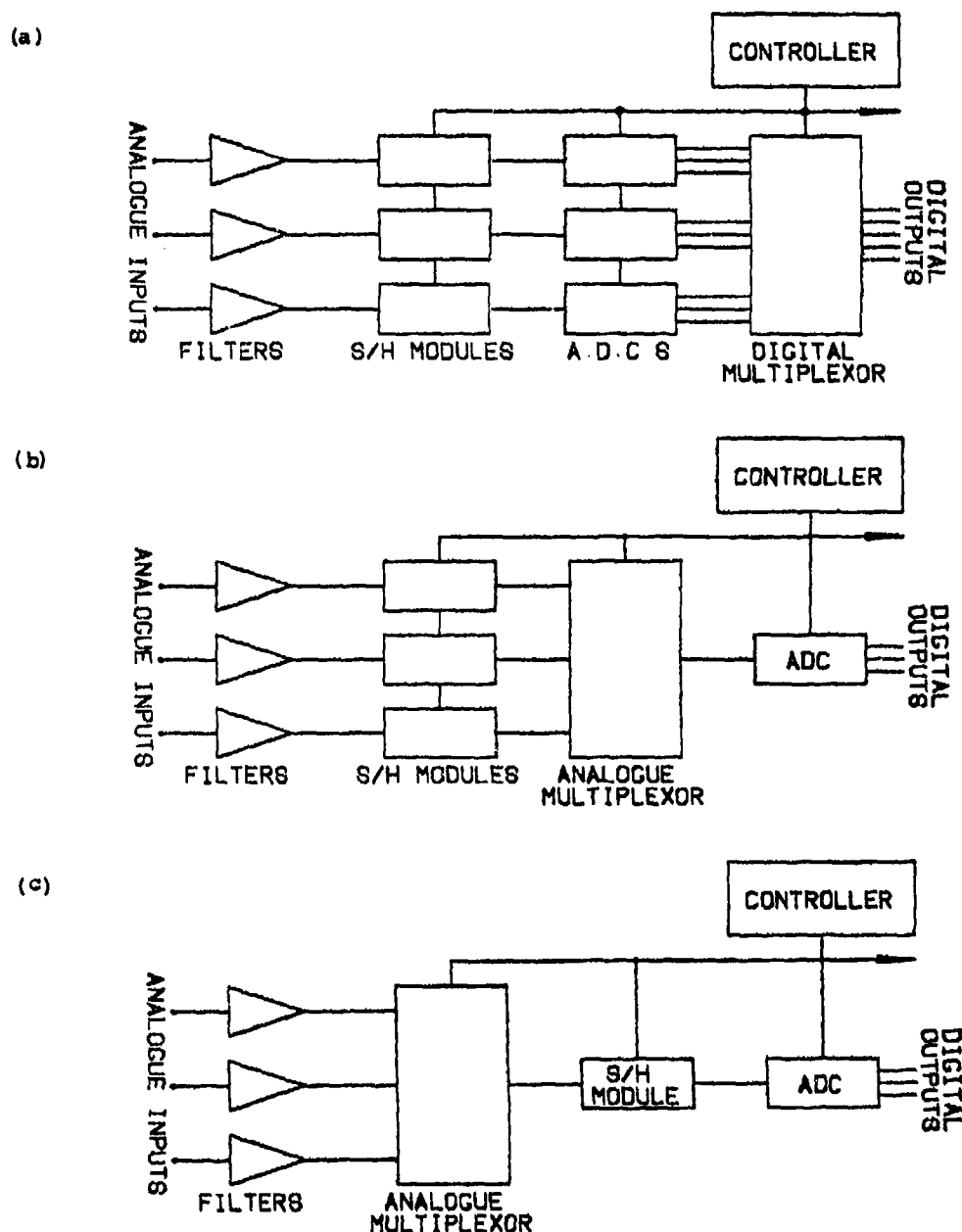


Figure 1.13. Schematics of Three Types of Sampling System.

1.4.5. Pre-sampling Filters

Pre-sampling filters are low pass filters which are used to prevent the introduction of aliasing by the sampling process (see Section 1.4.3.). All analogue low pass filters modify signals passing through them in three ways: they introduce

- frequency dependant attenuation or amplification
- frequency dependant phase lag
- possible "ringing" when excited by signal transients

The first is the characteristic required of a pre-sampling filter, the second and third are undesirable by-products of an analogue filter.

Frequency dependent phase lag translates to a frequency dependent delay in the time domain, and unless controlled it may seriously distort the analogue signal. An illuminating starting point to a discussion of filter design is to investigate the "ideal" phase lag versus frequency relationship for an analogue filter.

If it is accepted that an analogue low pass filter must introduce time delay, then an optimum filter is one which delays all signal frequency components by the same amount. Such a filter would pass a signal without distortion, provided that all frequency components of the signal lie within its pass band. This is the case when the action of the filter upon a signal component having an amplitude A and frequency f Hz may be expressed as

$$A \cdot \exp -i2\pi f \cdot (t + T) \quad 1.4.5.$$

where T is the (constant) time delay introduced by the filter. This represents a phase shift of

$$-2\pi f T = K \cdot f \quad \text{radians} \quad 1.4.6.$$

Thus one characteristic of an ideal low pass filter is that the rate of change of phase with frequency of its transfer function should be a constant, at least within the pass band of the filter.

The frequency-dependent gain (attenuation or amplification) of a low pass filter is characterised by its ultimate slope (the slope as frequency becomes very large). In a physically realisable filter this is given by:

$$H(f) = (f)^{-n} \quad 1.4.7.$$

where n is the order of the filter (and equals the number of "poles" or the order of the highest derivative in the equation of motion.) The gain characteristic of an ideal filter is shown in figure 1.14., which shows a unity gain up to the filter "break point" (also known as the cut-off frequency) and a gain inversely proportional to n thereafter.

In an "ideal" filter having a time delay which is constant at all frequencies, the delay T is related to f_0 by the relationship:

$$T = \frac{n}{8f_0} \quad \text{seconds.} \quad 1.4.8.$$

It can be seen from equation 1.4.8. that the time delay is directly proportional to the order of the filter. Thus a filter giving the optimum gain characteristic (an infinite value of n) would have an infinite time delay, i.e. would not work.

In practical terms, the number of poles used in a filter depends upon the application and an acceptable cost. A filter with a large number of poles is

- difficult to design (a computer is desirable to design a filter with more than three poles)
- Difficult to build (accurately matched components are necessary to control the characteristics)

Such a filter also has characteristics which can be undesirable, since it is not possible to achieve in one design both low time domain distortion and good frequency domain gain characteristics.

A third order low pass filter can be used to illustrate the problems. The equation of motion of such a filter is given by:

$$\frac{a_3}{(2\pi f_0)^3} \frac{d^3 y}{dt^3} + \frac{a_2}{(2\pi f_0)^2} \frac{d^2 y}{dt^2} + \frac{a_1}{2\pi f_0} \frac{dy}{dt} = f(t) - y \quad 1.4.9.$$

where y denotes the response of the filter to an input $f(t)$. Values of a_3 , a_2 , and a_1 , which result in filters optimised to satisfy different requirements, are shown in figure 1.15., and the corresponding frequency response functions are given in figures 1.16 and 1.17.

The synthesised responses of the three types of filter to a square wave input are shown in figure 1.18. The fundamental frequency of the input was, for the purposes of the illustration, chosen to be 1 Hz. and the natural frequencies of the filters (f_0) were chosen to be 3 Hz. The response of filter A does not overshoot the input and has the largest apparent time delay (compare the output values at 0.2 seconds). Filter C, on the other hand, overshoots the input by approximately eight percent (in this case) and distorts the input waveform to a considerable degree.

Thus the engineer who is interested in observing a sampled time history directly might conceivably select either filter A or filter B, and tolerate the rather poor frequency response in the pass band in return for a lower time domain distortion. On the other hand, the engineer who is interested in observing the data in the frequency domain would probably prefer the response characteristic of filter C.

The most commonly used filter for random data analysis, when constant time delay is not of paramount importance, is the Butterworth Filter. This has the gain characteristic given by:

$$H(if) = (1 + (f/f_0)^{2n})^{-1/2} \quad 1.4.10.$$

which is a relatively simple filter to realise and closely resembles the "ideal" of figure 1.14. when n is large.

When phase information between two channels is required, then it becomes necessary to control the phase characteristics of the filters. An eight pole filter has 360 degrees of phase shift at the cut-off frequency. If the relative phase shift between two filters is to be maintained to less than one degree, then individual circuit components must be controlled to better than 0.25 percent. Clearly this is close to the acceptable limit for component tolerances.

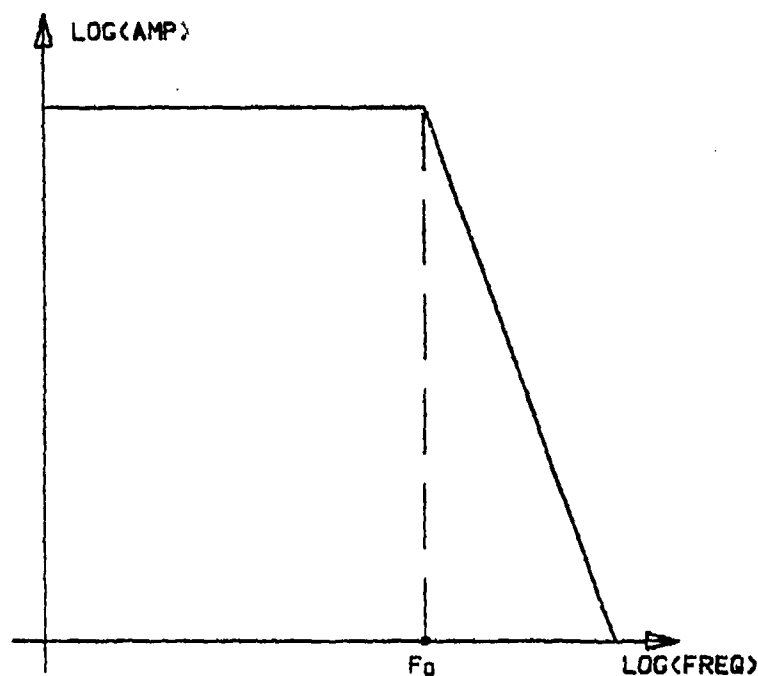


Figure 1.14. The Gain-Frequency Characteristic of an "Ideal" Low Pass Filter.

Ident. (Figure 1.16.)	a_1	a_2	a_3	$H(k), k = f/f_0$	Description.
A	2.4662	2.4329	1	$(1 + 1.216k^2 + 0.9865k^4 + k^6)^{-1/2}$	Lowest time domain distortion (Bessel.)
B	2.3562	2.2750	1	$(1 + k^2 + 0.46322k^4 + k^6)^{-1/2}$	Low frequency components have time delay given by equation 1.4.8.
C	2	2	1	$(1 + k^6)^{-1/2}$	"Best" frequency response characteristics (Butterworth).

Figure 1.15. Three Different Third-Order Filters.

(The Identifiers refer to the Frequency Response Functions shown in Figures 1.16. and 1.17.)

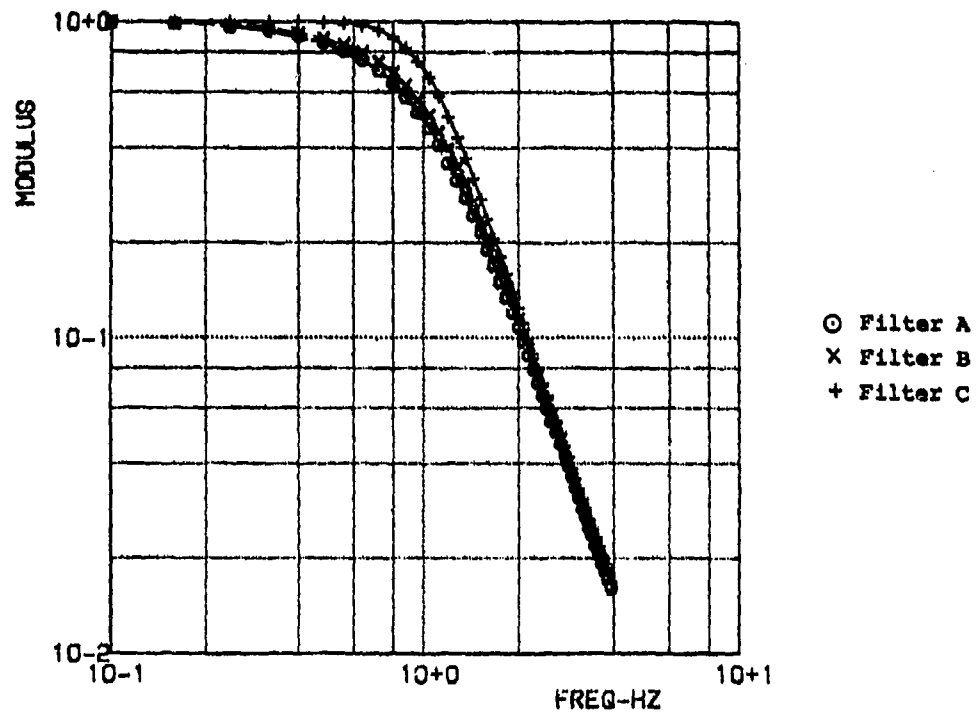


Figure 1.16. Gain Characteristic of Three Third-Order Filters.

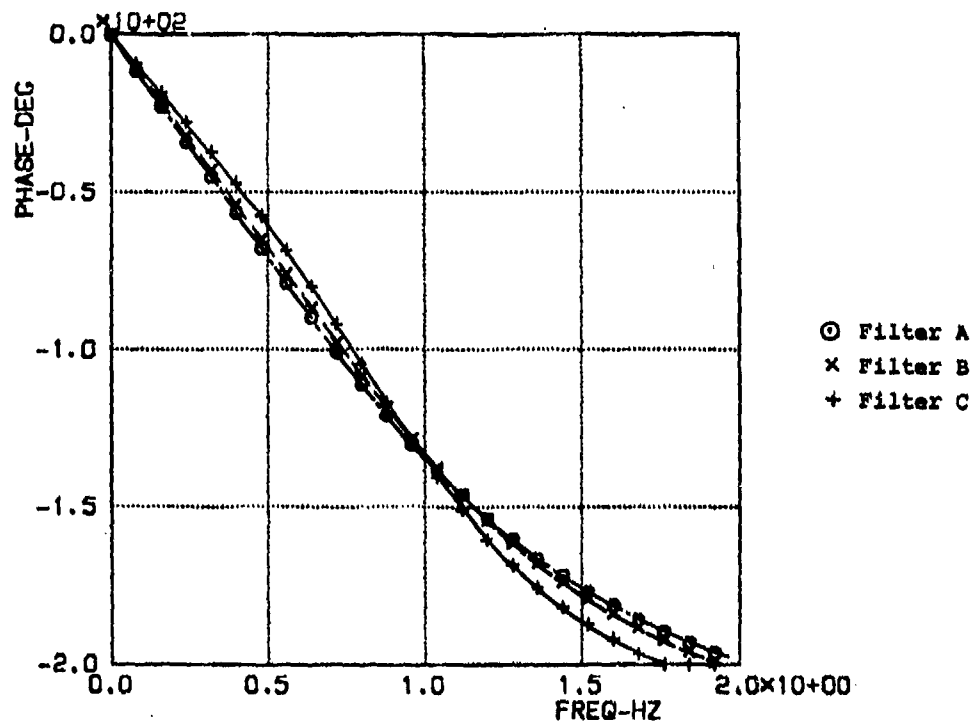
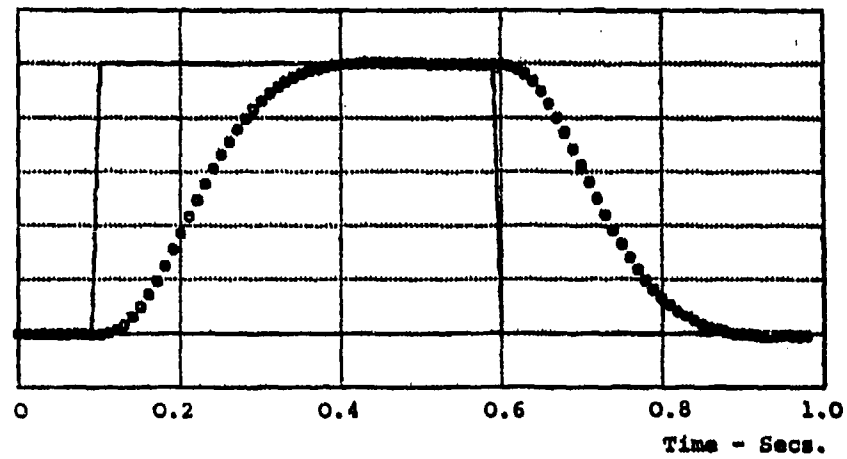
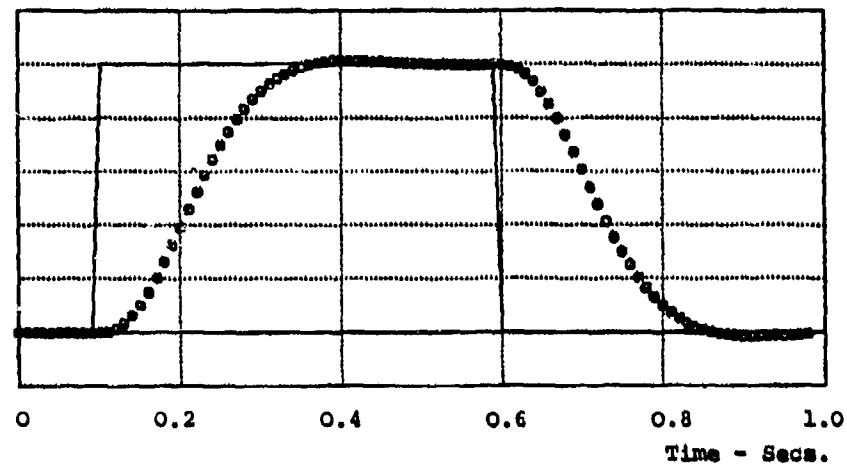


Figure 1.17. Phase Characteristic of Three Third-Order Filters.

Filter A.



Filter B



Filter C.

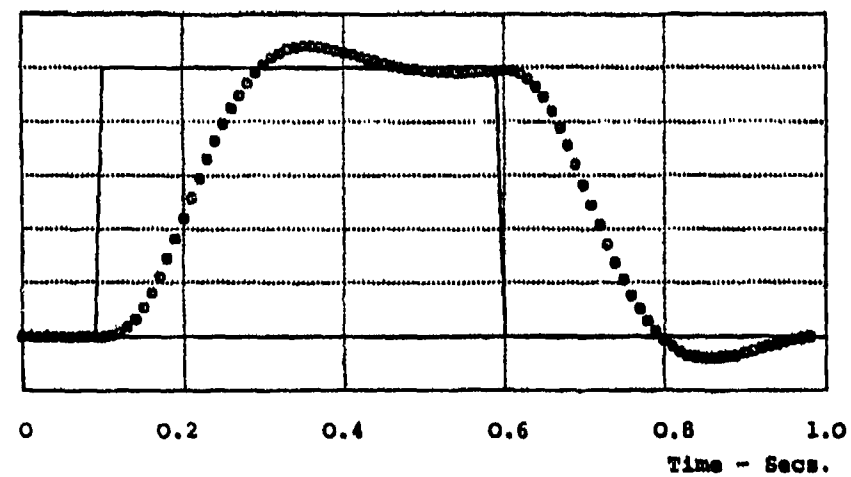


Figure 1.18. Step Responses of Three Third-Order Filters.

1.4.6. Sample/Hold Modules

A sample/hold module is a linear analogue device whose operation is controlled by a logic signal. When the logic signal is in one state the module is transparent, that is an analogue signal is transferred from input to output with no modification. When the logic signal changes state, however, the analogue output signal is "frozen" to the value obtaining at that time. The circuit diagram of a simple SH module is shown in figure 1.19, together with a diagram to illustrate its mode of operation.

Applications of the SH (Sample/Hold) module include:

- isolation of the input to an analogue-digital converter during conversion (see Section 1.4.7.)
- isolation of the output from a digital-analogue converter when the converter is settling to a new value
- aligning the inputs to an analogue multiplexor when inter-channel time delays are uncertain (see Section 1.4.4.)

It is presumed in the foregoing description that the operation of a SH module is ideal. Practical designs are characterised by a number of parameters:

- Settling time. This is the time required for the output voltage to mirror the input voltage within specified limits after the module has been released from the hold condition.
- Droop rate. The time rate of change of output voltage when the module is in the hold condition.
- Feed through. A measure of the proportion of the input voltage which is transmitted to the output voltage when the module is in the hold condition. Feed through is normally very small at low signal frequencies, but can be significant at higher frequencies.
- Aperture time. The time required for the switch to turn off after a hold command has been applied.
- Aperture Uncertainty time. The difference between maximum and minimum aperture times.

The first two parameters are of major importance when selecting (or designing) a SH module. This is because the requirements for reducing their values are conflicting: in general either a fast settling time or a low droop rate can be achieved. Thus a wide range of SH modules is available commercially to satisfy a variety of applications. It is, however, to be noted that the apparent settling time can be reduced to a very small value if two SH modules are used in parallel and the hold control is switched alternately between the two.

A suitable SH module may be selected for a particular application by estimating the maximum "hold" time, the minimum time between release of one hold condition and setting of the next, and the frequency bandwidth of the analogue signal(s). The required droop rate is then the required resolution divided by the maximum hold time. The required maximum settling time is the minimum sampling time, but this may be relaxed for low bandwidth analogue signals. In such cases, a pessimistic estimate of the maximum voltage change between one sample and the next may be obtained from:

$$(V_2 - V_1)_{\max} \propto 2.f_m.A.t_s \quad 1.4.11.$$

where f_m is the maximum frequency present in the signal

A is the maximum voltage of the signal

and t_s is the sum of the maximum hold time and the settling time.
(A good approximation for this time is often the sampling period.)

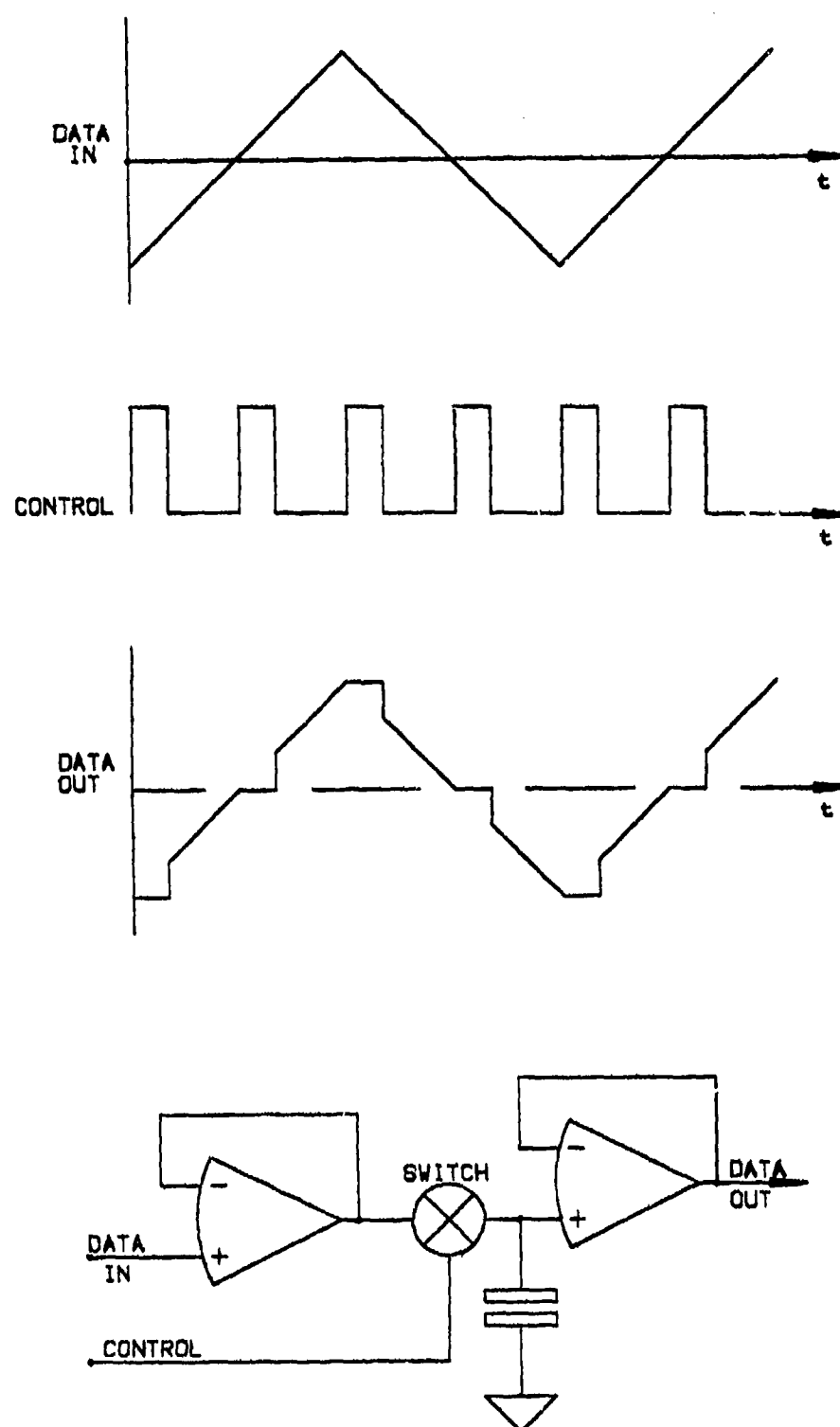


Figure 1.19. Schematic of a Sample-and-Hold Module, Illustrating its Method of Operation.

1.4.7. Analogue - Digital Converters

An analogue signal is converted to a sequence of digital samples by sampling the signal at regular intervals using a sample-and-hold module (see Section 1.4.6.) and by converting each sample into a digital value representing the voltage of the appropriate sample. The process of "digitising" an analogue sample is effected by an Analogue - Digital Converter (ADC).

The technique used for the conversion process depends upon the intended application, the parameters of interest being cost, accuracy and speed. Techniques which are used in ADCs which are available commercially include:

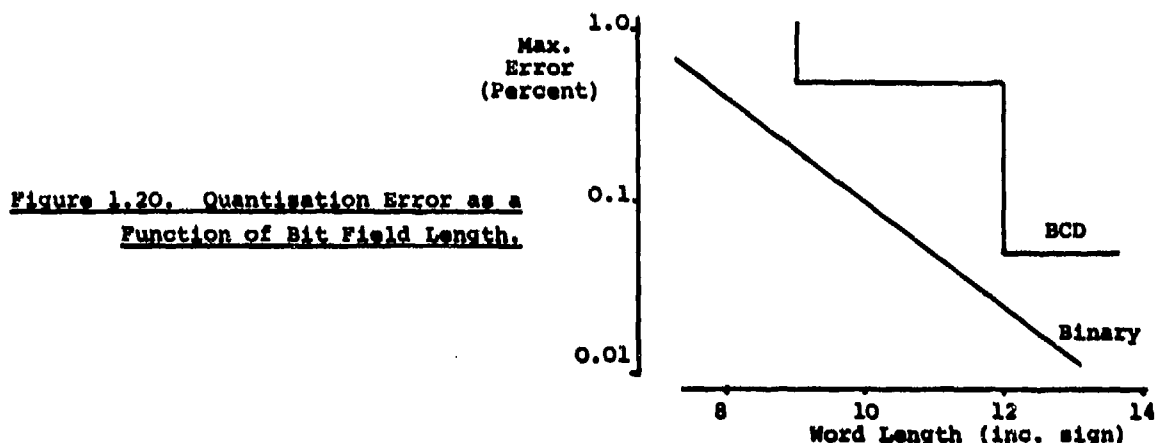
- Dual Slope Integration. Very slow, good noise rejection.
- Servo Counter. Cheap, relatively slow.
- Parallel. Very fast, expensive.

The dual slope integration technique (figure 1.21) compares the analogue signal with a reference voltage by integrating the input signal for a fixed period of time, and then switching a reference voltage to the integrator, counting the time required to reduce the output of the integrator to the start value. The time taken is proportional to the input voltage. The technique can be very accurate (it is used in digital voltmeters) but it is relatively expensive to implement and is very slow.

The servo counter (figure 1.22) compares the output of a digital-analogue converter (DAC) with the analogue signal and adjusts the input to the DAC until the two voltages are equal. The technique is the simplest and least expensive to implement but, unless a sophisticated strategy is used to determine the input to the DAC (successive approximation), it can be relatively slow.

An analogue signal may be digitised by comparing the signal with each of a number of reference levels. If each reference level corresponds to a transition point between one quantisation level and the next, then the conversion process can be effected very rapidly, the time required being the switching time of one comparator, and the time taken to convert the level thus determined to a digital output. This is the technique employed in a parallel (or "flash") ADC (figure 1.23). It is very fast and very expensive to manufacture, since a comparator is required for each quantisation interval.

The parameter of major importance in the selection of a suitable ADC is the resolution of the converted signal. The output of an ADC may take the form of a binary coded decimal (BCD) number or, more commonly, a "pure" binary number. Its resolution is limited by the number of bits (length) of the number, as shown in figure 1.20. It can be seen that an ADC producing a binary output always has a greater resolution than one producing BCD for a given number of bits.



It can be shown that the finite resolution of an ADC introduces an error into the sampled signal. This has an RMS value of:

$$\frac{\Delta V}{2\sqrt{3}} \quad \text{volts}$$

1.4.12.

where ΔV is the voltage interval between one quantisation level and the next. When a random signal is being sampled, then the RMS value of the signal should be no more than one third of the maximum voltage of the ADC. Under this favourable condition, the ratio of quantisation error to signal level has an RMS value:

$$\sqrt{3} \cdot 2^{-N}$$

1.4.13.

where N is the number of bits (including sign) output by a binary output ADC.

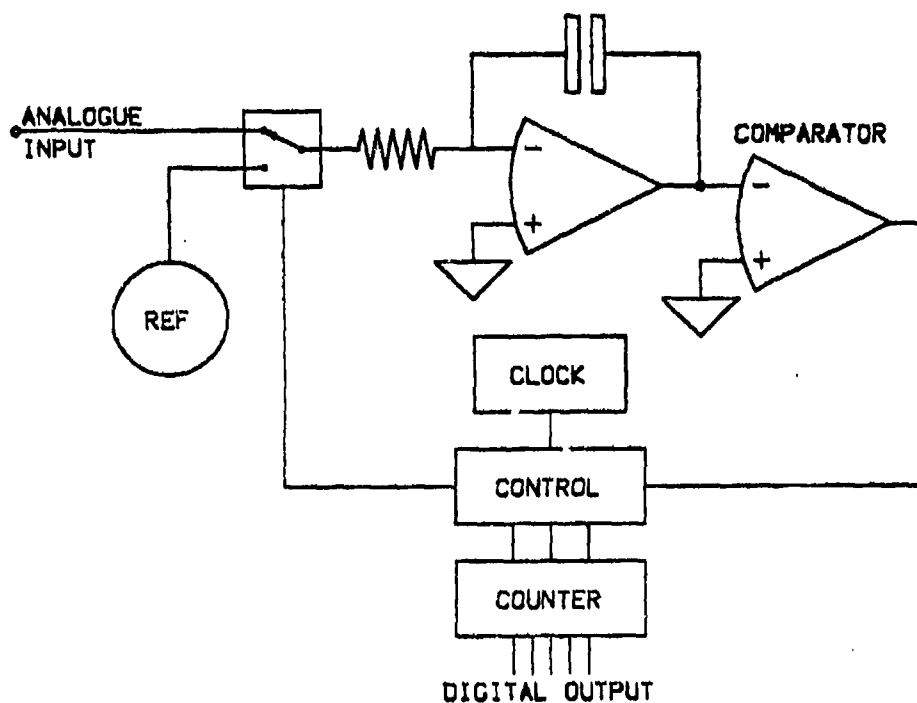


Figure 1.21. Schematic of a Dual Slope Integration ADC.

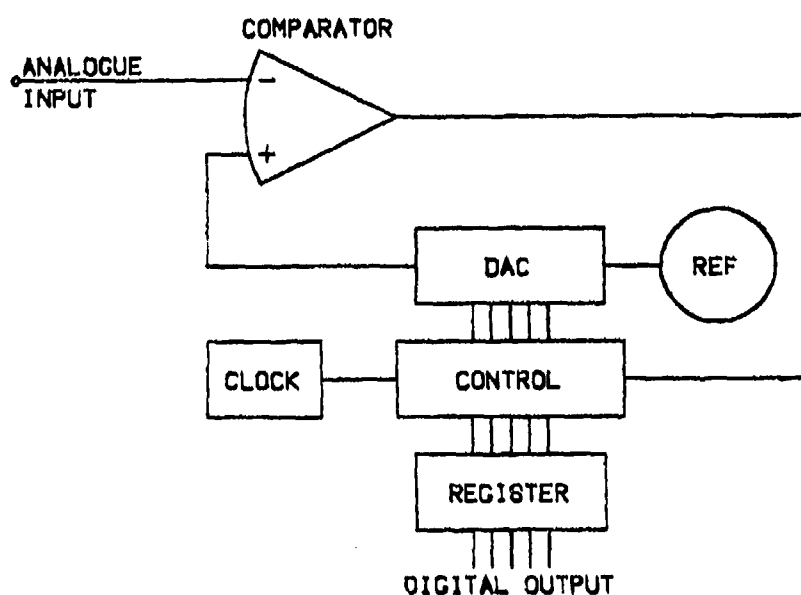


Figure 1.22. Schematic of Servo Counter ADC.

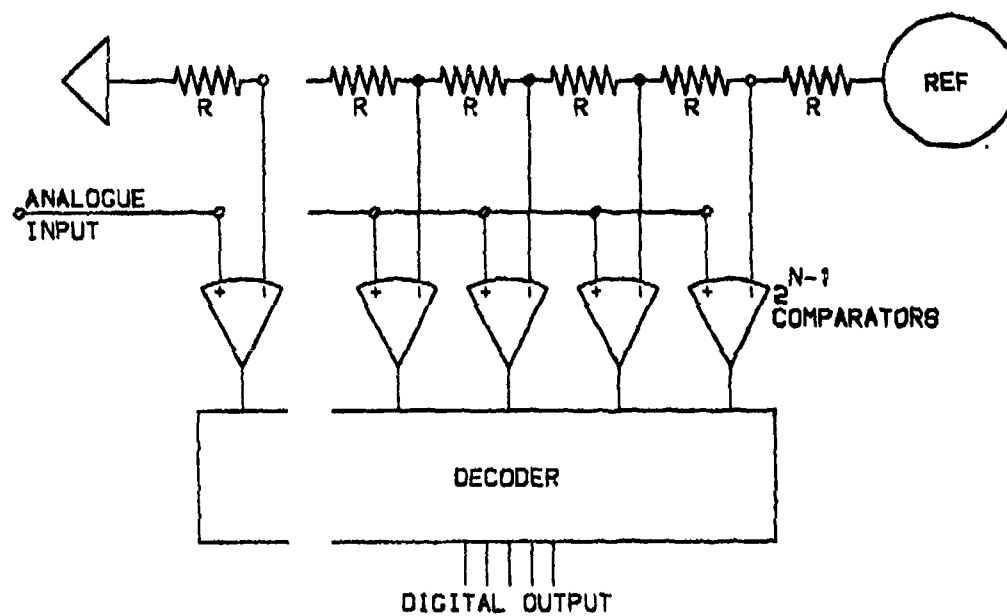


Figure 1.23. Schematic of a "Flash" Decoding ADC.

1.5 Digital Preprocessing

Digital sampled data sequences obtained from flight trials are sometimes characterised by:

- an inadequate sequence length
- an inappropriate acquisition rate
- the presence of unwanted signal components

An inadequate sequence length implies that no part of the sequence should be rejected on the grounds that unwanted components are present. An inappropriate acquisition rate may mean that analysing the sequence as it stands would be very expensive. The presence of unwanted signal components will contaminate processed results and may influence conclusions drawn from them.

Under these circumstances, it is imperative to have tools available for conditioning data sequences prior to processing proper. This Section describes certain commonly used processes for removing unwanted signal components. An introduction to the more general topic of digital filtering is included, and the Section ends by describing techniques for operating on a data sequence in order to reduce the time required for subsequent analysis.

1.5.1. Trend Removal

Random data sequences are sometimes superimposed upon slowly changing deterministic functions. It is normally essential to remove these functions prior to examining the nature of the random component in order to eliminate possible contamination of the results. An efficient method of removing a deterministic function is to identify its characteristics and then to subtract the function from the data sequence. This Section describes the techniques which may be used to identify the more important of these deterministic functions.

1.5.1.1. Removal of a Constant Level

It is assumed that the random component of the signal under investigation is stationary and has a zero average value. Then that random component may be estimated as:

$$\hat{y}(i) = y(i) - a \quad 1.5.1.$$

$$\text{Where} \quad a = \frac{1}{N} \sum_{i=1}^N y(i) \quad 1.5.2.$$

a is defined as the average value of y .

The estimate of the average value as defined by 1.5.2. will not, in general, be equal to the true average value $\bar{a}$, defined by 1.5.2. in the limit as $N \rightarrow \infty$. This is because its value will be affected by the presence of the random component. A detailed assessment of the uncertainty of mean value estimation is included in Appendix E. It is concluded that, if the random component can be approximated by band limited white noise (Section 2.3.) of bandwidth B Hz., then the standard error of the estimate of a is given by:

$$e_a = \frac{\sigma_y}{2BT} \quad 1.5.3.$$

where σ_y is the RMS (root mean square) value of the random component (see Section 2.3.) and T is the averaging time ($= Nh$). Equation 1.5.3. makes the logical statement that the averaging time required to achieve a given standard error is inversely proportional to the characteristic bandwidth of the random component.

It is important to consider the accuracy of the estimated average value a , because this will reflect upon the validity of the random component obtained from equation 1.5.1.

1.5.1.2. Removal of a linear trend

It is assumed that the data sequence has a zero average value (perhaps as a result of applying equation 1.5.1.) and that the random component under investigation is stationary and has no low frequency components which might be interpreted as a linear trend. Then a linear trend may be removed by computing:

$$\hat{y}(i) = y(i) - (i-m).b \quad 1.5.4.$$

$$\text{Where } b = \frac{3}{m(2m+1)(m+1)} \cdot \sum_{i=0}^{N-1} (i-m).y(i) \quad 1.5.5.$$

$$\text{and } m = \frac{N-1}{2} \quad 1.5.6.$$

As was the case for estimates of average value, estimates of the parameter defined in equation 1.5.5. will, in general, be subject to an error caused by the presence of the random component of the signal. The standard error of an estimate of b is given by:

$$\sigma_b = \left\{ \frac{3N}{2BT.m(2m+1)(m+1)} \right\}^{1/2} \cdot \sigma_y \quad 1.5.7.$$

where, as before, σ_y is the RMS value of the random component, B is the bandwidth of the (assumed band limited white) noise, and T is the averaging time.

1.5.1.3. Removal of an Exponential Function

It sometimes happens that the characteristics of a random signal are required when the measurement system is recovering from an overload condition. When the transducer is a piezo-electric accelerometer buffered by a charge amplifier, then the system recovery characteristic is exponential. For this reason the identification of an exponential function in the presence of a random signal has been chosen to illustrate techniques which may be used to obtain the characteristics of more general deterministic functions.

It is to be assumed that the random component of a signal under investigation is stationary and is defined as:

$$\hat{y}(i) = y(i) - A \cdot \exp(Bi) \quad 1.5.8.$$

The parameters A and B are to be determined. The first step is to calculate the rate of change of the function with respect to the parameters, that is:

$$\frac{\partial f}{\partial A} = \exp(Bi) = F(i) \quad , \text{ say} \quad 1.5.9.$$

$$\frac{\partial f}{\partial B} = iA \cdot \exp(Bi) = G(i) \quad , \text{ say} \quad 1.5.10.$$

The vector $y(i)$ is then minimised using the method of least squares by setting up the following linearised equations:

$$\begin{bmatrix} \sum F(i)^2 & \sum F(i).G(i) \\ \sum F(i).G(i) & \sum G(i)^2 \end{bmatrix} \cdot \begin{Bmatrix} A \\ B \end{Bmatrix} \approx \begin{Bmatrix} \sum F(i).(y(i) - A \cdot \exp(Bi)) \\ \sum G(i).(y(i) - A \cdot \exp(Bi)) \end{Bmatrix} \quad 1.5.11.$$

Equations 1.5.11. can be solved to determine ΔA and ΔB , incremental values which must be added to A and B to improve an initial estimate.

In order to use the technique, initial estimates for the parameters A and B are required and, since equations 1.5.11. use previous estimates of A and B to obtain updated values, the process is iterative. A test for convergence is to compare estimated mean square values of $y(i)$ (equation 1.5.8.) and to terminate the iteration when successive estimates differ by less than, say, one per cent.

Estimates for the standard errors of the parameters A and B can be obtained by computing the matrix:

$$\begin{bmatrix} \sum F(i)^2 & \sum F(i).G(i) \\ \sum F(i).G(i) & \sum G(i)^2 \end{bmatrix}^{-1} = \begin{bmatrix} K_1 & K_2 \\ K_3 & K_4 \end{bmatrix} \quad 1.5.12.$$

extracting the leading diagonal elements, and multiplying their square roots by the root mean square value of $y(i)$. Hence:

$$\sigma_A \approx \sigma_y \sqrt{K_1} \quad 1.5.13.$$

$$\text{and} \quad \sigma_B \approx \sigma_y \sqrt{K_4}$$

where, as before, σ_y is the RMS value of $\hat{y}(i)$.

The above process is quite general and may be used to define the parameters of any deterministic function, provided that the derivatives of the function with respect to the unknown parameters (equivalent to equation 1.5.9. and 1.5.10.) can be calculated.

It may be noted that the problem of identifying an exponential function could be linearised (and hence made non-iterative) by re-writing equation 1.5.8. as follows:

$$y(i) - \hat{y}(i) = A \cdot \exp(Bi) \quad 1.5.14.$$

and taking the logarithm of both sides to yield:

$$\ln(y(i) - \hat{y}(i)) = \ln A + Bi \quad 1.5.15.$$

Equation 1.5.15. may be solved to yield estimates of A and B by regression analysis. However difficulties are likely to be encountered when $(y(i) - \hat{y}(i))$ is close to zero, and the estimation problems described by equations 1.5.8. and 1.5.15. are different: the first implies that $\sum y^2(i)$ is to be minimised, and the second that $\sum (\ln(y(i) - \hat{y}(i)) - \ln A - Bi)^2$ is to be minimised. The first difficulty can be overcome by temporarily adding a known offset to all data points and restating the problem, but the second will give rise to biased estimates of the parameters A and B. Therefore, the formulation of equation 1.5.15. is not to be recommended as a technique for extracting an exponential function (at least for the present application.)

1.5.1.4. Removal of a Trend by Spline Fitting

When the "deterministic" (i.e. unwanted) component of a signal cannot easily be described by an explicit function, then it is tempting to consider using a polynomial function to describe the deterministic component. If the component has a complicated shape then high order polynomial functions may be necessary to describe the component adequately. Now, high order polynomials can be unstable unless their shape is controlled in some way. One method of stabilising a high order polynomial fit is to use splines.

The term spline originated from the engineer's tool which is widely used to interpolate smoothly between calculated points (eg constructing an aerofoil shape). The engineer's spline consists of a uniform flexible beam which is constrained to pass through the calculated points (nodes), but is not constrained (in general) in any other way (i.e. the slope of the beam at the nodes is not constrained). Clearly the shape taken up by the beam is stable, since it will obey the "minimum strain energy" principle.

Mathematicians have seized upon the principle of the engineer's spline and have generalised the concept, so that almost any function can now be defined and used as a "spline". The concept of "minimum energy" remains.

In this Section, the concept of spline fitting in the presence of noise is introduced by application of the traditional engineer's spline to the problem of removing a trend.

Consider a uniform beam, unloaded except by shear forces and moments applied to the ends as shown in figure 1.24.

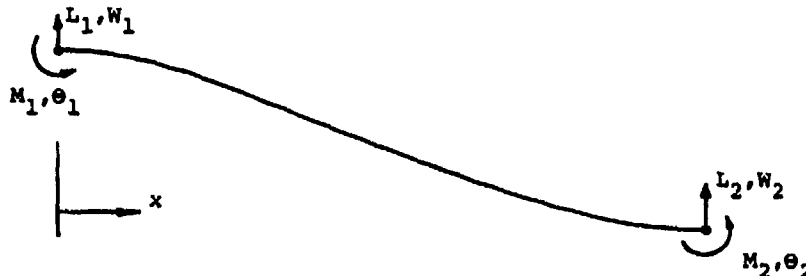


Figure 1.24.

The deflection anywhere along the beam may be determined if the applied loads (L_1, L_2) and moments (M_1, M_2) are known. Alternatively, the deflection may also be calculated if the deflections (W_1, W_2) and rotations (θ_1, θ_2) at either ends of the beam are known.

It may be observed that, with no applied loads except at the ends, the shear force (S) along the beam is constant. Since

$$S \propto \frac{d^3 w}{dx^3} = \text{a constant,}$$

it follows that:

$$w = a + bx + cx^2 + dx^3 \quad 1.5.16.$$

$$\text{and} \quad \theta = \frac{dw}{dx} = b + 2cx + 3dx^2 \quad 1.5.17.$$

Equations 1.5.16. and 1.5.17. define the deflection and the slope anywhere along the beam element once the coefficients have been determined. These can be expressed readily in terms of the node deflections, W_N , by applying suitable node conditions. These might be, for example,

$$\text{zero end slope (encastre):} \quad \theta = 0$$

$$\text{zero bending moment:} \quad \frac{d\theta}{dx} = 0$$

zero bending moment induced across the i th node:

$$i \left\{ \frac{d\theta}{dx} \right\}_{x=1} = i+1 \left\{ \frac{d\theta}{dx} \right\}_{x'=0}$$

where x refers to the i th beam element, and x' refers to the $(i+1)$ th element.

The first two conditions are relevant for the first and last nodes of a multi-node spline. The last can be applied at each of the intervening nodes. Thus, for example, if a spline having four equally spaced nodes and pin-jointed ends (zero moment) has deflections at each node W_1, W_2, W_3, W_4 and rotations $\theta_1, \theta_2, \theta_3, \theta_4$, then application of the appropriate end conditions gives:

$$\begin{bmatrix} -3 & 3 & 0 & 0 \\ -3 & 0 & 3 & 0 \\ 0 & -3 & 0 & 3 \\ 0 & 0 & -3 & 3 \end{bmatrix} \begin{Bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{Bmatrix} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{Bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{Bmatrix} \quad 1.5.18.$$

In general, the end rotations may always be expressed in terms of the node deflections using matrix equations having the form:

$$\begin{aligned} C.\theta_N &= D.W_N \\ \text{or} \quad \theta_N &= C^{-1}.D.W_N = E.W_N, \text{ say} \end{aligned} \quad 1.5.19.$$

The position of any point along the spline can be computed using equations having the form:

$$w(x) = A(x).W_N + B(x).\theta_N \quad 1.5.20.$$

where W_N and θ_N are the $(n \times 1)$ vectors of the n node deflections and rotations. $A(x)$ and $B(x)$ are $(1 \times n)$ row vectors in which it is understood that the elements are all null except for the two corresponding to the nodes at either end of the appropriate beam element (dependent upon the value of x .) Equation 1.5.20. may be evaluated for each of m locations along the beam, and the results can be listed as an $(m \times 1)$ vector:

$$w = F.W_N + G.\theta_N \quad 1.5.21.$$

Here F and G are $(m \times n)$ arrays, each row corresponding to one position along the spline.

Equation 1.5.21. can be used to interpolate between any set of coordinates. The validity of the result, however, depends directly upon the validity of the original information. An alternative application of equation 1.5.21. is to determine estimates of W_N given "noisy" values of w . This can be effected using the method of "least square errors", provided that the random component constituting the noise has a Gaussian distribution with a zero mean value, and provided that m is much greater than n . When this is the case, then equation 1.5.21. can be written:

$$w = (F + G.E).W_N \quad (\text{from equation 1.5.19.}) \quad 1.5.22.$$

$$\text{Hence: } (F + G.E)^T.w = (F + G.E)^T(F + G.E).W_N \quad 1.5.23.$$

$$\text{or} \quad W_N = [(F + G.E)^T(F + G.E)]^{-1}(F + G.E)^T.w \quad 1.5.24.$$

Once estimates of the node deflections have been determined using equation 1.5.24., then equation 1.5.21. can be used to isolate the random component of w .

Spline fitting can thus be used to identify the deterministic component embedded in random data by assuming that the random component is normally distributed about a zero mean, and, incidentally, that it has no significant energy in the bandwidth of the deterministic component. If these conditions are satisfied then the method of least squares can be used to estimate the node deflections which result in the "best" spline fit of the data once suitable positions for the nodes have been chosen. These may be chosen arbitrarily, or may be chosen to suite particular "events" in the deterministic component.

1.5.2. Digital Filtering

Where the characteristic of a digital data sequence is such that trend removal is not a practical proposition, then digital filtering can be used to remove undesirable signal components. Digital filtering is also used to develop practical integrators and differentiators, and in conditioning a data sequence prior to decimation (see Section 1.5.6.) and "zoom" spectral analysis (see Section 2.4.).

40

A great deal of theoretical work has been published on the subject of digital filtering (see, for example, references 1.6. and 1.7.), much of it involving complicated mathematical analysis. The purpose of this section is to introduce the main concepts of time domain digital filtering and to show how these concepts fit in with more esoteric design procedures.

1.5.2.1. First Order Low Pass Filter

An electrical "RC" first order low pass filter has the following characteristic equation:

$$T \dot{y} + y = f(t) \quad 1.5.25.$$

where $f(t)$ is the input signal

y is the output of the filter

and T is the Time Constant of the filter.

The task is to derive a simple transfer equation to implement 1.5.25. for a sampled data sequence $y(i)$.

Equation 1.5.25. can be re-written as follows:

$$\dot{y} = \frac{1}{T} \int_0^t (f(t) - y) . dt \quad 1.5.26.$$

This may be written in terms suitable for applying to a digital data sequence by presuming that the integration can be performed with sufficient accuracy using the "trapezoidal" rule, which states that the area contained within two sampled points is the average value of the two samples multiplied by the distance (or time) between them. Thus:

$$\int_0^T y . dt \approx \frac{h}{2} \left[y(0) + 2 \sum_{i=1}^{N-1} y(i) + y(N) \right] \quad 1.5.27.$$

where h is the sampling interval, and T translates to $h.N$ seconds. Hence equation 1.5.26. becomes

$$y(N) = \frac{h}{2T} \left[f(0) - y(0) + 2 \sum_{i=1}^{N-1} (f(i) - y(i)) + f(N) - y(N) \right] \quad 1.5.28.$$

$$\text{or } y(N) = \frac{h}{2T} \left[f(0) - y(0) + 2 \sum_{i=1}^N (f(i) - y(i)) - f(N) + y(N) \right] \quad 1.5.29.$$

$$\text{but } y(N-1) = \frac{h}{2T} \left[f(0) - y(0) + 2 \sum_{i=1}^{N-1} (f(i) - y(i)) - f(N-1) + y(N-1) \right] \quad 1.5.30.$$

subtracting 1.5.30. from 1.5.28. and rearranging gives:

$$y(N) = \frac{1 - h/2T}{1 + h/2T} y(N-1) + \frac{h/2T}{(1 + h/2T)} (f(N-1) + f(N)) \quad 1.5.31.$$

This is a recursion formula for implementing a low pass filter in the time domain. It is efficient in so far as only the current and one previous time period values are needed to compute the current output (information about time periods prior to those are contained implicitly in $y(N-1)$). The recursion formula given by 1.5.31. is identical in form to that obtained in the literature using very elegant and complicated transformations. It is not always obvious from these that the fundamental assumption has been made that trapezoidal integration is sufficient for computing the output of the filter (discussions on this aspect may be found in Reference 1.6.).

Equation 1.5.31. may be converted into a practical integration formula by multiplying the f terms by the time constant T (c.f. equation 1.5.25.). This then becomes:

$$y(N) = \frac{1 - h/2T}{1 + h/2T} y(N-1) + \frac{h/2}{(1 + h/2T)} (f(N-1) + f(N)) \quad 1.5.32.$$

In the limit as the time constant tends to infinity, equation 1.5.32. becomes:

$$y(N) = y(N-1) + \frac{h}{2} (f(N-1) + f(N)) \quad 1.5.33.$$

which is the standard result for trapezoidal integration. Equation 1.5.32. is better behaved than 1.5.33. when experimental data are to be integrated because of the problems introduced by spurious offsets in y (see section 1.5.1.1.). Any offset in the data will cause the integral as defined by 1.5.33. to "run away", whereas the offset will simply be multiplied by T if equation 1.5.32. is used, once the filter has stabilised.

1.5.2.2. Second-Order Low-Pass Filter

A recursive algorithm for implementing a second order low pass filter in the time domain may be developed using a method similar to that described for a first order low pass filter (and including the same assumptions).

Assuming the general characteristic:

$$\frac{\ddot{y}}{(2\pi f_0)^2} + \frac{2\zeta \dot{y}}{2\pi f_0} + y = f(t) \quad 1.5.34.$$

(i.e. a unity gain filter with natural frequency f_0 Hz. and damping factor ζ), then the following recursive algorithm may be derived:

$$y(N) = K_1 y(N-1) + K_2 y(N-2) + K_3 (f(N) + 2f(N-1) + f(N-2)) \quad 1.5.35.$$

Where:

$$\begin{aligned} K_1 &= \frac{2(1 - (2\pi f_0 h/2)^2)}{(1 + \zeta 2\pi f_0 h + (2\pi f_0 h/2)^2)} \\ K_2 &= \frac{-(1 - \zeta 2\pi f_0 h + (2\pi f_0 h/2)^2)}{(1 + \zeta 2\pi f_0 h + (2\pi f_0 h/2)^2)} \\ K_3 &= \frac{(2\pi f_0 h/2)^2}{(1 + \zeta 2\pi f_0 h + (2\pi f_0 h/2)^2)} \end{aligned} \quad 1.5.36.$$

Equation 1.5.35. may be converted into a practical double integration algorithm by dividing the factor K_3 by $(2\pi f_0)^2$ (see equation 1.5.34.) The function resulting from application of this algorithm is the true double integral of the data sequence (within the limitations imposed by the trapezoidal rule), but filtered by a second order high pass filter having a natural frequency f_0 Hz. and a damping factor ζ .

Equation 1.5.35. may be converted to a true double integration algorithm by dividing K_3 by $(2\pi f_0)^2$ and then setting $f_0 = 0$. This yields,

$$y(N) = 2y(N-1) - y(N-2) + \frac{h^2}{4} [\ddot{y}(N) + 2\ddot{y}(N-1) + \ddot{y}(N-2)] \quad 1.5.37.$$

which is the standard result for double integration assuming trapezoidal integration.

1.5.2.3. The General Form for Digital Filters

The results of section 1.5.2.1. and 1.5.2.2. are two examples of the more general algorithm:

$$y(N) = f(N) + \sum_{i=1}^Q \{ b(i) \cdot y(N-i) + c(i) \cdot f(N-i) \} \quad 1.5.38.$$

where Q is the order of the filter. The coefficients may be selected to yield low pass, band pass, or high pass filters to any order.

1.5.3. Transducer Correction Filters

Many types of transducer possess the secondary (but not necessarily undesirable) characteristic of filtering the signals being sensed. Examples are linear and angular accelerometers, pressure transducers and rate gyros. When this proves to be an embarrassment, and when the characteristics of the signal are to be described in the frequency domain, then the effects of the filter(s) can be removed most easily after the frequency domain description has been obtained. The situation is rather more complicated when other types of analysis are of interest. When such is the case, then it may be desirable to remove the effects of transducer (and any other) filters in the time domain prior to analysis.

In principle, the objective of time domain filter correction is to subject the signal in question to a synthesised filter having precisely the "inverse" properties of the unwanted signal filter. As an illustration, suppose that the transducer acted as a second order low pass filter having the form

$$\frac{x}{(2\pi f_0)^2} + \frac{2\zeta}{2\pi f_0} \cdot x + x = y(t) \quad 1.5.39.$$

where $y(t)$ represents the unfiltered signal, and x represents the signal as output by the transducer. It is clear that y can be reconstructed provided that the first and second derivatives of x can be computed, and provided also that the natural frequency (f_0) and damping coefficient (ζ) of the transducer are known.

One method of estimating the first and second derivatives of a sample data sequence, using only "past" data is to construct a parabola passing through the current point and the previous two points. The parabola may then be differentiated twice to yield

$$\dot{x}(i) = \frac{1}{2h} [3 \cdot x(i) - 4 \cdot x(i-1) + x(i-2)] \quad 1.5.40.$$

$$\text{and} \quad \ddot{x}(i) = \frac{1}{h^2} [x(i) - 2 \cdot x(i-1) + x(i-2)] \quad 1.5.41.$$

These can be substituted into equation 1.5.39. to give, after some rearrangement,

$$\begin{aligned} y(i) = & x(i) \cdot \left[1 + \frac{3\zeta}{2\pi f_0 h} + \frac{1}{(2\pi f_0 h)^2} \right] \\ & - 2 \cdot x(i-1) \cdot \left[\frac{2\zeta}{2\pi f_0 h} + \frac{1}{(2\pi f_0 h)^2} \right] \\ & + x(i-2) \cdot \left[\frac{\zeta}{2\pi f_0 h} + \frac{1}{(2\pi f_0 h)^2} \right] \end{aligned} \quad 1.5.42.$$

A few words of caution are necessary at this point. In the majority of cases, signals presented for processing are corrupted by "noise" (spurious signal components) which has been introduced after the transducer. When, as is usual, the bandwidth of the noise extends to frequencies where the transducer output is attenuated significantly, then components in this region will be amplified when the signal is corrected by the method outlined above.

When the presence of injected noise causes problems, then a less ambitious procedure may be adopted in which it is assumed that the acceleration term in equation 1.5.39. can be neglected (the term becomes insignificant at very low frequencies). Here, individual samples can be assumed to be connected by straight lines to yield:

$$\dot{x}(i) = \frac{1}{h} [x(i) - x(i-1)] \quad 1.5.43.$$

which, from equation 1.5.39., yields the following correction formula:

$$y(i) = x(i) \cdot \left[1 + \frac{2\zeta}{2\pi f_0 h} \right] - x(i-1) \cdot \frac{2\zeta}{2\pi f_0 h} \quad 1.5.44.$$

The examples outlined above are intended to illustrate the approach to be followed when it is necessary to correct data in the time domain for the effects of filtering introduced during the measurement process. It must be emphasised that the approach is fraught with difficulties, some of which have been noted, so that the technique should be considered only as a last resort.

1.5.4. Optimal Filters

The filters described in the preceding Sections, in general, make no explicit use of any knowledge which may exist about the processes which gave rise to the signals to be processed. When such knowledge does exist, and appropriate additional measurements are available, then the various components of a signal can be separated in a satisfying, and often highly reliable, way using the concept of optimal filters.

The simple example shown diagrammatically in figure 1.25. can be used to illustrate the principle of optimal filtering. It is assumed here that the measured output $y(t)$ consists of a random component $n(t)$ added to a deterministic component. The deterministic component is the function $u(t)$, filtered by the system having a transfer function H . If the function $u(t)$ can be measured at the same time as the signal $y(t)$, and the system transfer function is known precisely, then the deterministic component of $y(t)$ can be computed and subtracted directly from $y(t)$ to yield $n(t)$.

Practical complications arise when the parameters of the transfer function, H , are not known accurately. It is then necessary to apply the measurements to the task of estimating at least some of the parameters of H . This can be achieved, under certain circumstances, by minimising the mean square value of the random component estimated by subtracting the computed output of the system from the measured output $y(t)$.

The principle of optimal filters was originally published in 1960 (reference 1.8.) and was first applied to the solution of complex navigation problems. However it is now being used in a wide variety of applications, and publications on the subject are extensive. References 1.8. to 1.11. describe typical applications in aircraft flight testing.

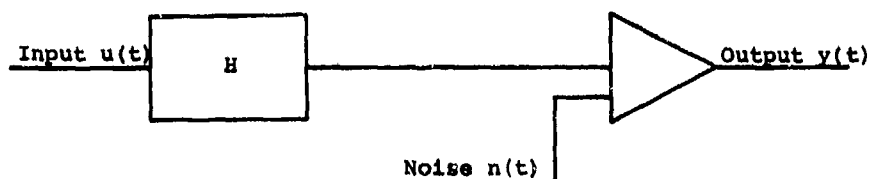


Figure 1.25.

1.5.5. Interpolation

A sequence of data samples is normally considered to represent discrete values of a function at corresponding values of an independent variable (for example, time or frequency.) The samples themselves may, or may not, be considered to be exactly representative of the function, but the independent variable is almost always assumed to be known exactly. The latter assumption is required for the majority of techniques described in this manual.

Interpolation is the process in which a number of consecutive data samples is used to estimate the value of the underlying function at one or more prescribed values of the independent variable. Examples in which interpolation may be required are:

- aligning samples obtained from the various channels of a sequentially sampled digital measurement system
- modifying the frequency interval between successive estimates of a frequency response function.

The process may involve the conscious removal of unwanted signal components (noise) by filtering. In any case, interpolation requires that a suitable function be chosen which describes the behaviour of a selected group of samples.

It may be demonstrated (see the comments made in Section 1.4.4.) that an infinite number of functions exist which describe precisely a particular sequence of samples. As a consequence no exact, or even "best", solution to the interpolation problem generally exists unless the form of the required function can be specified from physical considerations. A corollary to this statement is that the selection of an interpolation function will automatically impose constraints upon the results obtained from the process.

The purpose of this Section is to present several interpolation techniques and to discuss their relative merits. The selection is by no means exhaustive.

1.5.5.1. Linear Interpolation

It is assumed in this case that the signal alters its value linearly between each pair of samples. Clearly this is not a particularly satisfactory assumption because it implies that the derivative of the signal takes the form of a histogram, as shown in figure 1.26 with an abrupt change of value at each sample point.

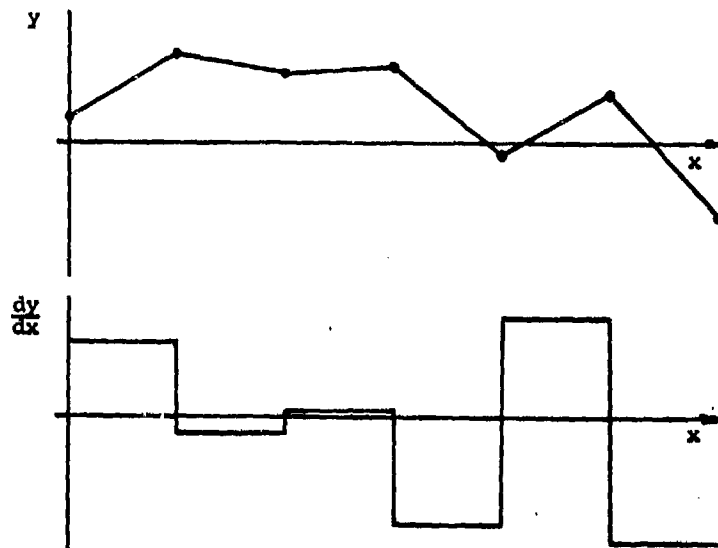


Figure 1.26. A Typical Signal and its Derivative Implied by the Assumption of Linear Interpolation.

It is required to find the value of the signal y at a point x given two samples y_0 at x_0 and y_1 at x_1 , as shown in figure 1.27, assuming that a straight line connects the two samples.

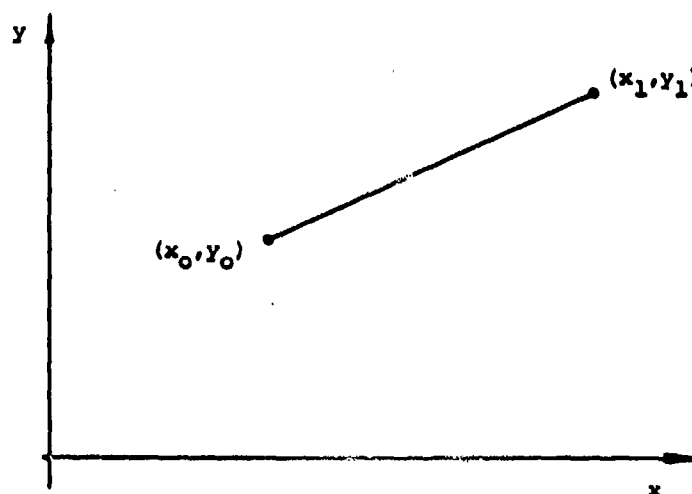


Figure 1.27.

The slope of the line is given by:

$$\frac{dy}{dx} = \frac{y_1 - y_0}{x_1 - x_0} \quad 1.5.45.$$

Hence:

$$\begin{aligned} y &= \frac{dy}{dx} \cdot (x - x_0) + y_0 \\ &= \frac{(y_1 - y_0) \cdot (x - x_0)}{(x_1 - x_0)} + y_0 \end{aligned} \quad 1.5.46.$$

When the interval between samples of a data sequence $y(i)$ is h , and it is required to generate a new data sequence $y(i)$, each element of which is delayed by $k \cdot h$ ($k-1$) with respect to the original sequence, then equation 1.5.46 can be used to derive:

$$y(i) = (1 - k) \cdot y(i) + k \cdot y(i + 1) \quad 1.5.47.$$

This is the relationship which is most commonly used to align multiplexed sequences in which the multiplexing process introduces known time delays between one channel and another. It should be noted, however, that the alignment can be effected in the frequency domain if the data are to be described in that way. For example, if the frequency response function (FRF) of one sequence relative to another has been estimated (see Section 2.4.3.), and the time delay of the first relative to the second is kh , then the phase of the N -length FRF should be corrected as follows (see Appendix A.2.2.):

$$\theta(i) = \theta(i) + k \cdot 360 \cdot \frac{1}{N} \quad i = 0(1)N-1 \text{ degrees} \quad 1.5.48.$$

The modulus of the FRF is unchanged.

1.5.5.2. Quadratic Interpolation

It is to be assumed here that the interpolation function has the general form:

$$y = a + bx + cx^2 \quad 1.5.49.$$

In order for this to be useful as an interpolation function it is necessary, in general, for the values of the parameters a , b , and c to vary between successive pairs of samples, or at least between groups of three samples, the minimum number required to compute the parameters.

A quadratic interpolation function which passes through three sets of coordinates (x_0, y_0) , (x_1, y_1) and (x_2, y_2) , as shown in figure 1.28., has parameters which are given by:

$$a = y_0 - b \cdot x_0 - c \cdot x_0^2 \quad 1.5.50.$$

$$b = \frac{(x_1^2 - x_2^2) \cdot (y_0 - y_1) - (x_0^2 - x_1^2) \cdot (y_1 - y_2)}{(x_1^2 - x_2^2) \cdot (x_0 - x_1) - (x_0^2 - x_1^2) \cdot (x_1 - x_2)} \quad 1.5.51.$$

$$c = \frac{(x_1 - x_2) \cdot (y_0 - y_1) - (x_0 - x_1) \cdot (y_1 - y_2)}{(x_1 - x_2) \cdot (x_0^2 - x_1^2) - (x_0 - x_1) \cdot (x_1^2 - x_2^2)} \quad 1.5.52.$$

The apparent complexity of these equations may be reduced considerably by assuming a constant interval of h between successive samples, and by replacing x by $(x - x_0)$. Then the expressions for the parameters become:

$$a = y_0 \quad 1.5.53.$$

$$b = \frac{1}{2h} \cdot (4 \cdot y_1 - 3 \cdot y_0 - y_2) \quad 1.5.54.$$

$$c = \frac{1}{2h^2} \cdot (y_0 + y_2 - 2 \cdot y_1) \quad 1.5.55.$$

so that:

$$y = y_0 + \frac{k}{2} \cdot (4 \cdot y_1 - 3 \cdot y_0 - y_2) + \frac{k^2}{2} \cdot (y_0 + y_2 - 2 \cdot y_1) \quad 1.5.56.$$

where

$$k = \frac{x - x_0}{h} \quad (0 \leq k \leq 2) \quad 1.5.57.$$

An alternative formulation may be derived by replacing x in equation 1.5.49. by $(x - x_1)$, rather than $(x - x_0)$. This yields the following expression for y :

$$y = y_1 + \frac{k}{2} \cdot (y_2 - y_0) + \frac{k^2}{2} \cdot (y_0 - 2 \cdot y_1 + y_2) \quad 1.5.58.$$

where

$$k = \frac{x - x_1}{h} \quad (-1 \leq k \leq 1) \quad 1.5.59.$$

The method of interpolation described above has the unsatisfactory characteristic that the derivative of the function is, in general, discontinuous at alternative samples, as shown in figure 1.29. (curve A.)

This characteristic may be improved somewhat by using a degenerate form of the spline described in Section 1.5.1.4. In this application the spline is to be fitted between 'n' successive samples, is to behave in a quadratic manner between each pair of samples, and is to have a derivative which is continuous for all values of the independent variable within boundaries implied by the group of selected samples. The form of the function is again to be:

$$y = a + bx + cx^2 \quad 1.5.60.$$

and $\theta = \frac{dy}{dx} = b + 2cx \quad 1.5.61.$

Here, the values of the parameters are defined if the deflection (y) and slope (θ) of the spline at each node (sample) are known.

The deflections are, of course, the values of the samples themselves, whilst, for a 5 node spline with samples spaced h apart, the slopes are given by:

$$h \cdot \begin{Bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{Bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & -1 & 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{Bmatrix} \quad 1.5.62.$$

An interpolated value may then be calculated as:

$$y = y(i) + (h \cdot \theta(i)) \cdot k + (\theta(i+1) - \theta(i)) \cdot \frac{hk^2}{2} \quad 1.5.63.$$

where $k = \frac{x - x(i)}{h}, \quad x(i) \leq x \leq x(i+1) \quad 1.5.64.$

The effect of using a quadratic spline is shown in figure 1.29. (curve B). It will be observed that, for the example presented, the quadratic spline generates a function which deviates considerably between samples.

1.5.5.3. Cubic Interpolation.

The rather unsatisfactory characteristics of quadratic interpolation may be improved considerably if cubic interpolation is employed, particularly if the spline described in Section 1.5.1.4. is used. For a time history, the cubic spline ensures that acceleration (the second derivative of the function) is continuous and varies linearly between one sample and the next. In the absence of other information, the cubic spline has considerable aesthetic appeal (see figure 1.29., curve C), but it should be recalled that the results produced by its application are only estimates, in general, and the additional computation required as compared with, say, linear interpolation may not be justified.

For the record, an interpolation algorithm based upon the cubic spline and interpolating between samples y(i) and y(i+1) using samples y(i-1) to y(i+2) inclusive can, for equally spaced samples, be obtained from the following:

$$y(k) = [1, k, k^2, k^3] \cdot [A] \cdot \begin{Bmatrix} y(i-1) \\ y(i) \\ y(i+1) \\ y(i+2) \end{Bmatrix} \quad 1.5.65.$$

where: $A = \frac{1}{15} \begin{bmatrix} 0 & 15 & 0 & 0 \\ -7 & -3 & 12 & -2 \\ 12 & -12 & 3 & -3 \\ -5 & 0 & 0 & 5 \end{bmatrix}$

1.5.66.

and where k is as defined in equation 1.5.64.

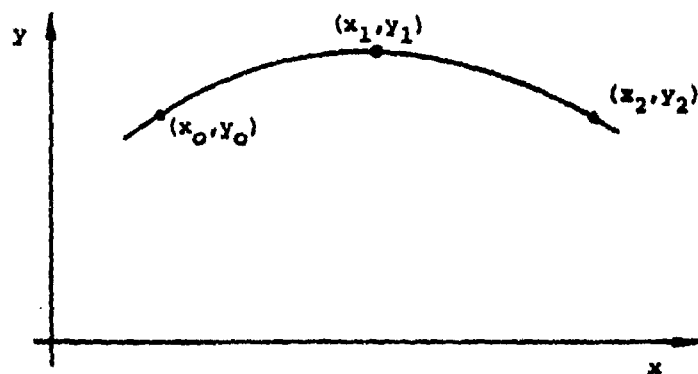


Figure 1.28

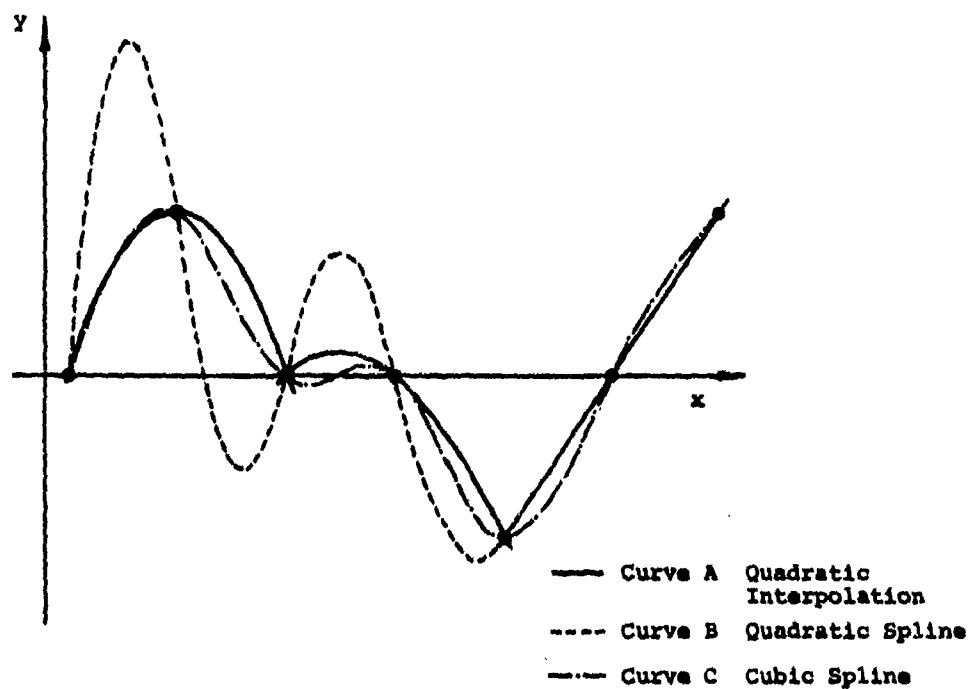


Figure 1.29 Illustration of Three Types of Interpolation

1.5.6. Decimation

The term decimation is used to describe the process in which the interval between adjacent samples of a data sequence is increased. The sequence resulting from decimation has fewer samples than the original. Decimation may, therefore, be used when it is required to:

- reduce a data sequence to a more manageable length,
- adjust the time interval between samples of a time domain sequence so that the interval between estimates is more appropriate after the sequence has been transformed into the frequency domain,
- perform "octave band" frequency domain analysis.

The effect of decimation is to reduce the folding frequency of the sequence (see Section 1.4.) It therefore follows that, if aliasing is to be avoided, any frequency components which might lie above the new folding frequency must be removed before decimation is attempted.

The tools required to effect the necessary filtering and, possibly, interpolation have been introduced in the preceding Sections.

CHAPTER 2 ANALYSIS TECHNIQUES

2.1. Introduction

The purpose of this Chapter is to introduce the techniques more commonly encountered when analysing random data. The techniques are explained at a heuristic level and, where this is possible, the mathematical specifications of the techniques are developed from this explanation. Mathematical development of the specifications from statistical principles is relegated to Appendices or, where this is considered to be beyond the scope of the Volume, references to publications which contain the appropriate development are included. Also included are expressions which allow estimates to be made of the reliability of results obtained from each technique.

2.2. Fundamental Concepts

Experimental data is frequently categorised as being either deterministic or random. In an ideal world there should be no problems in deciding whether data obtained from a trial lies in one category or in the other: deterministic data is obtained from experiments in which all the variables are measured and/or controlled, whereas random data is obtained from experiments in which certain variables cannot be measured and therefore cannot be controlled. In the former case it is possible to fit functional relationships to the data on a point-by-point basis, but useful information can be extracted in the latter case only by averaging the data in an appropriate way.

Of course, the real world is by no means ideal and data obtained from flight trials contains, in general, both deterministic and random components. The reasons for this are that it is not possible to control all the variables which may contribute to the measurements and that even if this were so, no element in a measurement system performs exactly in the manner it was designed to do. In this situation flight test data is often categorised as being deterministic or random as much according to the manner in which it is to be analysed as to the character of the data itself.

This Volume is devoted to the analysis of random data which, in the light of the above statements, means that it is concerned with ways by which the relevant characteristics of the data may be expressed in terms of averages computed from the data. Such averages may be more or less representative of the phenomenon under investigation, and, where this is in doubt, the trials should be organised so as to ascertain whether or not this is the case. Moreover, the precision to which averages can be estimated depends upon the averaging time and upon the properties of the signal being averaged. No average should be considered to be complete without an estimate of this precision. The remainder of this Section is devoted to the properties of random signals and, in general terms, to the implications of those properties.

2.2.1. Stationary Random Processes

A formal definition of a stationary process is included in Appendix D. The definition involves ensemble averages which are notionally computed from sample measurements extracted from a large number of independent trials records at a particular instant in time. The process is said to be stationary if the expected values of the averages do not depend upon the chosen instant in time. Two "degrees" of stationarity are identified:

- A weakly stationary process is one in which the ensemble mean value and the ensemble auto-correlation function are each independent of time.
- A strongly or strictly stationary process is one in which all ensemble moments and joint moments are independent of time.

Strictly stationary processes therefore represent a subset of weakly stationary processes. It is normal to presume that a process which is demonstrably weakly stationary is also strictly stationary.

An alternative, and subtly different, concept of stationarity involves sample averages (computed from a single trial) rather than ensemble averages. The definitions of sample stationary processes are similar to those for ensemble stationary processes, with the term "sample" replacing "ensemble" wherever it occurs. In engineering circles, the use of the term stationary normally implies sample stationarity. This is because the hypothesis of sample stationarity generally is easier to substantiate than the hypothesis of ensemble stationarity.

The ensemble view of stationarity is equivalent to the sample view for most practical purposes, although the values to be expected from estimates of the identifying averages may be different.

Pressure measurements made in the vicinity of the intersection of the horizontal and vertical tail surfaces of a T-tailed aircraft can be used to illustrate the concept of stationarity.

For the purposes of the illustration it is assumed that the pressure measurements exhibit a D.C. component (a non-zero mean value) which is proportional to tailplane setting angle (function of aircraft c.g. position, equivalent airspeed, Mach Number) and a random component caused by boundary layer pressure fluctuations. In straight and level flight the power spectral density (Section 2.4.1.) is assumed to be proportional to tailplane setting angle, equivalent airspeed, and Mach Number. Several possibilities can be considered:

- a) Samples obtained at arbitrary times during a number of flights. No constraints have been placed upon the controlling parameters, so that the process is non-stationary.
- b) Samples obtained only at nominated values of equivalent airspeed and Mach Number. No constraint has been placed upon tailplane setting angle, so that the process is non-stationary.
- c) Samples obtained only at nominated values of equivalent airspeed and Mach Number. It is further assumed that the c.g. position remains constant during a flight, but not between one flight and another. The process, for a particular equivalent airspeed and Mach Number, is stationary. However the sample statistics will vary between one flight and another, so that the identifying averages for the ensemble view will be different from those for the sample view.
- d) Samples obtained only at nominated values of equivalent airspeed, Mach Number, and tailplane setting angle. The process, for particular parameter values is stationary. Further, the statistical averages obtained from one flight will be representative of all flights.

It is interesting to note that, if the power spectral density of the random component of pressure did not depend upon tailplane angle, then cases b and c would have been equivalent to case d provided that the D.C. components of the measurements had been removed prior to analysis.

The practise of selecting samples of a random process according to the values of a number of controlling parameters frequently results in a stationary ensemble. When this occurs, the process is more properly referred to as conditionally stationary. The majority of random processes encountered in the real world fall into this category, provided that the appropriate controlling parameters can be identified and measured or controlled.

2.2.2. Ergodic Random Processes

Conditionally stationary random processes which have the characteristic that the expected values of sample averages are equal to the expected values of the corresponding ensemble averages are termed ergodic random processes or, more properly, conditionally ergodic random processes.

The example given in Section 2.2.1. illustrates that a conditionally stationary process can often be regarded as a conditionally ergodic process, but that it may be necessary to impose additional constraints upon sample selection. Incidentally, case d of that example is a conditionally ergodic process and, with the D.C. component removed from the measurements, so are cases b and c if the power spectral density of the random component is independent of tailplane setting angle.

The fact that a process is conditionally ergodic (and therefore conditionally stationary) means that ensemble characteristics can be estimated from sample characteristics, so that a complete description of the process can, in principle, be obtained from a single flight trial.

2.2.3. Normal Distribution

Many of the techniques developed in this Chapter are founded upon the assumption that the amplitude probability distribution is Normal (or Gaussian), the properties of which are given in Appendix D. In a great many practical applications this is a valid assumption, so far as can be ascertained by measurements, and is supported by the Central Limit Theorem (see Appendix D.). This states that, under quite general conditions, the sum of a number of mutually independent variables will have an amplitude probability distribution which tends to the normal distribution as the number of variables becomes large.

A signal which has a Normal amplitude probability distribution and is demonstrably weakly Stationary, is also strongly stationary because the amplitude probability distribution is completely described by the mean and RMS values.

2.3. Time Domain Analysis Techniques

Many descriptions of the characteristics of random data can be obtained by performing simple operations on the data. Such descriptions are generally of the single parameter variety (for example, mean and mean square values) but may include descriptions which are a function of signal amplitude (such as amplitude probability density functions). At a more complicated level, descriptions may be obtained which are functions of relative time (e.g. autocorrelation functions). All of these functions either can be computed relatively simply from the time histories of the data, or are themselves intrinsic functions of time. Descriptions of the techniques which may be considered to fall into this category are referred to as Time Domain Analysis Techniques and are the subject of this Section.

2.3.1. Mean and Mean Square Averages

The mean and mean square values represent the most rudimentary descriptions of a random signal.

The mean value is simply the average value of the signal, i.e.

$$\mu_y = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T y(t) \cdot dt \quad 2.3.1.$$

or, for an N length sampled data sequence,

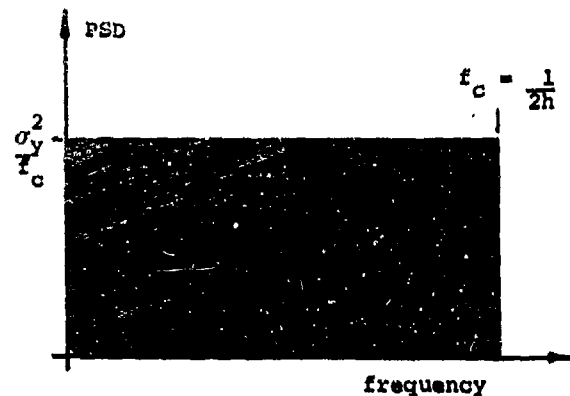
$$\mu_y = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N y_i \quad 2.3.2.$$

In practical terms, N must remain finite, so that equation 2.3.2. reduces to:

$$\hat{\mu}_y = \frac{1}{N} \sum_{i=0}^N y_i \quad 2.3.3.$$

The difference between 2.3.2. and 2.3.3. constitutes an estimation error. Such an error is, in its own right, a random signal whose properties depend upon the value of N and the way in which y_i varies as N is changed. In particular, the variance of the error depends upon the variance and the power spectral density of y_i and upon the averaging time (See Appendix E). An estimate of the error is often obtained by assuming that y_i represents a Normal distribution white noise sequence band limited at the folding frequency, and having the correct variance (see figure 2.3.1.).

Figure 2.3.1. Assumed
Distribution of Power for
Estimating Errors.



When this is the case, the variance of the error is given by:

$$\text{Var}(\hat{\mu}_Y) = \frac{\sigma_Y^2}{N} \quad 2.3.4.$$

where σ_Y^2 is defined by equation 2.3.6.

The mean square value is given by:

$$\Psi_Y^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Y^2(t) dt \quad 2.3.5.$$

Quite clearly, the quantity defined by 2.3.5. is dependent upon the mean value of the data sequence. It is often convenient to use a quantity related to 2.3.5., but with the mean value removed.

This is termed the variance and is defined by:

$$\sigma_Y^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (Y(t) - \mu_Y)^2 dt \quad 2.3.6.$$

or the sampled sequence equivalent:

$$\sigma_Y^2 = \lim_{N \rightarrow \infty} \frac{1}{(N-1)} \sum_{i=0}^N (Y_i - \mu_Y)^2 \quad 2.3.7.$$

which may be written, after some manipulation,

$$\sigma_Y^2 = \lim_{N \rightarrow \infty} \frac{1}{(N-1)} \sum_{i=0}^N Y_i^2 - \mu_Y^2 \quad 2.3.8.$$

The practical form of this, where N remains finite is given by:

$$\hat{\sigma}_Y^2 = \frac{1}{(N-1)} \sum_{i=0}^N Y_i^2 - \frac{N}{N-1} \mu_Y^2 \quad 2.3.9.$$

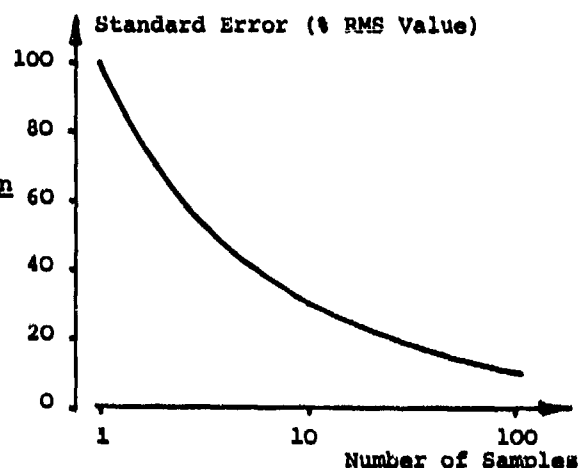
The variance of the difference between 2.3.9. and 2.3.8. is obtained from Reference 2.2., as:

$$\text{Var}(\hat{\sigma}_Y^2) = \frac{2\sigma_Y^4}{(N-1)} \quad 2.3.10.$$

A number of points are raised by these results:

1. The definition of mean square value contains the divisor $(N - 1)$. The latter could have used the divisor N , but this would have resulted in a "biased" estimate, see reference 2.3.
2. In certain of the equations namely 2.3.4., 2.3.9., and 2.3.10., true mean values, rather than estimates (denoted by the symbol " $\wedge$ ") are required. Of course routines implementing these relationships must use the estimated values, but this will result in higher uncertainties in the parameters concerned.
3. The variance of the estimates is, in each case, a function of the variance of the data and the power spectral density of the data. The simple relationships of 2.3.4. and 2.3.10. are based on the assumption of Normal probability distribution and a uniform power spectral density. If more than average energy in the data is concentrated at low frequencies, then the relationships as stated will underestimate the variance errors.
4. The variances of the estimated errors (equations 2.3.4. and 2.3.10.) indicate that in each case, the RMS of the error (standard error) falls increasingly slowly as averaging time is increased: doubling the averaging time decreases the standard error by only 30 percent (see, for example, figure 2.3.2.)

Figure 2.3.2. Standard Error of Mean Value Estimates as a Function of Number of Samples, N.



2.3.2. Amplitude Probability Distributions

Amplitude probability distributions are an important property of random signals. In practise they are usually computed in order to justify (or otherwise) the assumption of Normality.

The foundations of probability theory, as applied to random, stationary time histories are developed in Appendix C. The Appendix defines the amplitude probability and amplitude probability density functions for a single random signal, and shows how the various central moments of the signal can be used to describe its characteristics. The Appendix continues to discuss the properties of the Normal distribution, the transformation of probability densities, and ends with an introduction to joint probability density functions.

The central function of interest is the amplitude probability density distribution. This can be implemented very easily for an unscaled sampled data sequence, as follows:

- a) Set up a null $N \times 1$ array, where $N = 2^p$ and where p is the number of bits used to convert the signal.
- b) For the i^{th} sample in the data sequence y_i , compute:

$$k = y_i \left[+ \frac{N}{2} \right]$$

2.3.11.

where y_i is the value of the i^{th} sample. Dependent upon the hardware used, one of the following two statements will hold:

$$0 < y_i < N$$

2.3.12.

$$-\frac{N}{2} < y_i < \frac{N}{2}$$

2.3.13.

If, and only if, the latter statement is true then the component shown in brackets in equation 2.3.11. should be included. It therefore follows that

$$0 < k < N$$

2.3.14.

- c) Add one to the $(k + 1)$ th element of the $N \times 1$ array.
- d) Repeat b) and c) for each sample of the M -length data sequence.
- e) Divide the array by M .

The resulting $N \times 1$ array represents an amplitude probability density estimate normalised so that the value of the interval between one quantisation level and the next is unity. It can be converted to an actual APD by dividing each element in the array by the physical value represented by the quantisation interval. A flow diagram of the above is shown in figure 2.3.3.

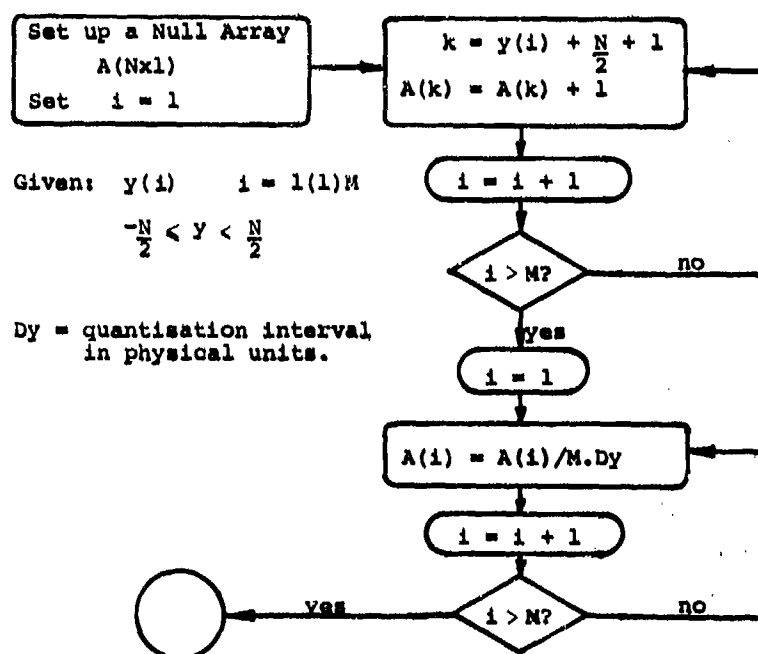


Figure 2.3.3. Flow Diagram for Amplitude Probability Analysis.

Estimates of APD functions obtained from the foregoing algorithm are subject to two types of error. The first is a statistical uncertainty which is a function of the character of the signal, the interval between estimates (quantisation interval in the "raw" state) and the averaging time. Specifically, the variance of this uncertainty when the signal has white noise characteristics over a limited bandwidth B Hz is approximately (Reference 2.2.):

$$\text{Var}(p(y)) \approx \frac{p(y)}{2B \cdot T \cdot Dy} \quad 2.3.15.$$

where Dy is the interval between successive estimates and T is the averaging time in seconds. It should be noted that the reliability of the estimates decreases as the value of the APD decreases.

The second source of error is a bias error which may be introduced because the probability density estimate for a particular level is obtained by grouping together values which fall into a finite interval around that level (typically, the quantisation interval). It is demonstrated in Reference 2.2. that an approximation to the bias error is given by:

$$b(p(y)) \approx \frac{(Dy)^2}{24} \cdot p''(y) \quad 2.3.16.$$

$$\text{where} \quad p''(y) = \frac{d^2 p(y)}{dy^2}$$

The following observations may be found to be useful:

- The quantisation interval may be too fine to give reliable estimates of probability density (depending upon the length of data sequence available). When this is the case, the statistical errors may be reduced (at the expense of increased bias error, see equation 2.3.16) by integrating across a number of intervals.
- An illuminating presentation of APD functions can be obtained when a Normal distribution is expected by plotting $\log(p(y))$ against $y \cdot \text{mod}(y)$. When the distribution is Normal, a triangular function is obtained (see figure 2.3.4.).
- The APD function can be used to estimate the RMS value of a signal even when the measurement/acquisition system is heavily overloaded. This is achieved by assuming that the original signal has a Normal distribution, that the mean value is low, and that the measurement system does not change the distribution (essentially, no latch-ups). Under these conditions the maximum achieved value of the APD is given by:

$$P_{\max}(y) = \frac{1}{\sigma_y \cdot \sqrt{2\pi}} \quad 2.3.17.$$

$$\text{or} \quad \sigma_y \approx \frac{0.4}{P_{\max}(y)} \quad 2.3.18.$$

- The amplitude probability function can be obtained by computing the running integral of the APD, viz:

$$P(k \cdot Dy) = Dy \cdot \left\{ p(0) + 2 \sum_{i=1}^{k-1} p(i) + p(k) \right\} \quad k = 0(1)N-1 \quad 2.3.19.$$

The statistical moments of the signal may be obtained by computing the weighted running integral of the APD, for example

$$M_n = (Dy)^{n+1} \left\{ \frac{N}{2} [(-1)^n \cdot p(0) + p(N-1)] + 2 \sum_{i=1}^{N-2} \left(\frac{k-N}{2} \right)^n \cdot p(i) \right\} \quad 2.3.20.$$

where M_n is the n th statistical moment. The results obtained for higher moments should be treated with caution, however, because the value of M_n depends increasingly upon less reliable information as n is increased (the outlying values of APD becoming increasingly important, see equation 2.3.20).

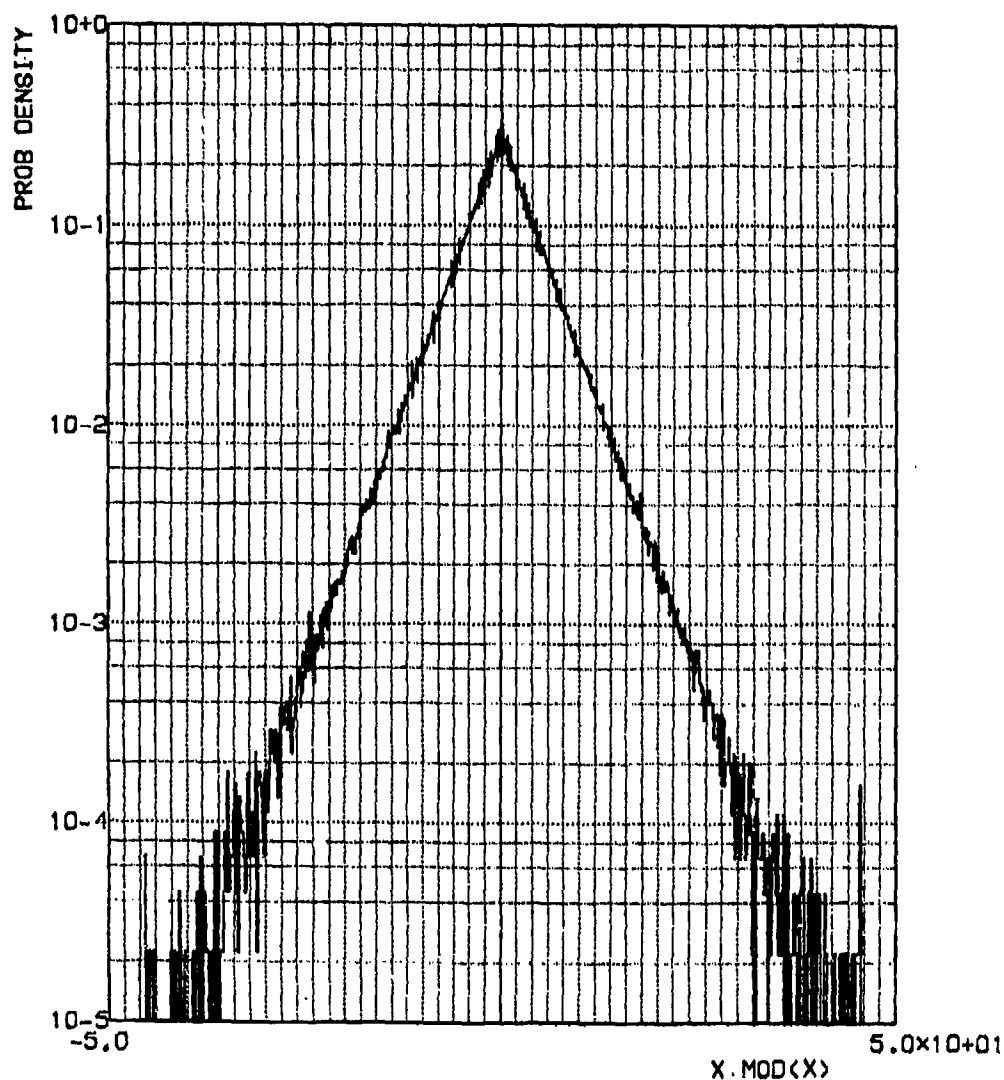


Figure 2.3.4. A typical Amplitude Probability Density Distribution

2.3.3. Probability Density Distribution of Peaks

The distribution of maxima (or minima) in a sampled data sequence may be of interest to structural engineers who are concerned with the problems of assessing the fatigue life of a structure. This is particularly so when the data represents structural strain or loads applied to a structure.

The usual algorithm for computing estimates of peak distributions assumes that each data point in the sampled sequence represents a maximum or minimum value. Thus the data is assumed to have been compressed by removing all data elements which do not constitute a turning point. The algorithm proceeds in a manner reminiscent of that used for computing amplitude probability density distributions:

- a) Set up a null $N \times 1$ array, where $N = 2^p$ and where p is the length of the converter field.
- b) For each even sample (if the sequence starts with a minimum) or for each odd sample (if the sequence starts with a maximum) compute:

$$k = y_1 \left[+ \frac{N}{2} \right] \quad 2.3.21.$$

where y_1 is the value of the appropriate sample. Observe that:

$$0 < y_1 < N \quad 2.3.22.$$

$$\text{or} \quad -\frac{N}{2} < y_1 < \frac{N}{2} \quad 2.3.23.$$

The bracketed component shown in equation 2.3.21. should be included only if the latter statement is true. Thus:

$$0 < k < N \quad 2.3.24.$$

- c) For each computed value of k , add one to the $(k + 1)^{\text{th}}$ element of the $N \times 1$ array.
- d) Divide the array so obtained by the total number of peaks to obtain the Peak Probability Density distribution.

A flow diagram of the algorithm is shown in figure 2.3.5. Estimates of the Peak probability density distribution are subject to both statistical and bias errors in just the same way as Amplitude Probability Density distributions. By comparison with 2.3.15. and 2.3.16., and under the same assumptions, the variance of the statistical error is given by:

$$\text{Var} [P_m(y)] = \frac{P_m(y)}{B.T.K.Dy} \quad 2.3.25.$$

where K is the "compression" factor.

whilst the bias error is given by:

$$b [P_m(y)] = \frac{Dy^2 \cdot \ddot{P}_m(y)}{24} \quad 2.3.26.$$

$$\text{where} \quad \ddot{P}_m(y) = \frac{d^2 P_m(y)}{d^2 y}$$

Here again, the statistical reliability of the estimates may be improved, at the expense of an increased bias error, by integrating across a number of intervals.

Given: $y(i)$, $i = 1(1)M$

$$-\frac{N}{2} < Y < \frac{N}{2}$$

Dy = Quantisation interval in physical units.

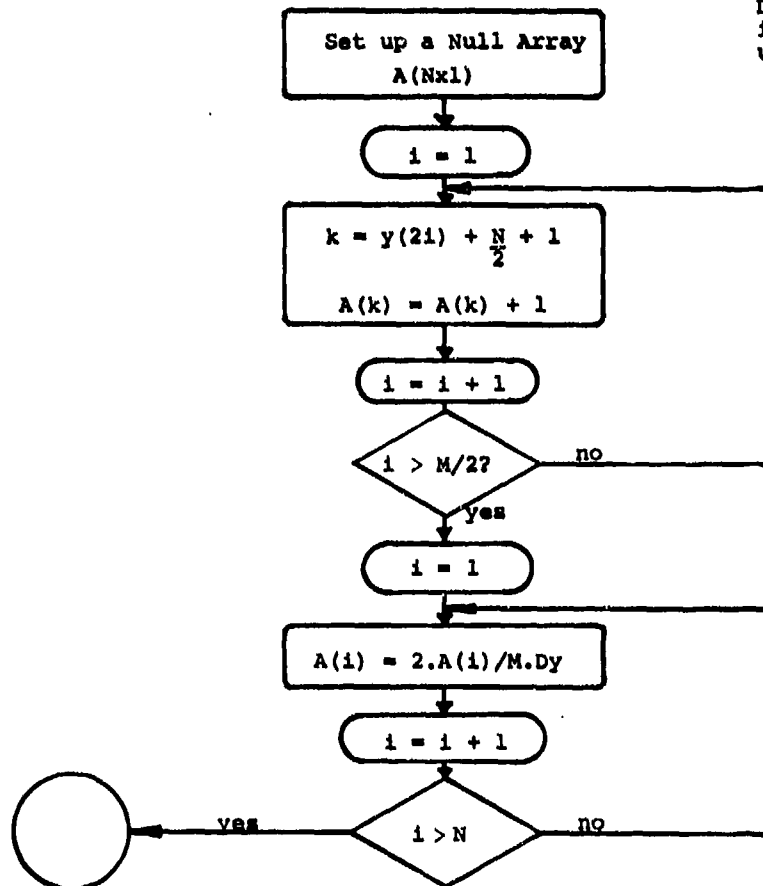


Figure 2.3.5. Flow Diagram for Peak Probability Density Analysis.

2.3.4. Range/Mean Distributions

Range/Mean Distributions are intended specifically to provide strain or load information for evaluating fatigue damage and/or estimating the life of a structure.

The algorithm for estimating the distribution is a direct extension of the one for estimating Probability Density distributions of peaks. In this case the Density array is an $N \times M$ matrix, with the N rows devoted to equally spaced values of range and the M columns devoted to equally spaced values of mean value. Here again the algorithm operates upon compressed data, each element indicating a maximum or minimum value. It works as follows, commencing with a null array of dimensions $N \times M$:

a) For the $2i^{\text{th}}$ sample compute:

$$j = |\hat{y}(2i) - \hat{y}(2i + 1)| \quad 2.3.27$$

$$k = \frac{\hat{y}(2i) + \hat{y}(2i + 1)}{2} \quad \left[+ \frac{N}{2} \right] \quad 2.3.28.$$

where the bracketed term is included if:

$$-\frac{N}{2} < \hat{y}_1 < \frac{N}{2} \quad 2.3.29.$$

- b) add one to the element in the $(j + 1)^{th}$ row, and $(k + 1)^{th}$ column of the range/mean matrix.
- c) repeat for all pairs of samples in the data sequence.
- d) divide the range/mean matrix by the number of pairs of samples to obtain the bivariate probability density distribution.

The statistical reliability of the estimates may be improved, at the expense of increasing the bias error by integrating across a number of intervals.

The results produced can be affected seriously by the presence of low amplitude noise. For this reason it is normal to reject reversals having a lower range than a nominated "gate" value. Even so, as pointed out in Reference 2.4., the Range/Mean Distribution provides a poor estimate of fatigue damage except in special circumstances.

2.3.5. Rainflow Analysis

Rainflow analysis is designed to provide a better estimate of Range/Mean Distributions than the method described in Section 2.3.4. The principal objective is to include the low frequency, large amplitude reversals as well as the more frequent low amplitude occurrences. This is achieved by "remembering" each peak until it is equalled or exceeded by another, and only counting a range (pair) when a closure (complete reversal) is obtained.

Algorithms for implementing Rainflow Analysis can be rather complicated and may require large quantities of computer storage. The most elegant implementation known to the author is detailed in Reference 2.5., to which the interested reader is referred.

2.3.6. Autocorrelation Functions

The concept of using the mean square value to describe a stationary signal can be extended logically to one in which the mean value is computed of the product of the signal and the same signal, but delayed by a specified time. The purpose of this is to characterise the manner in which the current value of the signal depends upon its values at previous points in time. If the relative time delay is varied, then a description of this characteristic can be compiled which is a function of the time delay. This function is known as the Autocorrelation Function. Specifically,

$$R_{yy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T y(t) \cdot y(t + \tau) \cdot dt \quad 2.3.30.$$

A more rigorous derivation of the autocorrelation function is included in Appendix A, which also states the connection between the Autocorrelation function and the Auto-covariance function.

It may be observed that, when the time delay τ is zero, then equation 2.3.30. degenerates to an expression for the mean square value, (c.f. equation 2.3.5.). Hence:

$$R_{yy}(0) = \sigma_y^2 \quad 2.3.31.$$

It turns out that $R_{yy}(\tau)$ is never greater than $R_{yy}(0)$.

Also, a simple change in variable ($t' = t + \tau$) in equation 2.3.30. leads to the fairly obvious, but important, conclusion that:

$$R_{yy}(\tau) = R_{yy}(-\tau) \quad 2.3.32.$$

Thus the Autocorrelation function can be said to be an even function of time delay.

It is reasonable to assert that, for a random signal having a zero mean, future values of the signal will depend to a decreasing extent upon the current value. More formally, the Autocovariance function,

$$K_{yy}(\tau) \rightarrow 0 \quad \text{as } \tau \rightarrow \infty \quad 2.3.33$$

Further, the Autocorrelation function of a periodic signal will also be periodic. This follows from the definition of a periodic signal, i.e.

$$y(t) = y(t + T) \quad 2.3.34.$$

where T is the period of the signal.

Thus the Autocorrelation function provides a mechanism for filtering out the periodic components in a signal (where the mean value is here seen as a periodic signal having an infinite period). It should be noted that the periodic component of a signal will not, in general, have the same waveform as the periodic component of its Autocorrelation function; the periodic component of the latter will always be such that its various frequency components will be in phase.

Estimates of the Autocorrelation function of a signal can be obtained by using an analogue analyser (figure 2.3.6.) or by operating upon an N -length sampled data sequence representing the signal. In the latter case, equation 2.3.30. is most commonly mechanised by using an approximation to the trapezoidal rule for integration, which gives:

$$R_{yy}(kh) = \frac{1}{(N-K)} \sum_{n=k}^N y(n) \cdot y(n-k) \quad 2.3.35$$

For reliable results, the maximum value of k should be less than ten percent of N .

Estimates of the Autocorrelation function may also be obtained using the Discrete Fourier Transform, a method which can be more efficient, computationally, than the "direct" method of equation 2.3.35. Details of this are presented in Reference 2.7.

When the Autocorrelation function of a signal is very complicated it is sometimes useful to elucidate matters by filtering the data selectively prior to computing the Autocorrelation function. This may be effected upon a sampled data sequence by digital filtering (section 1.5.), or, when the Discrete Fourier Transform is used, by weighting the Fourier coefficients. When it is required to filter the data a number of times, then the latter technique is by far the more efficient.

Autocorrelation estimates computed using equation 2.3.35 are subject to a statistical error which depends upon the characteristics of the data and the length of the data sequence (averaging time). The variance of this statistical error may be estimated by assuming the data to have white noise characteristics, band limited at a frequency of B Hertz. It is shown in Appendix E. that this results in:

$$\text{Var}\{R_{YY}(\tau)\} = \frac{1}{2.B.T} \cdot (\sigma_Y^4 + R_{YY}^2(\tau)) \quad 2.3.36.$$

It is also demonstrated in Appendix E that the bias error introduced by equation 2.3.35, is zero. This is not the case for estimates obtained using the Discrete Fourier Transform, and for this reason the direct method is to be recommended for estimating Autocorrelation functions despite its computational inefficiency.

2.3.7. Cross Correlation Functions

The Autocorrelation function provides a description of the way in which the current value of a signal depends upon its values at previous points in time. When the joint properties of the two signals are of interest, it is reasonable to extend this concept to obtain a description of the manner in which the current value of one signal depends upon the values of the other signal at various points in time. Formally, this idea may be expressed as:

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \cdot \int_0^T x(t) \cdot y(t - \tau) \cdot dt \quad 2.3.37.$$

This constitutes a definition of the Cross Correlation function. The expression is derived in a more rigorous way in Appendix D. It can be shown by making the substitution $t' = t - \tau$ in equation 2.3.37., that:

$$R_{xy}(\tau) = R_{yx}(-\tau) \quad 2.3.38.$$

It may be reasoned that the behaviour of one signal depends upon the other in a different way, in general, for negative time delays than it does for positive time delays. The Cross Correlation function is therefore an asymmetric function of time delay. It is sometimes useful to compute the symmetric and anti-symmetric components of the function separately. These are, respectively, as follows:

$$\begin{aligned} A(\tau) &= \frac{1}{2} \cdot [R_{xy}(\tau) + R_{xy}(-\tau)] \\ B(\tau) &= \frac{1}{2} \cdot [R_{xy}(\tau) - R_{xy}(-\tau)] \end{aligned} \quad 2.3.39.$$

or:

$$\begin{aligned} A(\tau) &= \frac{1}{2} \cdot [R_{xy}(\tau) + R_{yx}(\tau)] \\ B(\tau) &= \frac{1}{2} \cdot [R_{xy}(\tau) - R_{yx}(\tau)] \end{aligned} \quad 2.3.40.$$

As is the case for the Autocorrelation function, it can be shown that:

$$R_{xy}(\tau) \rightarrow 0 \quad \text{as } \tau \rightarrow \infty \quad 2.3.41.$$

for most cases when $y(t)$ is a random signal.

Estimates of the Cross Correlation function can be obtained by using an analogue analyser (figure 2.3.7.) or by operating upon an N-length sampled data sequence representing the signal. Since, however, the function is an asymmetric function of time delay, a complete description requires both positive and negative delays, or a reversal of the inputs. For a sampled data sequence, the most commonly used algorithm is based upon:

$$R_{xy}(kh) = \frac{1}{N-k} \cdot \sum_{n=k}^N x(n) \cdot y(n-k) \quad 2.3.42.$$

and

$$R_{yx}(kh) = \frac{1}{N-k} \sum_{n=k}^N y(n) \cdot x(n-k) \quad 2.3.43.$$

Once again, reasonably reliable results can be obtained if the maximum value of k is less than ten percent of N .

Estimates of the Cross correlation function may also be obtained using the Discrete Fourier Transform. This route is here even more efficient relative to the direct method than was the case for Autocorrelation functions. Details of the method are presented in Reference 2.7.

Statistical errors associated with estimates obtained by using equations 2.3.42. and 2.3.43. are similar to those for the Autocorrelation function. In this case, under the same assumption of white noise band limited at B Hertz, Appendix E gives the variance of error as:

$$\text{Var}\{R_{xy}(\tau)\} = \frac{1}{2 \cdot B \cdot T} (\sigma_y^2 \cdot \sigma_x^2 + R_{xy}^2(\tau)) \quad 2.3.44.$$

It is also shown that no bias is introduced by equations 2.3.42 and 2.3.43. This is not true for estimates obtained using the Discrete Fourier Transform and this method, despite its efficiency is not to be recommended.

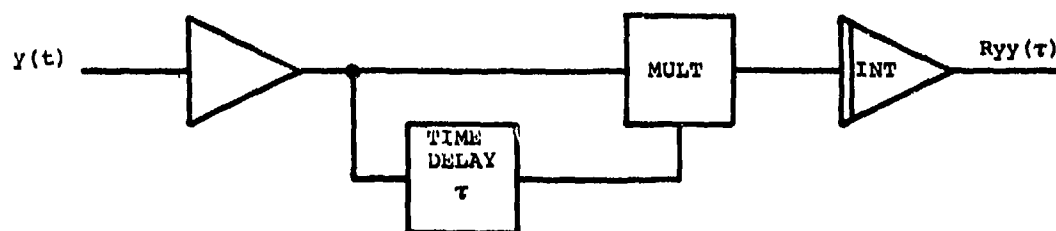


Figure 2.3.6. A Simple Analogue Autocorrelator.

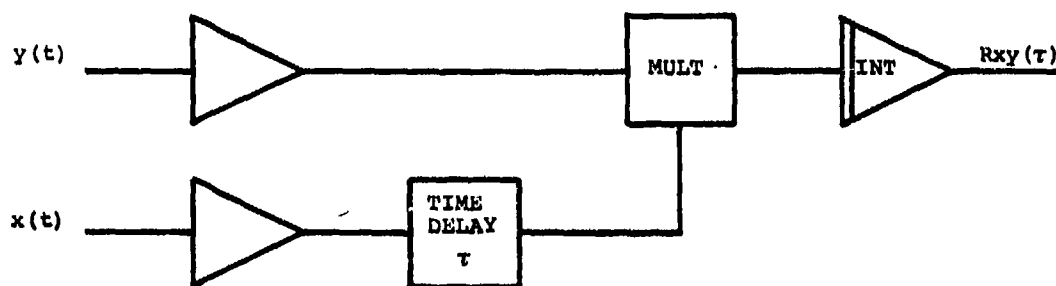


Figure 2.3.7. A Simple Analogue Cross Correlator.

2.4. Frequency Domain Analysis

Much of the interest in describing the characteristics of random data stems either from a requirement to express those characteristics by a functional relationship (e.g. the measurement of atmospheric turbulence) or from the need to minimise its effects by designing suitable filters (e.g. by installing delicate instruments on appropriate anti-vibration mounts). In either case, presentation of the detailed characteristics of the data in the frequency domain is normally much more immediately informative than the equivalent time domain descriptions, for two reasons. Firstly, engineers tend to think and to express themselves in frequency domain terms. Secondly data which is broad band in character (:) is characterised by relatively few points in the time domain. This means that quite large changes in the frequency domain characteristics can be almost indiscernable when, for example, those characteristics are described by the Autocorrelation function (see Chapter 4.).

This chapter is devoted to the two major frequency domain descriptions of random data, the power spectral density and the cross spectral density. A discussion on the derivation of frequency response functions is included, and the chapter is concluded by a Section relating frequency domain and time domain methods of presentation.

2.4.1. Power Spectral Analysis

The power spectral density of a random signal can be defined as the rate of change of mean square value of the signal with respect to frequency. One way to visualise this concept is to imagine that the signal is input to a narrow band unity gain bandpass filter. The output of the filter may be squared and averaged, (figure 2.4.1.). The mean square value so obtained is then divided by the bandwidth of the filter. The rate of change of mean square value (power spectral density) at the centre frequency of the filter is then obtained conceptually by reducing the bandwidth of the filter to zero.

Of course, it is not possible to create a filter with zero bandwidth, but the circuit shown in figure 2.4.1. is representative of one type of analogue power spectral analyser. The complete PSD function for a signal may be obtained from such a circuit either by arranging many similar circuits in parallel or, for reasons of economy, by varying the centre frequency of the band pass filter. In the latter case, the data must be re-analysed for each new selected value of centre frequency.

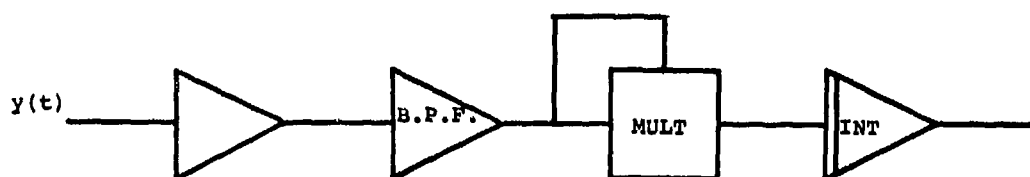


Figure 2.4.1. A Simple Analogue Power Spectral Analyser.

In fact the circuit shown in figure 2.4.1. is rarely used for analogue power spectral analysers. The technique most commonly implemented uses the signal being analysed to modulate a carrier. The modulated carrier is passed through a fixed narrow bandpass filter, and the output from this is squared averaged and scaled in the usual way. In this case the signal bandwidth is swept by altering the frequency of the carrier, rather than the filter. The principle of operation of the heterodyne analyser is shown in figure 2.4.2.

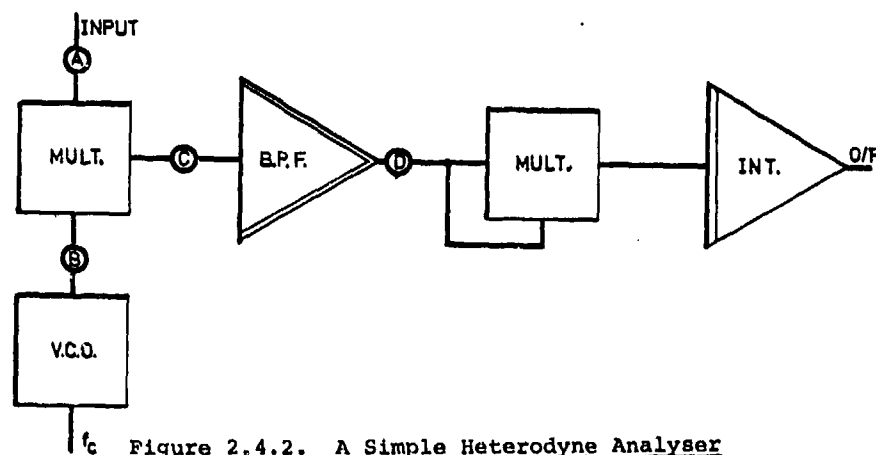


Figure 2.4.2. A Simple Heterodyne Analyser

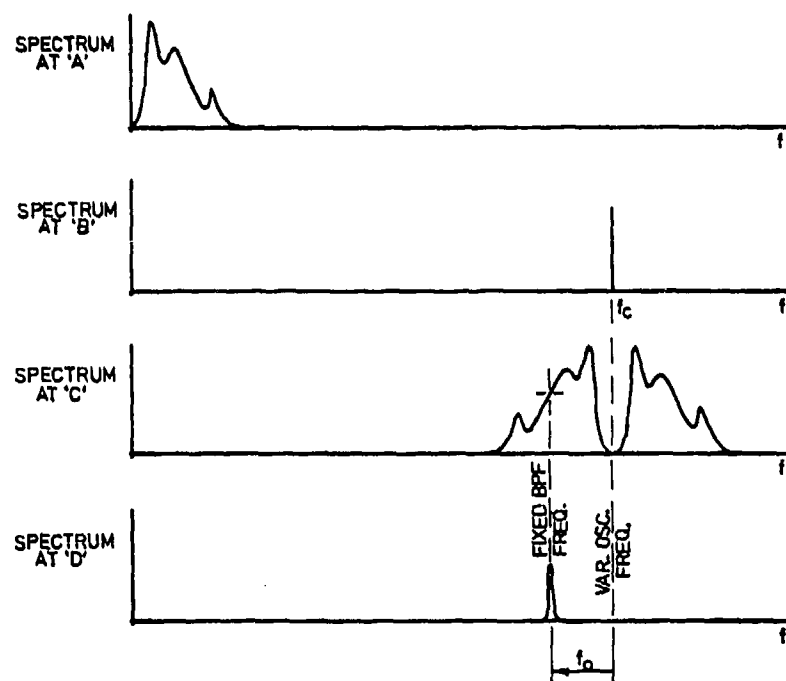


Figure 2.4.3. Principle of Operation of a Heterodyne Analyser.

The majority of digital power spectral analysers use the same principle as the analogue analyser shown in Figure 2.4.1. Here, however, the bandpass filter is simulated by computing the Fourier series of a fixed length portion of the sampled data sequence, calculating the mean square value of the modulus of the Fourier coefficients obtained from a portion of the data sequence, and dividing by the effective bandwidth of the 'filter' represented by the Fourier series estimator. An introduction to Fourier series is contained in Appendix A. The concept that each Fourier coefficient, as defined by equation A.1.12., represents the output of a digital bandpass filter is rendered plausible by equation A.1.10. This states that the original signal can be reconstructed by summing a series of sinusoids of different frequency (i.e. frequency components), each weighted by the Fourier coefficients appropriate to that frequency. For the justification that the coefficients can be used to estimate the power spectral density, the reader is referred to Section A.2.10. This, together with the results presented in Section A.3., provide an estimate for the power spectral density function as follows:

$$G_{yy}(k.\Delta f) = \frac{2h}{N} |C_k|^2 \quad 2.4.1.$$

$$k = 0(1)\frac{N}{2} - 1$$

$$\text{where,} \quad C_k = \sum_{n=0}^{N-1} y_n \cdot \exp \left[-i \frac{2\pi k n}{N} \right] \quad 2.4.2.$$

For reasons which will become apparent later, the estimate given by 2.4.1. is unreliable, and it is normal to improve this by computing the average of a number of independent estimates. This averaging process is equivalent to the integrator shown in figure 2.4.1.

The power spectral density estimator given by equation 2.4.1. contains $N/2$ frequency points equally spaced at an interval of $\Delta f = 1/hN$. It is therefore a constant bandwidth estimator, each 'filter' having a bandwidth Δf .

It should be observed that there is a subtle difference between the Fourier coefficients as defined in equation A.1.12. and the coefficients (defined by a similar equation) of 2.4.2. In the former, y_n is assumed to be periodic, and N is chosen so that the integration is evaluated over an exact number of periods. In the latter case N is chosen arbitrarily. It turns out that the band pass filter defined by equation 2.4.2. has excellent properties when the data happens to be periodic over the N -length sequence, but it is rather less good when this is not so. The reason for this disparity in performance is due to the shape of the pass band of the "filter" defined by 2.4.2., which can be shown to be (Reference 2.2.) of the form:

$$H(f) = \frac{\sin \pi f T}{\pi f T} \quad 2.4.3.$$

The shape of this function is shown in figure 2.4.4. The function has unity value at the centre frequency and has zeroes coinciding with the frequency interval 2.4.1., that is:

$$\Delta f = \frac{1}{h.N} \quad 2.4.4.$$

A sinusoidal signal y_n whose frequency coincides with the centre frequency of one filter ($=k.\Delta f$, say) will be interpreted by equation 2.4.2. as follows:

$$\begin{aligned} C_k &= A \cdot \frac{N}{2} & k &= K \\ &= 0 & k &\neq K \end{aligned} \quad 2.4.5.$$

where A is the amplitude of the sinusoid.

However if sinusoidal signal has a frequency exactly midway between two centre frequencies, equation 2.4.2. will give:

$$C_k = \frac{A.N. \sin \pi(K + \frac{1}{2} - k)}{2 \pi(K + \frac{1}{2} - k)} \quad 2.4.6.$$

where $(K + \frac{1}{2}) \Delta f$ is the frequency of the sinusoid and A is its amplitude.

Equation 2.4.5. clearly shows that equation 2.4.2. represents a perfect filter when a sinusoidal signal coincides with the centre frequency of one filter, ie. when the sample length contains an exact number of periods. When this is not the case, however, the filter performs rather poorly and, at worst, behaves like a second order bandpass filter. This is easily shown from equation 2.4.6., where the filter output values are proportional to the inverse of $|K + \frac{1}{2} - k|$. (It should here be noted that equation 2.4.6. describes the worst case frequency response function of the filter, since it can be viewed as describing the response of one filter as the sinusoid is stepped across its pass band).

A considerable amount of effort has been devoted to improving the shape of the 'filter' described by 2.4.2. This has resulted in a large number of different proposals which can leave the potential user in a state of some confusion. It is hoped that some clarification can be obtained from the following observations:

- Frequency domain averaging will not improve the characteristics of the filter implied by equation 2.4.1. and 2.4.2., though it may be useful for other reasons.
- It can be inferred from the above that the undesirable effects of the filter are caused by discontinuities (of amplitude, slope, etc.) between the end of one "period" and the start of the next - ie. between the end and beginning of the same data sequence, since this is assumed to be periodic.
- Time domain weighting which removes the discontinuities noted above also modifies the spectral shape of the data by distorting the data sequence. This distortion is least when the rate of change of the normalised weighting function with respect to time is small.
- The equivalent filter which results from time domain weighting followed by application of equation 2.4.2. may have additional undesirable characteristics in some applications. For example a sinusoidal function factored by the "80 - 20" cosine weight proposed in Reference 2.6. and shown in figure 2.4.5., results in a spectrum containing side lobes. The reason for this is that the modified form of the filter shape described by equation 2.4.3. has maxima which are not spaced Δf apart or a multiple of this.

In the light of the above observations, the most satisfactory "fix" for improving the characteristics of PSD function estimates obtained from 2.4.1. is to weight each N-length sample of y_n by:

$$W(n) = \frac{1}{2} \left(1 - \cos \frac{2\pi.n}{N} \right) \quad 2.4.7.$$

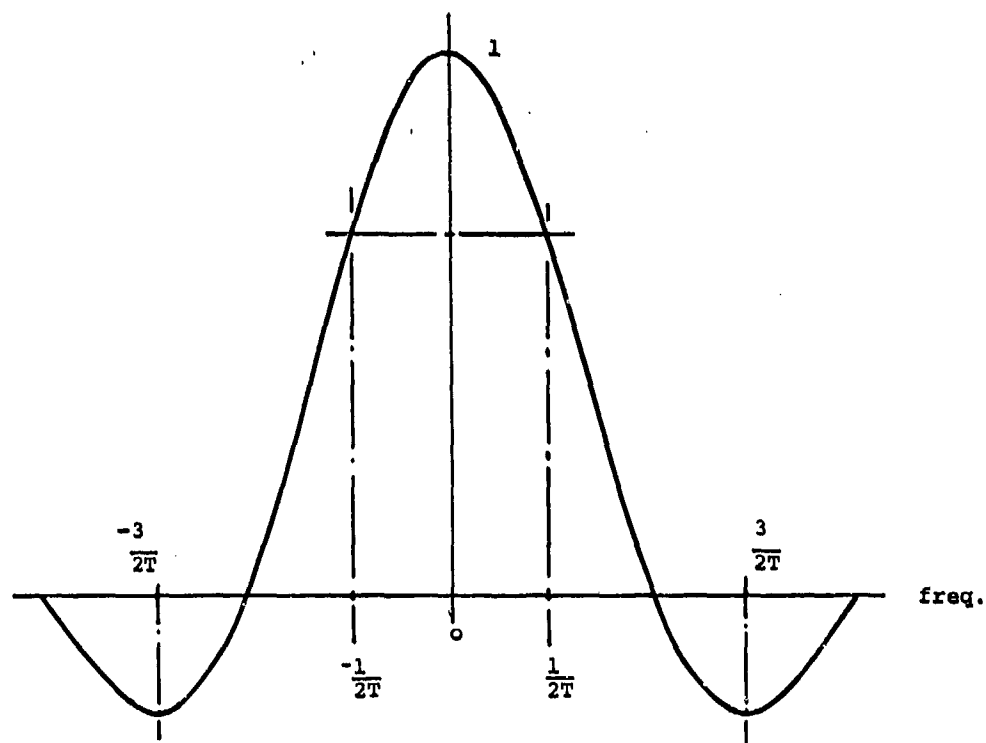


Figure 2.4.4. Filter Shape corresponding to a DFT of Unweighted Data.

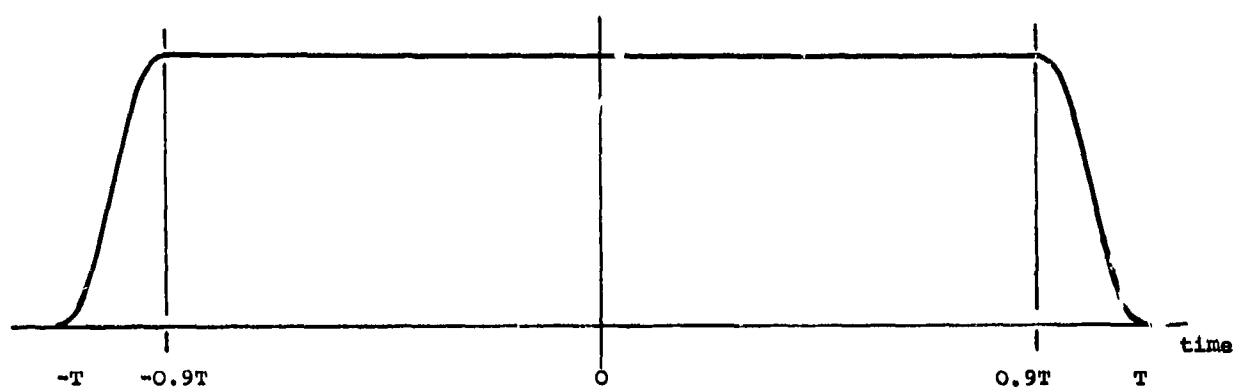


Figure 2.4.5. One Proposal for Weighting Time Domain Data Prior to Transformation.

When the weighting function defined by equation 2.4.7. is used, the power spectral density estimates obtained by 2.4.1. must be multiplied by 8/3.

The filtering of equation 2.4.2. is most efficiently achieved using the "Discrete Fourier Transform", also known as the "Fast Fourier Transform". The algorithm for this is described in Appendix B. The efficiency of this algorithm is remarkable, improving upon the obvious algorithm by a factor N/p where $N = 2^p$, the length of the data used in the transform.

An alternative digital method of estimating the power spectral density of a random function is to determine the autocorrelation function (Section 2.3.7.) as follows:

$$R_{YY}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T y(t) \cdot y(t-\tau) \cdot dt \quad 2.4.8.$$

or, using trapezoidal integration upon an N -length sampled data sequence:

$$R_{YY}(kh) = \frac{1}{(N-k)} \sum_{i=k}^N y(i) \cdot y(i-k) \quad 2.4.9.$$

where the maximum value of k should normally be less than ten percent of N .

The power spectral density can be obtained from the autocorrelation function using the Wiener-Khintchine relation:

$$G_{YY}(f) = \lim_{T \rightarrow \infty} 4 \cdot \int_0^T R_{YY}(\tau) \cdot \cos \pi f \tau \cdot d\tau \quad 2.4.10.$$

or, again using trapezoidal integration:

$$G_{YY}(n\Delta f) = 4h \cdot \sum_{k=0}^{mH} R_{YY}(k) \cdot \cos \frac{\pi nk}{m} \quad 2.4.11.$$

where $\Delta f = \frac{1}{2hm}$, and the double prime indicates that the first and last terms in the series must be halved.

Equation 2.4.11. may be looked upon as a bandpass filter acting upon the autocorrelation function in rather the same way as the Fourier series estimator of equation 2.4.2. Here again similar difficulties arise and, in order to improve the characteristics of the filter, the autocorrelation function should be weighted by the function:

$$W(k) = \frac{1}{2} \left[1 + \cos \frac{\pi k}{m} \right] \quad 2.4.12.$$

or by a similar function having the properties:

$$W(0) = 1, \quad \text{and} \quad W(m) \rightarrow 0 \quad 2.4.13.$$

This method for estimating the power spectral density function has fallen from favour to some extent, because the computational efficiency is greatly inferior to the "direct" method utilising the Discrete Fourier Transform.

Estimates of power spectral density are subject to two types of error. The first is a statistical error arising from the necessity to obtain the estimates from a finite length of data. The second is a bias error which stems from the finite bandwidth filter used to separate the various frequency components of a signal.

The variance of the statistical error is a function of the averaging time, as noted above, but it is also a function of the number of statements to be made about the data, which is equivalent to the bandwidth(s) of the analyser filters. Specifically, the variance of the statistical error is given by (Appendix E):

$$\text{Var}[G_{YY}(f)] = \frac{G_{YY}^2(f)}{B_e \cdot T} \quad 2.4.14.$$

where B_e is the "equivalent statistical bandwidth" and T is the averaging time.

For the digital analyser specified by equation 2.4.1., the bandwidth-time product is unity. This is equivalent to taking a single reading of a random phenomenon, a startling though not unreasonable conclusion, bearing in mind that the Fourier coefficients are simply a reversible transform of the particular sample block of data. Two methods are available to reduce the statistical error to sensible proportions. The first is to obtain M independent estimates of the PSD function by analysing different samples extracted from the data. If this method is used then the variance of the statistical error becomes:

$$\text{Var } G_{YY}(f) = \frac{G_{YY}^2(f)}{M} \quad 2.4.15.$$

The alternative method for effecting a reduction of statistical error is to average a number of consecutive spectral estimates, thereby effectively increasing the analyser bandwidth. Then if L spectral estimates are averaged, the variance of the statistical error becomes:

$$\text{Var } G_{YY}(f) = \frac{G_{YY}^2(f)}{L} \quad 2.4.16.$$

The two methods of improving the estimates noted above are sensibly interchangeable. That is, for a given length of record, the variance of the error can be halved either by averaging pairs of spectral estimates or by re-analysing, taking half the previous number of samples for each estimate, and doubling the number of independent estimates of the PSD function.

Incidentally, the use of "Hanning" to improve the filter characteristics does not affect equation 2.4.15. The increase in bandwidth so obtained is countered by a reduction in the effective integration time.

Bias errors can arise when the power spectral density changes value rapidly as the frequency is varied. The shape of the actual PSD function is averaged over the finite bandwidths of the filters, and this can cause "smearing" of that shape. Specifically, the bias error is approximated by (Reference 2.2.):

$$b G_{YY}(f) = \frac{B_e^2}{24} G_{YY}''(f) \quad 2.4.17.$$

where,

$$G_{YY}''(f) = \frac{d^2 G_{YY}(f)}{df^2}$$

Thus bias errors are significant only when the curvature of the PSD function differs significantly from zero.

A great many of the properties of a random signal may be deduced directly from its PSD function. For example:

The mean square value:

$$\sigma_y^2 = \int_0^{\infty} G_{yy}(f) \cdot df \quad 2.4.18.$$

The expected number of zero crossings per unit of time:

$$N_0 = 2 \cdot \left[\frac{\int_0^{\infty} f^2 \cdot G_{yy}(f) \cdot df}{\sigma_y^2} \right]^{1/2} \quad 2.4.19.$$

The expected number of level crossings per unit of time:

$$N_y = N_0 \cdot \exp \frac{-y^2}{2 \cdot \sigma_y^2} \quad 2.4.20.$$

The expected number of maxima per unit of time:

$$N_m = \left[\frac{\int_0^{\infty} f^4 \cdot G_{yy}(f) \cdot df}{\int_0^{\infty} f^2 \cdot G_{yy}(f) \cdot df} \right]^{1/2} \quad 2.4.21.$$

The relationships quoted above have been derived by various authors. Reference 2.1. contains explanations of these and other useful relationships.

2.4.2. Cross Spectral Analysis

The cross spectral density function represents a description of the joint properties of two random signals in the frequency domain, and can be considered to be a logical extension of the power spectral density function. Cross spectral density (CSD) functions are only rarely useful in their own right. They are, however, frequently computed en-route to estimating frequency response functions (see Section 2.4.3.). A simple analogue cross spectral analyser is shown in figure 2.4.6.

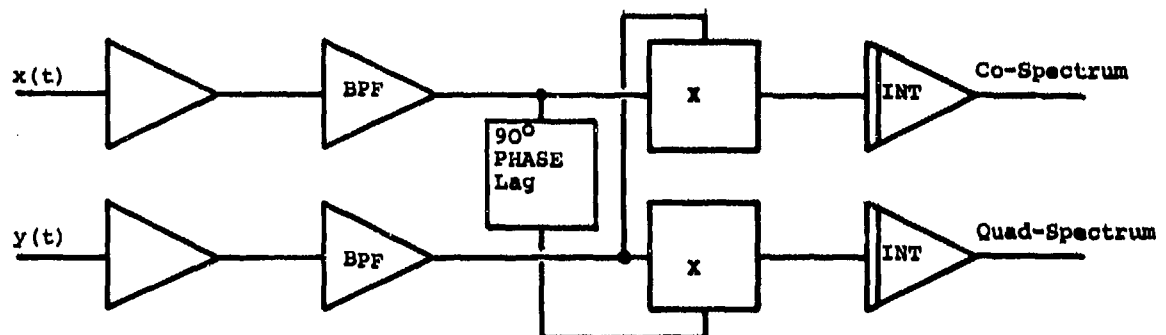


Figure 2.4.6. A Simple Cross Spectral Analyser.

The circuit shown in figure 2.4.6. is remarkably similar to the circuit for a simple power spectral analyser shown in figure 2.4.1., but is complicated by the fact that, on average, the relationship between $x(t)$ and $y(t)$ may not be in phase (i.e. have zero time delay). Therefore both in-phase (or coincident) and quadrature components must be computed. As was the case for the power spectral analyser the centre frequencies of the two narrow band pass filters must be varied over the frequency range of interest in order to compile a complete estimate of the CSD function.

Digital cross spectral analysers are similar in concept to the analyser shown in figure 2.4.6. However the bandpass filters are simulated by computing the Fourier series of fixed length blocks of the two sampled data sequences, in exactly the same way as for the power spectral analyser. (Note, however, that each pair of sample blocks must commence at the same point in time). An estimate for the CSD function may be deduced from Appendix A as:

$$G_{xy}(k\Delta f) = \frac{2h}{N} (x^C_k) \cdot (y^C_k)^* \quad , \quad k = 0(1)\frac{N-1}{2} \quad 2.4.22.$$

$$\text{where:} \quad x^C_k = \sum_{n=0}^{N-1} x(n) \cdot \exp\left(-j \frac{2\pi kn}{N}\right) \quad 2.4.23.$$

$$\text{and:} \quad y^C_k = \sum_{n=0}^{N-1} y(n) \cdot \exp\left(-j \frac{2\pi kn}{N}\right) \quad 2.4.24.$$

The similarity between equations 2.4.1. and 2.4.22. is obvious.

The estimate of cross spectral density given by equation 2.4.22. is normally made more reliable by averaging a number of such independent estimates.

The cross spectral density estimator given by equation 2.4.22. contains $N/2$ equally spaced frequency points at an interval of $\Delta f = 1/hN$.

An alternative method for arriving at an estimate of the CSD function for a stationary random signal is to determine first the cross correlation function (see Section 2.3.8.):

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) \cdot y(t - \tau) \cdot dt \quad 2.4.25.$$

or, using trapezoidal integration upon a pair of N -length sampled data sequences:

$$R_{xy}(kh) = \frac{1}{(N-k)} \sum_{i=k}^N x(i) \cdot y(i-k) \quad 2.4.26.$$

$$\text{and:} \quad R_{yx}(kh) = \frac{1}{(N-k)} \sum_{i=k}^N y(i) \cdot x(i-k) \quad 2.4.27.$$

where the maximum value of k should normally be less than ten percent of the value of N .

The cross spectral density can then be obtained by transforming the cross correlation function:

$$C_{xy}(n\Delta f) = 4h \cdot \sum_{k=0}^{m-1} A(k) \cdot \cos \frac{\pi n \cdot k}{m} \quad 2.4.28.$$

$$Q_{xy}(n\Delta f) = 4h \cdot \sum_{k=0}^{m-1} B(k) \cdot \sin \frac{\pi n \cdot k}{m} \quad 2.4.29.$$

where:

$$A(k) = \frac{1}{2} \left[R_{xy}(k) + R_{yx}(k) \right] \quad 2.4.30.$$

$$B(k) = \frac{1}{2} \left[R_{xy}(k) - R_{yx}(k) \right] \quad 2.4.31.$$

and where:

$$G_{xy}(if) = C_{xy}(f) - i \cdot Q_{xy}(f), \quad 2.4.32.$$

$$\Delta f = \frac{1}{2h \cdot m}$$

and the double prime indicates that the first and last terms in the series are to be halved.

The raw estimates defined by 2.4.28. and 2.4.29. may be refined by weighting the even and odd functions $A(k)$ and $B(k)$ of equations 2.4.30. and 2.4.31. in exactly the same way as for PSD estimates. Equation 2.4.12. is satisfactory for this purpose.

Cross spectral density estimates are subject both to variance errors and to bias errors. These are introduced by the finite averaging time and non-zero filter bandwidths. Reference 2.2. quotes estimates of these errors as:

$$\text{Var } C_{xy}(f) < \frac{G_{xx}(f) \cdot G_{yy}(f)}{B_e \cdot T} \quad 2.4.33.$$

$$\text{Var } Q_{xy}(f) < \frac{G_{xx}(f) \cdot G_{yy}(f)}{B_e \cdot T} \quad 2.4.34.$$

$$b \ C_{xy}(f) \pm \frac{B_e^4}{24} \cdot C_{xy}''(f) \quad 2.4.35.$$

$$b \ Q_{xy}(f) \pm \frac{B_e^4}{24} \cdot Q_{xy}''(f) \quad 2.4.36.$$

where:

$$C_{xy}''(f) = \frac{d^2 C_{xy}(f)}{d^2 f}, \quad \text{for example.}$$

B_e is the "equivalent statistical" bandwidth of the filter, and T is the averaging time, as before.

2.4.3. Frequency Response Functions

If a linear system having a single input is forced by a signal applied to that input, then a signal representing the response of the system will be dependent upon the character of the input signal and the properties of the system being excited. Specifically, if the input is sinusoidal at frequency f and has a unit amplitude, then the output will be sinusoidal, having an amplitude and phase relative to the input which is defined by $H(if)$, the frequency response function of the system. It will be noted that, in general, $H(if)$ is a complex function of frequency and is a property of the physical characteristics of the linear system. These ideas may be extended to the case where the input $y(t)$ is a stationary random signal. It turns out that three statements can be made about the characteristics of the output signal $x(t)$ (Reference 2.1.). These are:

- the output is a stationary random signal,
- the PSD function of the response is a function of the PSD function of the input and the system frequency response function,
- the CSD function of the output relative to the input is a function of the power spectral density of the input and the system frequency response function.

Specifically:

$$G_{xx}(f) = |H(if)|^2 \cdot G_{yy}(f) \quad 2.4.37.$$

and: $G_{xy}(f) = H(if) \cdot G_{yy}(f) \quad 2.4.38.$

The frequency response function of the system can be obtained from equation 2.4.38. as

$$H(if) = \frac{G_{xy}(if)}{G_{yy}(f)} \quad 2.4.39.$$

It was noted in Section 2.4.2. that the CSD function acts as a filter, rejecting those components in either signal which are not coherent (i.e. which are not related to one another). It seems reasonable, therefore, to use equations 2.4.37. and 2.4.38. to justify (or otherwise) the implicit assumption that $x(t)$ and $y(t)$ are related functions of time, and hence frequency. This is effected by computing the square of the modulus of the ratio of the two estimates of frequency response function, so:

$$\frac{G_{xy}(if)^2}{G_{xx}(f) \cdot G_{yy}(f)} = \gamma^2(f) \leq 1 \quad 2.4.40.$$

The quantity $\gamma^2(f)$ is known as the coherence function. If its value nowhere differs significantly from unity, then a linear frequency response function relating the output to the input certainly exists, and is defined by equation 2.4.39.

A set of experimental results for which the coherence function is everywhere close to unity is found only very rarely. Low values of coherence may be caused by:

- no frequency response function
- a "node" in the frequency response function
- "noise" contaminating either the input signal, or the output signal, or both

- a non-linear frequency response function.

A small amount of measurement system noise is normally to be found in experimental results. This is usually well distributed and causes the coherence function to "follow" roughly the square of the modulus of the frequency response function. Further investigation is called for when this does not happen, i.e. when the coherence remains low even when the modulus of the frequency response function is large.

Possible causes of a low coherence in this situation are:

- serious contamination of the signals,
- the system has another, independent, input,
- the input power spectral density has a low value in the region of the peak under investigation,
- the system is non-linear (if the system is a physical structure, does it rattle?)

The resolution of these matters can sometimes be helped by assuming that the low coherence is due entirely to the addition of noise to one or the other channel. Then under the assumption that the noise exists only in $x(t)$,

$$G_n(f) = G_{xx}(f) \cdot [1 - \gamma^2(f)] \quad 2.4.41.$$

and if all the noise is assumed to be contained in $y(t)$, then:

$$G'_n(f) = G_{yy}(f) \cdot [1 - \gamma^2(f)] \quad 2.4.42.$$

If either of these noise power spectral densities is a relatively smooth function of frequency, then it is usually safe to assert that the appropriate signal contains the majority of the noise.

Frequency response functions are computed from estimates of power and cross spectral densities, both of which will, in general, contain a statistical uncertainty and, possibly, a bias error. It therefore follows that estimates of frequency response functions will contain similar errors.

It can be reasoned that bias errors can arise from:

- bias errors in the estimates of power and cross spectral density functions,
- noise included in the measurement of the input signal,
- contributions from other inputs which are correlated with the measured input (it can be shown that other independent inputs do not cause bias).

A detailed study of the statistical uncertainty in estimates of frequency response functions is included in Reference 2.2. The result is neither concise nor positive, and is therefore not included here.

An assessment of bias errors is also necessarily inconclusive because of the range of possible causes noted above. One useful observation can, however, be made. The sections devoted to describing algorithms for estimating power and cross spectral density functions (2.4.1. and 2.4.2.) include proposals for improving the characteristics of the analysis filter. The effect of these improvements is to increase the bandwidth of the filter. Irrespective of whether they are to be preferred for power spectral analysis in particular, the use of these improved filters can increase the bias error of frequency response functions significantly, as shown in figure 2.4.7. It is, therefore, to be recommended that they be omitted when frequency response functions are of interest, particularly if the structure under investigation has lightly damped modes.

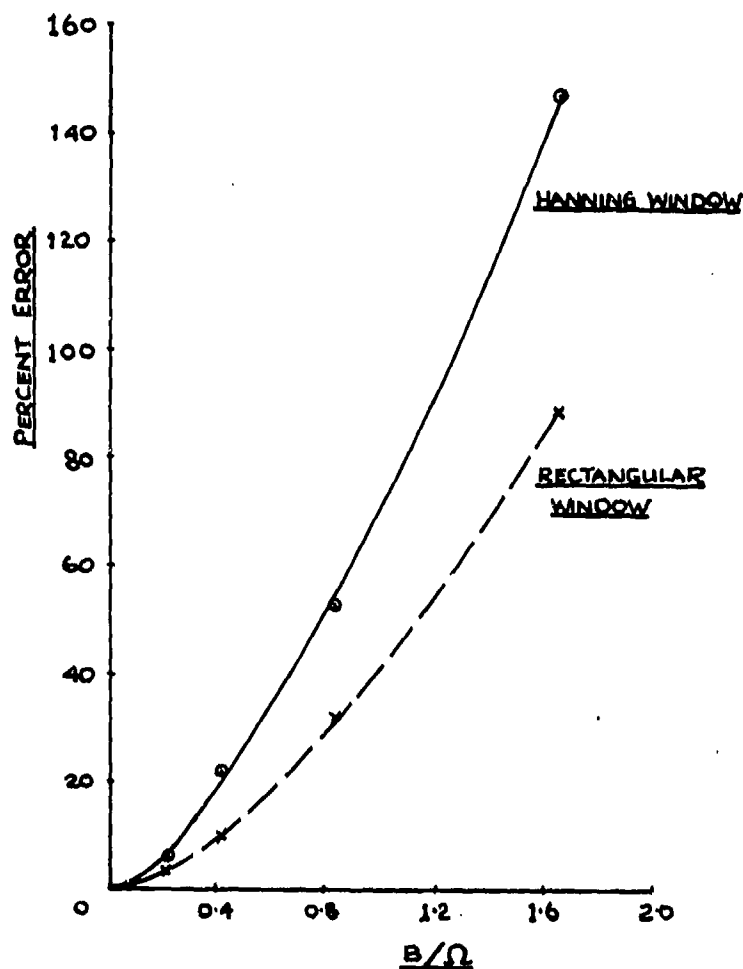


Figure 2.4.7.

Percent Error in the Estimation of Filter Bandwidth as a Function of Analyser to Filter Bandwidth Ratio.

2.4.4. Manipulations Involving a Number of Random Signals

This Section introduces techniques which can be used to carry out the equivalent of arithmetic operations on random signals. In general, the techniques are intended to yield frequency domain descriptions of the results of an arithmetic operation, using as operands frequency domain estimates which can be obtained directly from measurements.

No attempt has been made to include in this Section estimates of errors in the results obtained from such operations, although in some cases, bounds to such errors can be deduced directly from the results presented in Appendix E.

2.4.4.1. The Sum of Random Signals

The case under consideration here is shown diagrammatically in figure 2.4.8. That is:

$$x(t) = \sum_{k=1}^M y_k(t) \quad 2.4.43.$$

Equation 2.4.43. can be transformed into the frequency domain, when the linearity property of the Fourier transform (see Appendix A) gives:

$$F_x(if) = \sum_{k=1}^M F_{y_k}(if) \quad 2.4.44.$$

Each side of equation 2.4.44. can be multiplied by the complex conjugate of $F_x(if)$, denoted by $F_x^*(if)$:

$$F_x(if) \cdot F_x^*(if) = \sum_{k=1}^M F_{y_k}(if) \cdot \sum_{k=1}^M F_{y_k}^*(if) \quad 2.4.45.$$

It may be deduced from this expression that the PSD function of the Sum of Random Signals is given by:

$$G_{xx}(f) = \sum_{j=1}^M \sum_{k=1}^M G_{y_j y_k}(if) \quad 2.4.46.$$

which may be simplified to yield:

$$G_{xx}(f) = \sum_{j=1}^M G_{y_j y_j}(f) + 2 \operatorname{Re} \left[\sum_{j=2}^M \sum_{k=1}^{j-1} G_{y_j y_k}(if) \right] \quad 2.4.47.$$

The first term of equation 2.4.47. is the sum of the PSD functions of each signal; the second term is the real part of the sum of all possible CSD functions between one signal and the other. The latter has been simplified by making use of the symmetry properties of the CSD function.

Two extreme cases are worthy of note. When all of the signals to be summed are mutually independent, the CSD functions between one signal and another are everywhere zero, and equation 2.4.47. degenerates to:

$$G_{xx}(f) = \sum_{j=1}^M G_{y_j y_j}(f) \quad 2.4.48.$$

Thus the PSD function of the sum of a number of mutually independent signals is equal to the sum of their individual PSD functions.

The second extreme case worth noting occurs when all signals are identical. In this case equation 2.4.47. reduces to:

$$G_{xx}(f) = M \cdot G_{y_j y_j}(f) \quad 2.4.49.$$

as might be expected from equation 2.4.43.

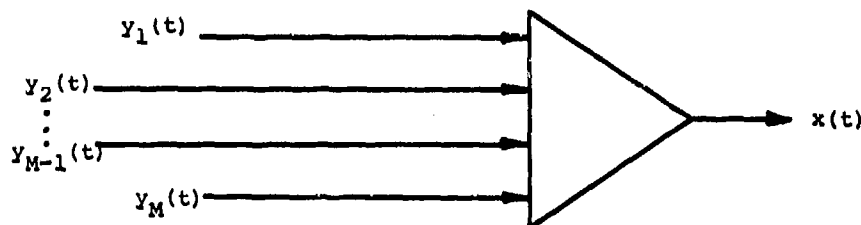


Figure 2.4.8. The Sum of a Number of Random Signals.

An alternative expression, which may be more convenient, can be obtained by multiplying equation 2.4.44. by the conjugate of the transform of one signal. This yields:

$$F_x(if) \cdot F_{y_j}^*(if) = \sum_{k=1}^M F_{y_k}(if) \cdot F_{y_j}^*(if) \quad 2.4.50.$$

or,

$$G_{y_j x}(if) = G_{y_j y_j}(f) + \left[\sum_{k=1}^M G_{y_j y_k}(if) \right]_{j \neq k} \quad 2.4.51.$$

This states that, for M mutually independent random signals, the cross spectral density function between the output and any input is real and is equal to the PSD function of that input.

/v

2.4.4.2. The Sum of Filtered Random Signals

The case shown diagrammatically in figure 2.4.9. can be stated, by comparison with the case described in Section 2.4.4.1., as:

$$F_x(if) = \sum_{k=1}^M H_k(if) \cdot F_{y_k}(if) \quad 2.4.52.$$

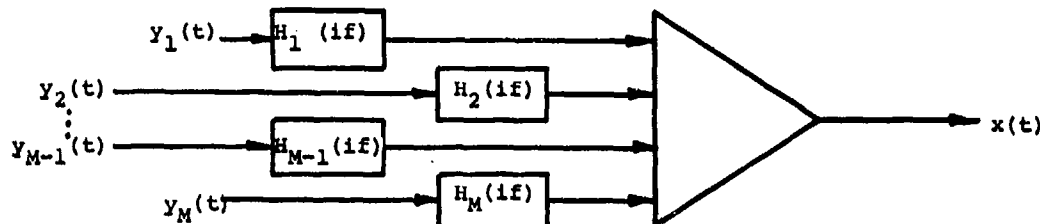


Figure 2.4.9. The Sum of a Number of Filtered Random Signals.

An argument similar to that followed in Section 2.4.4.1. can be used to yield:

$$G_{xx}(f) = \sum_{j=1}^M \sum_{k=1}^M H_j(if) \cdot H_k^*(if) \cdot G_{y_j y_k}(if) \quad 2.4.53.$$

which may be written, after some manipulation:

$$G_{xx}(f) = \sum_{j=1}^M |H_j(if)|^2 G_{y_j y_j}(f) + 2 \operatorname{Re} \left[\sum_{j=2}^M \sum_{k=1}^{j-1} H_k(if) \cdot H_j^*(if) \cdot G_{y_j y_k}(if) \right] \quad 2.4.54.$$

When the inputs, y_j , are mutually independent, the CSD functions between them are everywhere zero, and equation 2.4.54. degenerates to :

$$G_{xx}(f) = \sum_{j=1}^M |H_j(if)|^2 \cdot G_{y_j y_j}(f) \quad 2.4.55.$$

This result is equivalent to equation 2.4.48. A second interesting case occurs when the inputs are everywhere identical, as in figure 2.4.10. In this case, equation 2.4.54. becomes:

$$G_{xx}(f) = \sum_{j=1}^M |H_j(if)|^2 \cdot G_{yy}(f) \quad 2.4.56.$$

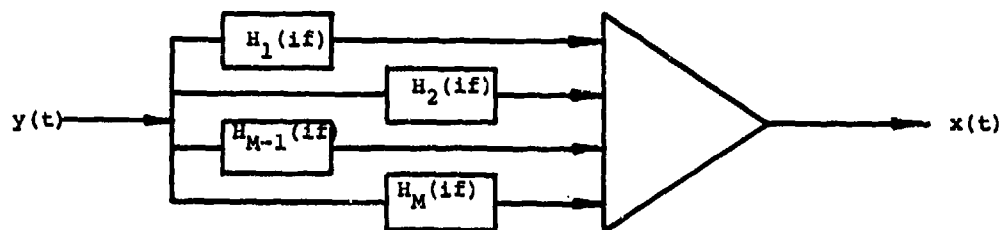


Figure 2.4.10.

The preceding development has been oriented towards describing the output signal in terms of the various input signals and the (assumed known) frequency response functions which describe the filters. When it is required to determine the frequency response functions of the individual filters, given measurements of the output and the various inputs, then it is more convenient to adopt a somewhat different development, multiplying equation 2.4.52. by the conjugate of the Fourier transform of any input, so:

$$F_x(if) \cdot F_{Y_j}^*(if) = \sum_{k=1}^M H_k(if) \cdot F_{Y_k}(if) \cdot F_{Y_j}^*(if) \quad 2.4.57.$$

from which it may be deduced that:

$$G_{Y_j x}(if) = \sum_{k=1}^M H_k(if) \cdot G_{Y_j Y_k}(if) \quad 2.4.58.$$

Equation 2.4.58. represents just one of M similar expressions which can be constructed. Organised in Matrix form, these can be written:

$$\{G_{Y_j x}(if)\} = [G_{Y_j Y_k}(if)] \cdot \{H_k(if)\} \quad 2.4.59.$$

which may be solved to yield an expression for the individual frequency response functions:

$$\{H_k(if)\} = [G_{Y_j Y_k}(if)]^{-1} \cdot \{G_{Y_j x}(if)\} \quad 2.4.60.$$

Equation 2.4.60. is valid whether or not the various inputs are mutually independent. However, when they are mutually independent, then the matrix of input CSD functions becomes a frequency dependent diagonal matrix, and equation 2.4.60. can be written:

$$H_k(if) = \frac{G_{Y_j x}(if)}{G_{Y_j Y_j}(if)} \quad 2.4.61.$$

This very important relationship states that, when all inputs are mutually independent, an estimate of a frequency response function can be obtained without bias from the PSD function of the appropriate input and the CSD function between that input and the summed output, regardless of the other inputs.

This conclusion confirms the assumption made in Section 2.4.3., which stated that additive noise which corrupted a measured output did not bias estimates of frequency response function, provided that the noise was independent of the input signal.

2.4 Frequency Response Function Estimates from Closed Loop Measurements

The arrangement to be discussed in this Section is shown in figure 2.4.11.

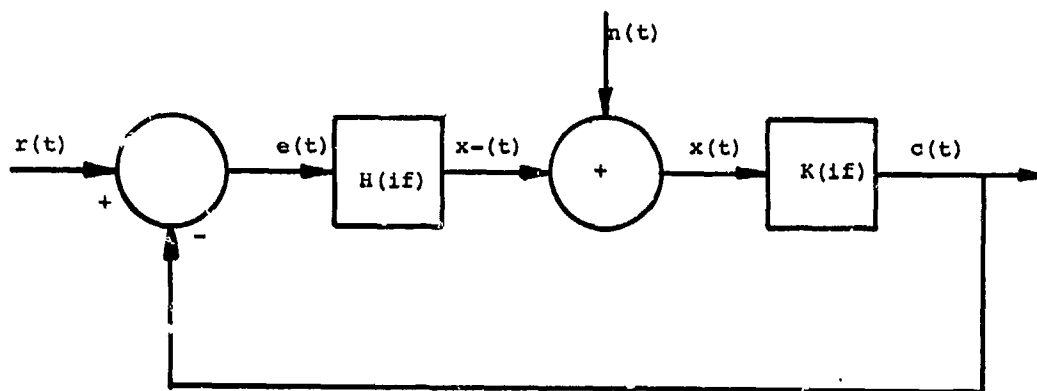


Figure 2.4.11.

The system described by $K(if)$ might represent a flight vehicle, and the system described by $H(if)$ might represent a Stability Augmentation System, an Auto-pilot, or even a human pilot. It is supposed that the two systems are arranged in a closed loop with two sources of input. The first, $r(t)$, represents a demand (speed or height demand, for example) and the second, $n(t)$ represents a system disturbance (e.g. atmospheric turbulence). The problem is to identify the system $H(if)$. A number of possible cases can be considered:-

- a) $e(t)$ and $x'(t)$ can be measured.

This constitutes the simplest case. The problem can be solved by direct application of equation 2.4.39.:

$$H(if) = \frac{G_{ex'}(if)}{G_{ee}(if)} \quad 2.4.62.$$

It is worth noting that:

$$F_{x'}(if) = H(if) \cdot F_e(if) \quad 2.4.63.$$

$$\text{or,} \quad F_{x'}(if) = H(if) \cdot (F_r(if) - F_c(if)) \quad 2.4.64.$$

Multiplying by the complex conjugate of $Fr(if)$ allows the following to be deduced (see equations 2.4.45. and 2.4.46.):

$$Grx'(if) = H(if).(Grr(if) - Grc(if)) \quad 2.4.65.$$

Equation 2.4.51. states, in this case, that:

$$Gre(if) = Grr(f) - Grc(if) \quad 2.4.66.$$

Hence, from equations 2.4.65. and 2.4.66.,

$$H(if) = \frac{Grx'(if)}{Gre(if)} \quad 2.4.67.$$

$$\text{or, } \frac{Gex'(if)}{Ges(f)} = \frac{Grx'(if)}{Gre(if)} \quad 2.4.68.$$

It can therefore be concluded that the frequency response function $H(if)$ can be estimated as the ratio of the CSD functions between one input and the output and between one input and the summed input.

b) $r(t)$ and $x'(t)$ can be measured.

The estimated value of the frequency response function may be written, from equation 2.4.67.

$$\hat{H}(if) = \frac{Grx(if)}{Gre(if)} \quad 2.4.69.$$

This may be expanded to give:

$$\hat{H}(if) = \frac{Grx'(if)}{Gre(if)} + \frac{Grn(if)}{Gre(if)} \quad 2.4.70.$$

$$= H(if) + \frac{Grn(if)}{Gre(if)} \quad 2.4.71.$$

For this case, provided that $r(t)$ and $n(t)$ are independent functions, equation 2.4.69. can be used to provide an unbiased estimate of the frequency response function $H(if)$, since $Grn(if)$ becomes zero.

c) $e(t)$ and $x(t)$ can be measured.

When the input signal $r(t)$ cannot be measured, an estimate of the frequency response function might be obtained by application of 2.4.62.:

$$\hat{H}(if) = \frac{Gex(if)}{Ges(f)} \quad 2.4.72.$$

$$\text{or,} \quad \hat{H}(if) = \frac{Gex'(if)}{Gee(f)} + \frac{Gen(if)}{Gee(f)} \quad 2.4.73.$$

$$\text{Hence:} \quad \hat{H}(if) = H(if) + \frac{Gen(if)}{Gee(f)} \quad 2.4.74.$$

Equation 2.4.74. demonstrates that estimates of the frequency response function $H(if)$ using only $e(t)$ and $x(t)$ are biased, since $Gen(if)$ will exist due to the action of the feedback loop. In fact the second term of equation 2.4.74. can be expanded to give:

$$\frac{Gen(if)}{Gee(f)} = \frac{1}{K(if)} + H(if) \cdot \frac{Gn'n'(f)}{Gee(f)} \quad 2.4.75.$$

where $n'(t)$ is the component of $e(t)$ which is due to the noise input $n(t)$. Clearly, when $n(t)$ is small compared with $r(t)$, then the component of equation 2.4.74. defined by 2.4.75. will also be small, and:

$$\hat{H}(if) \approx H(if) \quad 2.4.76.$$

However, when $n(t)$ is large compared with $r(t)$, then:

$$\hat{H}(if) \approx -\frac{1}{K(if)} \quad 2.4.77.$$

the inverse of the frequency response function of the known system!

The above example is designed to demonstrate not only the power of the technique and also the dangers which can be encountered if the measurement trial has not been designed properly.

The techniques and algorithms reviewed in Chapters 1 and 2 represent only the bare bones of part of a facility for processing random data. Such a facility may take many forms, ranging from a general purpose digital computer to a range of dedicated "black boxes", each designed to fulfil a particular task. It is, therefore, hardly relevant to discuss a particular, or even an ideal, facility for the processing of random data. It is, however, relevant to discuss in general terms the requirements which are to be sought after in any facility, regardless of the shape of the hardware and the precise analysis techniques employed within that hardware.

Practical matters such as how the data and results are archived, how quality assurance requirements are to be satisfied, and how the results of the various analyses are presented do not, in general, affect the precision and quality of analyses. However they assume major importance in a facility designed to produce such analyses of random data on a routine basis. These are the matters which are introduced in this Chapter.

3.1. Measurement System Calibrations

Each transducer and signal conditioning unit (SCU) system possesses a number of intrinsic properties. Those of direct interest to the analyst are:

- rest condition (zero offset)
- sensitivity (gain)
- dynamic behaviour (frequency response characteristic)

The first two properties may be determined from static calibrations in which the transducer or SCU is exercised by subjecting it to a series of known inputs (normally, though not necessarily, at steady levels), each value of the input and the associated unit output level being recorded. Static calibrations are normally relatively simple to execute, but there are exceptions - calibrating a set of strain gauges designed to deduce wing loads and moments, for example.

The third property is often much more difficult to determine with a reasonable degree of accuracy, but is just as important as the static properties if the measurements are expected to vary rapidly during flight trials.

Calibration information is likely to be gathered by a number of separate groups within an organisation and at widely differing times. Also, individual units may be replaced during the course of a flight trial. The task of maintaining reasonably reliable information about the overall characteristics of a measurement system can be formidable unless the information is handled and logged in a uniform manner. The requirements for such a logging system are:

- flexibility. The system must be capable of interpreting many different types of information and of converting it into a uniform format
- tracibility. Overall system characteristics may be deduced from a number of individual calibrations. It should always be possible to "back-track" to check suspect results
- reliability. Standard errors associated with individual calibrations should be propagated so that an estimate of the standard error of each measurement channel can be ascertained.

The information required from each individual calibration in order that these requirements may be satisfied is as follows:

- Calibration identifier
- Calibration date
- Transducer/SCU identifier

- Raw calibration data
- Reduction technique identifier
- Deduced properties and associated standard errors

A calibration data file for any trial may then be established by specifying calibration and unit identifiers for each measurement channel. An appropriate form for the calibration data file might be:

- Flight identifier
- Flight date
- Flight record identifier
- Measurement system specification)
- Total zero offset)
- Total static (or nominal) sensitivity) for each measurement
- Total dynamic characteristics) channel
- Associated standard errors)

The above system could become very unwieldy if it is used unwisely by breaking down each measurement channel into the greatest possible number of elements. A reasonable approach can be arrived at by dividing each measurement system into two parts. The transducer and its associated SCU is considered to constitute one unit, and the transmission and recording equipment to constitute the other. The last of these can be adjusted so that the dynamic characteristics of each channel are matched, and the overall system be organised so that calibration levels are recorded immediately prior to, or during, each trial (see Section 1.3.3.). Thus the characteristics of the transmission and recording equipment can be defined by a single dynamic calibration and by the calibration signals recorded at the time of the trial. This information must, of course, be reduced and logged in a manner similar to that described above.

The calibrations and associated accuracies of instrumentation systems have always been a source of irritation to users and quality assurance engineers alike. The above system can go some way towards alleviating this irritation, particularly if it is programmed onto a digital computer.

3.2. Data Qualification

The first requirement for data qualification is to select samples from a recording for further investigation. This implies that the analyst has access to a "quick look" display of the relevant parameters. It follows that an ability to produce such a display must be a feature of the analysis facility. It is the experience of the author that the ideal "quick look" output is one in which the complete recording is displayed onto no more than two metres of paper. The paper should include all the information channels which are necessary to make the selection, and the bandwidth of the presentation should be consistent with that of the data. The result does not allow the details of the waveform to be examined, but does highlight stationary areas of information, the distribution of dropouts, certain types of non-linearity (e.g. an asymmetric probability distribution), points where the recorder was switched on, etc., in fact the majority of the information required to select appropriate samples for analysis. One drawback of the presentation is the difficulty which may be experienced in defining the precise starting point for an analysis. Where random data is concerned, however, it is often not necessary to specify a starting point to a precision better than the second, so that the difficulty may not be important.

The second requirement for data qualification is to establish, in broad terms, the quality of the data. This can be achieved by computing from the nominated areas of interest trial RMS time histories and amplitude probability densities (Section 2.3.2.) for selected data channels. An RMS time history is obtained by computing the mean square value of a data channel (Section 2.3.1.) for a series of samples extracted from the nominated area of interest and plotting the square root of the results against the time corresponding to the mid-point of the appropriate sample. If such analyses give reasonable results, then this is normally justification enough for commencing detailed analysis of the nominated areas.

3.3. Data Handling

It has been observed in Chapter 2 that descriptions of the characteristics of random data are provided in the form of averages. The extraction of such averages may require the processing of large quantities of data. For example, a 14 track FM magnetic tape recording conforming to the Intermediate Band IRIG standard can contain the equivalent of over 10^5 data samples per foot of tape, and the analysis of a complete flight record can mean the handling of the equivalent of over 10^8 data samples. Coping with such quantities of data on a routine basis makes special demands on a processing facility. This section is devoted to a discussion of the requirement created by those demands.

The process of describing the characteristics of random data is inevitably a reduction process. This is because the number of reliable statements which can be made about a random signal is, by definition, very limited compared with the number of data points (or equivalent) contained within the data itself. The degree of reduction can be of the order 1000-1. In the majority of cases, random data is recorded prior to analysis, and when this is the case there seems to be no point in re-recording the raw data except to make it more readily accessible to the user. Indeed, re-recording is an expensive process in terms of the cost of effecting the transfer, the cost of the recording medium, and in terms of the floor area required to store the records. Clearly, when large quantities of data are involved, there is a case for making the original record accessible and using this directly whenever the data is to be processed.

The philosophy described above, of analysing the original record, implies that the data be analysed "on-line", that is the acquisition module should include the ability to reduce the data at the same rate that the data is acquired. The number of "on-line" reduction processes required for the majority of applications is, fortunately, limited and can be incorporated readily into a dedicated processor or can be provided by a small number of special purposes units. The routines primarily required are as follows:

- RMS and mean values as a function of time
- Amplitude probability density
- Sample acquisition
- Acquisition of peak values
- Power spectral analysis
- Power and cross spectral analysis.

Secondary routines which may be useful are:

- Autocorrelation
- Cross-correlation.

The first two routines are required primarily for data qualification purposes, and the results obtained are not normally required to be archived.

Sample acquisition is required for acquiring transient "events" and short samples of random data for which special purpose analyses are required. The process involves extracting a number of samples (or sampling an analogue signal) at constant time intervals for each of a specified set of data channels. The exact channels and the number of samples per channel (or acquisition time) are selected by the operator. The start of acquisition should be logic controlled so that it may be triggered by the operator, by an event (e.g. tape recorder switch on, or a particular recorded time), or by a nominated data channel exceeding a specified level (the start of a transient). Flexibility of operation is greatly enhanced if the start of the acquisition can be advanced or delayed relative to the sensed event. Acquired data is normally scaled to engineering units and filed in preparation for "off-line" analysis.

Peak value information may be required when fatigue-related calculations are to be performed upon random data. The acquisition of peak values involves sampling the data and computing the level and time of occurrence of each sensed maximum and minimum. In order to inhibit the acquisition of large numbers of peaks caused by noise, it is normal to include a "gate" which rejects peaks within a specified range of the previous trough, and vice versa. This type of acquisition is normally controlled by calling for a nominated number of peaks, but limited by a specified overall acquisition time. The data arrays for each channel contain, on completion, the levels of each peak and trough

and the time of occurrence relative to the start of the acquisition. These arrays are normally scaled to engineering units and filed in preparation for further "off-line" analysis.

Power and cross spectral analysis involves large amounts of processing. This activity will, in general, define the performance required of an on-line processor. A practical digital analyser can be designed so as to extract the cross spectral density of two channels, and the power spectral density of each channel from a single DFT. Alternatively the power spectral density of each of two channels can be computed, and the envelope of all individual estimates can be stored for each channel. The latter function, when divided by the estimated power spectral density, represents a sensitive indicator of the presence of periodic, or near periodic, components and of the presence of non-stationary occurrences in a signal.

In order to make the best use of the available data, the acquisition routine should be interleaved with the processing routine so that sampling can be continuous. This implies that the maximum acquisition rate of the system will be governed by the computing power of the processor.

A major disadvantage of on-line primary processing is that pre-processing activities (see Chapter 1) must also take place on-line prior to processing. When the data is analogue, then this may be effected by analogue filters prior to sampling. However, when the data is digital, then hardware must be available to simulate the analogue filters. This can be effected using the recursive digital filters described in Section 1.5.2. A system which can be used for all types of on-line processing is shown schematically in figure 3.1.

In all cases, processed data should be filed together with the relevant analysis parameters. Typically, these might include:

- Flight identifier
 - Flight data
 - Flight record identifier
 - Calibration data identifier
 - Acquisition rate
 - Number of samples per transform (if relevant)
 - Details of pre-processing filters (high pass, anti-aliasing, etc.)
 - Channel identifier
 - Acquisition start time
 - Length of acquisition time
 - Pre-sampling gain (if relevant)
-) for each data file

3.4. Data Presentation

However high the quality of the original data, and however precise the analysis tools used to operate upon the data, the usefulness of the results obtained from any trial depends upon the way the results are presented. The general rules for presentation of the results from any analysis are:

- the information should be assimilable
- should include traceability information
- should contain parameters necessary for assessing the quality of the data.

The first rule implies that the data presentation should be graphical, in engineering units and should include titles which identify the record and the measurement channel involved, bearing in mind that the presentation may be kept longer than the engineer responsible for it. Further, the presentation should be in a form which highlights those characteristics of interest. Power spectral density functions, for example, are normally

in a log-log format. This may be the right presentation in general terms, but the engineer interested in examining in some detail the spectral details at high frequency may be better served by a log-linear, or even a linear-linear presentation. Yet again, an engineer concerned with the modal damping characteristics of a structure might be more interested in a linear-square presentation.

Tracibility information is necessary to identify the following information with no ambiguity:

- The raw data record from which the presentation was produced
- The calibration information used to scale the raw data
- The suite of programmes used to reduce the raw data
- The location of the reduced data

The parameters for assessing the quality of the results vary from presentation to presentation. Broadly, they are those parameters which are necessary to estimate the statistical and, where relevant, the bias errors of the results.

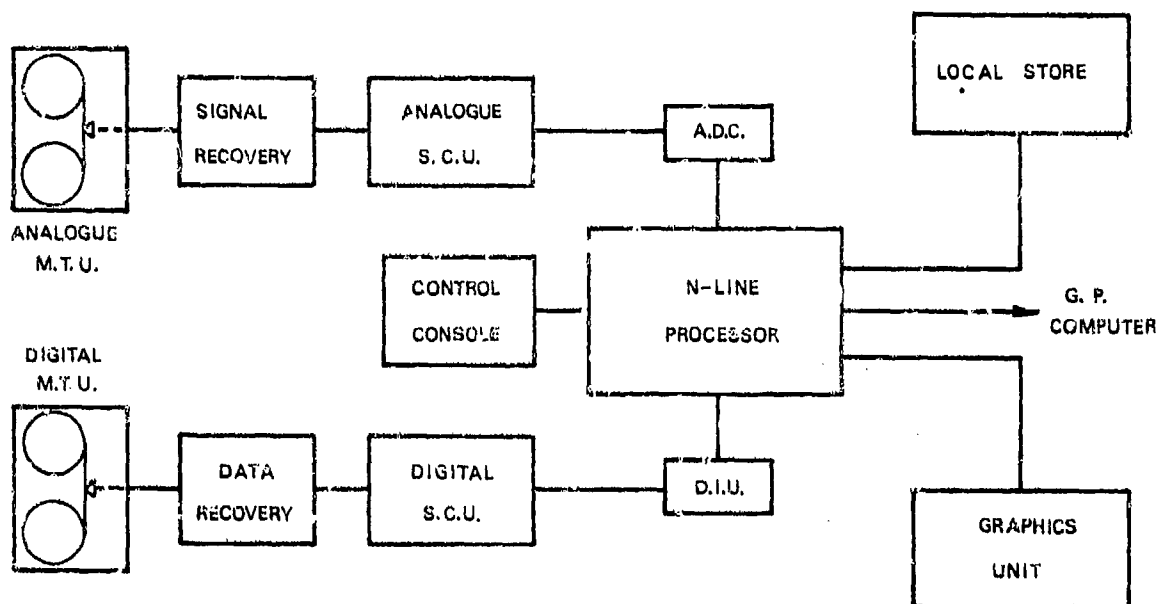


Figure 3.1. A Facility for Processing Random Data.

CHAPTER 4 SOME APPLICATIONS OF RANDOM DATA ANALYSIS

4.1. Examples of Spectral Analysis Computations.

The results which may be obtained from power and cross spectral analysis of random data depend, of course, upon the characteristics of that data, but also upon the various decisions the engineer responsible is called upon to make during the analysis. Those decisions relate to obtaining a satisfactory compromise between the detail which is to be extracted from the data and upon the reliability of that detail, i.e. between quantity and quality.

Sections 2.2 and 2.3 include relationships which are intended to assist the engineer in making the choice of analyser bandwidth which is appropriate for a given length of record. This Section contains results extracted from analyses of two types of signal. They are intended to provide a "feel" for the consequences of choosing any particular combination of analyser bandwidth and averaging time. The Section also contains results which demonstrate the effect of "Hanning" each block of data prior to analysis.

The signals used for the analysis were a "white" noise source band limited at 1 KHz., and a narrow band signal produced by passing the white noise through an analogue filter. The filter used was a second order bandpass filter centred at a nominal frequency of 159.2 Hz. with a damping factor of two percent of critical. The behavior of the filter is illustrated in figure 4.1.1., which shows the response of the filter in the time domain to a step input.

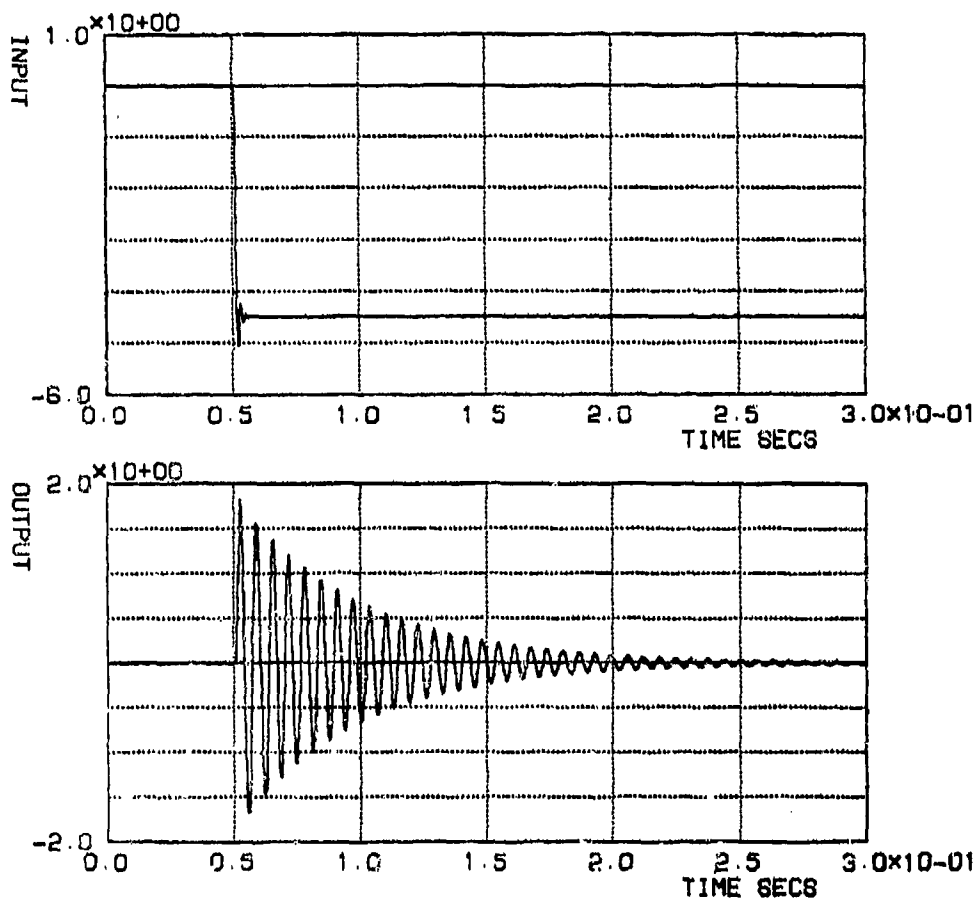


Figure 4.1.1. The Response of a Narrow Bandwidth Band Pass Filter to a Step Input.

The spectral analyser used for the analysis used a DFT algorithm for computing the power spectral density of each signal and the cross spectral density of one signal relative to the other. Auto and cross covariance functions were obtained by transforming the appropriate frequency domain results. Data were acquired at $10^4/7$ samples per second per channel, and the signals were conditioned prior to sampling by first order high pass filters having a time constant of one second, and by anti-aliasing filters cutting off at 675 Hz. with a roll-off rate of 72 dB per octave (signifying 12 poles.) The results were corrected to remove the effects of both filters. The parameters used during each analysis are identified on the figures by the following:

- N - the number of samples per channel used in each DFT transform
- M - the number of transforms used to compute the information presented
- H - a "Hanning" marker: 0 = No Hanning, 1 = Hanned raw data.

The remainder of this Section is devoted to a discussion of the results of spectral analysis and related computations presented in figures 4.1.2. to 4.1.11. inclusive.

4.1.1. The Effect of Varying Averaging Time.

Figures 4.1.2. and 4.1.3. demonstrate the influence of averaging time upon the reliability of the power spectral density estimates for each type of signal.

Figure 4.1.6. contains the inverse transforms of the above results.

Figure 4.1.8. demonstrates the effect of averaging time upon the reliability of frequency response functions computed from the cross spectral densities.

Figure 4.1.9. contains the corresponding inverse transforms of the cross spectral densities.

The figures show that the average of around one hundred sample blocks is required in order to describe the signals with a reasonable degree of reliability in the frequency domain, with power spectra apparently converging less rapidly than the corresponding frequency response functions. In practice, should the quantity of data available be insufficient to allow the average of approximately one hundred transforms to be obtained with the chosen length of transform, then it is to be recommended that either the transform length be reduced, or frequency domain averaging be employed. It should be assumed that the dissemination of unreliable results will lead to mis-interpretation of the data.

It will be observed that the auto-covariance function for the broad band signal contains information which is not significant after the first few time delays. The function, as presented in figure 4.1.6. would not be very useful to an analyst. A re-plot of the first twenty or so milliseconds would be justified if the information was required to be examined in detail.

The auto-covariance functions for the narrow band signal illustrate the problems of interpreting such information. The results contained in figure 4.1.6. are open to erroneous interpretation even after 100 sample blocks have been averaged - at first sight the plot indicates that there may be two modes present separated by approximately 2 Hz. Only after 1000 sample blocks have been averaged is it apparent that just one mode exists.

It should be noted that the computed RMS values for the broad band signal are more stable than those for the narrow band signal (figures 4.1.2. and 4.1.3.), a result anticipated by the expression for the statistical reliability of mean square value estimates contained in Section 2.3.1.

It may be seen that the power spectral density estimates for the narrow band signal, figure 4.1.3., do not conform to the expected shape at very low levels (a slope of 2 decades per decade, or 6 dB per octave, was anticipated.) This is the effect known as "side band leakage" and is caused by using a "boxcar" data window (individual data blocks were not weighted in any way prior to transformation.)

4.1.2. The Effect of Varying Bandwidth.

Figure 4.1.4. indicates the effect of analyser bandwidth on the power spectral density estimates for the narrow band signal. Corresponding information for the broad band signal has not been included because analyser bandwidth is not important for signals of this type.

Figure 4.1.7. contains the inverse transforms of the above results.

Figure 4.1.10. shows the effect of analyser bandwidth on the frequency response functions computed from the cross spectral densities.

All the results shown in the above figures were obtained using the average of 100 transforms. As the analysis bandwidth was increased, the averaging time was reduced accordingly to keep the bandwidth-time product constant.

Estimated power spectra for the narrow band signal (figure 4.1.4.) vary significantly over the range of bandwidths presented, the perceived peak amplitude reducing from approximately $0.43 \text{ units}^2/\text{Hz}$ at a bandwidth of 0.6975 Hz . to approximately $0.21 \text{ units}^2/\text{Hz}$ at a bandwidth of 11.1607 Hz . This effect is accompanied by an increase in amplitudes at off-peak frequencies to maintain roughly a constant RMS value. Thus as the analysis bandwidth was increased, the apparent damping of the mode increased.

The apparent increase in damping with increased analyser bandwidths is carried across to estimates of the frequency response function (figure 4.1.10.) Also apparent from these results is a decreased level of coherence as the analyser bandwidth is increased.

The increase in apparent damping is due to a bias error caused by the spectral averaging effect of individual analyser filters. The effect was predicted in Section 2.4.3. (The relationships shown in figure 2.4.7. were deduced experimentally from the results presented here by estimating the apparent system parameters using frequency domain identification.) The observed decrease in coherence is associated with the finite length of data used to calculate each DFT. The excitation of the narrow band filter by an event extends for a significant length of time. The result is that the cause and effect of a particular event could be in different transform blocks, thereby reducing the estimated coherence between the two signals.

The auto-covariance functions contained in figure 4.1.7. reflect the comment made with regard to the effect of analyser bandwidth upon estimates of power spectra. Here, however, the effect of decreased block length is demonstrated more explicitly: the maximum time delay which could be achieved using an analyser bandwidth of 11.1607 Hz . was reduced to 0.0448 seconds (compared with 0.7168 seconds for a bandwidth of 0.6975 Hz .)

4.1.3. The Effect of Hanning.

Figure 4.1.5., which should be compared with figure 4.1.4., demonstrates the effect on estimates of power spectral density of "Hanning" each block of raw data prior to analysis. The comparison shows the improved "side-band leakage" properties which are associated with Hanning (see Section 2.4.1.) Observe, however, the increase in apparent damping particularly at high analyser bandwidths: the perceived peak amplitude at a bandwidth of 11.1607 Hz . ($0.21 \text{ units}^2/\text{Hz}$ without Hanning) is reduced to approximately $0.16 \text{ units}^2/\text{Hz}$. The effect is reflected in the auto-covariance estimates contained in figure 4.1.7. which demonstrates the apparent increase in damping in a more dramatic way.

Figure 4.1.11. contains estimates of frequency response function using Hanned raw data. They should be compared with the equivalent estimates presented in figure 4.1.10. which were obtained without Hanning. The conclusions to be drawn are similar to those noted above for power spectra. Here, however, the effect is more serious with an obvious drop in the coherence function at frequencies close to the natural frequency of the filter. It should be noted that, at frequencies outside this problem area, the effect of Hanning was to increase estimates of coherence.

Thus the effect of applying a Hanning window to the raw data blocks is, apparently, to improve the stability of frequency response function estimates in areas which do not include resonances, but also to increase the bias error at frequencies close to a resonance. In general, an analyst would be most interested in the shape of a frequency response function at frequencies close to resonances: thus Hanning should not normally be used when frequency response functions are the primary objective of an analysis.

In Summary, the results provide an insight into the consequences of assuming particular values of averaging time and analyser bandwidth upon the resulting estimates of power spectral density and frequency response functions. They indicate that Hanning should not normally be adopted for computing cross spectra, and that the use of Hanning for power spectral density estimates does not improve the estimates significantly, provided that sufficient data is available to achieve a reliable average.

It is also observed that auto-covariance functions for narrow band signals converge only slowly to a result which is free from potentially misleading ambiguities.

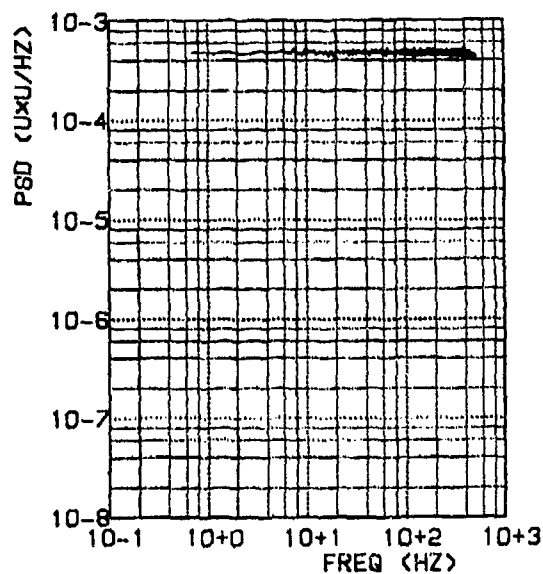
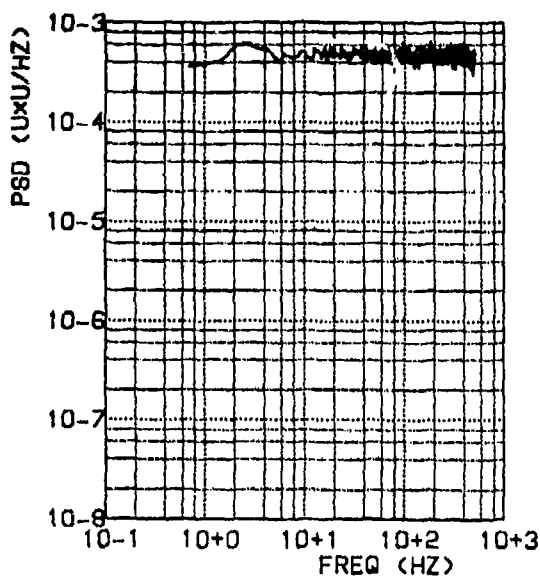
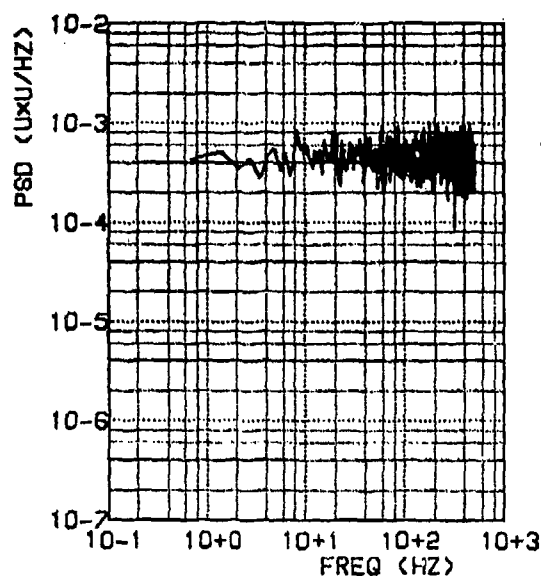
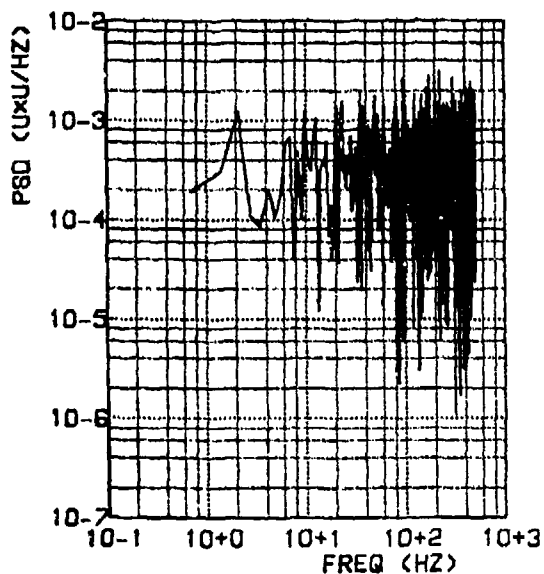
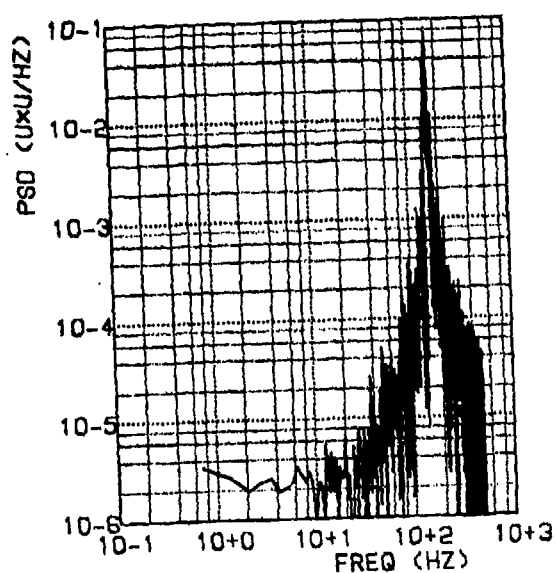
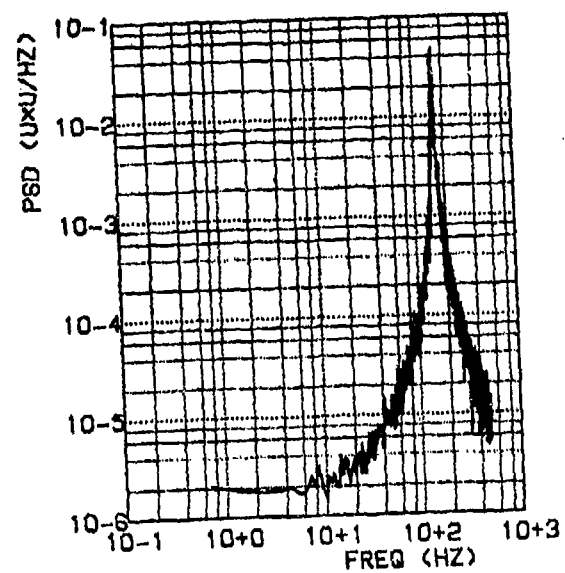


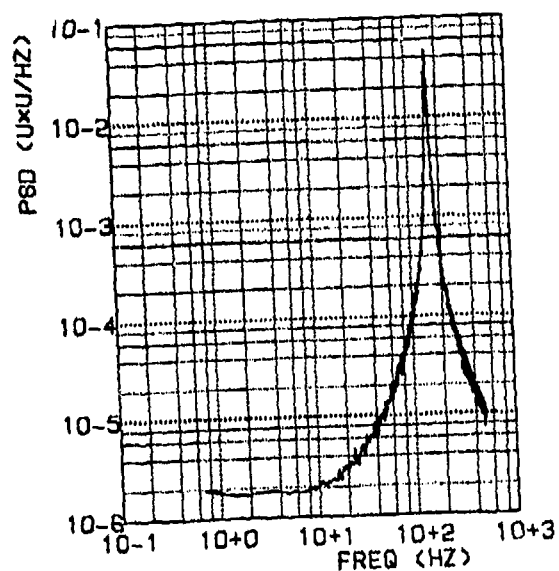
Figure 4.1.2. Effect of Averaging Time on PSD Estimates
(Broad Band Signal)



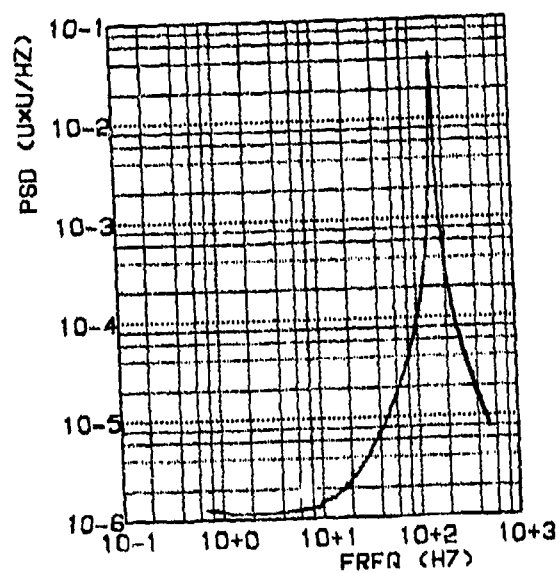
Bandwidth = 0.6975 Hz
 RMS Value = 0.6633
 N = 2048 M = 1 H = 0



Bandwidth = 0.6975 Hz
 RMS Value = 0.6867
 N = 2048 M = 10 H = 0

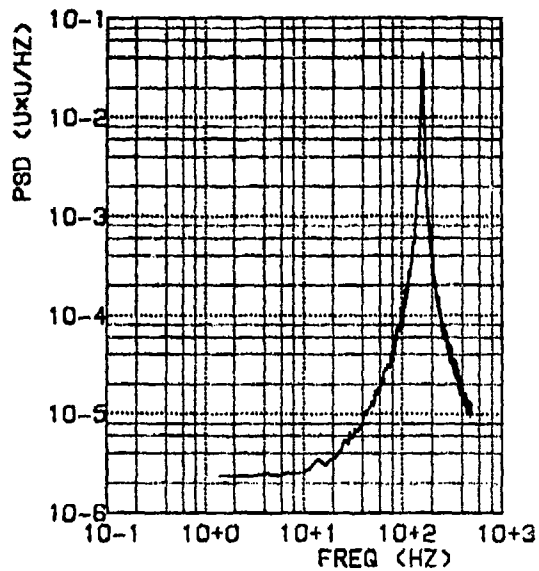


Bandwidth = 0.6975 Hz
 RMS Value = 0.6834
 N = 2048 M = 100 H = 0

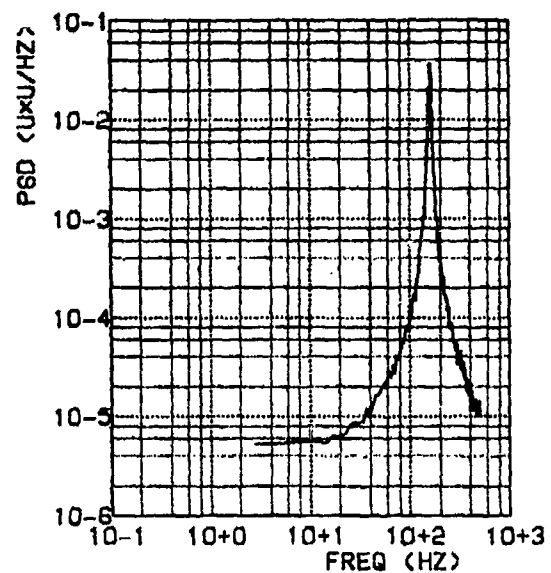


Bandwidth = 0.6975 Hz
 RMS Value = 0.6841
 N = 2048 M = 1000 H = 0

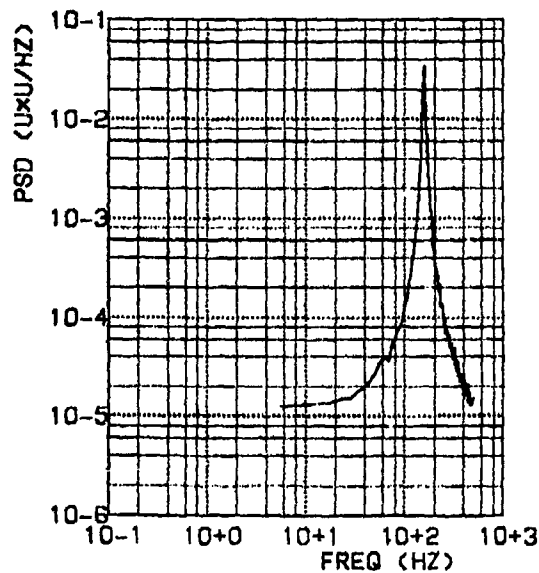
Figure 4.1.3. Effect of Averaging Time on PSD Estimates
 (Narrow Band Signal)



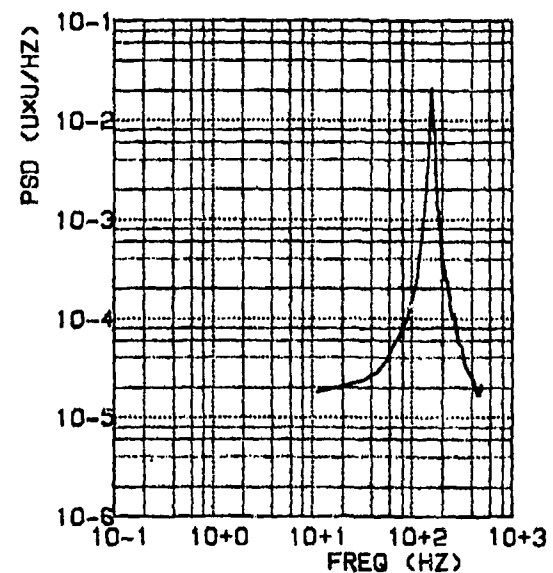
Bandwidth = 1.3951 Hz
 RMS Value = 0.6804
 N = 1024 M = 100 H = 0



Bandwidth = 2.7902 Hz
 RMS Value = 0.6727
 N = 512 M = 100 H = 0

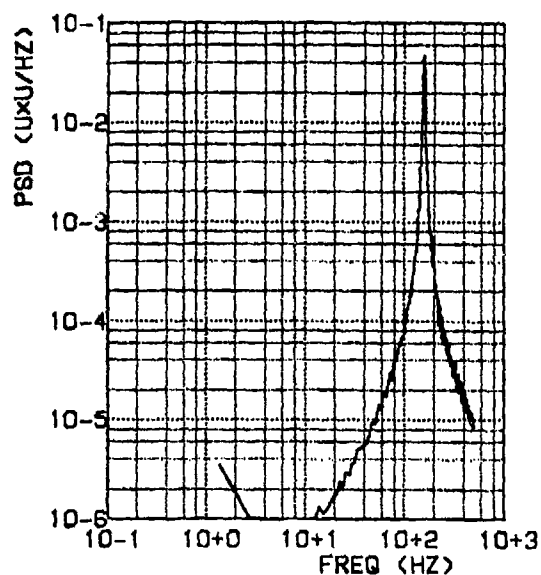


Bandwidth = 5.5804 Hz
 RMS Value = 0.6794
 N = 256 M = 100 H = 0

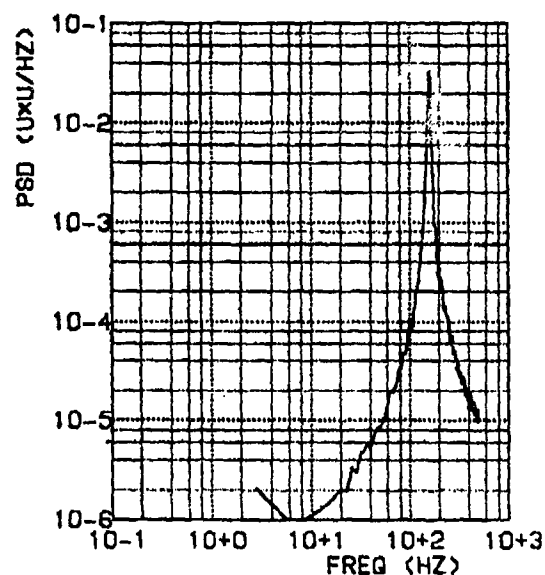


Bandwidth = 11.161 Hz
 RMS Value = 0.6508
 N = 128 M = 100 H = 0

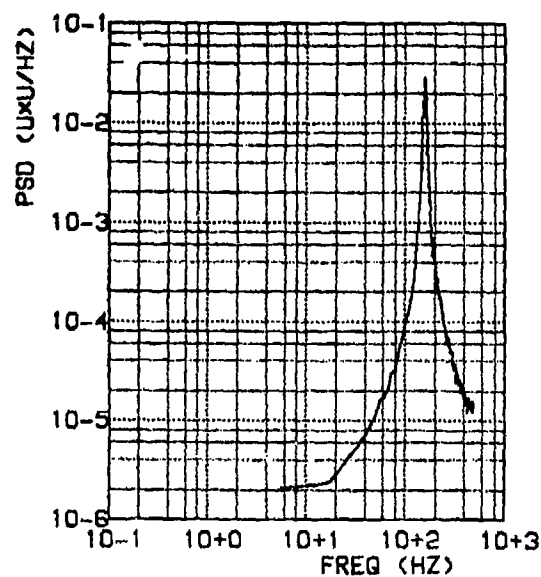
Figure 4.1.4. Effect of Analyser Bandwidth or PSD Estimates
 (Narrow Band Signal)



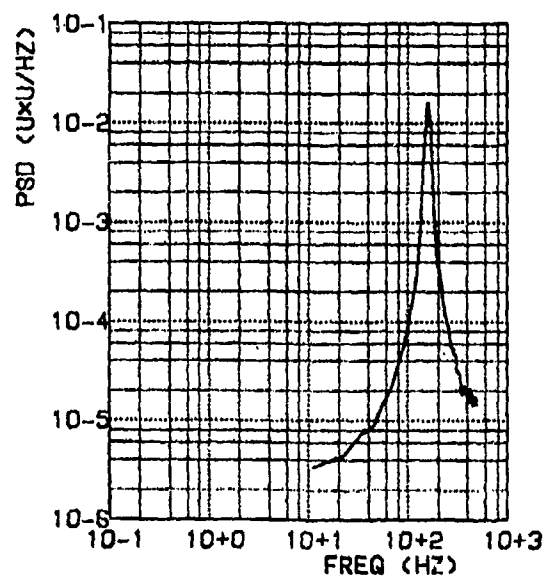
Bandwidth = 1.3951 Hz
 RMS Value = 0.6797
 N = 1024 M = 100 H = 1



Bandwidth = 2.7902 Hz
 RMS Value = 0.6765
 N = 512 M = 100 H = 1



Bandwidth = 5.5804 Hz
 RMS Value = 0.6783
 N = 256 M = 100 H = 1



Bandwidth = 11.161 Hz
 RMS Value = 0.6602
 N = 128 M = 100 H = 1

Figure 4.1.5. Effect of Hanning on PSD Estimates
 (Narrow Band Signal)

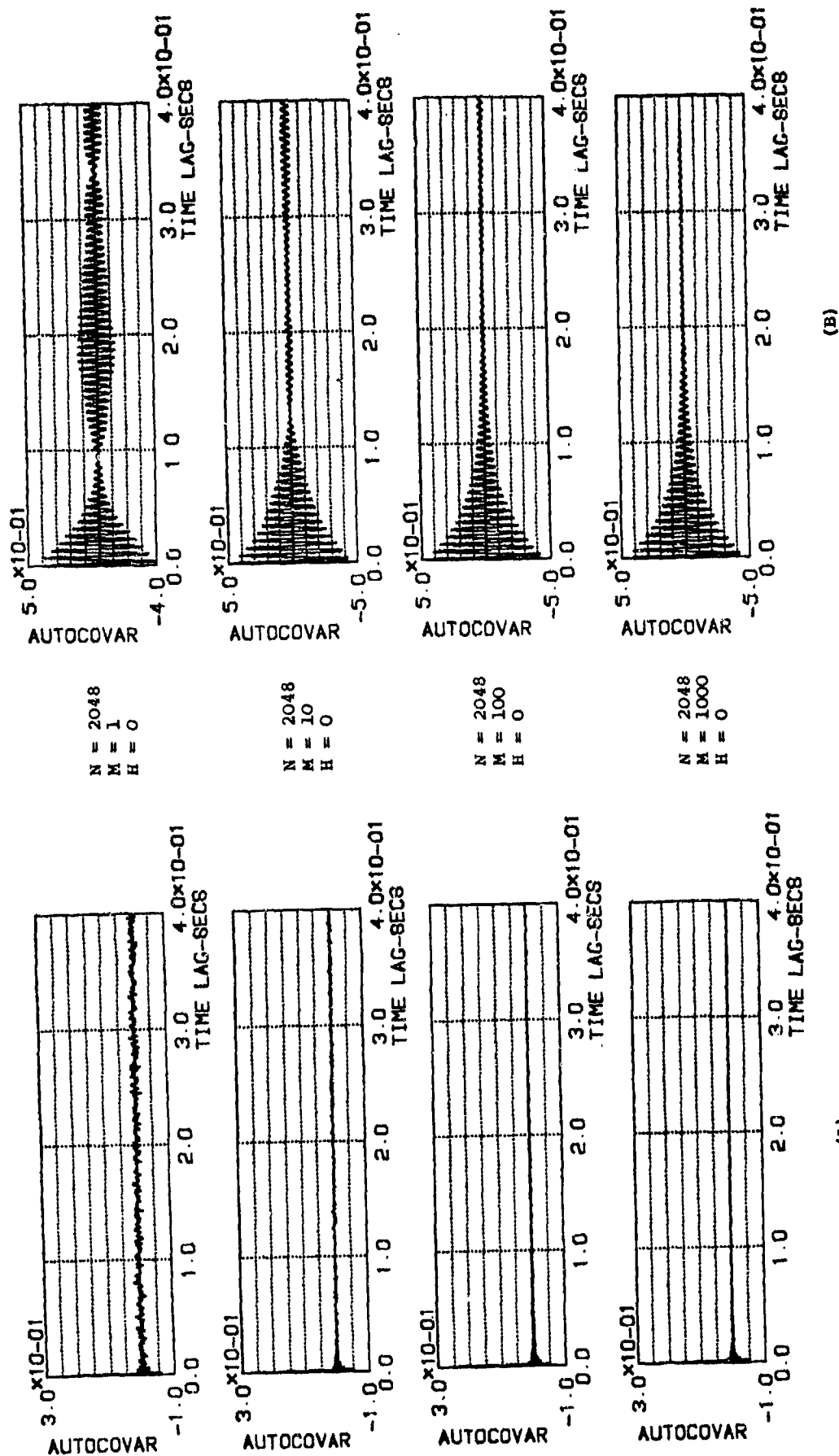


Figure 4.1.6. Effect of Averaging Time on Auto-Covariance Function Estimates for Both Broad (A) and Narrow (B) Band Signals

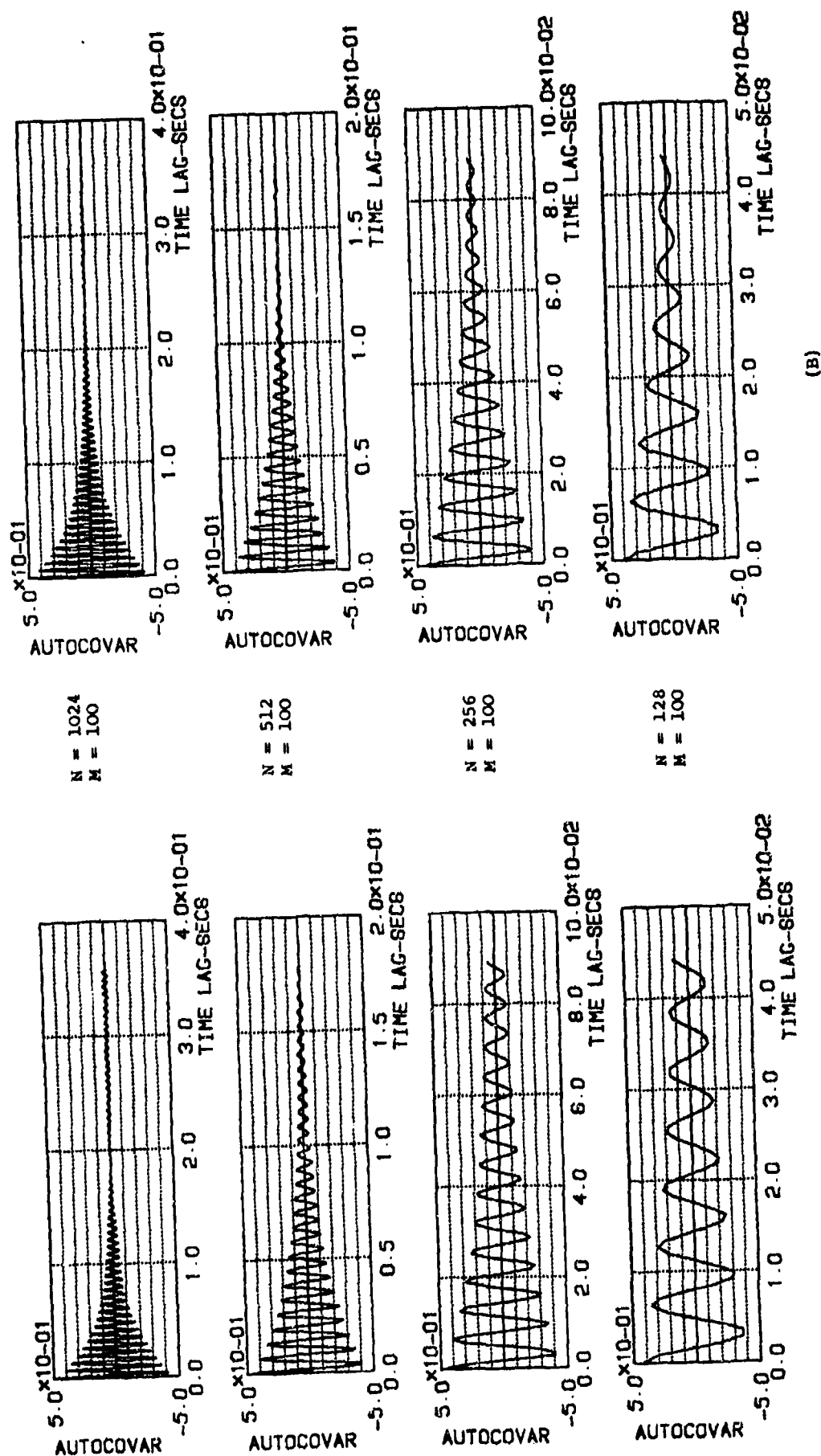
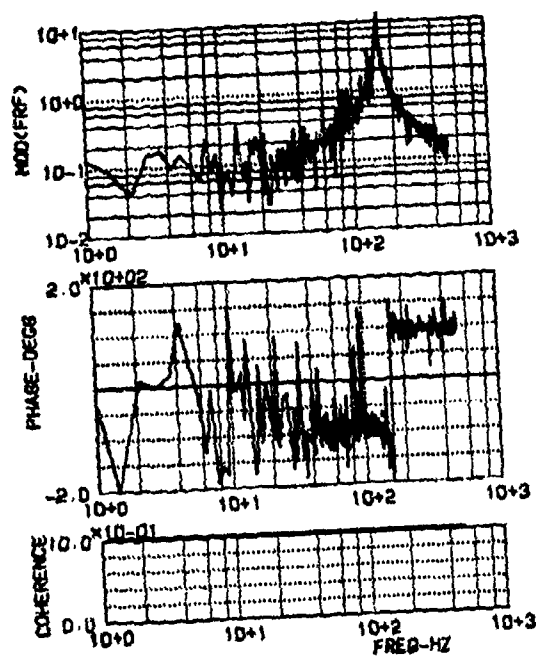
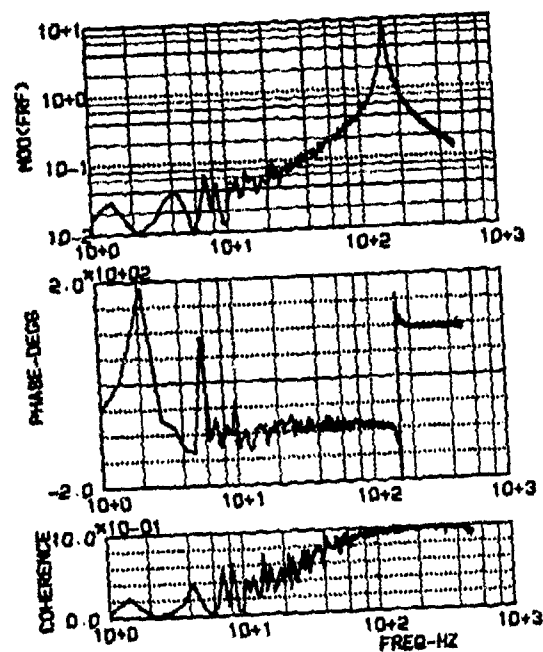


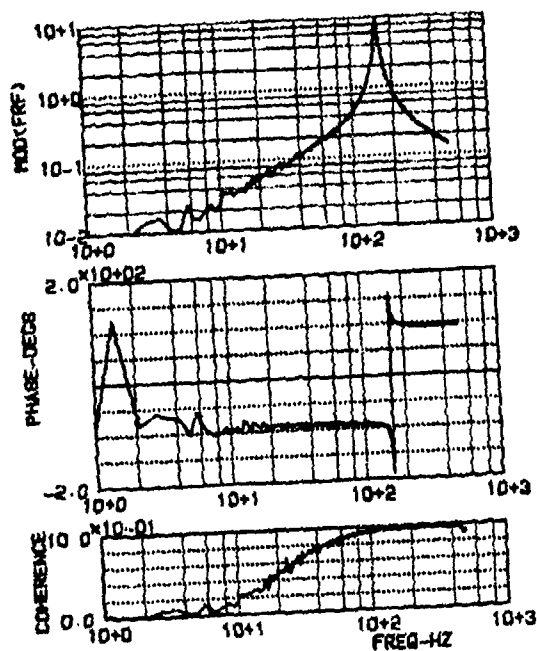
Figure 4.1.7. Effect of Analyser Bandwidth on Auto-Covariance Function Estimates for Both Unsmoothed (A) and Smoothed (B) Results



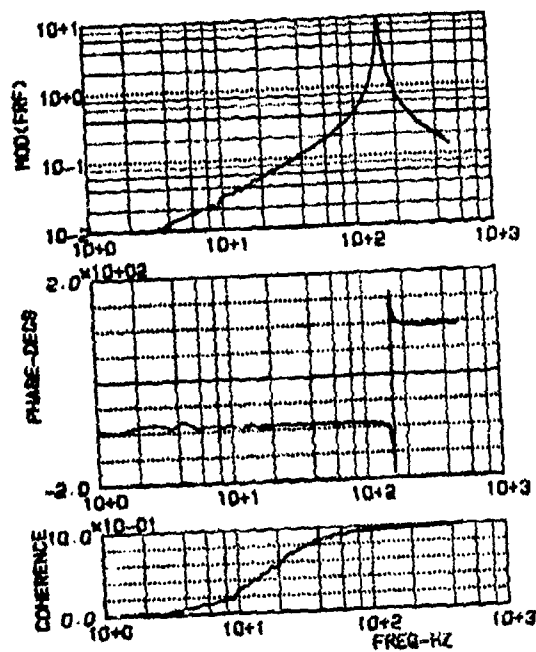
$N = 2048 \quad M = 1 \quad H = 0$



$N = 2048 \quad M = 10 \quad H = 0$



$N = 2048 \quad M = 100 \quad H = 0$



$N = 2048 \quad M = 1000 \quad H = 0$

Figure 4.1.8. Effects of Averaging Time on Estimates of Frequency Response Functions

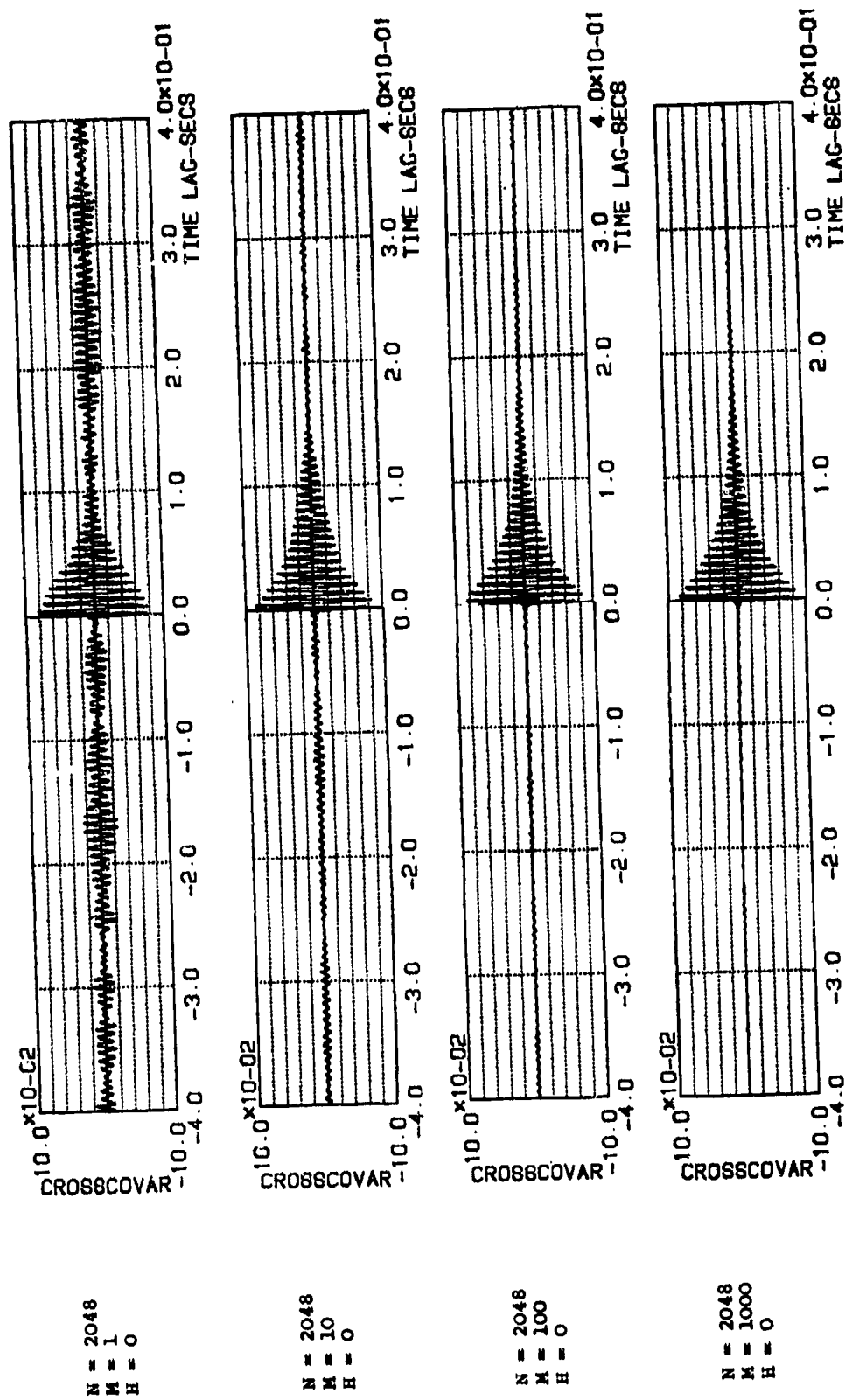
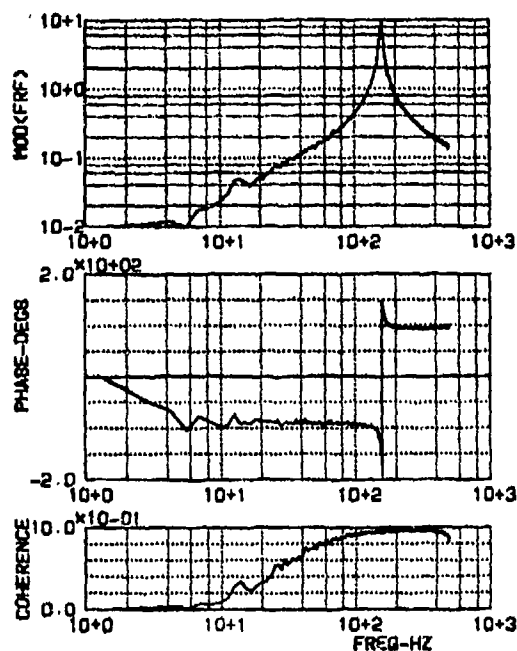
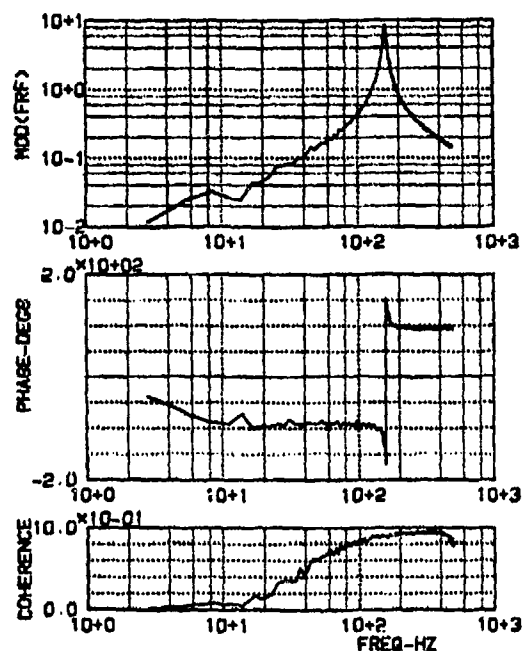


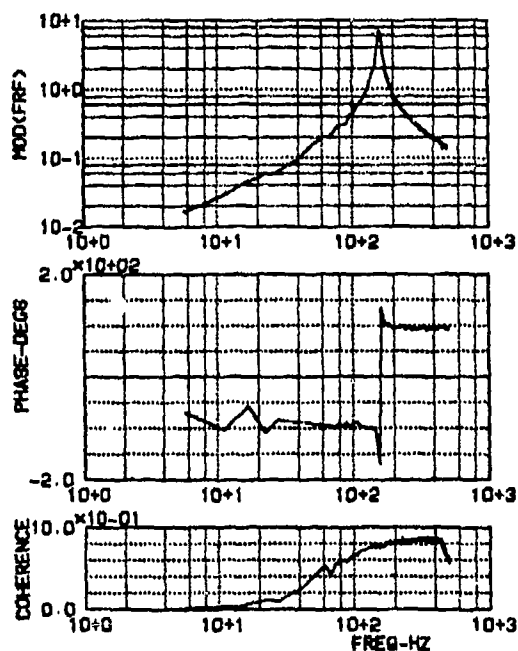
Figure 4.1.9. Effect of Averaging Time on Estimates of Cross-covariance Functions



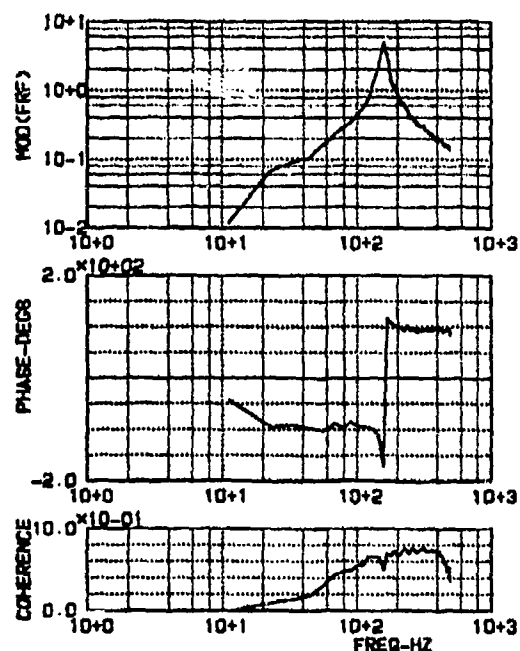
$N = 1024 \quad M = 100 \quad H = 0$



$N = 512 \quad M = 100 \quad H = 0$

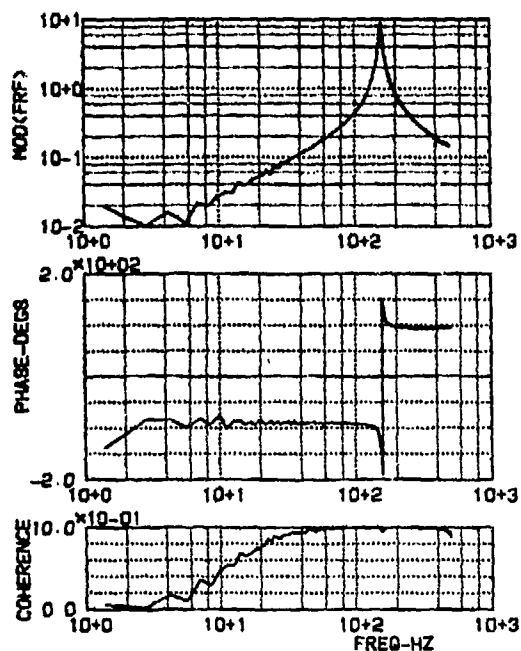


$N = 256 \quad M = 100 \quad H = 0$

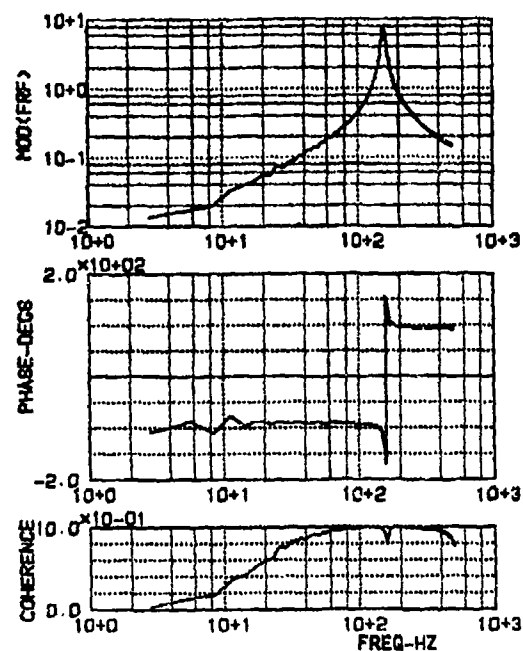


$N = 128 \quad M = 100 \quad H = 0$

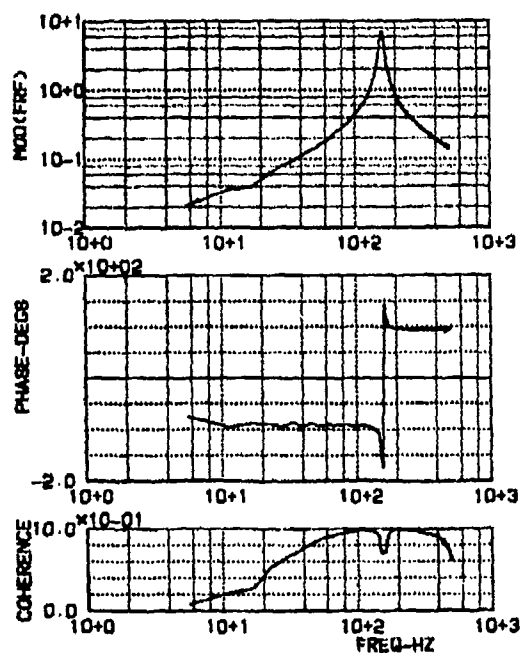
Figure 4.1.10. Effect of Analyser Bandwidth on Frequency Response Function Estimates (No Hanning)



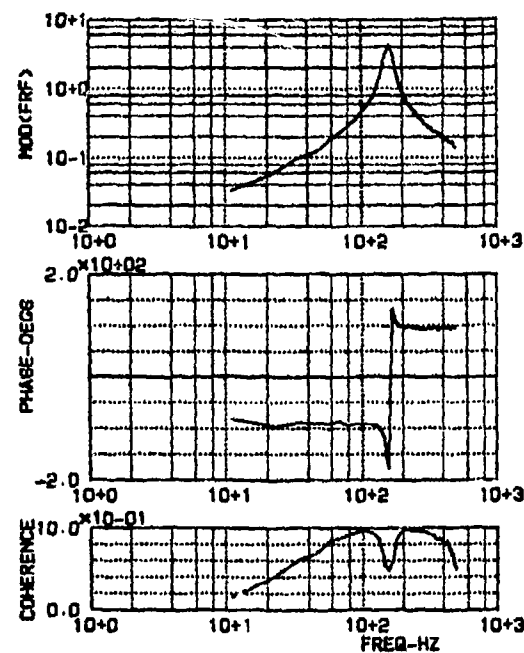
$N = 1024 \quad M = 100 \quad H = 1$



$N = 512 \quad M = 100 \quad H = 1$



$N = 256 \quad M = 100 \quad H = 1$



$N = 128 \quad M = 100 \quad H = 1$

Figure 4.1.11. Effect of Analyser Bandwidth on Frequency Response Function Estimates (With Hanning)

4.2. The Measurement of Atmospheric Turbulence.

Atmospheric turbulence is specified in terms of three orthogonal components of velocity. For the duration of particular atmospheric conditions, an adequate description of the turbulence is often considered to be provided by the amplitude probability density, the power spectral density and the spatial autocovariance function of each component of velocity. When viewed from an aircraft travelling through the turbulence, the description can usually be described by the amplitude probability density and power spectral density of the vertical and lateral aircraft axis-related components, since the scale of turbulence is usually large enough in relation to the size of the aircraft to justify an assumption of perfect spatial correlation. Further, gust velocities can be assumed to be Normally distributed everywhere except at altitudes less than approximately 100 metres. Under these conditions, atmospheric turbulence can be described by the power spectral density distributions of the vertical and lateral components of gust velocity. The methods by which these may be deduced from flight measurements are the subject of this Section. For the sake of simplicity the exposition is confined to the vertical component of gust velocity. However, results equivalent to those presented here can be deduced directly for the lateral component.

The measurement of the vertical component of gust velocity from a moving platform is complicated by the fact that the platform itself can translate relative to its mean trajectory. The measurement is normally made using an incidence vane mounted on the vehicle. These are then corrected using measurements obtained from locally mounted inertial transducers. Specifically, using the notation shown in figure 4.2.1.,

$$\frac{w}{V} = B + \int_0^t (\ddot{z} - \ddot{v}) dt + C \quad 4.2.1$$

The constant of integration, C, can be neglected since only the dynamic component of gust velocity is normally required. Hence,

$$w = VB + \int_0^t (\ddot{z} - \ddot{v}) dt \quad 4.2.2$$

The evaluation of equation 4.2.2. is difficult to implement because any offset whatsoever in either of the inertial measurements will result in the integral term "drifting" away without limit. A practical compromise can be obtained by replacing the integrator by a low pass filter (see Section 1.5) and setting the time constant of the filter to a value which is below the frequency range of interest.

The vertical component of gust velocity may thus be extracted from measurements of vane incidence, acceleration and pitch rate at the vane mounting point, by implementing equation 4.2.2. as shown in figure 4.2.2. The integrator is replaced by a low pass filter, and the vane incidence measurements are high pass filtered (using the same time constant as that used in the low pass filter) to maintain the validity of the reconstitution. The circuit can be described in the frequency domain, as follows,

$$\frac{w \cdot i f / f_0}{1 + i f / f_0} = \frac{V \cdot B \cdot i f / f_0}{1 + i f / f_0} + \frac{(\ddot{z} - \ddot{v})}{1 + i f / f_0} \quad 4.2.3.$$

It is clear from equation 4.2.3. that the reconstituted gust velocity is viewed through a high pass filter having a half power frequency of f_0 .

The circuit shown in figure 4.2.2. can be used to process the measurements either during flight prior to recording, or after replaying the recordings prior to processing. If it is used in flight before recording, then the circuit effectively reduces the number of parameters to be recorded, but difficulties can be caused by the need to factor two channels by the true airspeed of the flight vehicle. These difficulties can be eased somewhat by computing the gust angle, w/V , rather than the gust velocity, when only one parameter, vane mounting point acceleration, has to be scaled.

Typical results which have been obtained by applying the circuit shown in figure 4.2.2. to the task of reconstituting the lateral component of gust velocity are shown in figures 4.2.3. to 4.2.6.

An alternative method of extracting the vertical component of gust velocity is to process the measurements in the frequency domain using the relationships developed in Section 2.5 (equation 2.5.54.) to yield

$$G_{WV}(f) = V^2 |H_B(if)|^2 + \frac{G_{\ddot{z}\ddot{z}}(f)}{4.(\pi f)^2} + \frac{V^2 G_{\dot{\theta}\dot{\theta}}(f)}{4.(\pi f)^2} + \frac{V \cdot \text{Re} \left\{ i H_B(if) \cdot G_{\ddot{z}\dot{\theta}}(if) - i V H_B(if) \cdot G_{\dot{\theta}\ddot{z}}(if) - \frac{1}{2\pi f} \cdot G_{\ddot{z}\ddot{\theta}}(if) \right\}}{2\pi f} \quad 4.2.4.$$

Here the frequency response characteristics of the vane (H_B) have been included for the sake of generality. For a perfect vane H_B is, of course, unity.

It is interesting to note that equation 4.2.4. can be used to calibrate an incidence vane dynamically from measurements obtained during manoeuvres executed in still air conditions. Satisfactory manoeuvres are those which excite the short period pitching mode in a random fashion. The frequency response function of the vane may be estimated by computing:

$$H_B(if) = \frac{1.(G_{\ddot{z}\ddot{z}}(if) - V.G_{\dot{\theta}\ddot{z}}(if))}{2\pi f V.G_{\ddot{z}\dot{\theta}}(f)} \quad 4.2.5.$$

The results obtained from such a trial are likely to provide good results over the frequency range 0.5 to 1.5 Hz for a conventional aircraft.

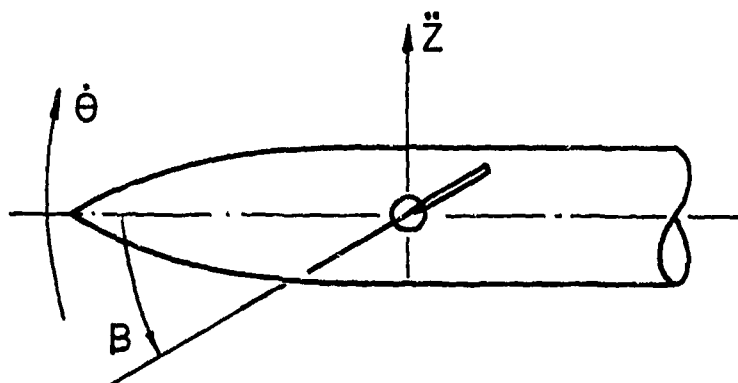


Figure 4.2.1. Sign Convention

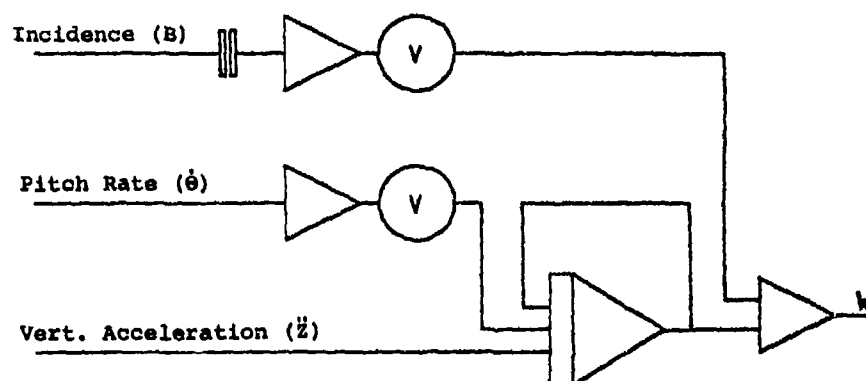


Figure 4.2.2. Analogue Gust Velocity Reconstitution Circuit

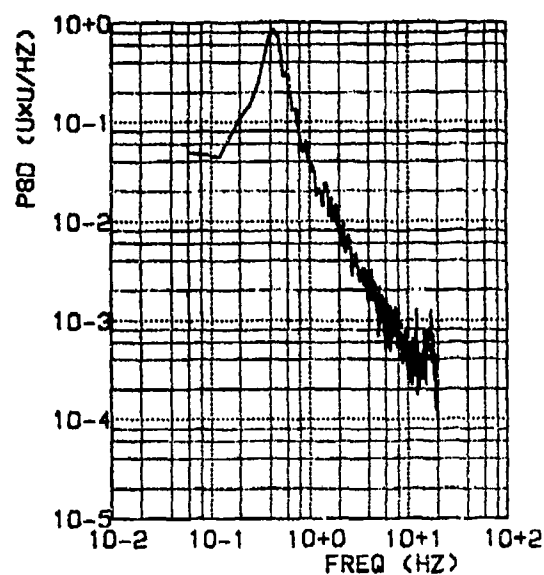


Figure 4.2.3. Vane Sideslip
(deg^2/Hz)

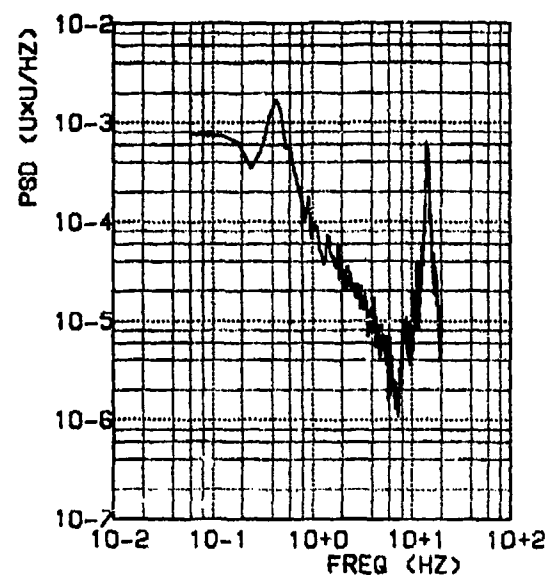


Figure 4.2.4.
Vane Root Lateral Acceleration
(g^2/Hz)

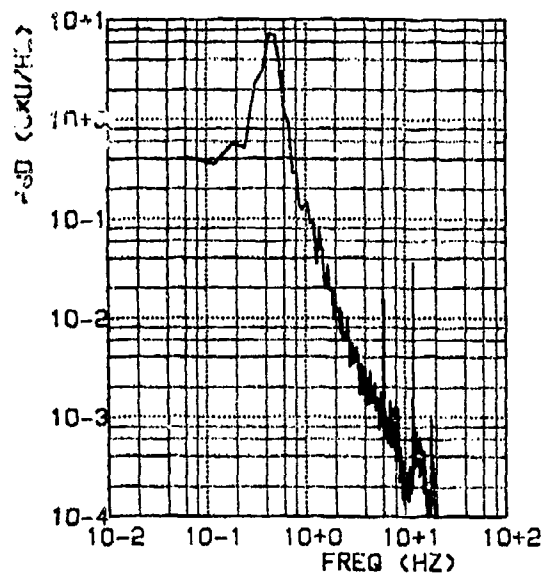


Figure 4.2.5. Vane Root Yaw Rate
($(\text{deg/sec})^2/\text{Hz}$)

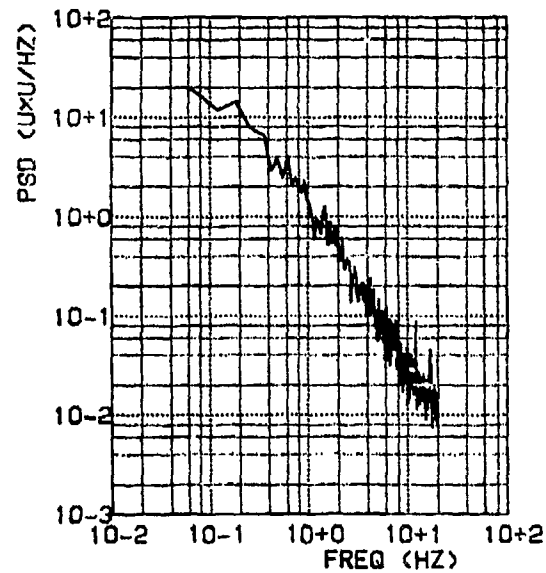


Figure 4.2.6. Lateral Gust Velocity
($(\text{ft/sec})^2/\text{Hz}$)

4.3. The Analysis of Records from Flight Flutter Trials.

Flight flutter trials are conducted in order to demonstrate that an aircraft is free from structural instabilities. Such instabilities can occur as a result of coupling between the flexible modes of an aircraft structure and the aerodynamic forces induced when the modes are excited.

The objective of a flight flutter trial is, therefore, to excite the flexible modes of the aircraft under investigation, or at least those modes which are considered to be critical. A trial is normally designed to cover a range of airspeeds and aircraft altitudes. During the periods of excitation, simultaneous recordings are taken of the excitation force (when appropriate) and the responses of accelerometers mounted at strategic points around the aircraft - normally close to the extremities where modal displacements are greatest.

The recordings made during a trial are required to yield the natural frequency and damping coefficient of each mode of interest. When sufficient accelerometer measurements are available, then estimates of the shape of each mode may be required.

A number of methods of excitation and techniques for analysing the measurements have been proposed and implemented in an effort to achieve estimates of parameters to the required accuracy. The difficulties in achieving this are that the estimation of natural frequency and, in particular, damping coefficient is an exacting task, and the measurements are often corrupted by the presence of one or more of the following sources of noise:

- control surface movement induced by the pilot or autostabiliser
- jet exhausts
- mechanical components (engines, pressurisation units, etc.)
- propellor or fan noise
- boundary layer pressure fluctuations
- atmospheric turbulence

Reference 4.1. contains a comprehensive survey of the techniques which are employed to execute, and to analyse the results of, flight flutter trials. However, the problems which may be encountered by the aeroelastician during the course of a series of flight flutter trials are faced by any engineer seeking to obtain information about the transfer function of a dynamical system from experimental measurements. Such measurements will contain, either by accident or design, random components which must be processed in order to elicit the required information. It was therefore considered to be relevant to include a discussion of the issues involved in the volume, with special reference to the analysis of flight flutter trials.

4.3.1. Excitation by Atmospheric Turbulence.

Atmospheric turbulence can excite an aircraft structure to a detectable level over a wide frequency bandwidth. It is for this reason, coupled with the fact that no special excitation equipment need be installed in the aircraft, that atmospheric turbulence has been considered as a means for exciting an aircraft for flight flutter trials.

Atmospheric turbulence can often be considered to be perfectly correlated spatially (see Section 4.2.) This implies that the wing span of the aircraft is small compared with the average wavelength of the turbulence, and that the aircraft passes through a point in space before the gust velocity obtaining at that point can change its value significantly. When this is the case, then the power spectral density of the response of a point on the aircraft can be described succinctly as,

$$G_r(f) = |H(if)|^2 G_{ww}(f) \quad 4.3.1.$$

where $G_r(f)$ is the PSD of the response

$G_{ww}(f)$ is the PSD of the gust velocity

and $H(if)$ is the transfer function of the aircraft

The transfer function includes components due to the aerodynamic forces in addition to those due to the characteristics of the aircraft structure. If the former, and the value of the PSD of gust velocity, change comparatively slowly with frequency, then their combined effect upon the response PSD can be assumed to be constant over the range of frequencies close to a resonance of the structure. The shape of the response PSD over the range of frequencies can then be used to identify the natural frequency and damping coefficient of the structural mode.

Several techniques have been used to identify the parameters. These include:

a) Half Power Frequencies.

When a mode is well separated from its neighbours, then its behavior resembles that of a single degree of freedom system. The natural frequency and damping coefficient of such a system can be estimated from the values of the two frequencies, f_1 and f_2 , at which the value of the PSD is half the peak value, provided that the system damping is low and the PSD of the excitation is substantially constant over the frequency range of interest. The parameters can be calculated using the following expressions.

$$f_0 = \frac{f_2 + f_1}{2} \quad (\text{natural frequency}) \quad 4.3.2.$$

$$z = \frac{f_2 - f_1}{2f_0} \quad (\text{damping coefficient}) \quad 4.3.3.$$

b) Direct Identification of the Parameters.

The simplest algorithm which can be used to identify the parameters is based (as above) on the assumption that the modes are well separated, and the mode under investigation can be represented by the transfer function of a single degree of freedom system. When this is the case, the response PSD in the region of the mode is given by

$$G_r(f) = \frac{K.(f/f_0)^2}{|1 - (f/f_0)^2 + 2isz/f_0|^2} + A + B(1 - (f/f_0)) \quad 4.3.4.$$

The parameters, including the dummy parameters A and B introduced to account for possible contamination from adjacent modes, can be identified using precisely the principle proposed in Section 1.5.1.3. It will be helpful, when deriving expressions for the rate of change of error with each parameter, to recall that

$$D(y) = y.D(\ln(y)) \quad 4.3.5.$$

where D is the derivative operator.

The idealisation represented by equation 4.3.4. requires some explanation. It was assumed for this that the response PSD was estimated from accelerometer data. A bandpass filter formulation was adopted to take approximate account of the shape of the PSD of gust velocity. Parameter B is optional, and may often be omitted without detriment to the results. Equation 4.3.4. can be extended to include two (or more) modes, thereby enhancing the applicability of the technique to situations where modes are grouped together. When this course of action is adopted, it is generally sufficient to assume uncoupled modes (i.e. assume no off-diagonal damping elements.)

c) Identification of Auto-covariance Functions.

The auto-covariance function of the response of a single degree of freedom system excited by white noise can be used directly to identify the natural frequency and damping coefficient of the system. After the first cycle or so (during which the function may be contaminated by noise and will contain information about the response of the system at frequencies well away from resonance), the reduction in amplitude of successive peaks or troughs in the response will follow, ideally, the law

$$\text{Peak Amplitude} = K.\exp(-2\pi gk) \quad 4.3.6.$$

where k is the peak (or trough) number.

Thus a plot of the logarithm of peak amplitude against peak number can yield the damping coefficient. The resonant frequency of the system can be computed simply by averaging the time between successive zero crossings having a positive (or negative) slope. The natural frequency can then be obtained as

$$f_0 = \frac{f_r}{(1 - s^2)^{1/2}} \quad 4.3.7.$$

The analysis of actual data is rendered rather more complicated by the fact that a number of modes will, in general, be present in the auto-covariance function of a measured response. When this occurs the normal procedure is to filter the function using a narrow bandpass filter (see Section 1.5.2.) centered upon the frequency of interest. The possibility of exciting transient responses of the filter is reduced by reversing the direction of time, so that the filter processes the auto-covariance function in the order of decreasing lag number. The technique was devised by Mazet for analysing measured responses obtained after a structure has been excited impulsively (see reference 4.1.)

The only advantage offered by the technique is that it is relatively simple to calculate the parameters manually. In all other respects the method described under b) is superior.

d) Random Decrement.

The principle of the Random Decrement technique is discussed in Reference 4.6. The results yielded when the technique is applied to the analysis of records obtained from flight flutter trials are similar to those derived from computing auto-covariance functions, and the comments which have been made with regard to that technique are equally applicable here. A possible advantage of the Random Decrement technique is, however, that slightly more stable results can be extracted from a given length of time history provided that suitable "gate" conditions are chosen.

Four methods of extracting natural frequencies and damping coefficients using atmospheric turbulence as the source of excitation have been outlined. Of these, the second method has proved to be reliable and economical when applied to PSD estimates. However, it should be remembered that the data has already been reduced and the results will, in general, contain bias errors as well as variance errors (see Section 2.4.) In fact it can be shown that only auto-covariance functions which have been computed directly and Random Decrement signatures are capable of yielding unbiased estimates for, in particular, damping coefficients. Even here, the use of filters to isolate particular modes will introduce a bias error.

The minimum averaging time required to obtain reliable PSD estimates can be deduced from the examples presented in Section 4.1. and the results presented in figure 2.4.7. From the latter, assuming that the raw data is not to be "Hanned" and allowing a ten percent error in estimates of damping coefficient, the required analyser-to-filter bandwidth ratio is 0.4. From equation 4.3.3., the required analyser bandwidth is thus

$$B = 0.8f_0 s \quad 4.3.8.$$

The averaging time is given by

$$T = M.N.h \quad 4.3.9.$$

where M is the number of transform blocks used to obtain the PSD. Equation 2.4.4. can be used to write

$$T = M/B \quad 4.3.10.$$

so that, from equation 4.3.8.,

$$f_0 s T = M/0.8 \quad 4.3.11.$$

The examples presented in Section 4.1. suggest that more than ten transforms should be used to obtain a reliable PSD estimate. Thus, if M is chosen to be twenty, the required bandwidth and averaging times are given by

$$B = 0.8f_0^2$$

$$T = 25/f_0^2$$

4.3.12.

The application of equations 4.3.12. to two typical examples yields:

Natural Frequency	Damping Coefficient	Analyser Bandwidth	Record Length
5.0	0.02	0.08	250
40.0	0.02	0.64	31.3

It is evident from the above that long runs at each nominated speed and altitude are required to realise stable PSD estimates. Such runs may be difficult to achieve, particularly at trans-sonic airspeeds. Furthermore, the existence of adequate levels of turbulence at the required altitude cannot be guaranteed. These facts mean that atmospheric turbulence is only rarely used as a primary source of excitation for flight flutter trials. However, the possession of, and confidence in, the techniques outlined above can permit flight flutter trials to proceed in conditions which would otherwise be unacceptable.

4.3.2. Excitation Source which can be Measured.

A number of the methods of excitation which are used in flight flutter trials permit the force applied to the structure to be derived from measurements or measured directly. These include inertial exciters and pyrotechnic sources.

When the input can be repeated (not necessarily exactly) then the methods described in Section 2.4. can be used to derive frequency response functions, even though the source is not necessarily random. This is an application in which a knowledge of techniques used for analysing random data can be used to reduce the effects of random components in an otherwise deterministic signal.

As an illustration of the procedure, suppose that an inertial exciter is used to drive the aircraft with a swept frequency sinusoidal force which can be measured. The response of an accelerometer mounted on the aircraft can, in the frequency domain, be expressed as

$$Y(if) = H(if).X(if) + N(if) \quad 4.3.13.$$

where X is the DFT (discrete Fourier transform) of the measured force (assumed to contain no error)

N is the DFT of all unwanted response signal components

Y is the DFT of the total accelerometer response (including unwanted components)

and H is the transfer function of the aircraft.

An estimate of the frequency response function is given by dividing equation 4.3.13. by the DFT of the input:

$$\frac{Y(if)}{X(if)} = H(if) + \frac{N(if)}{X(if)} \quad 4.3.14.$$

Quite clearly, the estimate of the frequency response function so obtained will be unbiased only if N(if) is zero at all frequencies. However, equation 4.3.13. can equally well be written:

$$Y(if).X^*(if) = H(if).X(if).X^*(if) + N(if).X^*(if) \quad 4.3.15.$$

It will be observed that:

$Y(f).X^*(f)$ is proportional to the cross spectral density of the response and the input force

$X(f).X^*(f)$ is proportional to the power spectral density of the input force

$N(f).X^*(f)$ is proportional to the cross spectral density of the "noise" and the input force.

If the unwanted signal components represented by $N(f)$ are independent (i.e. are not correlated with the input force), then the last term can be reduced by computing the power and cross spectral densities using, in each transform block, the results obtained from a different trial to obtain, ultimately,

$$G_{yx}(f) = H(f).G_{xx}(f) \quad 4.3.16.$$

from which an estimate of the frequency response function can be derived.

Note that, in this application, the filtering property of cross spectral analysis is used to reduce the unwanted components of the measured acceleration, even though the input force is deterministic. For best results, the force time history should be similar in each trial. When this can be arranged, the coherence function derived from the power and cross spectral density estimates can provide an insight into the levels of noise, the repeatability of the responses, etc.

The technique described above can, of course, be used equally well when random or impulsive force inputs are used. When a deterministic input is used, then the length of a transform block should be chosen so as to include the whole of the response to the input, or at least that part of the response which is discernable.

Once an estimate of the transfer (or, more properly, frequency response) function has been obtained, then identification methods can be used to estimate the natural frequency, damping coefficient, and response amplitude of each mode. In fact, if a number of accelerometers are used in the trials, then the frequency response function of each can be used collectively to obtain estimates of the natural frequency and damping coefficient, and individually to obtain estimates of the mode shape.

4.4. Applications in Evaluating Aircraft Handling Qualities.

The task of designing an aircraft so that it is capable of fulfilling its primary function and, at the same time, possesses the qualities necessary to enable the pilot to fly the aircraft accurately and safely, is a perennial problem which has attracted the close attention of aircraft designers since before the first manned flight. The problem can be described as assessing the stability of a non-linear, adaptive sensing and feedback mechanism (the pilot) situated in a "noisy" environment controlling a non-linear system (the aircraft) using a redundant set of control parameters. When aircraft were flown manually with few aids, the designer used heuristic criteria based, to a great extent, upon past experience.

The introduction of powered controls and stability augmentation systems (SAS) gave the designer some latitude, but the advent of automatic flight control systems (AFCS) has potentially removed most of the constraints within which the designer has, in the past, had to operate. However, the pilot of an aircraft equipped with AFCS is no longer flying the aircraft; it is more accurate to say that he is flying the AFCS. Thus the designer is presented with more flexibility but, at the same time, he is presented with a more complex control problem to solve.

4.4.1. Evaluation of Aircraft Ride Qualities.

Whilst it is not obviously connected with the aircraft handling problem, the ride quality of an aircraft can affect the overall safety of operation of, in particular, large transport aircraft. Indeed, several near accidents have been attributed to pilot disorientation caused by severe motion of the flight deck when the aircraft encountered atmospheric turbulence. Incidents of this type stimulated studies into the two factors which affect the ride quality of an aircraft: the response of the aircraft to atmospheric turbulence and the reaction of the pilot to that response.

The problem was investigated in some depth at (what was then) Hawker Siddeley Aviation Ltd., Hatfield in 1965. A program to compute the symmetrical response of a flexible aircraft to atmospheric turbulence was developed. The subjective "transfer function" of a pilot was assessed experimentally using twenty volunteers: an attempt at determining an objective transfer function by measuring the task performance of the volunteers was successful mainly in demonstrating the remarkable adaptability of the human servo system. The results of the subjective experiment were used to determine a frequency dependent weight function which could be used to weight calculated aircraft responses. The RMS values of the weighted spectra were used to assess the ride qualities of several existing and proposed designs. The study highlighted the need to keep the fuselage fundamental bending mode frequency away from the "heart-lung" body resonance which occurs normally at around 5 to 6Hz.

Studies similar to those outlined above have been reported elsewhere, notably in references 4.2 and 4.3.

4.4.2. Evaluation of the Pilot Describing Function.

It is often required to evaluate the performance, and particularly the stability, of an aircraft incorporating an AFCS before the aircraft has flown. This evaluation is often performed by qualified test pilots using a ground based flight simulator into which models of the aircraft and the AFCS have been incorporated. The value of such an exercise can be questionable in marginal situations particularly when the visual and motion cues presented to the pilot are not exact.

Attempts have been made to overcome the problems associated with subjective evaluation of the handling qualities of an aircraft by modelling mathematically the function of the pilot. Since a pilot is essentially a non-linear adaptive element producing an adequate model of his function is not an easy task. Those which have been produced to date have been, in the main, quasi-linear model estimated from the measurements made of his response to a range of tasks. The random measurements have been analysed using the techniques described in Section 2.4. to produce an equivalent transfer function applicable to the particular type of task. Reference 4.4. contains a description of the methods which can be used to derive models of pilot performance and presents results which have been achieved.

4.4.3. Evaluation of Handling Qualities.

The problem of optimising the performance and handling qualities of an aircraft/AFCS has received attention in flight test establishments. Difficulties arise because deficiencies in handling qualities may only become important during exacting phases of a flight, such as landing, combat manoeuvres, etc. Analysis techniques are therefore required which are capable of describing the performance of the various elements of the system from measurements obtained during such exacting phases.

The solution adopted at AFFTC is to require a pilot to execute specified manoeuvres, such as a precision tracking manoeuvre, whilst various aircraft parameters are being recorded. The measured parameters include steady state, or slowly changing, components and these are removed from the records by pre-processing (see Section 1) to leave a random remnant which can be used to obtain estimates of frequency response functions between various elements of the system. The frequency response functions are then examined in detail to evaluate the handling qualities of the aircraft.

The technique offers the attraction of providing linearised system characteristics whilst the aircraft is being exercised in a representative way, but it may be difficult to obtain sufficiently long records from one manoeuvre. Nevertheless, considerable insight into understanding the factors influencing the handling qualities of an aircraft has been gained from such flight trials (see reference 4.5.) and the technique promises to become a powerful tool for optimising the handling qualities of an aircraft during critical phases of a flight.

4.5. The Reduction of Random Measurements to a Form Suitable for Certification Trials.

A structure will normally be subjected to loads caused by a variety of phenomena during the course of its service life. The design organisation may be required to demonstrate that the structure is capable of withstanding the various loads for the duration of its service life without catastrophic failure. Such a requirement would normally be satisfied partly by calculation and partly by conducting ground trials on one or more specimens of the structure. Before either calculations or the design of ground trials can proceed, details of the characteristics of each source, and of its probability of occurrence must be obtained.

Certain types of sources produce a randomly varying load which is not predictable theoretically. Examples are runway and road surface undulations, boundary layer turbulence, control circuit loads induced by a pilot, jet exhausts, etc. In such cases the loads, or at least the response of the structure to the loads, must be measured and reduced to a form which can be used for predicting the life of the structure.

The techniques which are described in Chapters 2 and 3 are, in the main, applicable to stationary measurements, and sometimes only to measurements also having a Gaussian amplitude probability density distribution. Difficulties of interpretation can occur when these conditions do not apply. Two cases are considered in this Section. The first is concerned with the reduction of conditionally stationary data to a form which can be used for the specification of ground clearance trials. The second case is concerned with the more contentious situation when the assumption of conditional stationarity cannot be justified.

Case 1. Conditionally stationary measurements having a Gaussian APD.

When the RMS value of a random signal varies with parameters which can be controlled, but at any fixed set of parameter values the PSD is invariant, then a description of the signal can be compiled by computing a series of PSDs, each obtained from measurements recorded at constant values of the parameters. The PSDs, when associated with probabilities of occurrence of the appropriate values of the parameter sets, can be used to specify a suitable ground test. Examples of phenomena which can be described in this way are jet exhausts and boundary layer turbulence. The frequency bandwidths of these phenomena are often such that it may be assumed, at least for fatigue purposes, that the normalised PSD (PSD divided by the mean square value) is invariant. This assumption can permit a major simplification in the specification of a ground trial by using a knowledge of the fatigue properties of similar structures to compute a test time at one RMS value which will produce the same damage as that accumulated by application of each of the components for the appropriate time. As an illustration, suppose a set of measurements is reduced to the following:-

RMS value	Proportion of time (percent)
0.1	57
0.3	31
0.5	10
0.8	2

Suppose also that

$$\frac{T_1}{T_2} = \left(\frac{\text{RMS } 1}{\text{RMS } 2} \right)^{-3}$$

Then a test may be specified having a RMS value of 0.8, the appropriate PSD, and a time (as a proportion of real time) given by

$$\begin{aligned} 0.57 (0.1/0.8)^3 &= 0.001 \\ 0.31 (0.3/0.8)^3 &= 0.016 \\ 0.10 (0.5/0.8)^3 &= 0.024 \\ 0.02 (0.8/0.8)^3 &= 0.020 \\ &\underline{0.061} \end{aligned}$$

The resulting test would be conducted at an RMS level which is not unrealistic, but which would occupy only 6 percent of real time. The degree of conservatism of the test would depend upon the chosen value of the power index (a lower value giving a more severe test).

It is worth noting at this point that increasing the test RMS value to a level which is not encountered in service merely to reduce the test time, is somewhat hazardous for two reasons. The power index is normally a function of RMS value and can change sharply when a significant amount of plastic deformation occurs. Bolted and rivetted joints are non-linear elements; if they are made to "work" in an unrepresentative manner, the resultant "springing" of rivets and fretting of the lapping surfaces can dramatically alter the stress levels and cause premature failure of the structure.

Case 2. Non-Stationary Measurements.

When the RMS value of a random signal varies with parameters which cannot be controlled, then the method described above can be used only with difficulty (and with some uncertainty regarding the validity of the results obtained.)

An alternative method is suggested by the assumption, frequently invoked in the design of fatigue tests, that the order of application of load is immaterial. It seems reasonable to suppose that, if suitably short lengths of record could be selected and transposed, a signal of the type under consideration could be converted into one having a finite number of stationary segments. The method outlined above could then be used to compile a suitable fatigue test. It remains to devise a method of estimating RMS values for and the lengths of each stationary segment.

It is assumed that sufficient data exists to constitute a representative sample of the phenomenon (or response to the phenomenon) and to define the APD to a sufficiently low value of probability density. In the majority of cases, the estimated APD obtained from the data will have a leptokurtic form (the probability density of both very low and very high excursions will be greater than that expected from a Normal distribution.) When this is so, the APD can be idealised by the sum of N weighted Normal distributions:

$$\hat{p}(y) = \sum_{i=1}^N K_i \cdot \frac{1}{\sigma_i \sqrt{2\pi}} \cdot \exp \left(\frac{-y^2}{2\sigma_i^2} \right) \quad 4.5.1.$$

The success of the idealisation is measured by a "cost function" having the form

$$R = \lim_{Y \rightarrow \infty} \frac{1}{2Y} \int_{-Y}^Y [f(p(y)) - f(\hat{p}(y))]^2 dy \quad 4.5.2.$$

Where $f\{A\}$ is any suitable function of the argument A . A particular form of equation 4.5.2. which has proved to be successful is

$$R = \frac{1}{2Y} \int_{-Y}^Y [\log p(y) - \log \hat{p}(y)]^2 dy \quad 4.5.3.$$

Where Y is chosen to be suitably large. This may be written, when the estimated APD is a discrete function of y ,

$$R = \frac{1}{2N+1} \sum_{j=-N}^N [\log \{p(j\Delta y)\} - \log \{\hat{p}(j\Delta y)\}]^2 \quad 4.5.4.$$

In practice, equation 4.5.4. must be modified to take account of possible null estimates of probability density. These are omitted, and the divisor $2N+1$ adjusted accordingly.

In principle, equation 4.5.1. can be substituted into 4.5.4., and the cost function differentiated with respect to each of the unknown parameters to obtain estimation equations. However, difficulties can arise with this general formulation because the structure of each component of 4.5.1. is identical, so that very good initial estimates of the unknown parameters are required for the process to converge.

An implementation which has been used successfully, and is a drastic simplification of the method outlined above, assumes that

- a) Three components of equation 4.5.1. are sufficient to idealise an APD.
- b) The RMS values of the three components have the fixed relationship

$$\sigma_1 = 2\sigma_2 = 4\sigma_3$$
- c) The value of the greatest RMS value, σ_1 , is to be chosen by the user.

With these assumptions, the problem is reduced to estimating the values of the three weights K_1 , K_2 and K_3 .

If equation 4.5.1. is simplified to read

$$\hat{p}(y) = \sum_{i=1}^3 (K+\Delta K)_i \cdot \exp \frac{-y^2 \cdot 2^{(2i-2)}}{2\sigma^2} \quad 4.5.5.$$

Where σ is the greatest RMS value and $(K+\Delta K)_i$ are the required weights, then the linearised estimation equations are given by

$$\{\Delta K\} = [\sum H^T H]^{-1} \cdot [\sum H^T \{ \log p(y) - \log \hat{p}(y) \}] \quad 4.5.6.$$

Where H is a 1×3 vector, the i th element being given by

$$\frac{1}{\hat{p}(y_j)} \cdot \exp \frac{-y_j^2 \cdot 2^{(2i-2)}}{2\sigma^2} \quad 4.5.7.$$

The presence of $\hat{p}(y)$ in the H vector means that the estimates given by equation 4.5.6. are approximate only, and the estimation procedure must be repeated until the ΔK vector approaches zero. Suitable initial values must be chosen for K_i . It is suggested that the maximum value of the measured APD be used for K_3 , with $K_2 = K_3/10$ and $K_1 = K_3/100$.

The weights K_i are estimates of the proportion of the original signal having an RMS value equal to the appropriate σ_i : if they are multiplied by the averaging time, then they become estimates of time spent at each RMS value. It should be noted that, in general, the sum of the weights is not unity, the actual value being dependent upon the chosen Maximum RMS value and the shape of the measured APD. Normally, the sum turns out to be less than unity, implying that sections of the signal have an RMS value which has been idealised as zero.

The "best" maximum RMS value for a particular signal may be determined by repeating the procedure for various assumed RMS values, and using the cost function (equation 4.5.4.) to indicate the success of the resulting idealisation.

An example of the application of the technique is shown in figure 4.5.1. The data originated from measurements made during road transportation trials, and was reduced to an APD using the algorithm described in Section 2.3.2. This was idealised using the above technique for a range of assumed maximum RMS values. The idealised function for one assumed RMS value is shown in the figure as a dashed line. The derived test time (expressed as a proportion of real time) and the "fit" parameter (square root of the cost function) are shown as a function of maximum RMS value in figure 4.5.2.

A power spectral density of the process may be estimated by computing the "Normalised" PSD, that is, each PSD estimate is divided by its mean square value prior to averaging. The result will have a unit mean square value; thus a suitable test PSD can be obtained by multiplying the Normalised PSD by the square of the test RMS value.

A number of assumptions are implicit in the above analysis. They include:

- The measurements which are available are assumed to be representative of the service environment and of sufficient length to obtain a reliable estimate of the APD
- The spectral shape of the phenomenon is independent of the RMS value
- The shape of the APD is independent of frequency.

Within these assumptions, the technique provides a method of reducing non-stationary measurements to a form which can yield a specification for ground clearance trials which is compressed in time, but which is realistic in terms of fatigue damage and maximum induced excursions (loads.)

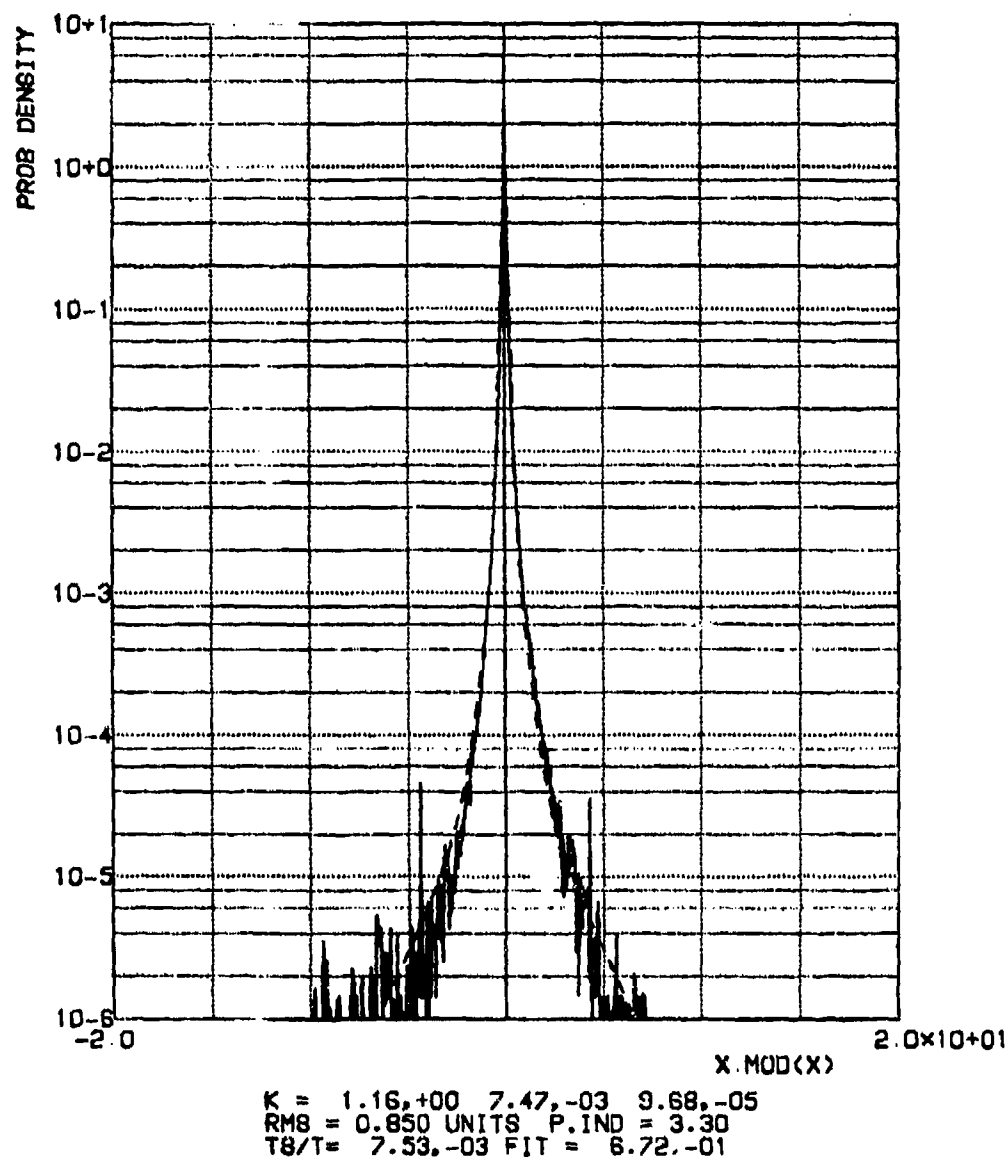


Figure 4.5.1. Measured APD and a Three Component Gaussian Idealisation.

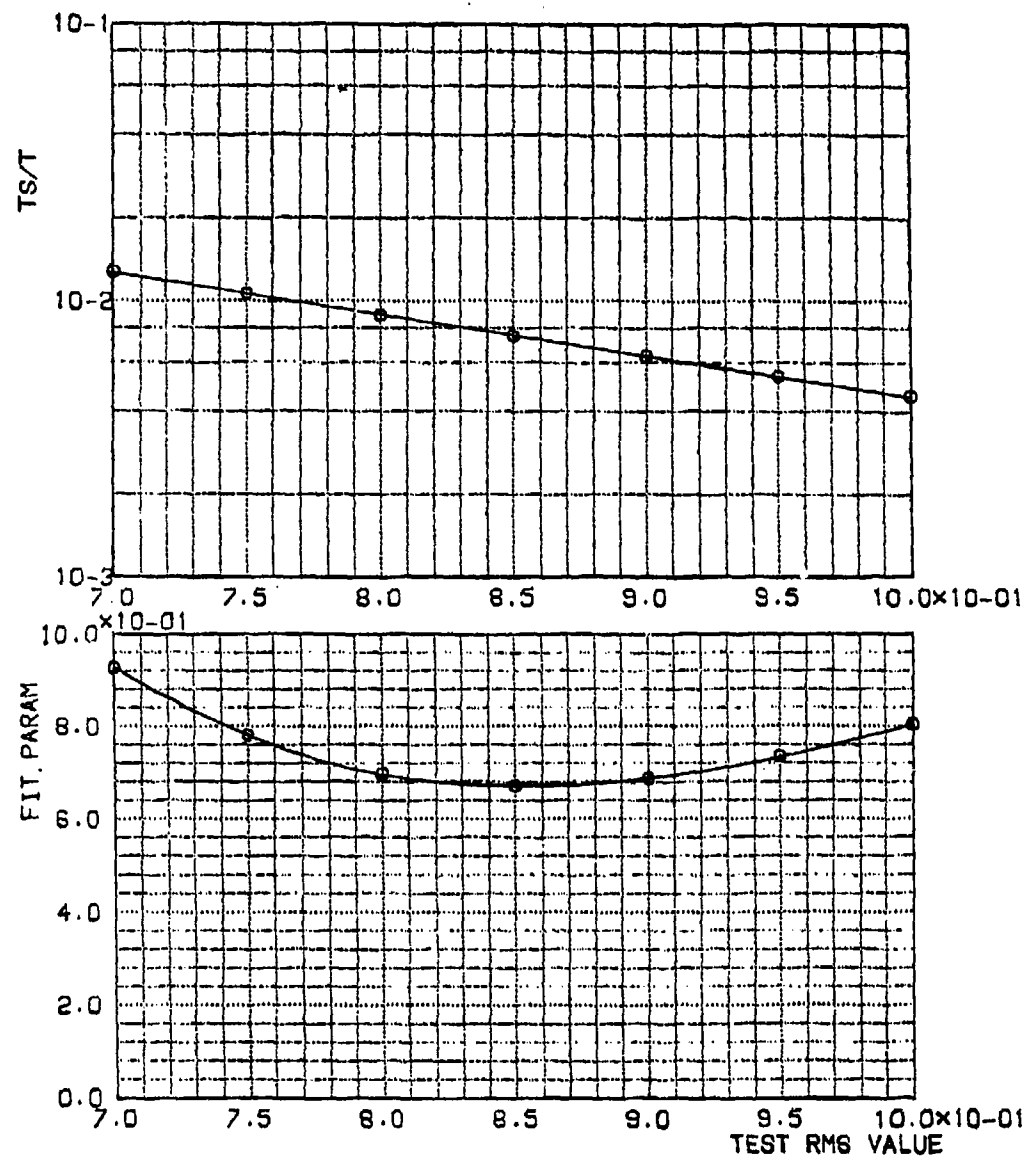


Figure 4.5.2. Variation of Test Time and Quality of Fit with Assumed Test RMS Value.

5. REFERENCES

- | | | | |
|-------|---|---|--|
| O.1 | Oppenheim R.
Schafer R. | Digital Signal Processing | Prentice Hall
1975 |
| O.2 | Rabiner L.
Gold B. | Theory & Application of Digital Signal
Processing | Prentice
1975 |
| O.3 | Newland D.E. | Random Vibration and Spectral Analysis | Longman, 1975 |
| 1.1 | | Modern Instrumentation Tape Recording | E.M.I., 1978 |
| 1.2. | | "Datatape" - High Density Digital Data Recording | Bell & Howell, 1976 |
| 1.3. | Nicola R.M.
Rauch L.L. | Radio Telemetry | Wiley, 1956 |
| 1.4. | Besserer C.W.
Floyd E.N.
(Editors) | Aerospace Telemetry | Prentice Hall
1961 |
| 1.5. | Blackman R.B.
Tukey J.W. | The Measurement of Power Spectra from the Point
of View of Communications Engineering | Dover, 1958 |
| 1.6. | Otnes R.K. | An Elementary Design Procedure for Digital
Filters | IEEE Trans. Audio
Electro-Acoustics
AU-16, p330, 1968 |
| 1.7. | Ormsby J.F.A. | Design of Numerical Filters with Applications
to Missile Data Processing | J. Assoc. Comp.
Mach. July, 1961 |
| 1.8. | Kalman R.E. | A New Approach to Linear Filtering and
Prediction Problems | Trans. ASME
J. Basic Eng.
Vol. 82, March 1960 |
| 1.9. | Mehra R.K.
Tyler J.S. | Case Studies in Aircraft Parameter
Identification | 3rd. IFAC Symposium
June, 1973 |
| 1.10. | Jonkers H.L. | Application of the Kalman Filter to Flight
Path Reconstruction from Flight Test Data
Including Estimation of Instrumental Bias
Error Corrections | Delft Univ of Tech.
Report VTH-162
Feb. 1976 |
| 1.11. | Klein V. | Aircraft Parameter Estimation in the
Frequency Domain | AIAA Paper No.
78-1344 Aug. 1978 |
| 2.1. | Bendat J.S. | Principles and Applications of Random
Noise Theory | Wiley, 1958 |
| 2.2 | Bendat J.S.
Piersol A.G. | Random Data: Analysis and Measurement
Procedures | Wiley, 1971 |
| 2.3. | Himmelblau
D.M. | Process Analysis by Statistical Methods | Wiley, 1968 |
| 2.4. | Schijve J. | The Analysis of Random Load-Time Histories
with Relation to Fatigue Tests and Life
Calculations | Fatigue of Aircraft
Structures
Macmillan 1963 |
| 2.5. | Richards F.D.
LaPointe N.R.
Wetzel R.M. | A Cycle Counting Algorithm for Fatigue
Damage Analysis | SAE Paper 740278
March 1974 |
| 2.6. | Bingham C.
Godfrey M.D.
Tukey J.W. | Modern Techniques of Power Spectrum
Estimation | IEEE Trans. Audio
Electro-Acoustics
AU-15 No.2 June 1967 |
| 2.7. | Otnes R.K.
Enochson L. | Digital Time Series Analysis | Wiley, 1972 |

- | | | | |
|------|---------------------------------|---|---|
| 4.1. | Piazzoli G.
Van Nunen J.W.G. | Aeroelastic Flight Test Techniques and
Instrumentation | AGARDograph No 160
Volume 9. |
| 4.2. | Rustenburg
J.W. | A Technique for the Evaluation of
Aircraft Ride Quality | TR-68-18, Aero.
Systems Div. AFSC
1968 |
| 4.3. | Rustenburg
J.W. | Development of Tracking Error Frequency
Response Function and Aircraft Ride Quality
Design Criteria | TR-70-18, Aero.
Systems Div. AFSC
1971 |
| 4.4. | McRuer D.T.
Krendel E.S. | Mathematical Models of Human Pilot Behavior | AGARDograph No. 188
1974 |
| 4.5. | Twisdale T.R. | A Mission Oriented Flight Test Technique for
Identifying Aircraft and Flight Control
System Transfer Function | AGARD CP223, 1976 |
| 4.6. | Cole H.A. | On-line Failure Detection and Damping Measurement
of Aerospace Structures by Random Decrement
Signatures | CR 2205, 1973 |
| A.1. | Lynn P.A. | An Introduction to the Analysis and
Processing of Signals. | Macmillan, 1973. |
| B.1. | Cooley J.W.
Tukey J.W. | An Algorithm for the Machine Calculation
Of Complex Fourier Series. | Math. Comp. 19
April, 1965. |
| B.2. | Gentleman W.M.
Sande G. | Fast Fourier Transforms - for Fun and Profit | AFIPS Conf. Proc.
29, Spartan Books
1966. |
| D.1. | Clenshaw | Chebyshev Series for Mathematical Functions. | N.P.L. Math. Tables
Volume 5. |
| D.2. | Rice S.O. | Mathematical Analysis of Random Noise. | Bell Systems Tech.
Journal Vol. 23,
1944. |
| D.3. | Bendat J.S. | Principles and Applications of Random
Noise theory | Wiley, 1958. |

APPENDIX A.

The Fourier Series and Fourier Transform.

A good working knowledge of the principles of the Fourier series and Fourier transform is essential for developing routines for analysing random data and for understanding fully the results produced by such routines. In recognition of this fact, Appendix A contains an outline of the principles of both the Fourier series and the Fourier transform. The strong relationship between the two when they are applied to sampled time sequences is demonstrated, and a list of properties is included which may be useful to the analyst who is engaged in applying the techniques to specific tasks.

The contents of this Appendix are an essential prelude to Appendices B and C, which contain, respectively, descriptions of the Fast Fourier Transform, and procedures which make use of the FFT.

A.1. The Fourier Series.

A signal $y(t)$ which is everywhere finite and is periodic over a time T can be represented by a Fourier series comprising the weighted sum of sine and cosine functions.

Specifically,

$$y(t) = a_0 + 2 \sum_{n=1}^{\infty} (a_n \cos 2\pi n f t + b_n \sin 2\pi n f t) \quad \text{A.1.1.}$$

$$\text{where } f = 1/T \quad \text{A.1.2.}$$

$$a_n = \frac{1}{T} \int_0^T y(t) \cos 2\pi n f t \, dt \quad \text{A.1.3.}$$

$$\text{and } b_n = \frac{1}{T} \int_0^T y(t) \sin 2\pi n f t \, dt \quad \text{A.1.4.}$$

Equation A.1.1. can be written more succinctly by recognising that

$$\cos \theta + i \sin \theta = \exp(i\theta) \quad \text{A.1.5.}$$

Thus

$$y(t) = \sum_{n=-\infty}^{\infty} c_n \exp(2\pi i n f t) \quad \text{A.1.6.}$$

$$\text{where } c_n = \frac{1}{T} \int_0^T y(t) \exp(-2\pi i n f t) \, dt \quad \text{A.1.7.}$$

$$= a_n - i b_n \quad \text{A.1.8.}$$

If $y(t)$ is represented by an N -length sampled sequence such that

$$y(j) = y(h \cdot j) \quad \text{A.1.9.}$$

where h is the interval between successive samples

$$\text{and } N = T/h$$

then equations A.1.6. and A.1.7. become, assuming trapezoidal integration,

$$y(j) = \sum_{n=-N/2}^{(N/2-1)} c(n) \exp\left(\frac{2\pi i n j}{N}\right) \quad \text{A.1.10.}$$

$$= \sum_{n=0}^{(N-1)} c(n) \exp\left(\frac{2\pi i n j}{N}\right) \quad \text{A.1.11.}$$

$$\text{and } c(n) = \frac{1}{N} \sum_{j=0}^{(N-1)} y(j) \exp\left(-\frac{2\pi i n j}{N}\right) \quad \text{A.1.12.}$$

Note that the summation limits of A.1.10. have been changed. The reason for this may be found by examining equation A.1.12. If kN is added to n in the exponential term, then it will be observed that

$$\exp\left(\frac{-2\pi i j(n+kN)}{N}\right) = \exp\left(\frac{-2\pi i jn}{N}\right) \quad \text{A.1.13.}$$

$$\text{since} \quad \exp(-2\pi i jk) = 1 \quad \text{for all } i, j, k$$

$$\text{Hence} \quad c(n+kN) = c(n) \quad \text{A.1.14.}$$

Thus no new information is obtained by setting $n < 0$ or $n > N-1$. This may be seen as a re-iteration of the Sampling theorem. The maximum frequency corresponding to $n = N/2-1$ is given by equation A.1.10., so that

$$f_{\max.} = \frac{(0.5N - 1)}{h.N} \text{ Hz.} \quad \text{A.1.15.}$$

The moduli of the coefficients $c(n)$ of equation A.1.12. define the magnitude of each sinusoidal component contributing to $y(j)$. Thus a sequence containing the moduli of $c(n)$ can be interpreted as a frequency "spectrum" of amplitudes spaced $1/hN$ Hz. apart, and equation A.1.12. can be looked upon as defining a "filtering" process since its action is to separate out the frequency domain components which, taken together, form the time domain sequence $y(j)$.

It is to be observed that the coefficients $c(n)$ completely describe $y(j)$ by virtue of equation A.1.11. They therefore constitute a definition of $y(j)$.

A.2. The Fourier Transform.

If a signal $y(t)$ is everywhere finite and of finite length, then the concept of the Fourier series can be extended to the situation where $y(t)$ is aperiodic. The relevant expressions, which are equivalent to equations A.1.6. and A.1.7. are:

$$y(t) = \int_{-\infty}^{\infty} Y(if) \cdot \exp(2\pi ift) \cdot df \quad \text{A.2.1.}$$

$$\text{where} \quad Y(if) = \int_{-\infty}^{\infty} y(t) \cdot \exp(-2\pi ift) \cdot dt \quad \text{A.2.2.}$$

A formal derivation of these expressions, which are known as a Fourier pair, may be found in Reference A.1. As was the case for the Fourier series coefficients $c(n)$, the function $Y(if)$ can be interpreted as a frequency spectrum completely describing the time domain function $y(t)$.

$Y(if)$ is a complex function which may be written:

$$Y(if) = R(f) + I(f) = |Y(if)| \cdot \exp(i\theta(f)) \quad \text{A.2.3.}$$

$$\text{where} \quad R(f) = \int_{-\infty}^{\infty} y(t) \cdot \cos 2\pi f t \cdot dt \quad \text{A.2.4.}$$

$$\text{and} \quad I(f) = \int_{-\infty}^{\infty} y(t) \cdot \sin 2\pi f t \cdot dt \quad \text{A.2.5.}$$

Thus, if $y(t)$ is real, $R(f)$ and $I(f)$ are, respectively, even and odd functions of frequency, i.e.

$$R(f) = R(-f) \quad \text{and} \quad I(f) = -I(-f) \quad \text{A.2.6.}$$

$$\text{or} \quad Y(if) = Y^*(-if) \quad \text{A.2.7.}$$

It therefore follows that

$$|Y(if)| = (R^2(f) + I^2(f))^{1/2} = Y(-if) \quad A.2.8.$$

$$\text{and } \theta(f) = \arctan \frac{I(f)}{R(f)} = -\theta(-f) \quad A.2.9.$$

A number of useful properties of the Fourier transform can be deduced by manipulating equations A.2.1. and A.2.2. The more important of these are listed below.

A.2.1. The Superposition Theorem.

The algebraic sum of two signals, $y_1(t)$ and $y_2(t)$, can be written, from equation A.2.1.,

$$\begin{aligned} a.y_1(t) + b.y_2(t) &= \int_{-\infty}^{\infty} a.Y_1(if).exp(2\pi ift).df \\ &\quad + \int_{-\infty}^{\infty} b.Y_2(if).exp(2\pi ift).df \\ &= \int_{-\infty}^{\infty} (a.Y_1(if) + b.Y_2(if)).exp(2\pi ift).df \end{aligned} \quad A.2.10.$$

$$\text{or } a.y_1(t) + b.y_2(t) \Leftrightarrow a.Y_1(if) + b.Y_2(if) \quad A.2.11.$$

The Fourier transform is therefore a linear operator.

A.2.2 The Shift Theorems.

If the time origin of a signal is moved by T , such that $y(t)$ becomes $y(t-T)$, then equation A.2.2. becomes

$$Y'(if) = \int_{-\infty}^{\infty} y(t-T).exp(-2\pi ift).dt \quad A.2.12.$$

$$\text{Write } t' = t-T$$

then equation A.2.12. becomes

$$Y'(if) = \int_{-\infty}^{\infty} y(t').exp(-2\pi if(t'+T)).dt \quad A.2.13.$$

$$\text{or, } Y'(if) = Y(if).exp(-2\pi ifT) \quad A.2.14.$$

Equation A.2.14. states that the introduction of a delay T in the time domain translates, in the frequency domain, to a constant rate of change of phase with frequency, given by

$$\frac{d\theta}{df} = -2\pi T \quad \text{radians per Hz.} \quad A.2.15.$$

The complementary relationship can be developed to yield:

$$y'(t) = \int_{-\infty}^{\infty} Y(1(f-p)).exp(2\pi ift).df \quad A.2.16.$$

$$\text{or, } y'(t) = y(t) \cdot \exp(2\pi i p t)$$

A.2.17.

A.2.3. The Multiplication Theorem.

When it is integrated over all time, the product of two signals, $y_1(t)$ and $y_2(t)$, can be written:

$$\int_{-\infty}^{\infty} y_1(t) \cdot y_2(t) \cdot dt = \int_{-\infty}^{\infty} y_1(t) \int_{-\infty}^{\infty} y_2(if) \cdot \exp(2\pi i f t) \cdot df \cdot dt \quad \text{A.2.18.}$$

$$= \int_{-\infty}^{\infty} y_2(if) \cdot df \cdot \int_{-\infty}^{\infty} y_1(t) \cdot \exp(2\pi i f t) \cdot dt \quad \text{A.2.19.}$$

$$= \int_{-\infty}^{\infty} y_2(if) \cdot y_1(-if) \cdot df \quad \text{A.2.20.}$$

$$= \int_{-\infty}^{\infty} y_2(if) \cdot Y_1^*(if) \cdot df \quad \text{A.2.21.}$$

$$\text{and, by symmetry, } = \int_{-\infty}^{\infty} Y_1(if) \cdot Y_2^*(if) \cdot df \quad \text{A.2.22.}$$

A.2.4. Parseval's Theorem.

When $y_1(t) = y_2(t) = y(t)$, then the multiplication theorem becomes:

$$\begin{aligned} \int_{-\infty}^{\infty} y(t)^2 \cdot dt &= \int_{-\infty}^{\infty} Y(if) \cdot Y^*(if) \cdot df \\ &= \int_{-\infty}^{\infty} |Y(if)|^2 \cdot df \end{aligned} \quad \text{A.2.23.}$$

This is known as Parseval's theorem.

A.2.5. Time Scaling.

If time is scaled, such that $y(t)$ is replaced by $y(kt)$, k greater than zero, then equation A.2.2. may be written:

$$Y'(if) = \int_{-\infty}^{\infty} y(kt) \cdot \exp(-2\pi i f t) \cdot dt \quad \text{A.2.24.}$$

$$\text{Let } t' = kt$$

then equation A.2.24. becomes

$$Y'(if) = \int_{-\infty}^{\infty} y(t') \cdot \exp(-2\pi i f t') \cdot \frac{dt}{k} \quad \text{A.2.25.}$$

$$\text{or, } Y'(if) = \frac{1}{k} \cdot Y\left(\frac{if}{k}\right) \quad \text{A.2.26.}$$

A.2.6. The Transform of Time Derivatives.

From equation A.2.1.,

$$y(t) = \int_{-\infty}^{\infty} Y(1f) \cdot \exp(2\pi i f t) \cdot df$$

$$\text{so that } \frac{dy(t)}{dt} = \int_{-\infty}^{\infty} 2\pi i f \cdot Y(1f) \cdot \exp(2\pi i f t) \cdot df \quad \text{A.2.27.}$$

It follows that, in general (provided that the transform exists),

$$\frac{d^n y(t)}{dt^n} \iff (2\pi i f)^n \cdot Y(1f) \quad \text{A.2.28.}$$

A.2.7. The Transform of Frequency Derivatives.

The above argument can be used to derive the transform of frequency derivatives, provided that the result exists, to yield:

$$\frac{d^n Y(1f)}{df^n} \iff (-2\pi i f)^n \cdot y(t) \quad \text{A.2.29.}$$

A.2.8. Conjugate Functions.

If $y(t)$ is a complex function of time, then, from equation A.2.1.,

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} Y(1f) \cdot \exp(2\pi i f t) \cdot df \\ &= \int_{-\infty}^{\infty} (R(f) \cdot \cos 2\pi f t - I(f) \cdot \sin 2\pi f t) \cdot df \\ &\quad + i \cdot \int_{-\infty}^{\infty} (R(f) \cdot \sin 2\pi f t + I(f) \cdot \cos 2\pi f t) \cdot df \end{aligned}$$

$$\begin{aligned} \text{then } y^*(t) &= \int_{-\infty}^{\infty} (R(f) \cdot \cos 2\pi f t - I(f) \cdot \sin 2\pi f t) \cdot df \\ &\quad - i \cdot \int_{-\infty}^{\infty} (R(f) \cdot \sin 2\pi f t + I(f) \cdot \cos 2\pi f t) \cdot df \end{aligned}$$

$$\text{or, } y^*(t) == Y^*(-1f) \quad \text{A.2.30.}$$

It follows from equation A.2.30. that $Y(1f)$ is an even function if $y(t)$ is real. Conversely, $Y(1f)$ is an odd function if $y(t)$ is imaginary.

A.2.9. The Convolution Theorem.

The convolution of two signals, $y_1(t)$ and $y_2(t)$, may be written:

$$y(t) = \int_{-\infty}^{\infty} y_1(x) \cdot y_2(t-x) \cdot dx \quad \text{A.2.31.}$$

From equation A.2.2.,

$$Y(if) = \int_{-\infty}^{\infty} \exp(-2\pi ift) \cdot \int_{-\infty}^{\infty} y_1(x) \cdot y_2(t-x) \cdot dx \cdot dt \quad \text{A.2.32.}$$

$$= \int_{-\infty}^{\infty} y_1(x) \cdot \int_{-\infty}^{\infty} \exp(-2\pi ift) \cdot y_2(t-x) \cdot dt \cdot dx \quad \text{A.2.33.}$$

Equation A.2.14. can be used to write

$$Y(if) = \int_{-\infty}^{\infty} y_1(x) \cdot \exp(-2\pi ifx) \cdot Y_2(if) \cdot dx \quad \text{A.2.34.}$$

Hence:

$$Y(if) = Y_1(if) \cdot Y_2(if) \quad \text{A.2.35.}$$

The converse expression may be developed in a similar way to yield:

$$y_1(t) \cdot y_2(t) \Leftrightarrow \int_{-\infty}^{\infty} Y_1(ix) \cdot Y_2(if-ix) \cdot dx \quad \text{A.2.36.}$$

A.2.10. Power Spectral Density.

The autocorrelation function is defined as:

$$R(r) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y(t) \cdot y(t+r) \cdot dt \quad \text{A.2.37.}$$

A convenient device to remove the limit in this expression is to presume that the signal $y(t)$ is contained within some time $2T$, say. Having made this assumption, equation A.2.37. becomes:

$$R(r) = \frac{1}{2T} \int_{-\infty}^{\infty} y(t) \cdot y(t+r) \cdot dt \quad \text{A.2.38.}$$

The shift theorem may then be used to obtain:

$$R(r) = \frac{1}{2T} \int_{-\infty}^{\infty} y(t) \cdot \exp(2\pi ift) \cdot \int_{-\infty}^{\infty} Y(if) \cdot \exp(2\pi ifr) \cdot df \cdot dt \quad \text{A.2.39.}$$

which, from equation A.2.7., may be written:

$$R(r) = \frac{1}{2T} \int_{-\infty}^{\infty} Y^*(if) \cdot Y(if) \cdot \exp(2\pi ifr) \cdot df \quad \text{A.2.40.}$$

$$= \int_{-\infty}^{\infty} \left| \frac{Y(jf)}{2T} \right|^2 \exp(2\pi jfr) \cdot df \quad \text{A.2.41.}$$

$$= \int_0^{\infty} \left| \frac{Y(jf)}{T} \right|^2 \exp(2\pi jfr) \cdot df \quad \text{A.2.42.}$$

when $y(t)$ is a stationary process.

The expression

$$\lim_{T \rightarrow \infty} \left| \frac{Y(jf)}{T} \right|^2 = G_y(f) \quad \text{A.2.43.}$$

is precisely the definition of the power spectral density function. The limit which has been inserted in equation A.2.43. can be justified by arguing that, although the signal $y(t)$ was presumed to be contained within a time of $2T$, no bound was placed upon the value of T .

A.3. The Discrete Fourier Transform (DFT).

If a signal $y(t)$ is sampled at discrete points in time, each h apart, then equation A.2.1. can be written:

$$y(jh) = \int_{-\infty}^{\infty} Y(if) \cdot \exp(2\pi i jfh) \cdot df \quad \text{A.3.1.}$$

The exponential function, $\exp(2\pi i jfh)$, is periodic in the frequency domain, having a fundamental period $F = 1/h$. Equation A.3.1. can therefore be written:

$$y(jh) = \sum_{k=-\infty}^{\infty} \int_{kF}^{(k+1)F} Y(if) \cdot \exp(2\pi i jfh) \cdot df \quad \text{A.3.2.}$$

$$= \int_0^F \sum_{k=-\infty}^{\infty} Y(if + ikF) \cdot \exp(2\pi i jfh) \cdot df \quad \text{A.3.3.}$$

The effect of replacing a signal by a sampled sequence is therefore equivalent to replacing the Fourier transform by:

$$Y(if) = \sum_{k=-\infty}^{\infty} Y(if + ikF) \quad \text{A.3.4.}$$

In the special case when the transform is zero everywhere except in the region corresponding to $k = 0$ (i.e. no aliasing), then equation A.3.3. becomes:

$$y(jh) = \int_0^F Y(if) \cdot \exp(2\pi i jfh) \cdot df \quad \text{A.3.5.}$$

When the transform $Y(if)$ is itself a sampled sequence, then trapezoidal integration can be used to write equation A.3.5. as follows:

$$y(jh) = \frac{1}{hN} \sum_{n=0}^{N-1} Y(inf') \cdot \exp\left(\frac{2\pi i jhn}{N}\right) \quad \text{A.3.6.}$$

$$\text{where, } f' = \frac{1}{hN}$$

Similarly, it can be demonstrated that:

$$Y(inf') = h \cdot \sum_{j=0}^{N-1} y(jh) \cdot \exp\left(-\frac{2\pi i jn}{N}\right) \quad \text{A.3.7.}$$

If equation A.3.7. is modified by multiplying the right hand side by f' , then it becomes:

$$Y'(inf') = \frac{1}{N} \sum_{j=0}^{N-1} y(jh) \cdot \exp\left(-\frac{2\pi i jn}{N}\right) \quad \text{A.3.8.}$$

and hence:

$$y(jh) = \sum_{n=0}^{N-1} Y'(inf') \cdot \exp\left(\frac{2\pi i jn}{N}\right) \quad \text{A.3.9.}$$

Equations A.3.8. and A.3.9. are identical to the discrete Fourier series equations A.1.12. and A.1.11., respectively, and are known as the DFT pair. It is therefore clear that the DFT is, in fact a Fourier series estimator, but with no constraints placed upon the selection of the fundamental period T.

The DFT pair have a number of properties which may be found to be useful. These have been extracted from Reference A.2. and are listed below. It is convenient to adopt a shorthand notation for these in which equations A.3.8. and A.3.9. are written:

$$Y(n) = \frac{1}{N} \sum_{j=0}^{N-1} y(j) \cdot W_N^{-jn} \quad n = 0(1)N-1 \quad \text{A.3.10.}$$

$$\text{and } y(j) = \sum_{n=0}^{N-1} Y(n) \cdot W_N^{jn} \quad j = 0(1)N-1 \quad \text{A.3.11.}$$

$$\text{where } W_N = \exp\left(\frac{2\pi i}{N}\right) \quad \text{A.3.12.}$$

$$\text{Hence } Y(n) \Leftrightarrow y(j) \quad \text{A.3.13.}$$

It is assumed that $y(j)$ is periodic such that:

$$y(-j) = y(N-j) \quad \text{A.3.14.}$$

A.3.1. The Superposition Theorem.

$$\text{Let } y_1(j) \Leftrightarrow Y_1(n)$$

$$\text{and } y_2(j) \Leftrightarrow Y_2(n)$$

$$\text{Then } Y(n) = a \cdot Y_1(n) + b \cdot Y_2(n) \quad \text{A.3.15.}$$

$$\text{where } y(j) = a \cdot y_1(j) + b \cdot y_2(j) \quad \text{A.3.16.}$$

A.3.2. The Symmetry Theorem.

The symmetry theorem states that:

$$y(-j) \Leftrightarrow Y(-n) \quad \text{A.3.17.}$$

Thus $y(j)$ is even if, and only if, $Y(n)$ is even. Conversely, $y(j)$ is odd if, and only if, $Y(n)$ is odd.

A.3.3. Conjugate Functions Theorem.

The conjugate functions theorem states that:

$$y^*(j) \Leftrightarrow Y^*(-n) \quad \text{A.3.18.}$$

$$\text{or, } y^*(-j) \Leftrightarrow Y^*(n)$$

Thus $y(j)$ is real (imaginary) if, and only if, $Y(n)$ is even (odd). Conversely, $Y(n)$ is real (imaginary) if, and only if, $y(j)$ is even (odd).

It can also be deduced that:

- $y(j)$ is real and even if, and only if, $Y(n)$ is real and even
 - $y(j)$ is real and odd if, and only if, $Y(n)$ is imaginary and odd
 - $y(j)$ is imaginary and even if, and only if, $Y(n)$ is imaginary and even
- and
- $y(j)$ is imaginary and odd if, and only if, $Y(n)$ is real and odd.

A.3.4. The Shift Theorems.

From Section A.2.2.,

$$y(j-k) \Leftrightarrow Y(n) \cdot W_N^{-nk} \quad \text{A.3.19.}$$

Conversely,

$$y(j) \cdot W_N^{mj} \Leftrightarrow Y(n-m) \quad \text{A.3.20.}$$

A.3.5. The Delta Function.

Let d represent the delta function. Then,

$$d(j) \Leftrightarrow \frac{1}{N} \quad \text{A.3.21.}$$

$$\text{and} \quad 1 \Leftrightarrow d(n) \quad \text{A.3.22.}$$

A.3.6. Integration.

$$y(0) = \sum_{n=0}^{N-1} Y(n) \quad \text{A.3.23.}$$

$$\text{and} \quad Y(0) = \frac{1}{N} \sum_{j=0}^{N-1} y(j) \quad \text{A.3.24.}$$

A.3.7. The Convolution Theorem.

$$\text{If} \quad y_1(j) \Leftrightarrow Y_1(n)$$

$$\text{and} \quad y_2(j) \Leftrightarrow Y_2(n)$$

Then,

$$\frac{1}{N} \sum_{k=0}^{N-1} y_1(k) \cdot y_2(j-k) \Leftrightarrow Y_1(n) \cdot Y_2(n) \quad \text{A.3.25.}$$

Similarly,

$$y_1(j) \cdot y_2(j) \Leftrightarrow \sum_{m=0}^{N-1} Y_1(m) \cdot Y_2(n-m) \quad \text{A.3.26.}$$

Hence,

$$\frac{1}{N} \sum_{j=0}^{N-1} |y(j)|^2 = \sum_{n=0}^{N-1} |Y(n)|^2 \quad \text{A.3.27.}$$

A.3.8. The Doubling Theorems.

Two doubling theorems may be deduced. These are:

$$a) \text{ Let } y_1(j) = y(j) \iff Y_1(n) \quad j, n = 0(1)N-1$$

$$\text{and } y_2(j) = y(N+j) \iff Y_2(n) \quad j, n = 0(1)N-1$$

Then,

$$\begin{aligned} Y(2n) &= \frac{1}{2N} \sum_{j=0}^{2N-1} y(j) \cdot W_N^{-2nj} \\ &= \frac{1}{2N} \sum_{j=0}^{N-1} y_1(j) \cdot W_N^{-nj} + \frac{1}{2N} \sum_{j=0}^{N-1} y_2(j) \cdot W_N^{-nj} \cdot W_N^{-2nN} \\ &= \frac{1}{2} (Y_1(n) + Y_2(n)) \end{aligned} \quad \text{A.3.28.}$$

Thus the even terms of a DFT are the average of the terms obtained by transforming separately the first and second halves of the time sequence.

Also,

$$\begin{aligned} Y(2n+1) &= \frac{1}{2N} \sum_{j=0}^{2N-1} y(j) \cdot W_{2N}^{-(2N+1)j} \\ &= \frac{1}{2N} \sum_{j=0}^{N-1} y_1(j) \cdot W_N^{-nj} \cdot W_{2N}^{-j} + \frac{1}{2N} \sum_{j=0}^{N-1} y_2(j) \cdot W_N^{-nj} \cdot W_{2N}^{-j} \end{aligned}$$

Hence, from equation A.3.20.

$$Y(2n+1) = \frac{1}{2} (Y_1(n+\frac{1}{2}) + Y_2(n+\frac{1}{2})) \quad \text{A.3.29.}$$

The odd terms of a DFT are, therefore, the average of the "half interval" terms theoretically obtained by transforming separately the first and second halves of the time sequence.

$$b) \text{ Let } y_1(j) = y(2j) \iff Y_1(n) \quad j, n = 0(1)N-1$$

$$\text{and } y_2(j) = y(2j+1) \iff Y_2(n) \quad j, n = 0(1)N-1$$

Then,

$$\begin{aligned} Y(n) &= \frac{1}{2N} \sum_{j=0}^{2N-1} y(j) \cdot W_{2N}^{-nj} \\ &= \frac{1}{2N} \sum_{j=0}^{N-1} y_1(j) \cdot W_{2N}^{-2nj} + \frac{1}{2N} \sum_{j=0}^{N-1} y_2(j) \cdot W_{2N}^{-2nj-n} \\ &= \frac{1}{2} (Y_1(n) + Y_2(n) \cdot W_N^{-n}) \end{aligned} \quad \text{A.3.30.}$$

Similarly,

$$Y(N+n) = \frac{1}{2} (Y_1(n) - Y_2(n) \cdot W_N^{-n}) \quad \text{A.3.31.}$$

A number of interesting facts emerge from a study of the doubling theorems:

- The effect of doubling the length of a data sequence by adding zeros to the end of the sequence is seen, by setting $Y_2(n) = 0$ in equations A.3.28. and A.3.29., to be to halve the frequency interval between estimates of $Y(n)$, with the $Y(2n)$ estimates being identical to $\frac{1}{2}Y_1(n)$.
- The effect of doubling the length of a data sequence by adding a zero between each $y(j)$ is seen, by setting $Y_2(n) = 0$ in equations A.3.30. and A.3.31., to cause estimates of $Y(n)$ to be halved, and to be repeated such that $Y(n) = Y(N+n)$.
- If N is a power of two, then equations A.3.30. and A.3.31. can be used as the basis for computing the DFT of a data sequence, starting from a unity value of N . It turns out that the number of operations required to compute a DFT is dramatically reduced if this principle is employed, Algorithms so constructed are termed "fast Fourier transforms", see Appendix B.

A.3.9. The LFT of Two Real Data Sequences.

Let $Y_1(j) \Leftrightarrow Y_1(n)$

and $Y_2(j) \Leftrightarrow Y_2(n)$

Write $Y(j) = Y_1(j) + iY_2(j)$

Then the superposition theorem states that

$$Y(j) \Leftrightarrow Y(n) = Y_1(n) + iY_2(n)$$

But, from equation A.3.18.,

$$Y^*(N-n) = Y_1(n) - iY_2(n)$$

Hence:

$$Y_1(n) = \frac{1}{2} (Y(n) + Y^*(N-n)) \quad \text{A.3.32.}$$

$$\text{and } Y_2(n) = \frac{1}{2i} (Y(n) - Y^*(N-n)) \quad \text{A.3.33.}$$

$$\text{or, } Y_2(n) = \frac{1}{2} (Y^*(N-n) - Y(n)) \quad \text{A.3.34.}$$

Equations A.3.32. and A.3.34. provide the means for computing the DFT of two real data sequences from a single entry into a DFT routine. Similar expressions may be evolved for imaginary data sequences.

APPENDIX B.

The Fast Fourier Transform.

A bewildering variety of FFT algorithms have been devised in order to attain a maximum of efficiency under particular circumstances, for example to make best use of the attributes of a particular type of computer. All such algorithms employ the strategy expounded in Section A.3.8., and all make assumptions regarding the length of the data sequence to be transformed. It is generally assumed that the length of the sequence is equal to some power of an integer number. There is evidence to suggest that the most efficient algorithm assumes a length which is equal to a power of eight. However, the majority of algorithms are based upon binary or "power of two" lengths, i.e. $N = 2^p$. The reasons for this accord are that binary algorithms are relatively simple to understand and encode, the resulting algorithms are compact, and they offer the user the greatest choice of N for a given size of computer store.

Two binary algorithms are developed in this Appendix. The Cooley-Tukey algorithm deserves attention because it was the first to be published in modern times (Reference B.1.), whilst the second, known as the Sande-Tukey algorithm (Reference B.2.) is a logical development of the first which may be preferred in particular circumstances.

B.1. The Cooley-Tukey Algorithm.

The Cooley-Tukey algorithm is based specifically upon the second doubling algorithm of Section A.3.8. Re-stating this as a starting point, but with $2N$ replaced by N :

$$\text{If } Y_1(j) = y(2j) \iff Y_1(n) \quad j, n = 0(1)N/2-1 \quad \text{B.1.1.}$$

$$\text{and } Y_2(j) = y(2j+1) \iff Y_2(n) \quad j, n = 0(1)N/2-1 \quad \text{B.1.2.}$$

$$\text{then } Y(n) = \frac{1}{2} (Y_1(n) + Y_2(n) \cdot W_N^{-n}) \quad \text{B.1.3.}$$

$$\text{and } Y\left(\frac{N+n}{2}\right) = \frac{1}{2} (Y_1(n) - Y_2(n) \cdot W_N^{-n}) \quad \text{B.1.4.}$$

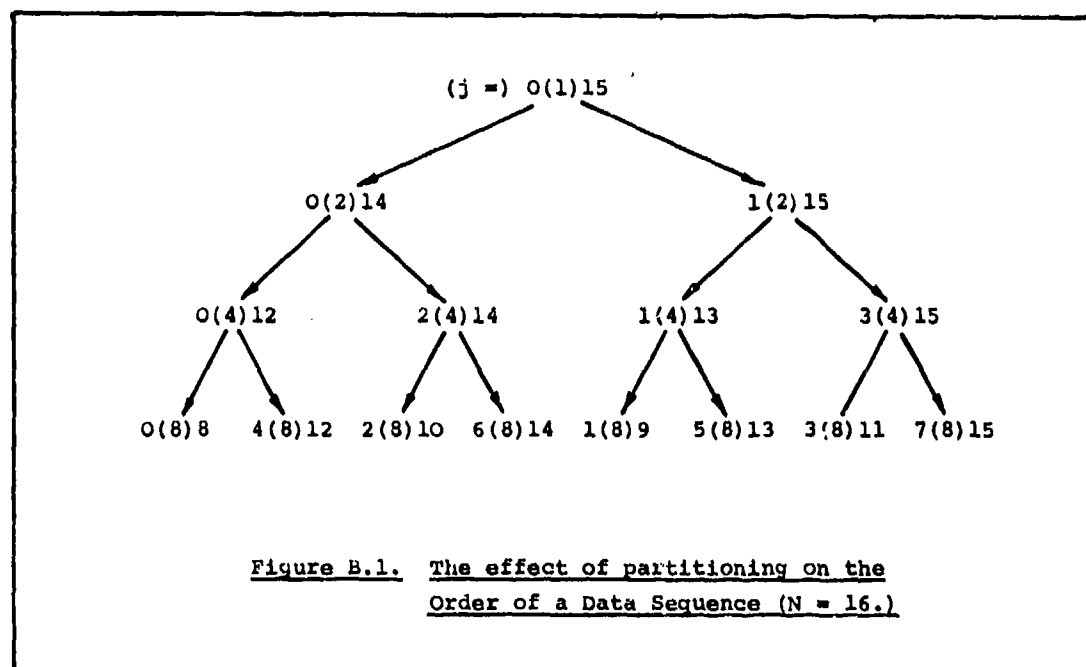
Before the transforms $Y_1(n)$ and $Y_2(n)$ can be calculated, the series $y(j)$ must be partitioned into two series $y_1(j)$ and $y_2(j)$ containing, respectively, the even and odd terms of $y(j)$ as defined in equations B.1.1 and B.1.2.

The transforms $Y_1(n)$ and $Y_2(n)$ may be calculated separately by again applying equations B.1.3. and B.1.4., but this time replacing N by $N/2$. Before this can be effected, the two series $y_1(j)$ and $y_2(j)$ must be further partitioned as defined in equations B.1.1. and B.1.2. (with $N/2$ replaced by $N/4$.)

The partitioning can be continued until N single element partitions have been obtained. At this point the argument is reversed, and the required DFT is obtained using successive applications of equations B.1.3. and B.1.4., it being noted that the DFT of a single element sequence is equal to the value of the element (see, for example, equation A.3.27.) i.e.

$$y(0) = Y(0) \quad N = 1 \quad \text{B.1.5.}$$

The partitioning process quite clearly alters the order of the data sequence $y(j)$. The precise nature of the alteration is indicated in figure B.1., which shows, for a 16 element sequence, the values of the suffices j after each partition.



It is not clear, at first sight, how a simple algorithm could be devised to arrive at the required order shown in figure B.1. However, if the initial and final orders are written in a binary format, as shown in figure B.2., it becomes clear that the partitioned sequence contains elements stored in "bit reversed" order, that is the relative address of the partitioned sequence can be determined by laterally inverting the binary address of the original sequence. Note also that the re-ordering involves sorting only one pair of elements at a time; it can therefore be effected in situ.

Original Order		Partitioned Order	
Decimal	Binary	Decimal	Binary
0	0000	0	0000
1	0001	8	1000
2	0010	4	0100
3	0011	12	1100
4	0100	2	0010
5	0101	10	1010
6	0110	6	0110
7	0111	14	1110
8	1000	1	0001
9	1001	9	1001
10	1010	5	0101
11	1011	13	1101
12	1100	3	0011
13	1101	11	1011
14	1110	7	0111
15	1111	15	1111

Figure B.2. Comparison of Original and Partitioned Data Sequence Orders (N = 16).

The final process in implementing the Cooley-Tukey algorithm involves applying equations B.1.3. and B.1.4. successively, as illustrated below:

First Step:

$$\begin{aligned} Y_1(2n) &= 0.5(y(2n) + y(2n+1)) \\ Y_1(2n+1) &= 0.5(y(2n) - y(2n+1)) \end{aligned} \quad n = 0(1)N/2-1$$

Second Step:

$$\begin{aligned} Y_2(4n+k) &= 0.5(Y_1(4n+k) + Y_1(4n+k+2) \cdot W_2^{-k}) \\ Y_2(4n+k+2) &= 0.5(Y_1(4n+k) - Y_1(4n+k+2) \cdot W_2^{-k}) \end{aligned} \quad \begin{aligned} k &= 0,1 \\ n &= 0(1)N/4-1 \end{aligned}$$

Third Step:

$$\begin{aligned} Y_3(8n+k) &= 0.5(Y_2(8n+k) + Y_2(8n+k+4) \cdot W_4^{-k}) \\ Y_3(8n+k+4) &= 0.5(Y_2(8n+k) - Y_2(8n+k+4) \cdot W_4^{-k}) \end{aligned} \quad \begin{aligned} k &= 0(1)2 \\ n &= 0(1)N/8-1 \end{aligned}$$

Etc.

Note that the weights can be obtained from the sequence W_N^{-j} , $j = 0(1)N/2-1$ by recognising that $W_P^{-k} = W_N^{-kN/P}$.

Note also that the first three steps shown above is, indeed, all that is required for the case when $N = 16$.

A flow diagram for the partitioning process is shown in figure B.3. and that for the final process leading to the DFT is shown in figure B.4.

B.2. The Sande-Tukey Algorithm.

The Sande-Tukey algorithm may be developed by reversing the argument used for developing the Cooley-Tukey algorithm. Referring once again to Section A.3.8., equations A.3.31. and A.3.32. can be manipulated to obtain the alternative expressions:

$$Y_1(n) = Y(n) + Y(N+n) \quad n = 0(1)N/2-1 \quad \text{B.2.1.}$$

$$Y_2(n) = (Y(n) - Y(N+n)) \cdot W_N^n \quad n = 0(1)N/2-1 \quad \text{B.2.2.}$$

where,

$$Y_1(n) \iff Y_1(j) = y(2j) \quad \text{B.2.3.}$$

$$\text{and } Y_2(n) \iff Y_2(j) = y(2j+1) \quad \text{B.2.4.}$$

Equations B.2.1. and B.2.2. are applied successively. At each step the number of partitions is doubled, and the length of each partition is halved, until $N = 1$. The result, noting equation B.1.5., is the IDFT of $Y(n)$, but the order of the result is "bit reversed." The sequence $y(j)$ may therefore be obtained by applying the algorithm shown in figure B.3. as the final stage.

A flow diagram for the Sandy-Tukey algorithm is shown in figure B.5.

Given: $y(j)$, $j = 0(1)2N-2$ (assumed to consist of an N -length complex sequence.)

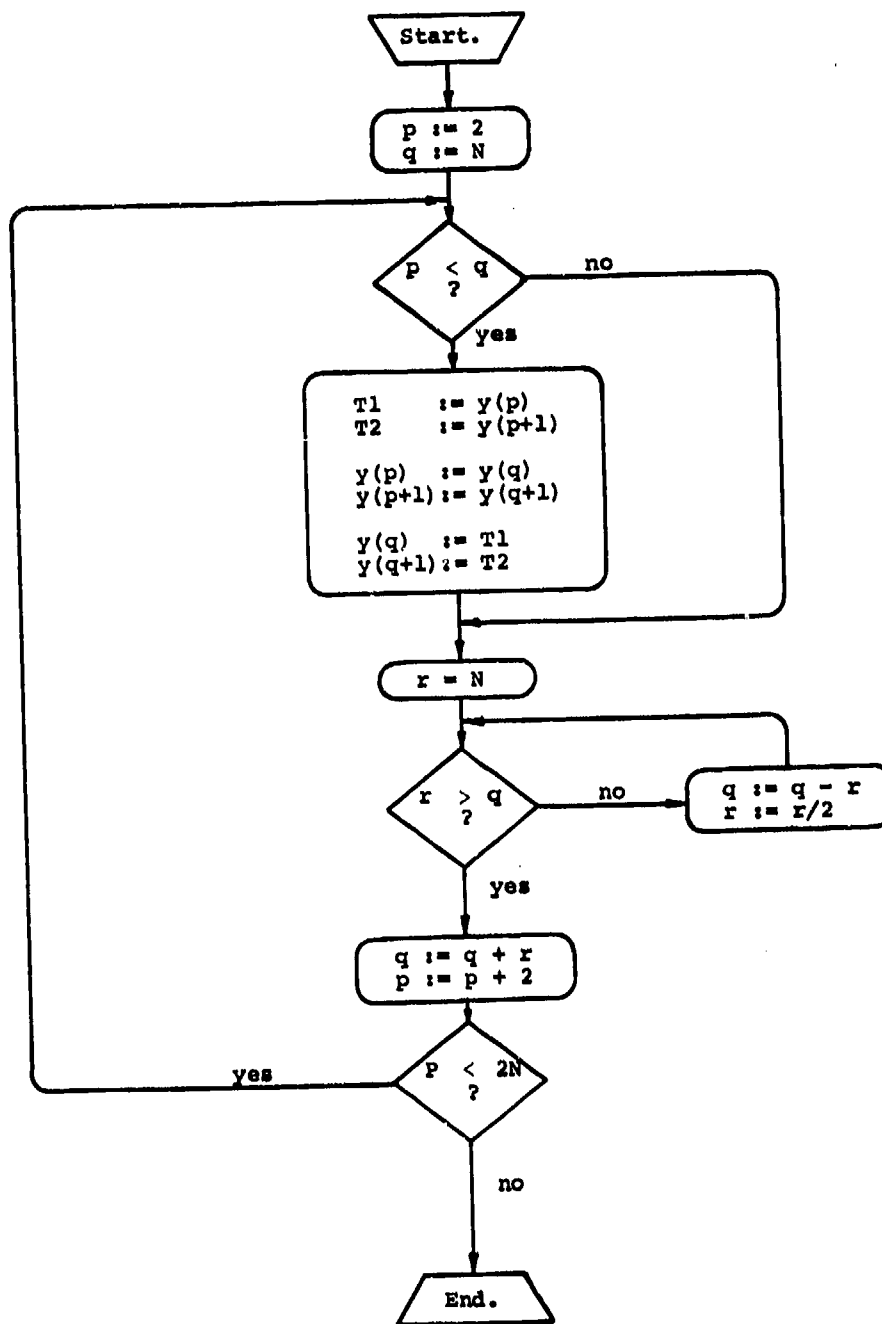


Figure B.3. A "Bit Reversing" Algorithm.

Given: $y(j)$, $j = 0(1)2N-2$ (assumed to consist of an N -length complex sequence.)

$$C(r) = \cos \frac{2\pi r}{N}, \quad r = 0(1)\frac{N-1}{2}$$

$$S(r) = \sin \frac{2\pi r}{N}, \quad r = 0(1)\frac{N-1}{2}$$

$W = -1$ for DFT, $+1$ for IDFT

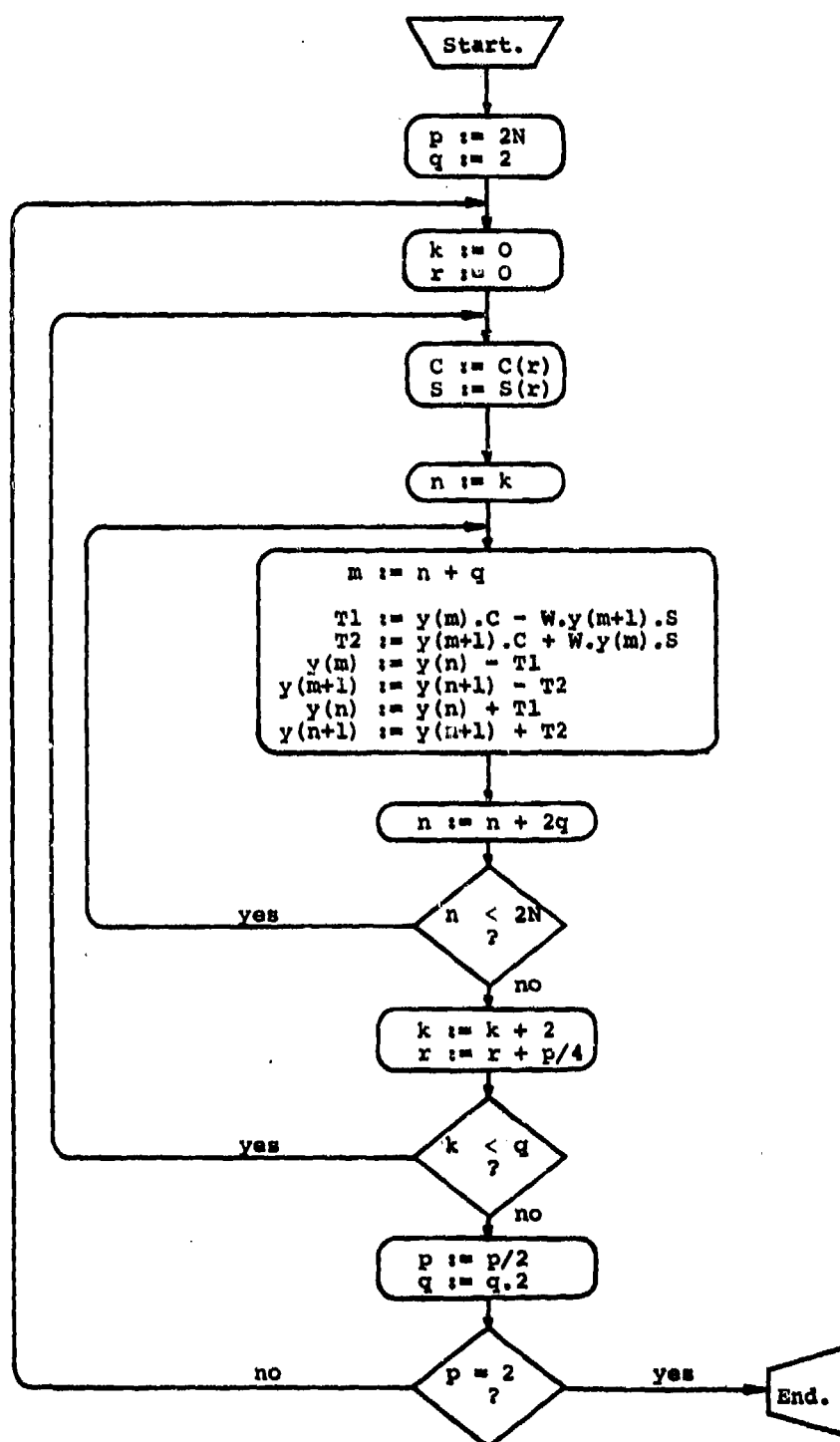


Figure B.4. The Cooley-Tukey FFT Algorithm.

Given: $y(j)$, $j = 0(1)2N-2$ (assumed to consist of an N -length complex sequence.)

$$C(r) = \cos \frac{2\pi r}{N}, \quad r = 0(1)\frac{N-1}{2}$$

$$S(r) = \sin \frac{2\pi r}{N}, \quad r = 0(1)\frac{N-1}{2}$$

$W = -1$ for DFT, $+1$ for IDFT.

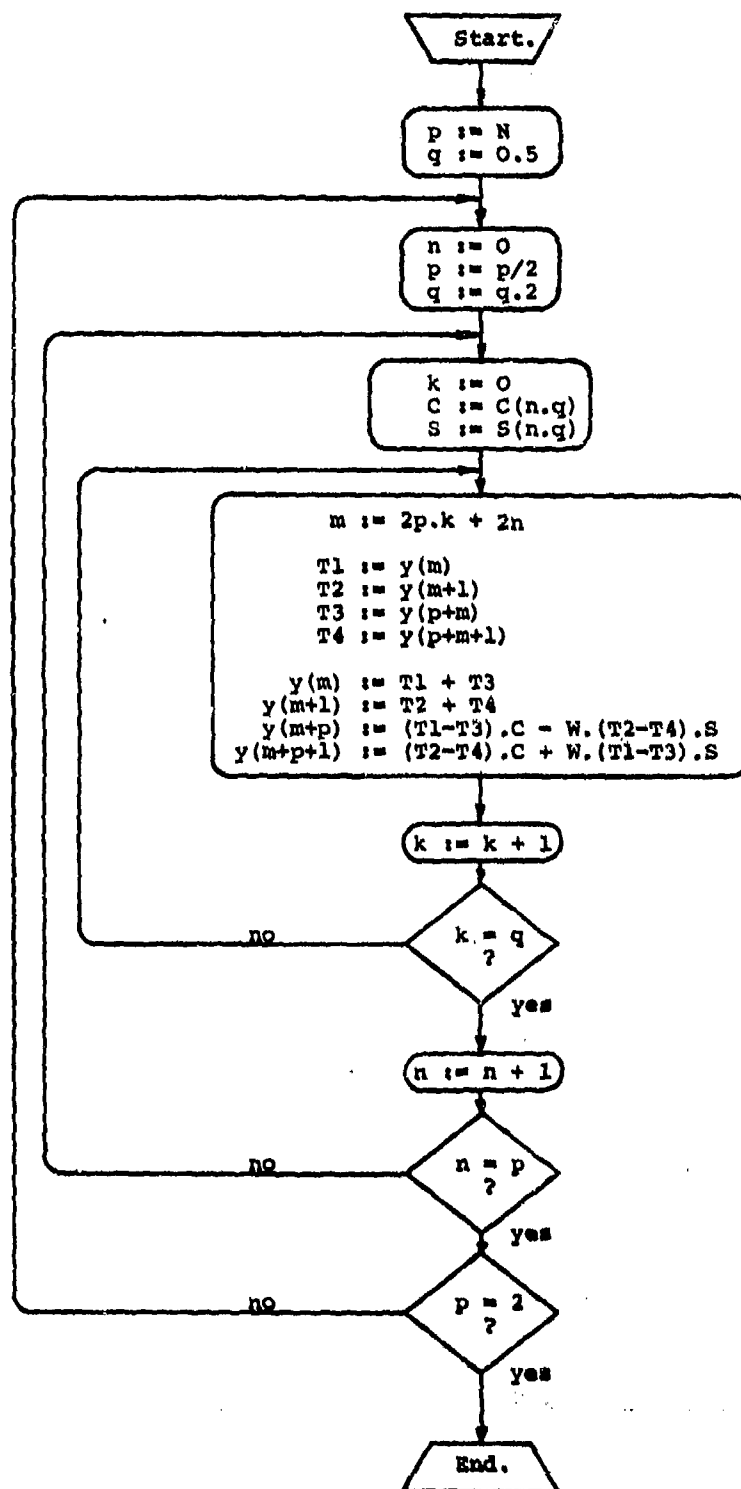


Figure B.5. The Sande-Tukey FFT Algorithm.

B.3. Comments.

The fast Fourier transforms which have been developed in this Appendix can each be used for effecting both the DFT and the inverse DFT (IDFT). The logic of this statement can be seen by examining equations A.3.8. and A.3.9. It can be seen from these that the DFT and the IDFT are identical except for a scale factor (1/N) and a minus sign included in the exponential function for the DFT. In practice, it is common for the scale factor to be omitted altogether from FFT routines, leaving the user to make the appropriate adjustment. When this is the case, the only difference between the DFT and its IDFT is the exponential term. The flow diagrams shown in figures B.4. and B.5. indicate the necessary adjustment required for each type of transform (and omit the scale factor.)

The algorithm to be preferred for a particular application will depend upon personal preferences and, perhaps, upon the facilities afforded by the computer instruction set, even though the number of arithmetic operations required by the two algorithms is identical. Certainly the Sande-Tukey algorithm has proved to be slightly more efficient on one 16 bit word-length computer. It should also be noted that when spectral analysis is to be effected in a "real time" environment, then the bit reversal routine can be entered "off-line" if the Sande-Tukey algorithm is used.

The bit reversal algorithm shown in figure B.3. can be implemented as stated in any high level language. It will, however, be very slow. It is to be recommended that the bit reversal calculation at least be executed in machine level code, taking advantage of logical instructions which are available at that level.

The efficiency of the FFT routines described in this Appendix is normally quoted in comparison with the time required to achieve the same result by implementing equation A.3.8. or A.3.9. directly. The exact speed ratio will depend upon the characteristics of the particular computer used, but will be of the order

$$\frac{\log_e N}{N} \quad \dagger \quad 0.01 \quad \text{for } N = 1024.$$

An additional benefit afforded by FFT routines is superior accuracy by virtue of the fact that fewer arithmetic operations are involved, leading to smaller "rounding" errors.

The restriction that the length of a data sequence must be a power of two for the FFT algorithms described may be overcome by adding a suitable number of zeros to a sequence of arbitrary length. The justification for this procedure can be implied from the results of Section A.3.8. The resulting transform must be weighted to account for the fact that data was present for only a proportion of the original sequence.

APPENDIX C.

Some Procedures which make use of the DFT.

This Appendix contains procedures for computing estimates of PSD and CSD from arrays of sampled data, and estimates of auto and cross correlation functions from estimates of PSD and CSD using the DFT as a filtering element.

It was decided to present the procedures in a formal high level language. The author recognised that Fortran is the scientific language, most widely used. However the clarity of Fortran code leaves something to be desired, particularly to engineers who are not fully conversant with the language. It was therefore decided to adopt Algol 60 as the language for this Appendix, partly because it is relatively easy to understand, and partly because it is probably the most popular scientific language apart from Fortran. The author apologises to adherents of other languages for their choice but, as no idiomatic tricks have been employed in the procedures, they anticipate that such enthusiasts will experience little difficulty in effecting a suitable translation.

The procedures have been designed to present the various steps in a logical manner and with a degree of clarity. It is worth noting, therefore, that the procedures would translate into grossly inefficient code if they were to be compiled exactly as they have been presented here.

C.1. A Procedure to Compute the PSD of a Single Data Sequence.

Procedure PSDA(y,N,M,H,h) Result:(Y); comment Note that semicolons are used to terminate every statement;

value y,N,M,H,h; integer N,M,H; real h; array y(0:2*N*M-1),Y(0:N-1);

begin comment This procedure computes the N-length PSD estimate of the 2NM-length data sequence y. The result is presented in Y. H is a hanning marker, and is set zero if hanning is not required. a 4N-length workspace array (yw) is used. The sampling interval is assumed to be h seconds, and N is assumed to be a power of two;

integer j,m; comment Declare counters used in the procedure;

real pi; pi:=3.14159265359; comment Declare and set pi;

array yw(0:4*N-1); comment Declare the workspace array;

Step 1: comment Clear the result array (it is used as an accumulator);

for j:=0 step 1 until N-1 do Y[j] := 0;

for m:=0 step 1 until M-1 do

begin comment Start of the block loop;

Step 2: comment Copy the next 2N data samples into the odd elements of the workspace array. Clear the even elements;

for j:=0 step 1 until 2*N-1 do begin
 yw[2*j] := y[j+m*N];
 yw[2*j+1] := 0;
end;

comment Halve the first and last elements;

yw[0] := 0.5*yw[0]; yw[4*N-2] := 0.5*yw[4*N-2];

Step 3: comment Weight the data if hanning is required;

if H = 0 then goto Step 4;

for j:=0 step 1 until 2*N-1 do yw[2*j] := 0.5*yw[2*j]*(1-cos(pi*j/N));

Step 4: comment Compute the DFT of yw(j), j = 0(1)4N-1 with W = -1 using the Sande-Tukey algorithm, figure B.5., followed by the sorting algorithm, figure B.3.;

FFT (yw,2*N,-1);

Step 5: comment Form the square of the moduli of the complex result and sum into the accumulator array;

for j:=0 step 1 until N-1 do

Y[j] := Y[j] + yw[2*j]*yw[2*j] + yw[2*j+1]*yw[2*j+1];

end; of the block loop. Steps 2,3,4 and 5 are executed M times.

Step 6: comment Correct the results if hanning was used;
 if H = 0 then goto Step 7;
 for j:=0 step 1 until N-1 do Y[j] := Y[j]*8/3;

Step 7: comment Compute the Energy, or mean square amplitude, spectrum;
 for j:=0 step 1 until N-1 do Y[j] := 2*Y[j]/((2*N-1)*(2*N-1)*M);

Step 8: comment Compute the PSD estimate by dividing by the analysis bandwidth;
 for j:=0 step 1 until N-1 do Y[j] := Y[j]*2*N*h;

end of the procedure;

Notes: Only half of the DFT array was used in step 5. This was because the original data was real (step 2) and the conjugate functions theorem (Section A.3.3) shows that the upper half of the DFT array contains the conjugate of the lower half (in reverse order.)

The factor $(2N-1)^2$ was included in step 7 because the FFT algorithm did not perform the necessary scaling (see Section B.3.) The factor $2N-1$, rather than $2N$, was used because the first and last elements of the data array were halved.

If the Sande-Tukey FFT algorithm is used to compute the DFT, then it is possible to effect the "bit reversal" sorting once only, outside the main loop. If this approach is adopted, then the addresses used in step 5 must be modified to suit.

The algorithm described above is simple to implement, but is expensive in terms of storage used for the FFT ($4N$ locations.) This may be reduced to $2N$ locations by replacing steps 2 to 5 inclusive by the following:

Step 2a: comment Copy the next $2N$ data samples into consecutive locations of the workspace array;
 for j:=0 step 1 until $2*N-1$ do yw[j] := y[j+m*N];
 comment Halve the first and last elements;
 yw[0] := 0.5*yw[0]; yw[$2*N-1$] := 0.5*yw[$2*N-1$];

Step 3a: comment Weight the data if hanning is required;
 if H = 0 then goto Step 4a;
 for j:=0 step 1 until $2*N-1$ do yw[j] := 0.5*yw[j]*(1-cos(pi*j/N));

Step 4a: comment Compute the DFT of yw(j), $j = 0(1)2N-1$ with $W = -1$ using the Sande-Tukey algorithm, followed by the sorting algorithm. Note that the length of the DFT is now one half the length previously used in step 4;
 FFT (yw,N,-1);

Step 5a: comment Use equations A.3.32. and A.3.34. to separate the transforms of the two sequences $yw(2j)$ and $yw(2j+1)$. Then use equation A.3.30. to compute the final transform. The two processes may be combined, but they cannot be performed in situ. Note also that the first element must be treated as a special case;

for $j:=1$ step 1 until $N-1$ do

begin

$S := \sin(\pi j/N); \quad C := \cos(\pi j/N);$

$T1 := 0.5*(yw[2*j]+yw[2*N-2*j]-S*(yw[2*j]-yw[2*N-2*j]))$
 $+C*(yw[2*N-2*j+1]+yw[2*j+1]);$

$T2 := 0.5*(yw[2*j+1]+yw[2*N-2*j+1]-S*(yw[2*j+1]+yw[2*N-2*j+1]))$
 $-C*(yw[2*j]-yw[2*N-2*j]);$

$Y[j] := Y[j] + T1*T1 + T2*T2;$

end;

comment Now for the first element;

$Y[0] := Y[0] + yw[0]*yw[0];$

The real declaration at the beginning of the procedure must now include the additional workspace locations $S, C, T1, T2$. Further, the workspace array yw may be declared with $2N$ elements, rather than the original $4N$.

Under normal circumstances, the exponential functions used in hanning the data, computing the FFT, and those used in step 5a would all be extracted from previously computed arrays in order to reduce the execution time.

C.2. A Procedure to Compute the PSD and CSD of Two Data Sequences.

Procedure PSDB(y1,y2,N,M,H,h) Result: (Y1,Y2,Y3,Y4);

value y1,y2,N,M,H,h; integer N,M,H; real h; array y1[0:2*N*M-1], y2[0:2*N*M-1];
array Y1[0:N-1], Y2[0:N-1], Y3[0:N-1], Y4[0:N-1];

begin comment This procedure computes the N-length PSD estimates of the 2NM-length data sequences y1 and y2. The results are presented in Y1 and Y2, respectively. The procedure also computes the CSD of the two sequences, assuming y1 contains the notional input. The real and imaginary parts of the CSD are presented in Y3 and Y4, respectively. H is a hanning marker, and is set zero if hanning is not required. A 4N-length workspace array (yw) is used. The sampling interval is assumed to be h seconds, and N is assumed to be a power of two;

integer j,m; comment Declare counters used in the procedure;

real T1,T2,T3,T4,pi; pi:=3.14159265359; comment Declare variables and set pi;

array yw[0:4*N-1]; comment Declare the workspace array;

Step 1: comment Clear the result arrays (they are used as accumulators);

for j:=0 step 1 until N-1 do

begin

 Y1[j] := 0; Y2[j] := 0; Y3[j] := 0; Y4[j] := 0;

end;

for m:=0 step 1 until M-1 do

begin comment Start of the block loop;

Step 2: comment Copy the next 2N data samples from y1 into the odd elements of the workspace array, and the next 2N data samples from y2 into the even elements of the workspace array;

for j:=0 step 1 until 2*N-1 do

begin

 yw[2*j] := y1[j+m*N]; yw[2*j+1] := y2[j+m*N];

end;

comment Halve the first and last elements;

 yw[0] := 0.5*yw[0]; yw[4*N-2] := 0.5*yw[4*N-2];

 yw[1] := 0.5*yw[1]; yw[4*N-1] := 0.5*yw[4*N-1];

Step 3: comment Weight the data if hanning is required;

if H = 0 then goto Step 4;

for j:=0 step 1 until 2*N-1 do

begin

 yw[2*j] := 0.5*yw[2*j]*(1-cos(pi*j/N));

 yw[2*j+1] := 0.5*yw[2*j+1]*(1-cos(pi*j/N));

end;

```

Step 4:  comment  Compute the DFT of  $y_w(j)$ ,  $j = 0(1)4N-1$  with  $W = -1$  using the
           Sande-Tukey algorithm, figure B.5., followed by the sorting
           algorithm, figure B.3.;

           FFT (yw,2*N,-1);

Step 5:  comment  Separate the two transforms and accumulate the square of the moduli
           in Y1 and Y2, and the product of the second and the conjugate of
           the first in Y3 (real) and Y4 (imaginary);

           for j:=1 step 1 until N-1 do
           begin
               T1 := 0.5*(yw[2*j] + yw[2*N-2*j]);
               T2 := 0.5*(yw[2*j+1] - yw[2*N-2*j+1]);
               T3 := 0.5*(yw[2*j+1] + yw[2*N-2*j+1]);
               T4 := 0.5*(yw[2*N+2*j] - yw[2*j]);
               Y1[j] := Y1[j] + T1*T1 + T2*T2;
               Y2[j] := Y2[j] + T3*T3 + T4*T4;
               Y3[j] := Y3[j] + T1*T3 + T2*T4;
               Y4[j] := Y4[j] + T1*T4 - T2*T3;
           end;
           comment  Compute the first elements;
           Y1[0] := Y1[0] + yw[0]*yw[0];  Y2[0] := Y2[0] + yw[1]*yw[1];
           Y3[0] := Y3[0] + yw[0]*yw[1];

           end; of the block loop. Steps 2,3,4 and 5 are executed M times;

Step 6:  comment  Correct the results if hanning was used;
           if H = 0 then goto Step 7;
           for j:=0 step 1 until N-1 do
           begin
               Y1[j] := Y1[j]*8/3;    Y2[j] := Y2[j]*8/3;
               Y3[j] := Y3[j]*8/3;    Y4[j] := Y4[j]*8/3;
           end;

Step 7:  comment  Compute the Energy spectra and cross spectrum;
           for j:=0 step 1 until N-1 do
           begin
               Y1[j] := 2*Y1[j]/((2*N-1)*(2*N-1)*M);
               Y2[j] := 2*Y2[j]/((2*N-1)*(2*N-1)*M);
               Y3[j] := 2*Y3[j]/((2*N-1)*(2*N-1)*M);
               Y4[j] := 2*Y4[j]/((2*N-1)*(2*N-1)*M);
           end;

Step 8:  comment  Compute the PSD and CSD estimates by dividing by the analysis
           bandwidth;

           for j:=0 step 1 until N-1 do
           begin
               Y1[j] := Y1[j]*2*N*h;    Y2[j] := Y2[j]*2*N*h;
               Y3[j] := Y3[j]*2*N*h;    Y4[j] := Y4[j]*2*N*h;
           end;

           end    of the procedure;

```

```
value y,N,df;   integer N;   real df;   array y[0:N-1];
```

```
integer j;      comment  Declare the counter used in the procedure;
real C,S,pi;    pi:=3.14159265359;  comment  Declare variables and set pi;
array yw[0:2*N-1];  comment  Declare the workspace array;
```

end of the procedure;

C.4. A Procedure to Compute the Cross-correlation Function from a CSD Estimate.

Procedure CROSS(y1,y2,N,df);

value y1,y2,N,df; integer N; real df; array y1[0:N-1],y2[0:N-1];

begin comment This procedure computes the 2N-length cross-correlation function from an N-length (complex) CSD estimate. The result is real. Positive time delays are returned in y1, and negative time delays are returned in y2, thereby overwriting the CSD estimate. The interval between each result element is $0.5/(N \cdot df)$ seconds. N is assumed to be a power of two. A 4N-length workspace array is used;

integer j; comment Declare the counter used in the procedure;

array yw[0:4*N-1]; comment Declare the workspace array;

Step 1: comment Clear the workspace array;

for j:=0 step 1 until 4*N-1 do yw[j] := 0;

Step 2: comment Copy the CSD from y1 (real) and y2 (imaginary), filling half the workspace;

for j:=0 step 1 until N-1 do

begin

 yw[2*j] := y1[j];

 yw[2*j+1] := y2[j];

end;

comment Halve the first and last elements of each array;

 yw[0] := 0.5*yw[0]; yw[1] := 0.5*yw[1];

 yw[2*N-2] := 0.5*yw[2*N-2]; yw[2*N-1] := 0.5*yw[2*N-1];

Step 3: comment Compute the IDFT of yw;

 FFT(yw,2*N,1);

Step 4: comment Complete the result as specified at the start;

for j:=1 step 1 until N-1 do

begin

 y1[j] := 0.5*df*(yw[2*j]-yw[2*j+1]+yw[2*N-2*j]-yw[2*N-2*j+1]);

 y2[j] := 0.5*df*(yw[2*j]+yw[2*j+1]+yw[2*N-2*j]+yw[2*N-2*j+1]);

end;

comment Compute the first two points separately;

 y1[0] := yw[0]; y2[0] := yw[0];

end of the procedure;

APPENDIX D.

Probability Measures.

Random processes yield data whose values cannot be predicted with any certainty. The best that can be hoped for is a description, which permits an estimate of the probability that a nominated event will occur using, as the basis for the estimate, averages obtained from data produced by the process at some time in the past. The concept of probability is therefore of primary importance in the study of random processes and the techniques used for describing data produced by such processes.

An introduction to those probability measures which are most frequently encountered when analysing random data is contained in this Appendix.

D.1. Fundamentals.

A collection of observations taken at various times of the instantaneous value of a function of time $y(t)$ is called a sample. An infinitely large number of such observations is termed a population.

If the number of possible values which can be adopted by the function is finite within a finite range of values, then the probability that any observation will adopt a particular value is equal to the limiting value of the ratio of the number of observations occurring at that value to the total number of observations, as the number of observations extends to the whole population. Specifically,

$$P(y_0) = \lim_{N \rightarrow \infty} \frac{N(y_0)}{N} \quad \text{D.1.1}$$

Probability is therefore a number to be associated with a chosen value of $y_0(t)$. It will have a value between zero and unity. Note also that, since every observation will have some value,

$$\sum_{y_0=-\infty}^{\infty} P(y_0) = 1 \quad \text{D.1.2.}$$

When the function of time is continuous, then the number of possible values which the function can adopt is infinite even when the range of possible values is finite. The probability of the function adopting a particular value is therefore zero, presuming that the function does vary with time.

This difficulty can be circumvented by defining a different probability measure. The probability that a function $y(t)$ takes a value which is less than, or equal to, a nominated value y_0 , say, is called the distribution function or cumulative function of y_0 . Thus:

$$D(y_0) = \text{Probability } (-\infty < y < y_0) \quad \text{D.1.3.}$$

The probability that the function adopts a value between y_0 and y_1 , say, is given by:

$$D(y_1) - D(y_0) = \text{Probability } (y_0 < y < y_1) \quad \text{D.1.4.}$$

The functions defined by equations D.1.3. and D.1.4. are capable of describing both discrete and continuous functions. Provided that $y(t)$ is everywhere finite, then they are continuous functions of, respectively, y_0 and y_1 and y_0 . When $y(t)$ is a continuous function, then the derivatives of the probability functions are also continuous.

Equation D.1.4. defines a function of y_0 and $(y_1 - y_0) = y$. Dependence upon the latter can be removed by postulating the function:

$$p(y) = \lim_{\Delta y \rightarrow 0} \frac{D(y + \Delta y) - D(y)}{\Delta y} \quad \text{D.1.5.}$$

$$= \frac{dD(y)}{dy} \quad \text{D.1.6.}$$

$p(y)$ is known as the probability density function. It is arguably the most useful probability measure for a single function.

The probability measures introduced above have a number of properties. The more useful of these are listed below:

$$D(y_0) = \int_{-\infty}^{y_0} p(y) \cdot dy \quad \text{D.1.7.}$$

$$0 < D(y_0) < 1. \quad \text{D.1.8.}$$

$$0 < p(y) < 1 \quad \text{D.1.9.}$$

$$\int_{-\infty}^{\infty} p(y) \cdot dy = 1 \quad \text{D.1.10.}$$

In general, the expected value of a function of y , $f(y)$, is given by:

$$E[f(y)] = \int_{-\infty}^{\infty} f(y) \cdot p(y) \cdot dy \quad \text{D.1.11.}$$

The moments of $y(t)$ characterize the shape of a probability density function. The n^{th} moment is defined as:

$$M_n = \int_{-\infty}^{\infty} y^n \cdot p(y) \cdot dy \quad \text{D.1.12.}$$

It can be deduced that:

$$M_0 = 1, \quad \text{from equation D.1.10.} \quad \text{D.1.13.}$$

$$M_1 = \bar{y}, \quad \text{the mean value of } y(t) \quad \text{D.1.14.}$$

$$M_2 = \psi_y^2, \quad \text{the mean square value of } y(t) \quad \text{D.1.15.}$$

$$M_3 \quad \text{is a measure of asymmetry, or "skewness"} \quad \text{D.1.16.}$$

$$M_4 \quad \text{is a measure of "kurtosis"} \quad \text{D.1.17.}$$

Formally, if $M_1 = 0$, then M_3^2/M_2^3 is a non-dimensional measure of skewness, and kurtosis, a measure of the "peakiness" of a probability density function, is given by M_4/M_2^2 . If the mean value is subtracted from $y(t)$, then the resulting moments are known as central moments. In general:

$$CM_n = \int_{-\infty}^{\infty} (y - \bar{y})^n \cdot p(y) \cdot dy \quad \text{D.1.18.}$$

It follows that, for example,

$$CM_0 = 1, \quad \text{as before} \quad \text{D.1.19.}$$

$$CM_1 = 0 \quad \text{D.1.20.}$$

$$CM_2 = \sigma_y^2, \quad \text{the variance of } y(t) \quad \text{D.1.21.}$$

The moments of $y(t)$ can be related to the conjugate of the Fourier transform of the probability density function, as follows:

$$\text{Let } P^*(r) = \int_{-\infty}^{\infty} p(y) \cdot \exp(ir y) \cdot dy \quad \text{D.1.22.}$$

$$= \int_{-\infty}^{\infty} p(y) \cdot (1 + ir y + \dots + \frac{(ir y)^n}{n!} + \dots) \cdot dy \quad \text{D.1.23.}$$

This is called the characteristic function. It may be written:

$$P^*(r) = \int_{-\infty}^{\infty} p(y) \cdot dy + ir \int_{-\infty}^{\infty} y \cdot p(y) \cdot dy + \dots + \frac{(ir)^n}{n!} \int_{-\infty}^{\infty} y^n \cdot p(y) \cdot dy \dots \quad \text{D.1.24.}$$

$$\text{or, } P^*(r) = M_0 + ir.M_1 + \dots + \frac{(ir)^n}{n!} M_n + \dots \quad \text{D.1.25.}$$

Conversely,

$$p(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P^*(r) \cdot \exp(-iry) \cdot dr \quad \text{D.1.26.}$$

D.2. The Probability Density of a Function.

If a function $x(t)$ which varies with time is transformed to obtain a new function:

$$y(t) = f(x) \quad \text{D.2.1.}$$

and $f(x)$ is single valued, then the probability density of $y(t)$ is given by:

$$p_y(y) = \frac{dx}{dy} \cdot p_x(x) \quad \text{D.2.2.}$$

When $y(t)$ corresponds to n values of $x(t)$, i.e. $x(t)$ is multi-valued, then the probability density function of $y(t)$ becomes:

$$p_y(y) = n \cdot \frac{dx}{dy} \cdot p_x(x) \quad \text{D.2.3.}$$

As an illustration of the above results, consider the function:

$$y = A \cdot \sin(x)$$

If the probability density function of $x(t)$ is uniform, that is,

$$\begin{aligned} p_x(x) &= \frac{1}{2\pi}, & 0 < x < 2\pi \\ &= 0, & 0 > x \text{ or } x > 2\pi \end{aligned}$$

then, because $x = \arcsin(y/A)$ is a double valued function (so that $n = 2$), the probability density of $y(t)$ is given by:

$$\begin{aligned} p_y(y) &= \frac{1}{2\pi} \cdot 2 \cdot (A^2 - y^2)^{-\frac{1}{2}}, & -A < y < A \\ &= 0, & -A > y \text{ or } y > A \end{aligned}$$

The above expression represents the probability density function of a sine wave.

It can be shown that, in general,

- A linear transformation of a Gaussian process is itself a Gaussian process.
- A non-linear transformation of a Gaussian process yields a non-Gaussian process.
- A linear transformation of a non-Gaussian process yields a process having a different probability density function from the original.

D.3. Joint Probability Functions.

The ideas presented in Section D.1. were concerned with describing the characteristics of a single variable. This Section is concerned with the extension of those principles to the situation where the joint characteristics of two variables are of interest.

Suppose that two functions of time, $x(t)$ and $y(t)$, exist. The functions may, or may not, be independent of one another. The probability distribution of each function considered separately is defined by equation D.1.3. A third probability distribution may also be defined, as follows:

$$D(x_0, y_0) = \text{Probability}(-\infty < y < y_0 \text{ and } -\infty < x < x_0) \quad \text{D.3.1.}$$

$D(x_0, y_0)$ is known as the joint probability distribution function. In a similar way,

$$p(x, y) = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{D(x + \Delta x, y + \Delta y) - D(x, y)}{\Delta x \cdot \Delta y} \quad \text{D.3.2.}$$

$$= \frac{\partial^2 D(x, y)}{\partial x \partial y} \quad \text{D.3.3.}$$

is known as the joint probability density function.

The joint probability density function has the properties:

$$\int_{-\infty}^{\infty} p(x, y) \cdot dx = p(y) \quad \text{D.3.4.}$$

$$\int_{-\infty}^{\infty} p(x, y) \cdot dy = p(x) \quad \text{D.3.5.}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \cdot dx \cdot dy = 1 \quad \text{D.3.6.}$$

In general, the expected value of a function of x and y , $f(x, y)$, is given by:

$$E[f(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot p(x, y) \cdot dx \cdot dy \quad \text{D.3.7.}$$

A definition of the cross-correlation function may be deduced from equation D.3.7., as follows:

$$R_{xy}(\tau) = E[x(t) \cdot y(t+\tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \cdot y(t+\tau) \cdot p(x(t), y(t+\tau)) \cdot dx \cdot dy \quad \text{D.3.8.}$$

The concept of joint probability can be extended to situations where the joint properties of many variables are of interest.

D.4. Conditional Probability Functions.

The two-variable situation considered in Section D.3. may be taken a step further by defining a subset of the joint probability distribution as:

$$F(x_0 | y_0, y_1) = \frac{\text{Probability}(-\infty < x < x_0 \text{ given } y_0 < y < y_1)}{\text{Probability}(y_0 < y < y_1)} \quad \text{D.4.1.}$$

The normalising factor has been introduced in order to ensure that the function has the property:

$$\int_{-\infty}^{\infty} F(x|y_0, y_1) \cdot dx = 1 \quad D.4.2.$$

The conditional probability function is derived from equation D.4.1. by letting the y interval tend to zero. That is:

$$D(x|y) = \lim_{\Delta y \rightarrow 0} \frac{F(x|y, y+\Delta y)}{D(y+\Delta y) - D(y)} \quad D.4.3.$$

Equation D.4.3. defines the conditional probability distribution. It follows, by comparison with the results of Section D.1., that:

$$\begin{aligned} p(x|y) &= \frac{dD(x|y)}{dx} \\ &= \frac{p(x,y)}{p(y)} \end{aligned} \quad D.4.4.$$

defines the conditional probability density function.

Note that if x is independent of y, then:

$$p(x|y) = p(x) \quad D.4.5.$$

and so

$$p(x,y) = p(x) \cdot p(y) \quad D.4.6.$$

Equation D.4.6. can be used to demonstrate the independence, or otherwise, of two functions x and y.

D.5. Stationarity and Ergodicity.

The preceding Sections of this Appendix contain definitions of a variety of probability measures, each measure, at least potentially, providing a description of one aspect of a random process. An estimate of each measure can always be obtained provided that an ensemble of sample time histories produced by the process are available for analysis. Whether, or not, such estimates are meaningful (i.e. contribute to an understanding of the process) depends upon the nature of the process and upon the behavior of the sample time histories yielded by the process.

Two important classes of random process can be identified for which probability measures yield meaningful information about the process. These are:

D.5.1. Stationary Processes.

A stationary random process is one in which the expected values of ensemble moments and joint moments do not vary with time. Specifically, a weakly stationary process is defined as one in which the mean value and the auto-correlation function, as defined by equations D.1.14. and D.3.8. respectively, are time invariant. A strictly stationary process is defined as one in which all moments and joint moments are time invariant.

D.5.2. Ergodic Processes.

An ergodic random process is a stationary process in which the expected values of ensemble moments and joint moments are equal to the expected values of the corresponding sample averages. The class is again sub-divided into weakly ergodic processes in which the sample mean value and auto-correlation function are representative of the ensemble, and strictly ergodic processes in which all sample averages are representative of the ensemble.

D.6. The Normal Distribution.

It can be shown that, if a process is the result of many independent random causes, none of which contribute significantly to the overall process, then the probability density function of data output by the process is Normal, or Gaussian. The probability density function of a Normal distribution is given by:

$$p(y) = \frac{1}{\sqrt{2\pi} \cdot \sigma_y} \cdot \exp\left(-\frac{(y - \bar{y})^2}{2\sigma_y^2}\right) \quad \text{D.6.1.}$$

The above statement is known as the Central Limit Theorem. It is employed to good effect in, for example, random noise generators which use a Geiger counter and a radioactive source.

It is worth noting that a Normal process which is demonstrably weakly stationary (ergodic) is also strictly stationary (ergodic) because the probability density function is completely defined by the ensemble mean value and variance.

The response of a linear system to an arbitrary input is given by:

$$y_2(t) = \int_{-\infty}^t h(t-s) \cdot y_1(s) \cdot ds \quad \text{D.6.2.}$$

where $h(t)$ is the impulse response function (IRF) of the system. When the IRF extends over a significant time (i.e. is long compared with the characteristic wavelength of the input), then equation D.6.2. can be considered to state that the output of the linear system comprises the weighted sum of a number of independent processes. It is therefore to be expected that, when the input is random, the output of a linear system satisfying this condition will have a probability density which is more nearly Normal than that of the input.

The properties of random processes noted above mean that stationary data collected from such processes very often exhibit a Normal probability density distribution. This, in turn, accounts for the importance attached to the Normal distribution.

The probability that a function $y(t)$ having a Normal distribution exceeds a nominated value, y_0 , is given by:

$$P(y > y_0) = \int_{y_0}^{\infty} p(y) \cdot dy \quad \text{D.6.3.}$$

$$= \frac{1}{\sqrt{2\pi} \cdot \sigma_y} \int_{y_0}^{\infty} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \cdot dy \quad \text{D.6.4.}$$

assuming that the function has a zero mean value. It is convenient to change the variable in equation D.6.4. to:

$$x = \sigma_y \cdot y \quad \text{D.6.5.}$$

With this substitution, equation D.6.4. becomes:

$$P(x > x_0) = \frac{1}{\sqrt{2\pi}} \int_{x_0}^{\infty} \exp\left(-\frac{x^2}{2}\right) \cdot dx \quad \text{D.6.6.}$$

The right hand side of equation D.6.6. is related to the Error Function, and is widely tabulated. Reference D.1. includes an algorithm for computing the Error Function.

The Normal distribution, and certain useful associated probabilities, are listed in Table D.1.

$\frac{y_0}{\sigma_y} = x_0$	$p(x_0)$	$P\{x > x_0\}$	$P\{ x > x_0\}$	$P\{ x < x_0\}$
0	0.39894	0.50000	1.00000	0.00000
0.5	0.35207	0.30854	0.61708	0.38292
1.0	0.24197	0.15866	0.31732	0.68268
1.5	0.12952	0.06681	0.13362	0.86638
2.0	0.05399	0.02275	0.04550	0.95450
2.5	0.01753	0.00621	0.01242	0.98758
3.0	0.00443	0.00135	0.00270	0.99730
3.5	0.00087	0.00023	0.00047	0.99953
4.0	0.00013	0.00003	0.00006	0.99994
4.5	0.00002	0.00000	0.00001	0.99999
5.0	0.00000	0.00000	0.00000	1.00000

Table D.1. The Normal Distribution and Related Probability Distributions.

D.7. Properties of Normally Distributed Processes.

It is to be assumed, for the purpose of this Section, that a stationary, random function of time, $y(t)$, has a Normal distribution and a zero mean value. It is also to be assumed that, when they are specified, the various time derivatives of $y(t)$ exist.

D.7.1. The Expected Number of Level Crossings per Second.

The probability that the function adopts a value between y and $y + dy$ and, at the same time, a velocity between $\dot{y}$ and $\dot{y} + d\dot{y}$ is, if dy and $d\dot{y}$ are sufficiently small,

$$p(y, \dot{y}) \cdot dy \cdot d\dot{y} \quad \text{D.7.1.}$$

But, $dy = \dot{y} \cdot dt$ (distance = speed \cdot time) D.7.2.

The average number of times y is exceeded in time dt is equal to the probability that the function lies between y and $y + dy$ and also has a positive velocity. Hence, if $\bar{N}_y$ is the average number of times per second that y is exceeded,

$$\bar{N}_y \cdot dt = dt \cdot \int_0^{\infty} \dot{y} \cdot p(y, \dot{y}) \cdot d\dot{y} \quad \text{D.7.3.}$$

or, $\bar{N}_y = \int_0^{\infty} \dot{y} \cdot p(y, \dot{y}) \cdot d\dot{y}$ D.7.4.

This result is true for any probability distribution. For a Normal distribution, the expected value of the product $y \cdot \dot{y}$,

$$E[y \cdot \dot{y}] = \left. \frac{d}{dx} (R_{yy}(x)) \right|_{x=0} = 0. \quad \text{D.7.5.}$$

so that y and $\dot{y}$ are independent of one another. Equation D.4.6. can therefore be invoked to write:

$$p(y, \dot{y}) = p(y) \cdot p(\dot{y}) \quad \text{D.7.6.}$$

Hence, $\bar{N}_y = p(y) \cdot \int_0^{\infty} \dot{y} \cdot p(\dot{y}) \cdot d\dot{y}$ D.7.7.

$$= p(y) \cdot \frac{\sigma_{\dot{y}}}{\sqrt{2\pi}} \quad \text{D.7.8.}$$

$$= \frac{1}{2\pi} \cdot \frac{\sigma_{\dot{y}}}{\sigma_y} \cdot \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \quad \text{D.7.9.}$$

It is noted that:

$$\sigma_y^2 = \int_0^{\infty} G_{yy}(f) \cdot df$$

and $\sigma_{\dot{y}}^2 = (2\pi)^2 \cdot \int_0^{\infty} f^2 \cdot G_{yy}(f) \cdot df$

These results may be substituted into equation D.7.9. to give:

$$\bar{N}_y = \left[\frac{\int_0^{\infty} f^2 \cdot G_{yy}(f) \cdot df}{\int_0^{\infty} G_{yy}(f) \cdot df} \right]^{\frac{1}{2}} \cdot \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \quad \text{D.7.10.}$$

Equation D.7.10. is known as Rice's formula (Reference D.2.) It states that, for a Normally distributed process, the average number of positive going level crossings per second can be obtained directly once the power spectral density of the process is known. It can be inferred that, for this to have a finite value, the ultimate "roll-off" rate of the PSD must be greater than 6 dB per octave.

Equation D.7.10. is often written:

$$\bar{N}_y = \bar{N}_0 \cdot \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \quad D.7.11.$$

where, $\bar{N}_0$ is the average number of positive going zero crossings per second.

A "feel" for the implications of the above can be obtained by considering a process which has an ideal band limited white noise PSD centered at f_0 Hz with a bandwidth f_1 Hz, as shown in figure D.1. This has the property:

$$\bar{N}_0 = f_0 \cdot (1 + K^2/12) \quad D.7.12.$$

where $K = \frac{f_1}{f_0}$

Hence,

$$\bar{N}_0 \rightarrow f_0 \text{ as } f_1 \rightarrow 0$$

and $\bar{N}_0 > f_0$ if $f_1 > 0$

If $f_1 = 2.f_0$, the PSD becomes that shown in figure D.2., and

$$\bar{N}_0 = 0.5773.f_1$$

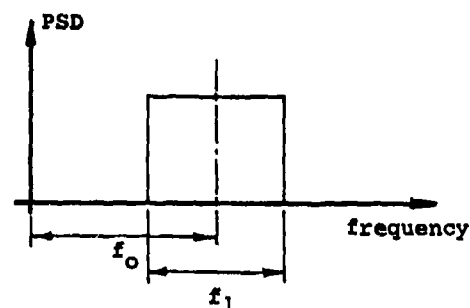


Figure D.1. PSD of a band-limited white noise process.

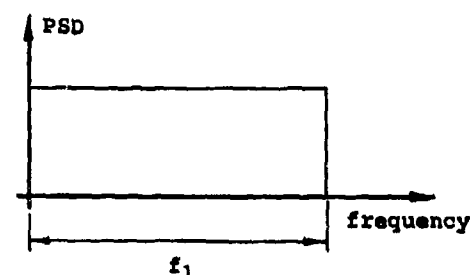


Figure D.2. PSD of a low-pass filtered white noise process.

D.7.2. The Expected Number of Maxima per Second.

The probability distribution of maxima is governed by the joint probability density:

$$p(y, \dot{y}, \ddot{y})$$

with a maximum being defined as $\dot{y} = 0$ and $\ddot{y} < 0$. It can be demonstrated (Reference D.3.) that the probability density of maxima for a Normal process is given by:

$$p_p(y) = \frac{(1-z^2)^{1/2}}{\sigma_y (2\pi)^{1/2}} \cdot \exp\left[\frac{-y^2}{2\sigma_y^2(1-z^2)}\right] + \frac{z \cdot y}{2\sigma_y^2} \left[1 + \operatorname{erf} \frac{y \cdot z}{2\sigma_y (1-z^2)}\right] \cdot \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \quad D.7.13.$$

where $z = \frac{\bar{N}_0}{\bar{N}_m} \quad D.7.14.$

and $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-r^2) \cdot dr \quad D.7.15.$

$\bar{N}_m$ is the expected number of maxima per second. Using an argument similar to that employed in Section D.7.1., it may be deduced that,

$$\bar{N}_m = \frac{1}{2\pi} \frac{\sigma_y}{\sigma_f} = \left[\frac{\int_0^\infty f^4 \cdot G_{yy}(f) \cdot df}{\int_0^\infty f^2 \cdot G_{yy}(f) \cdot df} \right]^{1/2} \quad \text{D.7.16.}$$

For the ideal band limited white noise process shown in figure D.1., the expected number of maxima per second becomes:

$$\bar{N}_m = f_0 \cdot \frac{(1 + K^2/2 + K^4/80)^{1/2}}{(1 + K^2/12)^{1/2}} \quad \text{D.7.17.}$$

where $K = \frac{f_1}{f_0}$

When $f_1 = 2 \cdot f_0$, as shown in figure D.2., this reduces to:

$$\bar{N}_m = 0.7746 \cdot f_1 \quad \text{D.7.18.}$$

Equations D.7.11. and D.7.17. indicate that, for a narrow band process,

$$f_1 \rightarrow 0, \text{ so that } Z \rightarrow 1$$

that is, there is one maximum for each positive going zero crossing. In this particular case, the probability density of maxima becomes, from equation D.7.13.,

$$p_p(y) = \frac{y \cdot \exp(-y^2)}{\sigma_y^2 \cdot 2\sigma_y^2} \quad (y > 0) \quad \text{D.7.19.}$$

This is known as the Rayleigh Distribution.

Conversely, for a broad band process,

$$f_1 \rightarrow \infty, \text{ and } Z \rightarrow 0$$

and the probability density of maxima becomes:

$$p_p(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \cdot \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \quad \text{D.7.20.}$$

Thus, for a broad band process, the probability density distribution of maxima is equal to the amplitude probability density distribution.

APPENDIX E.

Error Analysis.

The results obtained from an analysis of random data are never exact. Thus two selections extracted from a recording made under nominally steady conditions will always yield different results. The magnitude and the character of the differences, or likely differences, between two such analyses are just as important as the results themselves if the results are to be interpreted with any degree of confidence.

This Appendix contains an outline of the method by which the errors associated with the analysis of random data may be estimated. The method is then applied to the task of estimating the errors for several of the more familiar analysis techniques.

E.1. Principles of Error Analysis

The result yielded by an attempt to determine a characteristic of a random process is an estimate of that characteristic. A number of similar estimates of a characteristic will form another random sequence having a mean value and a variance. The precision of an individual estimate can be described by the variance of the estimates. The accuracy of an individual estimate, on the other hand, can be described by the expected value of the square of the difference between an estimate and the true value, i.e.

$$\begin{aligned}
 & E [(\hat{f}(y) - f(y))^2] \\
 &= E [(\hat{f}(y) - E[\hat{f}(y)] + E[\hat{f}(y)] - f(y))^2] \\
 &= E [(\hat{f}(y) - E[\hat{f}(y)])^2] \\
 &+ 2.E [(\hat{f}(y) - E[\hat{f}(y)]) . (E[\hat{f}(y)] - f(y))] \\
 &+ E [(E[\hat{f}(y)] - f(y))^2]
 \end{aligned}
 \tag{E.1.1.}$$

The first term represents the variance of the random sequence of estimates. The second term is zero because:

$$E [\hat{f}(y) - E[\hat{f}(y)]] = E[\hat{f}(y)] - E[\hat{f}(y)] = 0 \tag{E.1.2.}$$

The third term will be zero if, and only if, the average value of estimates tends to the true value as the number of estimates increases without limit. The third term is therefore referred to as a bias error.

The two types of error, namely variance error and bias error, are normally evaluated separately since the first can be considered to be a shortcoming of the experiment, whilst the second can be considered to be a shortcoming of the analysis technique.

It is normal, when presenting estimates for the two types of error, to normalise the square root of the quantities derived above. Thus:

$$\text{Normalised variance error, } \epsilon_v = \frac{\{E [(\hat{f}(y) - E[\hat{f}(y)])^2]\}^{\frac{1}{2}}}{f(y)} \tag{E.1.3.}$$

$$\begin{aligned}
 \text{Normalised bias error, } \epsilon_b &= \frac{\{E [(E[\hat{f}(y)] - \hat{f}(y))^2]\}^{\frac{1}{2}}}{f(y)} \\
 &= \frac{E[\hat{f}(y)] - f(y)}{f(y)}
 \end{aligned}
 \tag{E.1.4.}$$

$$\text{The Normalised RMS error, } \epsilon = \{ \epsilon_v^2 + \epsilon_b^2 \}^{\frac{1}{2}} \tag{E.1.5.}$$

E.2. Variance Errors

The variance error term of equation E.1.1. can be expanded to give:

$$\begin{aligned}
 E [(\hat{f}(y) - E[\hat{f}(y)])^2] &= E[\hat{f}(y)^2] - 2.E[\hat{f}(y) . E[\hat{f}(y)]] + E[E[\hat{f}(y)]]^2 \\
 &= E[\hat{f}(y)^2] - E[E[\hat{f}(y)]]^2
 \end{aligned}
 \tag{E.2.1.}$$

Write
$$\hat{f}(y) = \frac{1}{T} \int_0^T g(t) \cdot dt \quad \text{E.2.2.}$$

where $g(t)$ is some function of time based upon $y(t)$. Then the first term of equation E.2.1. becomes:

$$E[\hat{f}(y)^2] = \frac{1}{T^2} \int_0^T \int_0^T E[g(a) \cdot g(b)] \cdot db \cdot da \quad \text{E.2.3.}$$

Substituting $b = a + r$, equation E.2.3. becomes:

$$E[\hat{f}(y)^2] = \frac{1}{T^2} \int_0^T \int_{-a}^{T-a} E[g(a) \cdot g(a+r)] \cdot dr \cdot da \quad \text{E.2.4.}$$

$$= \frac{1}{T^2} \int_0^T \int_{-a}^{T-a} R_g(r) \cdot dr \cdot da \quad \text{E.2.5.}$$

Equation E.2.5. can be evaluated to yield:

$$E[\hat{f}(y)^2] = \frac{2}{T} \int_0^T \left(1 - \frac{r}{T}\right) \cdot R_g(r) \cdot dr \quad \text{E.2.6.}$$

Note that $R_g(r)$ is the autocorrelation function of $g(t)$. It may be written:

$$R_g(r) = K_g(r) + \bar{g}^2 \quad \text{E.2.7.}$$

Where $K_g(r)$ is the autocovariance function of $g(t)$ and $\bar{g}$ is the mean value of $g(t)$. Thus, equation E.2.6. may be written:

$$E[\hat{f}(y)^2] = \frac{2}{T} \int_0^T \left(1 - \frac{r}{T}\right) \cdot (K_g(r) + \bar{g}^2) \cdot dr \quad \text{E.2.8.}$$

$$= \frac{2}{T} \int_0^T \left(1 - \frac{r}{T}\right) \cdot K_g(r) \cdot dr + \bar{g}^2 \quad \text{E.2.9.}$$

This may be substituted into Equation E.2.1. to give, from equation E.1.3.,

$$(\hat{f}(y) - \bar{f})^2 = \frac{2}{T} \int_0^T \left(1 - \frac{r}{T}\right) \cdot K_g(r) \cdot dr \quad \text{E.2.10.}$$

This is a general result which can be used to estimate the variance error of any function of y .

E.2.1. Mean Value Estimates

The mean value of a function $y(t)$ is estimated by:

$$\hat{\bar{y}} = \frac{1}{T} \int_0^T y(t) \cdot dt \quad \text{E.2.11.}$$

Comparison with E.2.2. shows that $g(t) = y(t)$ so that equation E.2.11. becomes:

$$(\bar{y}(t) \cdot \epsilon_v)^2 = \frac{2}{T} \int_0^T (1 - \frac{r}{T}) \cdot K_{yy}(r) \cdot dr \quad \text{E.2.12.}$$

Note that the autocovariance function $K_{yy}(r)$ tends to zero at large T for an ergodic process, so that:

$$\epsilon_v \rightarrow 0 \quad \text{as} \quad T \rightarrow \infty \quad \text{E.2.13.}$$

Equation E.2.12. is often difficult to evaluate for a particular application, if only because the true autocovariance function (as opposed to an estimate of it) cannot be determined experimentally. As a result, it is normal to assume that the power of the random component of a data sequence is uniformly distributed over a bandwidth B Hz. The autocovariance function for a sequence of this type is:

$$K_{yy}(r) = \sigma_y^2 \cdot \frac{\sin 2\pi Br}{2\pi Br} \quad \text{E.2.14.}$$

When $Br \gg 1$, equation E.2.12. can be evaluated for this function to yield:

$$\bar{y} \cdot \epsilon_v = \sigma_y \cdot \frac{1}{(2Br)^{1/2}} \quad \text{E.2.15.}$$

E.2.2. Mean Square Value Estimates

The mean square value of a function $y(t)$ is estimated by:

$$\hat{\psi}^2 = \frac{1}{T} \int_0^T y^2(t) \cdot dt \quad \text{E.2.16.}$$

Comparison with E.2.2. shows that $g(t) = y^2(t)$.

The autocovariance function included in equation E.2.10. may be written for this case:

$$R_g(r) - \bar{g}^2 = E[y(t) \cdot y(t) \cdot y(t+r) \cdot y(t+r)] - \bar{g}^2 \quad \text{E.2.17.}$$

An important general relationship which is relevant here is:

$$\begin{aligned} E[y_1 y_2 y_3 y_4] &= E[y_1 \cdot y_2] \cdot E[y_3 \cdot y_4] + E[y_1 \cdot y_3] \cdot E[y_2 \cdot y_4] \\ &\quad + E[y_1 \cdot y_4] \cdot E[y_2 \cdot y_3] - 2 \cdot \bar{y}_1 \cdot \bar{y}_2 \cdot \bar{y}_3 \cdot \bar{y}_4 \end{aligned} \quad \text{E.2.18.}$$

Hence, equation E.2.17. can be written:

$$K_g(r) = R_y^2(0) + 2 \cdot R_y^2(r) - 2 \cdot \bar{y}^4 - \bar{g}^2 \quad \text{E.2.19.}$$

But $R_Y^2(0) = \bar{g}^2$, so that E.2.19. becomes:

$$\begin{aligned} K_g(r) &= 2.R_Y^2(r) - 2.\bar{y}^4 \\ &= 2.(K_Y(r) + \bar{y}^2)^2 - 2.\bar{y}^4 \\ &= 2.K_Y^2(r) + 4.\bar{y}^2.K_Y(r) + 2.\bar{y}^4 - 2.\bar{y}^4 \\ &= 2.K_Y(r).(K_Y(r) + 2.\bar{y}^2) \end{aligned} \quad \text{E.2.20.}$$

Finally, therefore, from equation E.2.10.,

$$(\psi^2.\epsilon_v)^2 = \frac{4}{T} \int_0^T (1 - \frac{r}{T}).K_Y(r).[K_Y(r) + 2.\bar{y}^2].dr \quad \text{E.2.21.}$$

When the PSD of the noise is assumed to be uniform over a bandwidth B Hz., and $BT \gg 1$,

$$K_Y(r) = \frac{\sigma^2 \sin 2\pi Br}{2.\pi Br} \quad \text{E.2.22.}$$

from equation E.2.14., equation E.2.21. becomes, for $\bar{y} = 0$,

$$(\psi^2.\epsilon_v)^2 = \frac{K_Y^2(0)}{BT} \quad \text{E.2.23.}$$

$$\text{or, } \epsilon_v = \frac{1}{(BT)^{\frac{1}{2}}} \quad \text{E.2.24.}$$

E.2.3. Correlation Function Estimates

The cross-correlation function of two random processes $y_1(t)$ and $y_2(t)$ is estimated by

$$\hat{R}_{12}(r) = \frac{1}{T} \int_0^T y_1(t).y_2(t+r).dt \quad \text{E.2.25.}$$

Thus, for this case, $g(t) = y_1(t).y_2(t+t_1)$. The autocovariance function included in equation E.2.10. may then be written:

$$R_g(r) - \bar{g}^2 = E[y_1(t).y_2(t+t_1).y_1(t+r).y_2(t+t_1+r)] - \bar{g}^2 \quad \text{E.2.26.}$$

This may be written, from equation E.2.18.,

$$\begin{aligned} K_g(r) &= K_{11}(r).K_{22}(r) + 2[y_1^2.K_{22}(r) + y_1^2.K_{11}(r)] \\ &+ K_{12}(t_1+r).K_{12}(t_1-r) + 2.y_1.y_2[K_{12}(t_1+r) + K_{12}(t_1-r)] \end{aligned} \quad \text{E.2.27.}$$

This may be substituted into equation E.2.10. to provide a general expression for the variance error of cross-correlation function estimates. When both processes have a zero mean value and $BT \gg 1$, then E.2.10. gives:

$$(R_{12}(r).\epsilon_v)^2 = \frac{2}{T} \int_0^T (1 - \frac{r}{T}) [K_{11}(t).K_{22}(t) + K_{12}(t_1+t).K_{12}(t_1-r)].dt \quad \text{E.2.28.}$$

$$\approx \frac{1}{(BT)^{\frac{1}{2}}} [K_{11}(0).K_{22}(0) + K_{12}^2(r)]^{\frac{1}{2}} \quad \text{E.2.29.}$$

or, when $K_{12} \neq 0$,

$$\epsilon_v = \frac{1}{(2BT)^{\frac{1}{2}}} \left[1 + \frac{\sigma_1^2 \cdot \sigma_2^2}{K_{12}^2(r)} \right]^{\frac{1}{2}} \quad \text{E.2.30.}$$

It may be noted that the Normalised variance error for autocorrelation function estimates is, when $K_{yy}(r) \neq 0$, given by:

$$\epsilon_v = \frac{1}{(2BT)^{\frac{1}{2}}} \left[1 + \frac{\sigma_y^4}{K_{yy}^2(r)} \right]^{\frac{1}{2}} \quad \text{E.2.31.}$$

When $r = 0$, equation E.2.31. reduces to equation E.2.24.

Note that the expressions for the Normalised variance error for auto and cross correlation functions are, unlike previous results, dependent upon the value of the appropriate function. In fact

$$\epsilon_v \rightarrow \infty \quad \text{as} \quad K_{yy}^2 \rightarrow 0 \quad \text{E.2.32.}$$

Therefore it is not really appropriate to normalise the error in this case.

E.2.4. Power Spectral Density Estimates

Estimates of power spectral density are subject to a variance error which may be evaluated from first principles. Precisely the same result may be obtained by recognising that PSD estimates are obtained by computing the mean square value of the original signal after it has been selectively filtered (see Section 2.4.1.). The normalised variance error for mean square value estimates is given by equation E.2.24.,

$$\epsilon_v = \frac{1}{(BT)^{\frac{1}{2}}} \quad \text{E.2.33.}$$

This expression describes the normalised variance error for a PSD estimate, provided that B is interpreted as the effective analyser, rather than the data, bandwidth.

When PSD estimates are obtained from discrete Fourier transforms (see Appendix C), equation E.2.33. takes a particular form since:

$$B = \frac{1}{h.N} \quad \text{and} \quad T = h.N.M.$$

where M is the number of transform blocks used to obtain the PSD estimate. Thus the normalised variance error can be written:

$$\epsilon_v = \frac{1}{M^{\frac{1}{2}}} \quad \text{E.2.34.}$$

This surprisingly simple result states that no useful information is obtained if only one transform is used to complete the PSD estimate. This is not unreasonable when it is considered that a DFT is just another way of describing the original data. The result also indicates the large quantity of data required in order to obtain reasonably accurate PSD estimates. In order to illustrate the point, suppose that the PSD is required from data which has been sampled at a rate of 100 per second. Suppose also that 2048 samples are used for each transform. Then the normalised variance error can be reduced to 10 percent if around 34 minutes of data are available. However, approximately 57 hours of data are required to obtain 1 percent normalised variance error.

Finally, it is worth noting that the technique of "looping", (widely used in early analogue analysers) or overlapping data does nothing to reduce the variance error of PSD estimates, even though the results may be made to look more convincing.

REPORT DOCUMENTATION PAGE			
1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-160 Volume 14	ISBN 92-835-1405-X	UNCLASSIFIED
5. Originator	Advisory Group for Aerospace Research and Development North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly sur Seine, France		
6. Title	THE ANALYSIS OF RANDOM DATA		
7. Presented at			
8. Author(s)/Editor(s)	D.A. Williams		9. Date November 1981
10. Author's/Editor's Address	Cranfield Institute of Technology Cranfield Bedfordshire UK		11. Pages 168
12. Distribution Statement	This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.		
13. Keywords/Descriptors	Flight tests Test equipment Data processing Random inputs Signal processing Aircraft		
14. Abstract	This AGARDograph is the 14th of the AGARD Flight Test Instrumentation Series and discusses the analysis of random data. The availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment. Further, the techniques described in this Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. This Volume is not intended to be a reference document for the specialist analyst who is required to generate the software for analyzing random data. Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD.		

AGARDograph No.160 - Volume 14 Advisory Group for Aerospace Development, NATO THE ANALYSIS OF RANDOM DATA by D.A.Williams Published November 1981 168 pages This AGARDograph is the 14th of the AGARD Flight Test Instrumentation Series and discusses the analysis of random data. The availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment. Further, the techniques described in this P.T.O.	AGARD-AG-160 Vol.14 Flight tests Test equipment Data processing Random inputs Signal processing Aircraft	AGARDograph No.160 - Volume 14 Advisory Group for Aerospace Development, NATO THE ANALYSIS OF RANDOM DATA by D.A.Williams Published November 1981 168 pages This AGARDograph is the 14th of the AGARD Flight Test Instrumentation Series and discusses the analysis of random data. The availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment. Further, the techniques described in this P.T.O.	AGARD-AG-160 Vol.14 Flight tests Test equipment Data processing Random inputs Signal processing Aircraft
AGARDograph No.160 - Volume 14 Advisory Group for Aerospace Development, NATO THE ANALYSIS OF RANDOM DATA by D.A.Williams Published November 1981 168 pages This AGARDograph is the 14th of the AGARD Flight Test Instrumentation Series and discusses the analysis of random data. The availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment. Further, the techniques described in this P.T.O.	AGARD-AG-160 Vol.14 Flight tests Test equipment Data processing Random inputs Signal processing Aircraft	AGARDograph No.160 - Volume 14 Advisory Group for Aerospace Development, NATO THE ANALYSIS OF RANDOM DATA by D.A.Williams Published November 1981 168 pages This AGARDograph is the 14th of the AGARD Flight Test Instrumentation Series and discusses the analysis of random data. The availability of powerful computing facilities, both on-line and off-line, for processing experimental data means that the flight test engineer has great flexibility in choosing the dividing line between "hard wired" and "soft" signal conditioning equipment. Further, the techniques described in this P.T.O.	AGARD-AG-160 Vol.14 Flight tests Test equipment Data processing Random inputs Signal processing Aircraft

Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. This Volume is not intended to be a reference document for the specialist analyst who is required to generate the software for analyzing random data. Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD. ISBN 92-835-1405-X	Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. This Volume is not intended to be a reference document for the specialist analyst who is required to generate the software for analyzing random data. Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD. ISBN 92-835-1405-X
Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. This Volume is not intended to be a reference document for the specialist analyst who is required to generate the software for analyzing random data. Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD. ISBN 92-835-1405-X	Volume are being used increasingly to extract meaningful information in situations where more conventional test and analysis techniques are inappropriate. This Volume is not intended to be a reference document for the specialist analyst who is required to generate the software for analyzing random data. Rather it is intended to introduce the non-specialist both to the possibilities and to the fundamental limitations of those techniques which are most frequently encountered. This AGARDograph has been sponsored by the Flight Mechanics Panel of AGARD. ISBN 92-835-1405-X

8431

9

AGARD

NATO OTAN

7 RUE MICHELLE 92200 MUILLY SUR SEINE
FRANCE

Telephone 74000000 - Telex 810178

DISTRIBUTION OF UNCLASSIFIED
AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Coordination AGARD - VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA

Defence Science Information Services
Department of National Defence
Ottawa, Ontario K1A 0K2

DENMARK

Danish Defence Research Board
Oesterbrogades Kaserne
Copenhagen Q

FRANCE

O.N.S.M.A. (Direction)
29 Avenue de la Division Lesciere
92320 Châtillon sous Bagneux

GERMANY

Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungsanstalt
D-7514 Eggenstein-Lepoldshafen 2

GREECE

Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens

ICELAND

Director of Aviation
c/o Flugrad
Reykjavik

ITALY

Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG

See Belgium

NETHERLANDS

Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
3600 A.C. Delft

NORWAY

Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL

Direcção do Serviço de Material
da Força Aérea
Rua da Escola Politécnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY

Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM

Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES

National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche

Space Documentation Service
European Space Agency
10, rue Mario Villis
75015 Paris, France

Microfilm

European Space Agency
10, rue Mario Villis
75015 Paris, France
Attn: Microfilm Section

Requests for microfiche or photocopies of AGARD documents should include the AGARD report number, title, report or series, and publication date. Requests to NTIS should include the NASA accession report number. Full details of the AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Aeronautics (GRA)
published by the National Technical
Information Service, Springfield
Vols. 22161, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St, London W1P 1RD

ISBN 92-835-1405-X