Non-uniform sampling and the zero crossing distribution

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:


Metadata Record: [https://dspace.lboro.ac.uk/2134/34527](https://dspace.lboro.ac.uk/2134/34527)

Publisher: © M.O. Al-Nuaimi

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
NON-UNIFORM SAMPLING

AND

THE ZERO CROSSING DISTRIBUTION.

By

M.O. Al-Nuaimi, B.Sc.

A Thesis submitted to the Electronic and Electrical Engineering Department, Loughborough University of Technology for the degree of Doctor of Philosophy.

Supervisors:  Prof. J.W.R. Griffiths

Mr. A. Pratt  October 1971.
ACKNOWLEDGEMENTS

The foremost acknowledgements go to Professor J.W.R. Griffiths and Mr. A. Pratt for their guidance and helpful discussions.

The author wishes to thank Mrs. L. Al-Nuaimi for typing the manuscript.

Thanks are due to the University of Technology, Loughborough for supporting this research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF PRINCIPAL SYMBOLS</td>
<td></td>
</tr>
<tr>
<td>GENERAL INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td><strong>PART I</strong></td>
<td></td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1 Problems associated with Non-Uniform Sampling</td>
<td></td>
</tr>
<tr>
<td>1.2 Purpose of Investigation</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 2: THEORETICAL CONSIDERATION</td>
<td></td>
</tr>
<tr>
<td>2.1 Recurrent Non-Uniform Sampling</td>
<td></td>
</tr>
<tr>
<td>2.1.1 Theorem</td>
<td></td>
</tr>
<tr>
<td>2.2 Frequency Spectrum of Sampled Signals</td>
<td></td>
</tr>
<tr>
<td>2.3 The Recovery Scheme</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 3: THEORETICAL ANALYSIS</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td></td>
</tr>
<tr>
<td>3.2 Evaluation of the Cross Correlation Function</td>
<td></td>
</tr>
<tr>
<td>3.3 Mean Square Error</td>
<td></td>
</tr>
<tr>
<td>3.4 Signal/Error Ratio</td>
<td></td>
</tr>
<tr>
<td>3.5 Discussion of theoretical results and Optimisation</td>
<td></td>
</tr>
<tr>
<td>3.6 Optimisation with respect to the delay</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 4: COMPUTER SIMULATION</td>
<td></td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td></td>
</tr>
<tr>
<td>4.2 Pseudo Random Noise Generators</td>
<td></td>
</tr>
<tr>
<td>4.3 Direct Simulation on a Digital Computer</td>
<td></td>
</tr>
<tr>
<td>4.4 Computation of the Mean Square Error and the S/E ratio</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 5: EXPERIMENTAL WORK AND MEASUREMENT...</td>
<td>Page No.</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>5.1 The Mean Square Error Measurement</td>
<td>42</td>
</tr>
<tr>
<td>System</td>
<td></td>
</tr>
<tr>
<td>5.1.1 The Function Generator f(t)</td>
<td>43</td>
</tr>
<tr>
<td>5.1.2 The Code Sequence Pulse Generator</td>
<td>43</td>
</tr>
<tr>
<td>5.1.3 The Sample and Hold Circuit</td>
<td>46</td>
</tr>
<tr>
<td>5.1.4 Sample and Hold Circuits</td>
<td></td>
</tr>
<tr>
<td>in Cascade</td>
<td>49</td>
</tr>
<tr>
<td>5.1.5 Sample and Hold Analog Delay Line</td>
<td>49</td>
</tr>
<tr>
<td>5.2 System Calibration and Experimental Procedure</td>
<td>51</td>
</tr>
<tr>
<td>CHAPTER 6: REVIEW OF RESULTS</td>
<td>53</td>
</tr>
<tr>
<td>6.1 Statistical Properties of the Computer</td>
<td></td>
</tr>
<tr>
<td>simulated Function</td>
<td>53</td>
</tr>
<tr>
<td>6.2 Comparison of Results obtained by the</td>
<td></td>
</tr>
<tr>
<td>three methods</td>
<td>53</td>
</tr>
<tr>
<td>6.3 The Cross Correlation Function</td>
<td>54</td>
</tr>
<tr>
<td>6.4 Power Spectrum of the Error Function</td>
<td>55</td>
</tr>
<tr>
<td>6.5 Approximate relationships between the</td>
<td></td>
</tr>
<tr>
<td>maximum value of the S/E Ratio, the optimum delay and the given distribution.</td>
<td>61</td>
</tr>
<tr>
<td>6.6 Summary of Results</td>
<td>63</td>
</tr>
<tr>
<td>CHAPTER 7: A SIMPLE METHOD OF CHANNEL CODING WITH AN EXACT RECONSTRUCTION SCHEME.</td>
<td>64</td>
</tr>
<tr>
<td>CHAPTER 8: CONCLUSIONS</td>
<td>66</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>71</td>
</tr>
<tr>
<td>APPENDIX B: THE MEAN AND VARIANCE OF g(t)</td>
<td>74</td>
</tr>
<tr>
<td>PART II</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>77</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>81</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.2</td>
<td>Expected number of Zeros per unit time</td>
</tr>
<tr>
<td>2.3</td>
<td>The Statistical Distribution of zeros</td>
</tr>
<tr>
<td>2.4</td>
<td>The In and Exclusion Method</td>
</tr>
<tr>
<td>2.5</td>
<td>Series Expansions for ( P_n(\tau) )</td>
</tr>
<tr>
<td>2.6</td>
<td>The Asymptotic behaviour of ( P_n(\tau) )</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Asymptotic behaviour at small interval lengths</td>
</tr>
<tr>
<td>2.6.1.1</td>
<td>The Singular Case</td>
</tr>
<tr>
<td>2.7</td>
<td>Behaviour of ( P_0(\tau) ) at large</td>
</tr>
<tr>
<td>2.8</td>
<td>Approximations over the range of large intervals</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Behaviour of ( P_0(\tau) ) at large ( \tau )</td>
</tr>
<tr>
<td>2.8.2</td>
<td>The Graphical Extrapolation Method</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Analytical method</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Comparison between the Graphical and Analytical Method</td>
</tr>
<tr>
<td>2.9</td>
<td>Approximate methods utilising integral equations</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Preliminary Results</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Relationships between ( P_n(\tau) ) and ( P(n,\tau) )</td>
</tr>
<tr>
<td>2.9.3</td>
<td>Derivation of integral equations based on statistically independent intervals</td>
</tr>
<tr>
<td>2.9.4</td>
<td>Markov Chain of Intervals</td>
</tr>
<tr>
<td>2.9.5</td>
<td>Zero Crossings as a Renewal Process</td>
</tr>
<tr>
<td>2.9.6</td>
<td>The advantages of the integral equation approach</td>
</tr>
<tr>
<td>2.10</td>
<td>Relationship between ( P_0(\tau) ) and the probability integral of Multivariate Distribution</td>
</tr>
</tbody>
</table>
CHAPTER 3: THE MULTIVARIATE NORMAL INTEGRAL...

3.1 Introduction ...

3.2 Methods of evaluating the Multivariate Normal Integral ...

3.3 The Geometric Interpretation of the Multivariate Normal Integral ...

3.4 Series Expansions for the Multivariate Integral ...

3.5 McFadden's Urn Model Approximation ...

3.5.1 Polya's Urn Scheme ...

3.5.2 The Generalised Urn Scheme ...

3.5.3 Comparison with the multivariate Normal Integral ...

3.6 Approximations for $P_0(\tau)$ based on the Schlaffli Function $S_n$ ...

3.6.1 The Approximation $P_{on}(\tau)$ ...

3.6.2 The Approximation $P^{*}_{on}(\tau)$ ...

3.7 Final note on the Multivariate Normal Integral ...

CHAPTER 4: THE MARKOVIAN MODEL ...

4.1 Introduction ...

4.2 Introduction to the theory of Markov Processes ...

4.2.1 Linear systems ...

4.2.2 Components constituting the vector process $\tilde{x}(t)$ ...

4.3 Derivation of the generalised integral equation ...
4.4 The relationship between the Zero Crossing distribution $P_0(t)$ and the solution of the Generalised Integral equation...

4.5 Approximations based on the integral equation...

4.6 Approximations based on the second order Markovian model...

4.7 Derivation of expressions for $Q(\dot{y}, t/t_0)$ and $P(\dot{y}, t/\dot{y}_0, t_0)$ for a Gaussian Process...

4.8 Asymptotic behaviour of $Q(t, \dot{y}/t_0)$ and $P(t, \dot{y}/t_0, \dot{y}_0)$...

4.9 The solution of the second order Integral equation...

4.10 An expression for $P_n(t)$ based on the Markovian model...

4.11 Summary...

CHAPTER 5: EXPERIMENTAL MEASUREMENT OF THE DISTRIBUTION $P_0(t)$...

5.1 Introduction...

5.2 Basic system of axis crossings Histogram Measurement...

5.3 System Implementation...

5.4 Alternative method of probability Histogram Measurement...

5.4.1 Departmental Computer...

5.4.2 Outline of measurement system...

5.4.3 The White Noise Generator...

5.4.4 Digital filtering...

5.4.5 Zero detection and Histogram Measurement...
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>Accuracy of computer method</td>
<td>191</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Histogram class overlap</td>
<td>193</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Effect of class overlap</td>
<td>196</td>
</tr>
<tr>
<td>CHAPTER 6: REVIEW AND DISCUSSION OF RESULTS</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>200</td>
</tr>
<tr>
<td>6.2</td>
<td>Auto Correlation Function</td>
<td>203</td>
</tr>
<tr>
<td>6.3</td>
<td>Review of experimental results</td>
<td>206</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Experimental results</td>
<td>207</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Theoretical results based on the second order Markovian equation</td>
<td>209</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Theoretical results based on the Schlaffli Function</td>
<td>212</td>
</tr>
<tr>
<td>6.4</td>
<td>General discussion of results</td>
<td>214</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Effects of spectrum variations on $P_0(\tau)$</td>
<td>215</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Relationship between the power spectrum and the correlation of successive intervals</td>
<td>220</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Effects of spectrum variations on the MK2 approximation</td>
<td>228</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Effects of spectrum variations on the multivariate Approximation</td>
<td>232</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Behaviour of $P_0(\tau)$ at large interval length</td>
<td>234</td>
</tr>
<tr>
<td>6.4.6</td>
<td>Comparisons of various approximations</td>
<td>236</td>
</tr>
<tr>
<td>CHAPTER 7: CONCLUSIONS</td>
<td></td>
<td>239</td>
</tr>
<tr>
<td>APPENDIX 1: DERIVATION OF THE IN AND EXCLUSION EQUATION</td>
<td></td>
<td>243</td>
</tr>
<tr>
<td>APPENDIX 2: DERIVATION OF THE PROBABILITY $V(\tau)$ FOR FIRST ORDER MARKOV PROCESS</td>
<td></td>
<td>246</td>
</tr>
<tr>
<td>APPENDIX 3: NUMERICAL INTEGRATION AND THE SOLUTION OF THE SECOND ORDER MARKOVIAN INTEGRAL EQUATION</td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>Section</td>
<td>Page No.</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>A.3.1 Numerical Integration</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>A.3.2 The Gauss Laguerre Quadrature</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>A.3.3 Numerical computation of the solution of the integral equation</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>APPENDIX 4: A BRIEF NOTE ON DIGITAL FILTERING</td>
<td>259</td>
<td></td>
</tr>
<tr>
<td>REFERENCES: PART I</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>REFERENCES: PART II</td>
<td>264</td>
<td></td>
</tr>
</tbody>
</table>
In order to increase the security of the normal Time Division Multiplex system, it is proposed that the channels should be multiplexed in accordance with a predetermined code sequence which can easily be altered and which is known only to the sender and the receiver. One possible method of achieving this form of coding is to employ non-uniform sampling of the signal sources prior to the transmission. In Part I of the thesis we discuss the theoretical aspects of non-uniform sampling and assess its practical implications in the light of the above application.

The points at which a random function intersects the zero level finds many applications in applied science and Engineering. This constitutes what is known as the 'Zero Crossing Problem' which remains today one of the most intractable problems in the theory of noise, for which no satisfactory solution has yet been found. In Part II a study of this problem is presented. The study surveys some of the past research carried out on the subject and discusses some new ideas and techniques.
List of Principal Symbols

Part I

\( f(t) \)  
Band Limited Signal in \((0-W)\).

\( g(t) \)  
Output of Approximate Recovery Scheme.

\( E_t(t) \)  
Error Function \( f(t+t) - g(t) \).

\( W \)  
Highest Frequency present in the spectrum of \( f(t) \).

\( \omega_0 \)  
Highest Radian Frequency present in the spectrum of \( f(t) \).

\( M \)  
Number of frames in one sequence period.

\( N \)  
Number of channels in each frame.

\( T \)  
Sampling (Nyquist) Interval.

\( \sigma^2 \)  
Variance of \( f(t) \).

\( \psi_{11}(t) \)  
Auto-Correlation Function of \( f(t) \).

\( \psi_{12}(t) \)  
Cross-Correlation Function of \( f(t) \) and \( g(t) \).

\( \Delta_k \)  
Deviation of the \( k \)th sample point from its regular position \( kT \).

\( S/E \)  
Signal to Error Power Ratio.

\( \tau_0 \)  
Optimum value of Delay at which Signal to Error Ratio is Maximum.

Part II

\( f(t) \)  
Random Stationary Function (Process) of Mean Zero and Unity Variance.

\( P(0,\Delta t) \)  
Probability Density of a Zero Crossing occurring in the infinitesimal interval \( \Delta t \).

\( P_-(0,\Delta t) \)  
Probability Density of a Downward Zero Crossing occurring in the infinitesimal interval \( \Delta t \).

\( P_+(0,\Delta t) \)  
Probability Density of an Upward Zero Crossing occurring in the infinitesimal interval \( \Delta t \).

\( P_0(\tau) \)  
Probability density of an interval between two successive zero crossings having a length in the range \( \tau, \tau + d\tau \).
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_n(\tau)$</td>
<td>Probability Density of the interval between a given zero crossing and the $(n+1)^{th}$ succeeding one having a length in the range $\tau, \tau+d\tau$.</td>
</tr>
<tr>
<td>$P(n,\tau)$</td>
<td>Probability that an interval of length $\tau$ contains exactly $n$ zero crossings.</td>
</tr>
<tr>
<td>$Q(\tau)$</td>
<td>Probability Density of a downward crossing in $t_0+\tau, t_0+\tau+d\tau$ occurring, given an upward crossing at the origin $t_0$.</td>
</tr>
<tr>
<td>$U(\tau)$</td>
<td>Same as $Q(\tau)$, except that the crossing in $t_0+\tau, t_0+\tau+d\tau$ can be any crossing.</td>
</tr>
<tr>
<td>$P_{ODD}(\tau)$</td>
<td>Probability Density of an upward crossing occurring at $t_0+\tau, t_0+\tau+d\tau$, given an upward crossing at $t_0$.</td>
</tr>
<tr>
<td>$P_{-\text{1-}}(\tau)$</td>
<td>Same as $P_{ODD}(\tau)$.</td>
</tr>
<tr>
<td>$V_n(t_1 \ldots t_n)$ also $V_n$</td>
<td>$n^{th}$ dimensional multivariate integral.</td>
</tr>
<tr>
<td>$V(\tau)$</td>
<td>Probability that the random function $f(t)$ is positive throughout the interval $\tau$.</td>
</tr>
<tr>
<td>$P(\dot{y}, t/\dot{y}_0, t_0)$</td>
<td>Conditional probability that $f(t)$ goes through zero at $t, t+dt$ with a slope $\dot{y}, \dot{y}+d\dot{y}$, given that it crossed the zero level at $t_0$ with a slope $\dot{y}_0$.</td>
</tr>
<tr>
<td>$Q(y, t/t_0)$</td>
<td>Conditional probability density of $f(t)$ going through zero at $t, t+dt$ with a slope $\dot{y}, \dot{y}+d\dot{y}$ given that it had an upward crossing at $t_0$.</td>
</tr>
<tr>
<td>$N_0$</td>
<td>Expected number of zero crossings per unit time.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Average length of interval between successive crossings.</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Variance of the interval length.</td>
</tr>
<tr>
<td>$\psi(\tau)$</td>
<td>Auto-Correlation Function of $f(t)$.</td>
</tr>
<tr>
<td>$\rho(\tau)$</td>
<td>Normalised auto-correlation function of $f(t) = \frac{\psi(\tau)}{\psi(0)}$.</td>
</tr>
<tr>
<td>$r(\tau)$</td>
<td>Normalised auto-correlation function of the infinitely clipped version of $f(t)$.</td>
</tr>
</tbody>
</table>
\( r'(0^+) \) The value of the first derivative of \( r(\ ) \) at \( =0^+ \).

\[ \rho(t) \] Correlation matrix of \( f(t) \).

\[ M \] Inverse of Correlation matrix.

\( \Gamma(n) \) Gamma Function.

\( X(t) \) Markov Vector Process.

\( x(t), \dot{x}(t), \ldots, x^{(\rho)}(t) \) \( \rho \) components of \( X(t) \).

\( h_n(x) \) The \( n \)th degree Hermite Polynomial in \( x \).

\( L_n(x) \) The \( n \)th degree Laguerre Polynomial in \( x \).

\( T \) Sampling Interval.

\( X(nT) \) Present value of Input Sample.

\( Y(nT) \) Present value of Output Sample.

\( \omega_0 \) Reference Frequency.

**Note.**

In addition to the above symbols, other symbols, particularly those representing probabilities, are fully defined in the text. While the letter \( P \) is common to various probability quantities, these probabilities are distinguished by the symbols in the arguments of \( P \).
GENERAL INTRODUCTION.

This thesis consists of two parts. Part I deals with the sampling theorem, in particular non-uniform sampling of band limited signals. Part II deals with the statistical distribution of the zero crossings of random functions.

One of the fundamental requirements of modern communication theory is the matching of the information source to the communication channel in order to achieve optimum efficiency in transmission. Since sources and channels can either be continuous or discrete, the need for conversion from one form to the other often arises in practical situations. The conversion is normally accomplished with the aid of the sampling theorem, which states that a function \( f(t) \) band-limited in \((0, W)\) is uniquely determined by a set of uniform samples taken at the Nyquist rate \( \left( \frac{1}{2W} \right) \). However, the samples need not be amplitudes or indeed have to be taken uniformly. For the sampling theorem has been generalised to include non-uniform sampling; sampling of the signal and its derivatives; and 'zeros' sampling. The reverse process, i.e. the reconstruction of the original signal from the samples, is more complex when the samples are taken non-uniformly or include parameters, other than amplitudes.

Part I of the thesis deals mainly with a particular process of non-uniform sampling, used in conjunction with a Time Division Multiplex, pulse modulation system, in order to introduce an element of security in the channel. The recovery scheme required at the receiver to reconstitute the signal is discussed in detail. The complexity of the recovery scheme involved led to the de-
development of a far simpler scheme, but at the expense of some signal distortion. This was estimated in terms of the mean square error and the crosscorrelation between the original signal and its reconstituted version. Results for these quantities obtained from theoretical analysis, computer simulation and experimental measurements were found to be in close agreement.

The sampling of zeros of band limited signals is another form of non-uniform sampling where the characteristics of \( f(t) \) determines the location of the sampling points \( t_n \), instead of these being normally independent of \( f(t) \). The discovery of the small effect infinite clipping has on the intelligibility of speech, suggested that a great deal of information is contained in the zero crossings of the speech waveform. Using the concept of analytic signals which are band limited in \( (0,\omega) \), one can show that zeros (both real and complex) occur at the rate of \( 2\omega \) per second and are sufficient to specify the signal uniquely. Real zeros are those represented by the crossing points of the zero-axis and are preserved in the operation of infinite clipping. Complex zeros, on the other hand, are a mathematical concept which arises in the theory of analytic signals. Infinite clipping destroys the information contained in these zeros. A sinusoid of repeated differentiations and the addition of a suitable amplitude and frequency converts complex zeros to real ones and thus serve to increase the information contained in the clipped waveform. To recover the original signal an operation known as the real zero interpolation is required.

Zero crossings or real zeros occur on average at a rate less than the Nyquist rate. For random signals, the
occurrence of zeros and the intervals between them are both random. The statistical distribution of zeros is part of the more general problem which also includes the statistics of the points at which random processes cross a particular threshold level. This problem arises in many fields, such as the statistical communication theory, oceanography and control engineering. For example in the recovery of the original signal from its real zero samples, knowledge of the distribution of intervals between zeros is useful in order to estimate the size of peak amplitudes that are likely to result after the (RZI) operation. This is necessary in order to prevent overloading.

In the field of statistical communication theory, for example when radar or sonar detection is concerned, the probability of errors, represented by 'false alarm' or 'missed target', is directly related to the distribution of threshold crossings. Therefore the knowledge of these distributions can greatly aid the devising of methods aimed at minimising these errors.

Part II of the thesis presents a study of the problem of the statistical distribution of zero crossings. It surveys some of the past research carried out on the subject and discusses some new ideas and techniques.
PART I
CHAPTER 1

Introduction.

The well known sampling theorem enables a band limited signal to be expressed in terms of a set of sample values taken regularly at the Nyquist rate. The theorem came into prominence as a tool of interpolation after Whittaker\(^{(1)}\), and later was rediscovered, as an important aspect of communication theory, by Nyquist, Gabor and Shannon\(^{(2)}\). Since then, the sampling theorem has played an important role in meeting the fundamental requirement of matching the information source to the communication channel in order to achieve optimum efficiency of transmission. Its value becomes apparent in practical situations involving information sources and channels, which can either be continuous or discrete, thus necessitating conversion from one form to the other, in order to satisfy the matching requirement.

A very important application of the sampling theorem has been the development of pulse modulation systems incorporating time division multiplex. These systems involve the transmission of discrete information or data from a number of independent sources, on a shared time basis, over the transmission medium; and so offer the principal advantage of time conservation, over the CW systems used hitherto.

The normal pulse modulation system allocates a proportion of the time between samples to each channel in a regular manner. In addition, the multiplexed signal
so formed contains synchronous information in the form of periodic pulses defining the beginning and end of each frame of samples. At the receiver, the decoding process is straightforward and simply requires periodic pulses to be generated in unison with the incoming synchronous pulses. The resulting de-multiplexed data can then be processed in a manner complementary to that undergone at the transmitter, in order to display the original information.

Because of the simplicity of the decoding process, anyone who wishes to gain access to the information transmitted in a regularly multiplexed fashion can do so easily, and the system can hardly be described as secure. In some applications where security of the channel is the chief requirement, the normal pulse modulation system will have to be modified and coded in a manner known only to the sender and the receiver. One possible method would be to alter the sequence of channel multiplexing in each frame in a pre-determined fashion. To simplify matters, the whole sequence can be made to repeat, say every $M$ frames. In an $N$-channel system therefore, there exists $N^M$ possible different sequences of channel multiplexing, out of which only $M$ sequences are used at one time. Providing $N$ and $M$ are reasonably large, decoding without the prior knowledge of the particular $M$ sequences used would be extremely difficult. Decoding by trial and error would be tedious and time consuming in view of the large number of possible different sequences.

The above method of channel multiplexing gives rise to a much more secure system than the normal pulse mod-
ulation system. To introduce this form of channel coding, signal sources can be sampled regularly and then multiplexed in a manner determined by the M sequences known and agreed upon by the sender and the receiver. Some form of buffer storage would be required to store, and thus delay, the incoming uniform samples in order to regulate the multiplexing process. Alternatively, a direct method which requires no buffer storage would be to perform the sampling process on the signal sources in accordance with the M multiplexing sequences. The sampling process is thus non-uniform, repeating every M samples and this poses certain problems which are discussed below.

1.1 Problems associated with Non-uniform Sampling.

If a band limited signal is sampled uniformly at a rate higher than or equal to, the Nyquist rate, then it can be reconstituted from its samples through an ideal low pass filter. Any physical low pass filter would produce a near exact replica of the signal. If the sampling process is non-uniform performed at an average rate higher than, or equal to, the Nyquist rate, then in general the signal can be reconstituted exactly from its sampled values. The recovery scheme however becomes far more complex than in the uniform case and its degree of complexity cannot be estimated in a general manner. Black\(^3\) discussed non-uniform sampling with the aid of a few examples, and stated that their application can at most cause some loss in accuracy and the simplicity of the reconstruction procedure. Yen\(^4\) examined some special non-uniform sampling
processes in detail and deduced some interesting properties of band limited signals. His results were contained in four generalised sampling theorems giving the reconstruction formulae required to reproduce the signal from its sampled values. Yen also stated that there is no way of judging complexity of reconstruction in a general manner. Therefore when one is confronted with a non-uniform sampling process it is necessary to judge its complexity according to the particular requirements and this can be done by detailed investigation of the reconstruction formulae.

One of the interesting sampling processes studied by Yen is the so-called "Recurrent non-uniform sampling distribution", in which the sample points are divided into groups of say M points and the groups have a recurrent period of MT seconds, T being the Nyquist interval. From our application point of view, this is the most relevant sampling process discussed in Yen's article in that it can be applied directly to the signal sources in order to achieve the form of channel coding required for security. The reconstruction scheme necessary to reproduce the signal from its sampled values was shown by Yen to be unique and complex, involving time-variant filtering processes.

This means that any other alternative scheme which might be used for interpolation would yield a different signal from the original one. The output of such a system however, can be made to approximate and indeed preserve some of the basic features of the original so that it would be capable of representing it for most practical purposes.
1.2 **Purpose of Investigation.**

The purpose of this study therefore is to investigate the three possible alternative methods discussed above, namely:—

1. Uniform sampling of signal sources plus buffer storage.
2. Non-uniform sampling involving an exact reconstruction which is complex and time variant.
3. Non-uniform sampling involving a simpler reconstruction scheme, capable only of yielding an approximate version of the original signal.

The form of buffer storage required in the first method is shown to be quite simple as far as this application is concerned, (see Chapter 7). Nevertheless, it is thought useful to consider the other two methods, and to study and investigate the process of non-uniform sampling applied to band limited signals, in order to ascertain whether it has any practical significance regarding this and other applications.

The major part of this study is therefore devoted to non-uniform sampling and its effects on band limited signals. Starting from Yen's exact solution, pointing out its disadvantages and shortcomings, a much simpler, but approximate recovery scheme is developed (Chapter 2). The performance of this scheme is evaluated in terms of the mean square error and the cross correlation function between its output and the original signal (Chapter 3). In Chapter (4), a method of simulating the approximate scheme on the computer, is discussed and in Chapter (5), some experimental work is presented. Review and comparison of various results is
included in Chapter (6). Finally, the study is concluded in Chapter (8).
Theoretical Consideration.

In this chapter, we look at the process of recurrent non-uniform sampling applied to band limited signals, and discuss the basic problems associated with it; namely the degree of accuracy of the sampling process and the complexity of the reconstruction scheme required to reproduce a near exact replica of the signal. We then consider the sampling process in the frequency domain by viewing it as an amplitude modulation process, in the light of which we conclude by introducing an approximate but much simpler, recovery scheme than the exact one.

2.1 Recurrent Non-Uniform Sampling.

As pointed out in the last chapter, the sample points in this distribution are divided into groups of \( M \) points each and the groups have a recurrent period of \( MT \) seconds. Fig. 1 shows an example of such a distribution. In one period the points are denoted by \( t_p, p = 1, 2, \ldots, M \). The complete set of sample points are written

\[
\tau_{pm} = t_p + \frac{mM}{2W},
\]

where \( m \) is an integer, and \( W \) is the highest frequency present in the signal. For this distribution Yen showed that the following theorem is valid.

2.1.1 Theorem.

A band limited signal \( f(t) \) limited in frequency to \( WHZ \) is uniquely determined by its values at a set of recurrent sample points \( t = \tau_{pm} = t_p + \frac{mM}{2W} \), the recon-
Fig. (1) - A Recurrent Non-Uniform Sample Distribution.
struction is:

\[ f(t) = \sum_{m=\infty}^{\infty} \sum_{p=1}^{M} f(\tau_{pm}) \psi_{pm}(t) \quad \ldots (2.1.1) \]

where

\[ \psi_{pm}(t) = \frac{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t-t_q) (-1)^{mM}}{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t_p-t_q) \frac{2\pi W}{M} (t-t_p-mM)} \quad \ldots (2.1.2) \]

\( \psi_{pm}(t) \) is referred to as the composing function which actually defines the kind of impulse response the reconstruction scheme is expected to exhibit at time \( t \) in the event of receiving the \( pm \)-th sample of the distribution.

In the proof of the theorem, Yen established both the possibility as well as the uniqueness of the recovery of \( f(t) \) from \( f(\tau_{pm}) \), but did not in fact, show how (2-1-2) is derived. This however can be done by inspection, as shown below.

If \( \psi_{pm}(t) \) is going to be the desired composing function then it will have to satisfy the following requirements:

a) For integer values of \( r \) and \( s \),

\[ \psi_{pm}(t = \tau_{rs}) = \delta_{pr} \delta_{ms} \quad \ldots \quad (2-1-3) \]

where \( \delta_{mn} \) is the Kronecher delta.

In other words, (2-1-3) implies that

\[ \psi_{pm}(\tau_{rs}) = 1 \text{ if } r=p \text{ and } s=m \quad \ldots \quad (2-1-4) \]
\[ = 0 \text{ if } r \neq p \text{ and } s \neq m \quad \ldots \quad (2-1-4) \]

b) The highest frequency contained in \( \psi_{pm}(t) \) must be \( W_HZ \).
Consider the form:

\[ \Upsilon_{pm}(t) = \frac{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t - MmT - t_q)}{K \frac{2\pi W}{M} (t - MmT - t_p)} \]  \hspace{1cm} (2.1.5)

where \( K \) is a constant. Clearly, the Fourier spectrum of this expression is governed by the \( M \) sine function terms of the numerator. Since each term is of frequency \( \frac{W}{M} \), their product is of bandwidth \( W \). Hence the expression satisfies requirement (b). Further, the right hand side of the equation vanishes for all \( t = MmT + t_q (q = 1, 2, \ldots, M) \) (except when \( t = MmT + t_p \)). To ensure that the expression takes on the value unity at \( t = MmT + t_p \) it is only necessary that:

\[ K = \frac{M}{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t_p - t_q)} \]  \hspace{1cm} (2.1.6)

Hence the desired composing function is as follows:

\[ \Upsilon_{pm}(t) = \frac{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t - MmT - t_q)}{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t_p - t_q)} \]  \hspace{1cm} (2.1.7)

which is the same as (2.1.2)

(2.1.7) can also be written as:

\[ \Upsilon_{pm}(t) = \frac{\sum_{q=1, \neq p}^{M} \sin \frac{2\pi W}{M} (t - MmT - t_q) \sin \frac{2\pi W}{M} (t - MmT - t_p)}{\sum_{q=1}^{M} \sin \frac{2\pi W}{M} (t_p - t_q)} \]  \hspace{1cm} (2.1.8)

thus if \( r \) is an integer, then by inspection.
This implies that for constant $p$, \( \psi_{pm}(t) \) possesses the same form irrespective of what value \( m \) takes. On the other hand, owing to the presence of sine functions containing the factors \((tp-tq)\) in the denominator of (2-1-8), the composing function \( \psi_{pm}(t) \) has different forms for different \( p \). It therefore follows that given a particular recurrent sample sequence, the reconstruction scheme must be capable of producing \( M \) different time responses corresponding to the \( M \) different values of \( tp \). In a dynamic situation similar to the one our application demands where the scheme is expected to deal with a large number of different sequences namely \( N^M \), the scheme must include means of analysing the incoming sequence to determine the actual values of \( tp \)'s characterising it in order that the appropriate responses are produced. Hence the scheme must be complex and time variant.

Another disadvantage of this scheme is the accuracy requirement. The degree of accuracy required in the sampling process for a reasonably accurate recovery, varies with respect to the different sequences and usually should be higher than that required in the uniform case. This is because the maximum value of the composing function which is unity in the uniform case, is generally greater than unity, occurring somewhere between the sample points. In cases where bunching of sample points occurs, thus leaving relatively large gaps in the distribution it is possible for the composing function to attain much larger values than unity, due to the presence of the
factors $\sin \frac{2\pi W}{M} (t_p - t_q)$ in the denominator of (2-1-8). This implies that if the signal is to be reproduced within reasonable limits of accuracy, then the degree of accuracy in the sampling process must be higher by a factor equal to the maximum value of the composing function.

From the above discussion, we can conclude that recurrent non-uniform sampling as a means of introducing the desired form of channel coding, is highly unsuitable owing to the complexity of the recovery scheme and the somewhat high degree of accuracy required in the sampling process. This conclusion should not however rule out the consideration of other reconstruction schemes which because of the uniqueness of the solution represented by (2-1-8) cannot reproduce the signal exactly, but rather an approximation of it. In looking for such a scheme we aim to satisfy the following requirements:

1) The scheme should be simple and sufficiently flexible to deal with any incoming sequence.

2) The output reproduced by such a scheme should be highly correlated to the original signal so that it can represent it for most practical purposes.

Before describing such a scheme we shall consider the frequency spectrum of the sampled signal and the effects of the recurrent non-uniform sampling process on this spectrum, as this would be very helpful in developing and understanding the proposed recovery scheme.

2.2 Frequency Spectrum of Sampled Signals.

The sampling process can be viewed as an amplitude
modulation system in which the amplitudes of the sampling pulses, acting as a carrier wave, are modulated by the signal \( f(t) \). If the sampling process is periodic, occurring at or above the Nyquist rate, the frequency spectrum of the sampled signal can be shown to consist of a DC component and line spectra corresponding to the sampling frequency and its harmonics. Each line spectra has two bands on either side corresponding to the original signal frequency spectrum as shown in Fig. 2. To reconstitute the signal, it is sufficient to recover the DC component sideband, which simply involves the use of a low pass filter of the appropriate bandwidth.

If however, the sampling process is of the recurrent non-uniform type discussed above, the frequency spectrum can be shown to consist of a DC component plus line spectra corresponding to the recurrence frequency \( 1/MT \) and its harmonics. The spacing between the line spectra is thus \( M \) times smaller than that shown in Fig. 2. A situation therefore arises whereby \( M \) sidebands overlap totally and so it would no longer be possible for a simple low pass filter to recover the signal spectrum. It is possible to show that, by carefully choosing the times \( t_p \), some harmonics can be eliminated and consequently the number of overlapping sidebands is reduced. However, the number of sampling sequences possessing such qualities are only a fraction of the total \( N^M \) sequences and this fraction decreases rapidly as the number of eliminated harmonics approaches \( M \). When this happens, there will be only one sequence and this corresponds to the uniform
Fig(2) – Frequency spectrum of Uniformly Sampled Signals.
For an arbitrary sampling sequence, the recovery scheme required must therefore be capable of extracting the signal spectrum from the overlapped sidebands. Yen's solution (2-1-8), provides the Impulse response of such a scheme whose uniqueness makes it the only one that can reconstruct the signal exactly from its samples. Any other form of filtering is only capable of producing an approximate version of the signal. This is the subject of the following section.

2.3 The Recovery Scheme.

The sidebands overlap is a factor of irregularity in the sampling process and can be removed at the expense of introducing some distortion, by removing the irregularity in the incoming sequence at the receiver. The sample points of the incoming sequence are shifted along the axis in order to make the sequence uniform, similar to that which would have been obtained had the sampling process been periodic. The amplitudes of the corresponding samples constituting these two uniform sequences are different but nevertheless correlated owing to the band limitation property of the signal. The degree of sample correlation being naturally dependent on the time shifts involved. Interpolation of the resulting uniform sequence by a simple low pass filter would then yield a correlated version of the original signal.

Fig. 3 shows such a recovery scheme in block diagram in which the incoming non-uniform sample sequence is first stored and then released as a uniform sequence, having the same amplitude information. This process can easily be
Non-Uniform Pulse sequence synchronised to incoming data

Regular Pulse sequence

Fig. (3) - The Approximate Recovery Scheme.
carried out by using a sample and hold circuit to store one sample at a time. The sample process is controlled by a non-uniform pulse waveform synchronised to the incoming sequence. The output of this circuit, resembling a staircase waveform, is then sampled uniformly at the end of each frame period and the uniform sample sequence is fed into a low pass filter whose output provides an approximate version of the original signal.

Because this system inherently ignores the time information contained in the relative time displacements of the sample points in the original sequence, its output is normally different from the original signal, but somewhat related to it. Offsetting this disadvantage are the simplicity and flexibility of the system. The first requirement outlined above thus appears to be satisfied.

As far as the second requirement is concerned, an estimate of the error and the degree of correlation between the output of the recovery scheme, say \( g(t) \), and the original signal \( f(t) \), is necessary to establish the extent to which the system is capable of satisfying this requirement. This is discussed in the following chapters. It is useful to mention here that the above system was used to reconstitute speech signals from their non-uniform sample values. The speech outputs were of good quality and high intelligibility and were almost identical to the original extracts. This should not be surprising in view of the large amount of redundancy and high degree of self-correlation that
characterises the speech signals.

In theoretical analysis as well as in experimental work, speech signals are not a convenient class of test signals because of their non-stationary properties. Instead, low pass stationary random noise was used in the analysis and assessment of the overall performance of the approximate recovery scheme.
CHAPTER 3.

Theoretical Analysis.

3.1 Introduction.

In this chapter, some analytical treatment is given in order to estimate in quantitative terms, the extent to which \( g(t) \), the output of the recovery scheme described in the previous chapter, resembles or more precisely approximates to the original signal \( f(t) \). For theoretical purposes therefore \( f(t) \) is assumed to satisfy the following conditions:

1) \( f(t) \) is a stochastic (random) stationary process of mean value zero and variance \( \sigma^2 \).

2) It has a low pass power spectrum which is defined by the following relationships:

\[
\phi(\omega) = \text{Constant} \begin{cases} |\omega| < \omega_0 & = 0 \\ |\omega| > \omega_0 & \cdots \cdots (3.1.1) \end{cases}
\]

This last condition implies that the autocorrelation function of \( f(t) \) is given by

\[
\gamma_{11}(\tau) = \sigma^2 \sin \frac{\omega_0 \tau}{\omega_0} \cdots \cdots (3.1.2)
\]

Since \( f(t) \) is assumed to be a stochastic process, therefore by definition the sample values \( f(t_k), k = -1, 0, 1 \ldots \) are random variables with the same mean and variance as those of the process \( f(t) \). Moreover, because \( f(t) \) is band limited, \( f(t_k) \) are generally correlated random variables. It therefore follows that \( g(t) \), the output of a linear filtering process on the random variables \( f(t_k) \) should also be a stochastic stationary process, the mean and
variance of which can be shown to be the same as those of \( f(t) \), see Appendix (B), where a proof is given.

Quantitatively speaking, what is needed in comparing two random stationary processes, in addition to their mean and variance, is the evaluation of their cross-correlation function which is defined as

\[
\gamma_{12}(\tau) = \langle f(t+\tau)g(t) \rangle \quad \ldots\ldots\ldots (3.1.3)
\]

where angular brackets are used to denote ensemble averages.

This quantity is very useful since it provides us with some measure of resemblance between the two random processes. Another useful quantity which is related to the cross-correlation function is the mean square value of the error \( E_\tau(t) \), defined as

\[
\langle E_\tau(t) \rangle = f(t+\tau) - g(t) \quad \ldots\ldots\ldots (3.1.4)
\]

This value is related to the cross-correlation function by the following relationship:

\[
E_\tau^2(t) = \langle f^2(t+\tau) \rangle - 2\langle f(t+\tau)g(t) \rangle + \langle g^2(t) \rangle \quad \ldots\ldots\ldots (3.1.5)
\]

Therefore

\[
\langle E_\tau^2(t) \rangle = 2(\sigma^2 - \gamma_{12}(\tau)) \quad \ldots\ldots\ldots (3.1.6)
\]

Since \( \langle f^2(t+\tau) \rangle = \langle f^2(t) \rangle = \sigma^2, \langle g^2(t) \rangle = \sigma^2 \). Equation (3.1.6) thus shows a linear relationship between the mean square error and the cross correlation function.

### 3.2 Evaluation of the Cross-Correlation Function

Fig. 4 shows a recurrent non-uniform sample distribution in which the sample points are denoted by \( t_k = kT + \Delta_k \), where \( k \) is an integer, and where \( \Delta_k \) can be defined as the time deviation of the \( k^{th} \) sample point from the end of
Fig. (4) - An Example of a Recurrent Sample Distribution.
the \( k \)th Nyquist interval measured relative to an arbitrary time origin. Since the sampling distribution has a group periodicity of \( MT \); then if \( l \) is an integer (positive or negative) we have

\[
\Delta_k = \Delta_{k+1} \cdots \cdots \cdots \ (3.2.7)
\]

(3.2.7) will be referred to as the 'group periodicity equation'. In the proposed recovery scheme the sample points \( t_k \) are shifted to new positions \((k+1)T\) such that the \( k \)th sample \( f(t_k) \) would now be appearing at \( t=(k+1)T \). Theoretically speaking however, it makes no difference if we assume that the sample points \( t_k \) are shifted to positions \((t=kT)\) instead. Bearing this in mind, the output \( g(t) \) can be expressed in terms of the sample values \( f(t_k) \) as

\[
g(t) = \sum_{k=-\infty}^{\infty} f(t_k) \frac{\sin(\omega_0(t-kT))}{\omega_0(t-kT)} \cdots \cdots \ (3.2.8)
\]

If \( f(t) \) had been sampled regularly, at \( t=kT \), then by virtue of the sampling theorem \( f(t) \) could be expressed in terms of its sampled values \( f(kT) \) as

\[
f(t) = \sum_{k=-\infty}^{\infty} f(kT) \frac{\sin(\omega_0(t-kT))}{\omega_0(t-kT)} \cdots \cdots \ (3.2.9)
\]

Therefore replacing \( t \) by \( t+\tau \) in (3.2.9), the cross-correlation function is given by

\[
\psi_{12}(\tau) = \left\langle \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(kT+\Delta_k) f(nT) \frac{\sin(\omega_0(t-kT))}{\omega_0(t-kT)} \right. \\
\left. \left\{ \sin(\omega_0(t+\tau-nT)) / \omega_0(t+\tau-nT) \right\} \right\rangle \cdots \cdots (3.2.10)
\]
The ensemble average need only be performed on the product of the sample values since these are the only random variables in (3.2.10). Using (3.1.2) we have

\[ \langle f(kT + \Delta_k) f(nT) \rangle = \sigma^2 \sin \frac{\omega_0(kT+nT+\Delta_k)}{\omega_0(kT-nT+\Delta_k)} \ldots \] (3.2.11)

Substituting (3.2.11) in (3.2.10) and rearranging we obtain

\[ \psi_{12}(\gamma) = \sigma^2 \sum_{k=-\infty}^{\infty} \frac{\sin \omega_0(t-kT)}{\omega_0(t-kT)} \sum_{n} \frac{\sin \omega_0(t+\gamma-nT)}{\omega_0(t+\gamma-nT)} \cdot \frac{\sin \omega_0(kT+\Delta_k-nT)}{\omega_0(kT+\Delta_k-nT)} \ldots \ldots \] (3.2.12)

In appendix (A) it is shown that the following identity

\[ \sum_{k=-\infty}^{\infty} \frac{\sin \omega_0(t_1-kT)}{\omega_0(t_1-kT)} \cdot \frac{\sin \omega_0(t_2-kT)}{\omega_0(t_2-kT)} = \frac{\sin \omega_0(t_1-t_2)}{\omega_0(t_1-t_2)} \ldots \ldots \] (A6)

is valid for all \( t_1 \) and \( t_2 \) provided that

\[ T \leq \frac{\pi}{\omega_0} \]

Applying this identity with \( t_1 = t_+ \gamma, t_2 = kT + \Delta_k \), the summation over \( n \) can be expressed in a closed form and (3.2.12) becomes

\[ \psi_{12}(\gamma) = \sigma^2 \sum_{k=-\infty}^{\infty} \frac{\sin \omega_0(t-kT)}{\omega_0(t-kT)} \cdot \frac{\sin \omega_0(t+\gamma-\Delta_k-kT)}{\omega_0(t+\gamma-\Delta_k-kT)} \ldots \ldots \] (3.2.13)

The same identity however, cannot be applied directly to evaluate the summation over \( k \). This is because \( t_2 \) would be...
\[(t + \xi - \triangle_k)\] which is not constant but varies with respect to the summing index \(k\).

As a first step towards overcoming this problem, we remove this dependency on \(k\) by making use of the group periodicity equation (3.2.7). Thus replacing \(k\) by \(Mn+r\). The infinite summation over \(k\) is now split into two summations: an infinite summation over \(n\) and a finite summation of \(M\) terms over \(r\). This, together with (3.2.7) results in

\[
\psi_{12}(\xi) = \varphi^2 \sum_{r=1}^{M} \sum_{n=0}^{\infty} \sin \omega_0(t - (Mn+r)T) \cdot \frac{\sin \omega_0(t + \xi - \triangle_r - (Mn+r)T)}{\omega_0(t + \xi - \triangle_r - (Mn+r)T)}
\]

Even though \(\Delta_r\) is now independent of the summing index \(n\) it is still improper to apply the identity (A6) to sum over \(n\). This is because in (3.2.14), the effective period is now the group period \(MT\) thus violating the condition A2 (see Appendix A) which is necessary for the identity to hold true. Physically speaking, the summation over \(n\) in (3.2.14) represents the output of an ideal low pass filter fed with a series of impulses taken from a band limited function at the \(1^{th}\) critical Nyquist rate. The output therefore cannot be written in a closed form owing to the insufficient number of samples. To be able to express (3.2.14), in a closed form requires more samples and this we can obtain by considering more equations of the form (3.2.14).

Since both of \(f(t)\) and \(g(t)\) are in fact stationary processes the cross-correlation function \(\psi_{12}(\xi)\) is
independent of the time at which the ensemble average is performed. This means

$$\gamma_{12}(\tau) = \left\langle g(t) f(t + \tau) \right\rangle = \left\langle g(t-a) f(t-a+\tau) \right\rangle$$

...(3.2.15) where $a$ is an arbitrary constant.

Letting $a=\tau T$ and replacing $t$ by $t-\tau T$, in (3.2.14) for $l=1,2,\ldots M$, we can obtain $M$ equations identical to (3.2.14). Adding these equations yields

$$M \gamma_{12}(\tau) = \sigma^2 \sum_{r=1}^{M} \sum_{r=1}^{M} \sum_{k=-\infty}^{\infty} \frac{\sin {\omega_0}(t-\tau T-(Mn+r)T)}{\omega_0(t-\tau T-(Mn+r)T)} \cdot$$

$$\sin {\omega_0}(t+\tau -\Delta \tau -rT-(Mn+r)T)$$

$$\omega_0(t+\tau -\Delta \tau -rT-(Mn+r)T) \ldots (3.2.16)$$

Rearranging (3.2.16) and replacing $Mn+1$ by $k$ so that the summation over $n$ and $l$ are combined into one single summation over $k$ we obtain

$$M \gamma_{12}(\tau) = \sigma^2 \sum_{r=1}^{M} \sum_{k=-\infty}^{\infty} \frac{\sin {\omega_0}(t-rT-kT)}{\omega_0(t-rT-kT)} \cdot$$

$$\sin {\omega_0}(t+\tau -\Delta \tau -rT-kT)$$

$$\omega_0(t+\tau -\Delta \tau -rT-kT) \ldots (3.2.17)$$

Now with $t_1=t-rT$ and $t_2=t+\tau -\Delta \tau -rT$, the identity A6 can be applied to evaluate the summation over $k$. Hence the cross-correlation function is finally given by

$$\gamma_{12}(\tau) = \frac{1}{M} \sigma^2 \sum_{r=1}^{M} \frac{\sin {\omega_0}(\tau - \Delta \tau)}{\omega_0(\tau - \Delta \tau)} \ldots \ldots (3.2.18)$$
3.3 Mean Square Error.

Substituting (3.2.18) in (3.1.6) we obtain the mean square error

$$\left\langle \hat{E}_{C}^{2}(t) \right\rangle = 2 \sigma^{2} \left( 1 - \frac{1}{M} \sum_{r=1}^{M} \frac{\sin \omega_{0} (\zeta - \Delta r)}{\omega_{0} (\zeta - \Delta r)} \right) \quad (3.3.19)$$

3.4 Signal/Error Ratio.

A useful quantity which is more meaningful than the mean square error, is the signal to error ratio which is defined in analogy with the familiar quantity 'signal/noise ratio' commonly used in assessing the performance of communication systems. Therefore using (3.3.19) this quantity is given by

$$\frac{S}{E} = \frac{\left\langle f^{2}(t) \right\rangle}{\left\langle \hat{E}_{C}^{2}(t) \right\rangle} = \frac{1}{2(1 - \frac{1}{M} \sum_{r=1}^{M} \frac{\sin \omega_{0} (\zeta - \Delta r)}{\omega_{0} (\zeta - \Delta r)} )} \quad (3.4.20)$$

The basic difference between S/E ratio and S/N ratio is the fact that the former expresses a ratio of the average powers contained in two statistically correlated processes whereas in the latter, the two processes are statistically independent. In the latter case it is possible to increase the S/N ratio by increasing the signal power or reducing the noise power; doing this would have no effect on the S/E ratio since it is independent of $\sigma^{2}$ as can be seen from (3.4.20).

3.5 Discussion of theoretical results and optimisation.

From the above analyses it is obvious that the three
quantities, namely the cross-correlation function, the mean square error and the signal/error ratio are related quantities. The knowledge of one of the quantities plus the knowledge of the mean square value of the signal is sufficient to determine the other two. In practice therefore, determination of any one of these quantities can become a matter of convenience and relative ease of measurement. Theoretically speaking it is important to establish whether these quantities can be optimised with respect to one or more of the parameters of the system. In order to do that, we notice in the course of theoretical analysis, that the magnitude of these quantities depends on

a) The autocorrelation function of the signal which is explicitly included in (3.2.11) when performing ensemble averages on the sample values. This is not surprising since higher self correlation means more redundancy and this leads to a smaller error when one sample is substituted by another, occurring a little time later, or earlier, as in the recovery scheme.

b) The particular sample sequence used in the sampling process. This is explicitly included in (3.2.18), (3.3.19) and (3.4.20) in the parameters $\Delta_t$.

c) The parameter $\tau$ whose significance lies in the fact that due to the action of the sample and hold circuit included in the recovery scheme, the signal would inevitably suffer some delay. Because of this inherent delay, the maximum value of the cross correlation function and therefore the minimum value of the mean square error do not necessarily occur at $\tau = 0$, but at some other value $\tau$ characteristic of the sample distribution. This may
be shown by differentiating either (3.2.18) or (3.3.20) with respect to $\tau$, setting the derivative to zero and solving the resulting equation.

For a given signal and sample distribution therefore the cross correlation function and the $S_E$ ratio can only be optimised with respect to the delay $\tau$. Since delay introduces no distortion whatever on the signal, the maximum values of these quantities with respect to $\tau$ should provide the criteria in evaluating the performance of the recovery scheme. The optimum value of this delay $\tau_0$ is a characteristic of the sample sequence used in the sampling process and it can be defined as the effective time delay introduced by the recovery scheme on the signal. This definition becomes clear when we consider a uniform sample distribution being interpolated by such a recovery scheme. The output would simply be a delayed version of the signal which would be best correlated to an equally delayed version.

3.6 Optimisation with respect to the delay.

By differentiating (3.2.18) with respect to and setting the derivative to zero we obtain

$$\sum_{r=1}^{M} \frac{\omega_0 \sin \omega_0 (\tau - \Delta_r) - \omega_0^2 (\tau - \Delta_r) \cos \omega_0 (\tau - \Delta_r)}{\omega_0^2 (\tau - \Delta_r)^2} = 0$$

Solving this equation by analytical method is extremly tedious and difficult, if not impossible. Instead Numerical method may be used to solve the equation, or alternatively
the optimum value of $\mathcal{C}$ may be obtained by graphical methods. By plotting the signal/error ratio for a few values of $\mathcal{C}$, in the range $0-T$ the optimum value is simply the value of the delay at which the S/E ratio is maximum.

The relation between the various sample distributions and their characteristic values of delay will be further discussed in Chapter 6 where a general review of results is given.
CHAPTER 4.

Computer Simulation

4.1 Introduction.

In Chapter 3, expressions for the cross correlation function, the mean square error and the signal/error ratio were derived for a low pass stationary input in terms of the sample distribution parameters. This was done by performing ensemble averages on the random quantities involved and evaluating the infinite summations using the identity $A_6$ derived in Appendix A.

In this chapter, a method by which the mean square error and the signal/error ratio may be calculated is described. It involves direct simulation of the system with random stationary input on a digital computer. The output statistics, namely, the mean square error and the mean square value of the random input can thus be calculated by performing finite time averages. This method is basically a Monte Carlo simulation process which is a powerful computer technique that has found wide applications in the study of random processes. Indeed in many situations direct repeated simulation of systems with random inputs is frequently the only possible method of conducting realistic random process studies.

This method can therefore provide an alternative way of evaluating the mean square error and the signal/error ratio, and if the ergodicity hypothesis applies to the random process $f(t)$, i.e. the time and ensemble averages are identical, then the results produced by this method can be used to check the theoretical results.
obtained in Chapter 3.

In simulating the recovery scheme on the computer it was found necessary to simulate the interpolating action of the ideal low pass filter, since it can be shown that the mean square value of a continuous band limited signal is equal to the arithmetical average of the squares of its sampled values provided these samples are taken at the Nyquist rate and that a sufficiently large number of them are considered. This is shown below.

If \( m(t) \) is a random function, then its average power \( P \) is given by

\[
P = \overline{m^2(t)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} m^2(t) \, dt \quad \ldots (4-1-1)
\]

(bars denoting time averages.) and if \( m(t) \) is band limited, then by virtue of the sampling theorem, \( P \) can be written as

\[
P = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left[ \sum_{k=-\infty}^{\infty} m(kT) \frac{\sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)}}{\omega_0(t-kT)} \right]^2 \, dt \quad \ldots (4-1-2)
\]

Reversing the order of summation and integration (4-1-2) becomes

\[
P = \lim_{T \to \infty} \frac{1}{2T} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} m(kT) m(nT) \int_{-T}^{T} \frac{\sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)}}{\omega_0(t-kT)} \cdot \frac{\sin \frac{\omega_0(t-nT)}{\omega_0(t-nT)}}{\omega_0(t-nT)} \, dt \quad \ldots (4-1-3)
\]

Because of the orthogonality property which implies that

\[
\int_{-\infty}^{\infty} \frac{\sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)}}{\omega_0(t-kT)} \frac{\sin \frac{\omega_0(t-nT)}{\omega_0(t-nT)}}{\omega_0(t-nT)} \, dt = \begin{cases} \frac{\pi}{\omega_0} \text{ if } k=n \quad (4-1-4) \\ 0 \text{ if } k \neq n \end{cases}
\]
The average power \( P \) is given by

\[
P = \lim_{T_1 \to \infty} \frac{T}{4 \pi} \sum_{k=-\infty}^{\infty} m^2(kT) \ldots \ldots \quad (4-1-5)
\]

If however, the infinite series is truncated at a large value of \( k \) say, \( N \) such that (4-1-4) still approximately holds, then the average power \( P \) will be given by

\[
P = \frac{1}{2N} \sum_{k=-N}^{N} m^2(kT) \ldots \ldots \quad (4-1-6)
\]

The average power of a band limited random process is thus approximately equal to the arithmetical average of the squares of its sampled values providing these are taken at or above the Nyquist rate and that sufficiently large numbers are considered.

For the error function, which is defined by (3-1-4), we obtain by substituting (3-2-8) and (3-2-9) in (3-1-4),

\[
E_{c}(t) = \sum_{k=0}^{\infty} \left[ f(kT+\zeta) - f(kT+\Delta_k) \right] \sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)} \ldots \ldots (4-1-7)
\]

This may be expressed as

\[
E_{c}(t) = \sum_{k=0}^{\infty} e_{c,k}^{\Delta_k} (kT) \frac{\sin \omega_0(t-kT)}{\omega_0(t-kT)} \ldots \ldots (4-1-8)
\]

where

\[
e_{c,k}^{\Delta_k} (kT) = f(kT+\zeta) - f(kT+\Delta_k) \ldots \ldots (4-1-9)
\]

This quantity may be referred to as the 'local sample error.' The error function is thus band limited to \( \omega_0 \). Therefore by using the result obtained in (4-1-6), the mean square error is given (approximately) by

\[
\frac{E_{c}^2(t)}{E_{c}^2} = \frac{1}{2N} \sum_{k=-N}^{N} e_{c,\Delta_k}^2 (kT) \ldots \ldots \quad (4-1-10)
\]
and the signal/error ratio is given by

\[
\frac{S}{E} = \frac{\sum_{k=-N}^{N} f^2(kT + \zeta)}{\sum_{k=-N}^{N} e^2(kT + \Delta_k)} \quad \ldots \quad \ldots \quad (4-1-11)
\]

Equations (4-1-9) and (4-1-10) thus enables us to calculate the mean square error and the signal/error ratio. The simulation process simply involves setting a low pass stationary function on the computer and calculating the finite time averages of the relevant quantities in accordance with the appropriate relationships given above.

The main part of the computer programme deals with the generation of the random function. This part is a direct simulation of the physical system described below, which is normally used to generate pseudo-random test signals.

4.2 Pseudo-Random Noise Generators.

Random noise makes a good test signal because it is broad band and realistically simulates naturally occurring disturbances. Its randomness however, is not very helpful to the experimenter since it introduces an element of statistical variance into the results of the experiment. The statistical variance can be minimised if the measurements are taken over a very long period of time, usually inconvenient in practice.

Pseudo-random signals, which can be designed to have probability density functions and power spectra similar to those of a truly random noise, are more convenient test signals because they possess the following advantages:
1) Pseudo random noise is produced by reliable and easily maintained digital circuits. The analogue noise output is thus unaffected by the physical environments and its statistical properties are much easier to control.

2) The binary noise bandwidth is simply proportional to the clock pulse rate.

3) The pseudo random noise generator can be reset at any time to repeat a sequence of random events exactly. For example we can reset once per computation run. This permits repetitive computations and if computations or measurements are performed over a whole period of the sequence, the statistical variance of results can be truly eliminated. This is extremely valuable when parameters of the system being tested are varied so that it is important to know that changes in results are caused by parameter manipulation and not by statistical variance.

The most useful and widely used pseudo random test signals are of two types: the pseudo random binary (two-level) signals, and the pseudo random gaussian (multi-level) signals which are more suitable in testing analogue systems and can be obtained by sending a random binary signal through a low pass filter.

Many binary wave-forms have properties of pseudo random sequences; one family called the 'maximal length' sequences can be generated by a shift register with an appropriate feedback. In generating such sequences, the shift register is made to operate in a closed loop condition where the input to the first stage is supplied via a feedback path derived from later stages. The
feedback path is determined by the so-called 'generating polynomial' (5). Fig. 5 shows a 65-stage shift register whose generating polynomial is given by

\[ x^{65} + x^4 + x^3 + x + 1 = 0 \quad \ldots \quad (4-2-12) \]

This implies that the input to the first stage is supplied by a feedback path derived from the output of a modulo-2 adder, whose inputs are directly derived from the outputs of the 1st, 3rd, 4th, and the 65th stages.

This gives a maximum length sequence of pattern length \(2^{65} - 1\). The power spectrum of the generated sequence consists of a line spectra, the envelope of which is a \((\sin x/x)^2\) curve, where most of the power is contained in the first lobe and the nulls occur at intervals of \(f_c\), the clock frequency. The harmonics (line)spacing is a function of the sequence length and the clock frequency, and is given by \(f_c/2^{65} \cdot n\). The upper 3-db point of the spectrum occurs at approximately 0.45\(f_c\). Hence by adjusting the clock period the nominal bandwidth can be changed accordingly.

To convert the binary output to a multi-level Gaussian output, a low pass filter is needed. To get a reasonably the Gaussian probability density function, bandwidth of the filter is nominally required to be 1/20th of the clock frequency. It is also more suitable to use a digital filter than a conventional analogue one, especially when dealing with relatively low clock pulse rates. A digital filter can be easily synthesised from a delay line, a number of multipliers or weighting networks, chosen suitably to give the desired frequency response,
Fig. (5) - A Binary Maximal length sequence Generator.
and a summing amplifier. The delay line can take the form of a shift register when the input to the filter is a binary signal.

Fig. 6 shows the binary waveform generator shown previously in Fig. 5 where the shift register also performs the duty of the delay line needed for the digital filter. The weighting networks or multipliers are simply resistors whose values are selected so that the contributions of the outputs of successive stages to the current at the summing points are graded to follow the $(\sin X/X)$ curve, (the impulse response of an ideal low pass filter shown in Fig. 7). The output of the digital filter is not smooth but a series of steps like any waveform that has been generated digitally. These steps can be removed in practice by further analogue filtering so that a smooth Gaussian signal is obtained.

The bandwidth of the digital filter is inversely proportional to the time of the first null in the impulse response in Fig. 7. This null occurs after eight clock pulse periods, consequently the cut off frequency is $1/16$ of the clock frequency. The bandwidth of the filter can thus simply be adjusted by varying the clock frequency. The shape of the frequency response is only approximately rectangular owing to the truncation of the $(\sin X/X)$ curve. Since the upper frequency limit of the unweighted binary sequence power spectrum occurs at $0.45 f_c$, the spectrum of the multi-level output should be reasonably flat over the range $0 \rightarrow \frac{f_c}{16}$.
Fig. (6) - Pseudo-Random (Multi-Level) Signal Generator

Fig. (7) - Quantised impulse response of a low pass filter.
4.3 Direct Simulation on a Digital Computer.

The system described above lends itself readily to straightforward simulation on a digital computer. The actions of the shift register and the digital low pass filter can easily be implemented with the help of a number of simple instructions. The computer output then gives a series of numbers or sample values which if displayed in time would form the desired quantised pseudo random signal.

Since the interval between the nulls of \((\sin X/X)\) curve is inversely proportional to the bandwidth of the filter and also equals the Nyquist interval, then scale multiplication of the time axis by the constant \(\omega_0\) leads to time and frequency normalisation such that the nulls now occur at integer multiples of \(\pi\) instead of \(T\). The time axis is thus replaced by a normalised time axis which is more meaningful and suitable to use in the computer.

With the impulse response of the filter shown in Fig. 7 being represented by 65 co-ordinates spaced by \(\pi/8\) normalised time units, the computer output gives random samples at regular intervals \(\pi/8\) units. The amplitude of the \(i^{th}\) sample occurring at the normalised time being given by

\[
A_i = \sum_{j=1}^{65} a_j \ x^i(j) \quad \ldots \quad (4-3-13)
\]

where \(a_j\) is the weighting coefficient of the \(j^{th}\) stage of the shift register, given by

\[
a_j = \frac{\sin \left[ \frac{\pi}{2} \frac{(j-33)}{(j-33) \pi/8} \right]}{(j-33) \pi/8} \quad \ldots \quad (4-3-14)
\]

and \(x^i(j)\) is the binary state of the output of the \(j^{th}\)
stage caused by the $i^{th}$ clock pulse. The binary state can either be 1 or 0. If however, the mean value of the function is desired to be zero, then the 0 can be replaced by -1 and the modulo 2 addition by multiplication. Because the +1s and the -1s occur with equal probability, the mean value of the function will be very approximately zero.

The sample values thus produced occur at a rate of eight samples per $\pi$ interval, i.e. at eight times the Nyquist rate. The quantised signal they constitute would be sufficiently close to the smooth analogue version.

4.4 Computation of the Mean Square Error and the S/E Ratio.

The recurrence period of the sampling distributions is taken to be $8\pi$. This implies that $M=8$, so that the number of possible different sample distributions is $8^8$. Out of these distributions 8 are uniform and are used to represent the signal $f(t)$ and its delayed versions $f(t+\tau)$ on the computer. This becomes clear when we discuss the following simple method of designating a sample distribution.

Since any sequence is assumed to repeat every $8\pi$ units then it is only necessary to consider one group period. The 64 sample points contained in this $8\pi$ long period are divided into 8 groups of 8 points each. These points are numbered 1-8, within one group, for identification purposes. The number attached to a particular sample point has the significance of defining its position within the group. For example if the number 4 is mentioned, then the 4th sample point occurring at $4\pi/8$ within the appropriate group is meant. The groups need not be labelled since they are taken in an ascending order. A set of eight integer numbers lying in the range 1-8 is thus sufficient to specify a particular sample distribution. A uniform sample
distribution results when all the numbers in the set are equal. Fig. 8 shows two sample distributions. Fig. 8a shows a uniform distribution in which the sample points occur at the second position (\( \frac{2\pi}{8} \)) within each group and so it is denoted by 2 2 2 2 2 2 2 2. Fig. 8b shows a non-uniform distribution in which the sample values points occur at \( \frac{\pi}{8}, \frac{3\pi}{8}, \frac{5\pi}{8}, \frac{7\pi}{8}, \frac{9\pi}{8}, \frac{11\pi}{8}, \frac{13\pi}{8}, \frac{15\pi}{8} \) within groups 1, 2, ..., 8 respectively. Consequently the distribution is denoted by 1 5 7 6 3 4 8 2.

Any uniform sample distribution when interpolated by an ideal low pass filter would yield the original signal or a delayed version of it depending on where the time origin is taken. In this case the eight uniform distributions can be assumed to represent eight successively delayed versions of the original signal such that the \( r^{th} \) uniform sampling sequence can be assumed to represent the delayed version \( f(\Theta + r\cdot\frac{\pi}{8}) \) where \( \Theta \) is the normalised time and \( r \) lies in the range 1-8.

Hence for any non-uniform sampling sequence the local sample error \( \varepsilon_k(\Delta_k^{(kT)}) \) can be calculated for all of the eight values of the normalised delay using (4-1-9). The mean square value of the error and the signal/error ratio can be computed for each value of the delay in accordance with (4-1-10) and (4-1-11).
Fig. (8) - Pictorial Representations of Sample distributions.

(a) - A Uniform Distribution 2 2 2 2 2 2 2 2 2

(b) - A Non-Uniform Distribution 1 5 7 6 3 4 8 2
CHAPTER 5.

Experimental Work and Measurements.

Experimental work was carried out on the approximate recovery scheme in order to verify, and thus establish, the theoretical results obtained in Chapter 3. It was also desired that experiments should be carried out to test the performance of the scheme under more realistic conditions than those assumed theoretically. This involved the use of speech signals which are more realistic than pseudo-random noise, in order to ascertain whether or not the quality and intelligibility of the reproduced signal is sufficiently high.

As far as the first point is concerned, it was found easier, and for reasons which will become clear later, more rewarding to measure the mean square value of the error function, rather than cross-correlation function. The system we are about to describe, was used to measure the former quantity at different values of the delay \( \tau \) in the range of 0-T and for a number of different sample distributions, using pseudo-random noise as a test signal.

5.1 The Mean Square Error Measurement System.

Fig. 9 shows a schematic representation of the system used in measuring the mean square error. The input to the system is the signal \( f(t) \) which is sampled non-uniformly by a recurrent pulse sequence derived from a special purpose pulse generator referred to as the 'code sequence pulse generator', which is designed to give all the possible
Fig. (9) - Block Diagram of the Error Measurement System.
sequences $N^M$. The non-uniform sample values are stored on a capacitor, forming the 'hold' part of the 'sample and hold' circuit, and the staircase-like waveform is then subsequently sampled by a regular pulse sequence also derived from the pulse generator. The output of the second sampling circuit is stored on a capacitor just before low pass filtering, in order to restore the considerable amount of signal power lost as a result of the second sampling process. The output of the lowpass filter $g(t)$, is now fed into one input of a difference amplifier, whose other input is a delayed version of the signal $f(t)$. A true R.M.S. Meter is then used to average the error function produced at the output of the amplifier, and so gives a reading proportional to the R.M.S. of the error. The mean square error was measured at various values of the delay $T$. In particular, the optimum value of the delay $T^*$, was noted. This is the value at which the reading on the meter is minimum. The system will now be described in detail.

5.1 The Function Generator $f(t)$.

$f(t)$ is a pseudo-random Gaussian test signal obtained by analogue low pass filtering of a maximal length binary sequence which has been generated in a manner similar to that described in the previous chapter, but by hardware rather than software. It has a Gaussian pdf and a low pass power spectrum cutting off at 4KHz.

5.1.2 The Code Sequence Pulse Generator.

The Pulse generator was designed to produce recurrent
non-uniform pulse sequences which repeat every eight pulses. On eight channel design basis, the generator can produce $8^8$ different pulse sequences of which eight are uniform sequences. The generator consists of two 8-stage shift register delay lines, a number of gates and a set of eight (one pole, eight ways) wafer switches for sequence selections. With the aid of the method explained at the end of Chapter 4, and used to designate a particular sample sequence, the action of the generator is described below.

Fig. 10 shows a shift register of eight stages. The first stage is set initially at the binary state of 1, whereas the rest of the stages are set initially at the binary state of 0. The register is operated in a closed loop condition where the output of the last stage is fed directly to the input of the first stage. When the register is clocked, the 1 propagates down the line until it reaches the eighth stage, where a further clock pulse causes the 1 to appear at the first stage again. The cycle repeats every eight clock pulses. Looking at the outputs of successive stages, it is easy to see pulse waveforms of frequency which is $1/8^{th}$ that of the clock frequency successively delayed from each other by one clock period. If the output of the eighth stage is now allowed to clock another shift register of eight stages having the same initial settings (see Fig. 11), then the outputs of the successive stages of the second register will be pulse waveforms of frequency $1/64^{th}$ that of the clock frequency, delayed successively by eight clock pulse periods; (see Fig. 12 for waveforms).
Fig. (10) - A Shift Register Delay Line.
Fig. (12) - Pulse Waveforms produced by system in Fig. (11).
Now considering the sequence period which consists of eight frames, each frame (group) containing eight sample point positions, the position of the 1 in the shift register B, can be used to define the relative positions of the frames, whereas the 1 in register A, can be used to define the positions of the sample points within each frame. This means that in order to generate sampling pulse to occur at the \( r^{th} \) position of the \( n^{th} \) group (both \( r \) and \( n \) are in the range 1-8), simply requires an And-type Gating of the waveforms appearing at the outputs of the \( r^{th} \) and the \( n^{th} \) stages of registers A and B respectively. Since there are eight frames in each sequence period, eight such gating operations are necessary, whose outputs are then Or-Gated in order to generate the complete sampling sequence. For example, the logic relationship defining the gating operations required to produce the sequence 1 5 7 6 3 4 8 2 is given by:

\[
\begin{align*}
S &= Q \cdot Q + Q \cdot Q + Q \cdot Q + Q \cdot Q + Q \cdot Q + Q \cdot Q + Q \cdot Q \\
&= \left( Q_1 \right) \left( Q_2 \right) \left( Q_3 \right) \left( Q_4 \right) \left( Q_5 \right) \left( Q_6 \right) \left( Q_7 \right) \left( Q_8 \right) \\
&= (5.1.1)
\end{align*}
\]

where \( Q \) and \( Q \) are the binary states of the outputs of the \( j^{th} \) stages of registers A and B respectively. Note that the stages of register \( A \) are taken in accordance with the numbers specifying the sequence. This is in agreement with the method of sequence designation explained in Chapter 4. Fig. 13 shows the complete pulse generator where integrated circuits are used in constructing the shift register and providing the necessary gate circuits. In addition, the outputs of register A are gated to those of register B via wafer switches (one-pole eight way). The eight slider
Fig. (13)
Code Sequence Pulse Generator.
positions in each switch are marked 1-8, and the positions bearing the same identification numbers in all the eight switches are directly connected to the output of the corresponding stage of the register. To select a particular sequence therefore, simply requires setting the sliders of the switches at the positions determined in accordance with the set of numbers identifying the sequence. In Fig. 13, the sliders are set at the necessary positions to generate the sequence 1 5 7 6 3 4 8 2.

The system can also be developed further, to act as a sequence multiplexer so that by adding seven more sets of gates and switches, eight multiplexed pulse sequences can be generated and used in the multiplexing process. At the receiver, prior knowledge of the multiplexed sequences is essential for the decoding process, which would otherwise be extremely difficult. This is the essence of security in this type of pulse modulation system.

5.1.3 The Sample and Hold Circuit.

An electronic switch, followed by a capacitor of a suitable size constitutes a sample and hold circuit. The switch normally actuated by a short pulse, allows a capacitor to charge or discharge quickly in order to attain the instantaneous value of the input voltage to the switch. The charge is then held on the capacitor until the next actuating pulse arrives. In control systems this circuit is usually used as a simple interpolator of sampled data where the capacitor holds or remembers the value of the sample until the arrival of the next; thus giving a staircase
approximation of the original signal. This is a crude form of low pass filtering and only satisfactory when the sampling rate is many times higher than the Nyquist rate. Where this is not the case, further analogue filtering is required. In Pulse Amplitude Modulation systems, sample and hold circuits are sometimes used just before low pass filtering in order to extend the short samples into rectangular waves filling the whole intervals between the samples, thus increasing considerably the average signal power received by the channel. Offsetting this advantage however, there is a small loss of high frequency response in the signal band.

In our recovery scheme a sample and hold circuit was used to store the incoming non-uniform sequence, one sample at a time, so that by sampling the output of the capacitor regularly it was possible to transform the non-uniformly occurring data into a uniform sequence more suitable for direct low pass filtering.

Fig. 14 shows a sample and hold circuit where the electronic switch function is performed by a field effect transistor (FET). The FET couples a capacitor of a suitable size to an emitter follower. The application of a short pulse at the gate of the FET causes it to conduct fully in either direction, and the capacitor charges or discharges to attain the value of the input voltage. In the absence of a pulse, the FET acts as an open circuit and the charge on the capacitor is held until the arrival of the next pulse at the gate, when the capacitor acquires the new value.
Fig. (14) A Sample and Hold Circuit.
In designing the circuit the following design criteria were followed:

1) To ensure fast charging and discharging of the holding capacitor \( C \), its value must satisfy the following inequality

\[
(R_o + R_{on})C \leq \tau \quad \ldots \ldots (5.1.2)
\]

where \( R_o \) is the output impedance of the emitter follower, \( R_{on} \) is the on-resistance of the FET when it is fully conducting, and \( \tau \) is the sampling pulse width.

2) To ensure full conduction of the FET the drain current \( I_D \) should be maximum. This occurs when the gate potential relative to that of the source is zero. Conversely, the FET cuts off completely when this potential is negative and below a critical value called the pinch off voltage \( V_p \).

Summarising:

- \( I_D \) is maximum when \( V_{GS} = 0 \) \( \ldots (5.1.3) \)
- \( I_D \) is zero when \( V_{GS} \leq -V_p \) \( \ldots (5.1.4) \)

Hence, if a pulse waveform switching between zero and a positive level 'A' is desired, it would be necessary to connect a diode between the gate and the source.

To satisfy the condition in (5.1.3), \( A \) should be

\[
A \geq V_E + V_1 \quad \ldots \ldots (5.1.5)
\]

where \( V_E \) is the emitter of the FET dc bias, and \( V_1 \) is the maximum positive level of the input signal. In the absence of a pulse, \( V_{GS} \) is given by

\[
V_{GS} = -V_E - V_2 \quad \ldots \ldots (5.1.6)
\]

where \( V_2 \) is the maximum negative level of the signal.

Therefore, to satisfy the condition implied in (5.1.4), \( V_E \) should be determined by

\[
-V_E - V_2 \leq -V_p \quad \ldots \ldots (5.1.7)
\]

(5.1.5) and (5.1.6) thus give the values of \( A \) and \( V_E \).
necessary to achieve optimum switching operation. In the circuit shown in Fig. 14, these quantities have the following values:

\[ V_E = 7 \text{volts}. \]
\[ A = 9 \text{volts}. \]
\[ V_p \text{ for the FET used} = -6 \text{volts}. \]

The maximum allowable swing of the input voltage is thus \( \pm 1 \text{ volt}. \)

5.1.4 Sample and Hold circuits in cascade.

When a sample and hold circuit is cascaded to a similar circuit or any other system, care must be taken to ensure that the input impedance of the second circuit is sufficiently high, otherwise the charge on the capacitor might decay considerably during the holding period. Where this is not possible, a high input impedance buffer such as a source follower must be used to couple the two circuits. Fig. 15 shows two cascaded sample and hold circuits coupled via a FET source follower. This circuit forms the essential part of the recovery scheme. The first FET switch is controlled by a recurrent pulse sequence derived from the code sequence pulse generator. The resulting staircase version of \( f(t) \) was then sampled by a second FET switch controlled by a uniform pulse sequence also derived from the sequence generator. The output of the circuit was filtered to produce \( g(t) \).

5.1.5 Sample and Hold Analogue Delay Line.

As a result of the storage action of the holding capacitor in the sample and hold circuit, an active
Fig. (15) - A Cascade of two Sample and Hold Circuits.

Reciprocal Pulse Sequence

Regular Pulse Sequence

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
<th>Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>6.8 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>91 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>220 Ω</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>1 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R8</td>
<td>6.8 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R9</td>
<td>91 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R10</td>
<td>220 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R11</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R12</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R13</td>
<td>10 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R14</td>
<td>1 kΩ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>250 μF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>4700 pF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>5 μF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>250 μF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>4700 pF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>5 μF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>250 μF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1, Q4</td>
<td>ZTX 302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2, Q3</td>
<td>TIS 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5, Q6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9 volts
analogue delay line can be constructed by cascading a number of these circuits whose switching operations are controlled by uniform successively delayed pulse waveforms of the same frequency. This delay line works on the sampling principle and so is suitable to use with band limited signals. The pulse waveforms required for switching the successive stages can be provided by a pulse waveform of sufficiently high frequency, and the same waveform delayed by an amount which is equal to half its period, say $T_1$. If these two waveforms are applied alternately to the gates of the successive FET switches, a total delay would be produced, given by $(n-1)\frac{T}{2}$, where $n$ is the number of sample and hold circuits cascaded to form the delay line. This total delay can be varied by simply adjusting the switching frequency, provided this is always kept above the Nyquist rate. This is therefore an attractive form of analogue delay line which requires no inductances and is capable of providing variable delay. To produce the delay between the switching waveforms a monostable circuit whose time constant is $\frac{T}{2}$ may be used.

The circuit of Fig. 15 can be seen in this light as a delay line of two stages where the maximum possible delay that can be achieved, is $T$ seconds. In measuring the mean square error, this circuit was used to introduce variable delay on the signal $f(t)$. The variable delay between the sampling pulse waveforms was produced by using a uniform pulse waveform to trigger a monostable delay circuit whose delay was made adjustable by controlling the time constant of the circuit.
The use of this circuit in the error measurement scheme has the following advantages over other delay circuits:

1) Because the delay is produced digitally, it can easily be made variable. In this case, it was achieved by adjusting the time constant of a monostable circuit.

2) In making the delay circuit, identical to that used in producing the output $g(t)$, any distortion suffered by the signal $f(t)$ apart from that due to the non-uniform sampling process, would almost be identical in both circuits, and so cancels out at the output of the differential amplifier. It can therefore be said with confidence that the resulting error function is mainly due to the non-uniform sampling process.

If the cross-correlation function was to be measured instead of the mean square error, then for obvious reasons no advantage would have been gained by making the delay circuit identical to the recovery scheme circuit. This is why it is more rewarding to measure the mean square error since more accurate results can be expected.

5.2 System Calibration and Experimental Procedure.

Fig. 16 shows the complete system used in error measurements where the circuit of Fig. 15 was used for both the recovery scheme and the delay circuit. To calibrate the system, so that it would be ready for error measurement the gain of the recovery scheme must be equal to the gain of the delay circuit. This was achieved by selecting one of the eight uniform pulse sequence and adjusting the delay until the reading on the R.M.S. meter is minimum. The gains of both circuits are then adjusted by means of the pot-
Fig. (16) - Error Measurement System (circuit component values are the same as in Fig. 15).

Recovery Circuit

Delay Circuit
entiometers located at their outputs to minimise the meter reading still further until a near zero reading on the most sensitive scale is obtained.

The system is now calibrated and ready for future error measurements. The minimum error reading obtained in the calibration process can be regarded as a measure of accuracy which can be expected from the system. In our case the minimum reading was 40 db below the signal level, giving a 1% degree of accuracy, which was thought satisfactory. Having calibrated the system, the mean square error was measured from a number of sampling distributions at various values of the delay $\tau$. In particular, the minimum value of the error and the value of delay at which this occurred were noted; the results obtained are discussed in Chapter 6.
CHAPTER SIX.

Review of Results.

In this chapter, a general review of results obtained by the three methods described in the previous three chapters, is given and their implications are discussed.

6.1 Statistical properties of the Computer Simulated Function.

Tests were conducted on the computer simulated function to see whether its statistical property, namely its probability density and autocorrelation functions conform to those desired in the simulation process. Strictly speaking these tests should be performed over a whole period of the function which contains $2^{65} - 1$ samples. This of course is impractical, in view of the extremely large number of samples which require computations lasting for years on the fastest computers. Instead, 6,400 samples constituting a small portion of this period, were considered sufficient for these tests. The pdf and the autocorrelation function were calculated using the computer and the results are shown plotted in Figs. (17&18); in Fig. (17), the pdf is super-imposed on a true Gaussian distribution with mean and variance identical to those of the simulated function. It can be seen from this figure that the pdf of the simulated function is very approximately Gaussian. Figure (18) shows the autocorrelation function super-imposed on a $\frac{\sin X}{X}$ curve.

The agreement between the two curves is very good except for a slight departure which begins after the null 311. This is expected, because of the truncation of the impulse response curve in the simulation process. The simulated function is therefore approximately Gaussian having a low pass power spectrum and in this respect is almost identical to the idealised function assumed in the theory given in Chapter 3.

6.2 Comparison of Results obtained by the Three Methods.

It can be generally assumed that the ergodicity hypothesis applies to white or low pass noise functions. Starting from this point we can therefore expect some form of agreement between the theoretical results and those obtained in the simulation process. In addition, it was hoped that these results would be verified and confirmed by those
Fig. (17) - Probability density function of the simulated random function.
Fig. (18) - The Auto-Correlation function of the simulated function.
obtained from experiment. For this purpose and the purpose of delay optimisation, the signal/error ratio was evaluated by the three methods for a number of sample distributions and the results are plotted against the delay in Figs (19-23).

In the experimental measurements, the sampling frequency was set at 8.3KHz which corresponds to a Nyquist interval of 120/μsec. The S/E ratio measurements were conducted at 15/μsec. delay intervals corresponding to T/8 normalised delay units. Looking at these graphs we see a very close agreement between the three curves. Moreover, the three curves reach maximum values at the same delay point. This is the optimum value which can be found directly from the graphs.

6.3 The Cross Correlation Function.

The close agreement between the signal to error ratio curves obtained by the three methods implies that the same applies to the cross correlation function curves, since it is related linearly to the mean square error. Therefore, it is sufficient to plot the theoretical expression given in (3.2.18) to show its variation with respect to . The normalised cross correlation function ( ) for a number of sample distributions, is shown plotted in Figs (24-27). The curves appear similar to the curve with two basic differences.

1) The maximum values of these curves are different for different sample distributions and in all cases are less than unity. This is due to the fact that the outputs can only be partially correlated to the signal .

2) The curves are shifted along the delay axis, due to
Theoretical
Computer Sim.
Experimental

Fig.(19) - Error results for the sample distribution (1 2 6 5 4 8 1 7)
Fig. (20) - Error Results for the Sample Distribution

(NORMALISED DELAY)
Fig. (22) - Error results for the Sample Distribution

NORMALISED DELAY

Theoretical
Computer Sim.
Experimental
Fig. (23) - Error results for the Sample Distribution

(NORMALISED DELAY)

(1 8 1 8 1 8 1 8)
the inherent delay of the recovery scheme. The shift is naturally equal to the optimum delay.

If a uniform sample sequence had been reconstituted by the recovery scheme then \( g(t) \) would have been identical to \( f(t + \tau) \) and the cross correlation function would have been of the form \( \frac{\sin X}{X} \) delayed by an amount \( \tau \). Therefore in order to give a pictorial representation of the distortion introduced by the recovery scheme, a \( \frac{\sin X}{X} \) curve delayed by an amount \( \tau_0 \), is shown superimposed on the cross correlation curves. These curves give a rough measure of the distortion, and can also provide a comparison between the distortion resulting from using different sample distributions. For example, the distortion is more pronounced in the case of the distribution 1 8 1 8 1 8 than in the case of other distributions, since the departure between the two curves in the case of the former distribution is more than that in the latter cases.

6.4 Power Spectrum of the Error Function.

The power spectrum of the error function is band limited spectrum whose shape depends largely on the particular sample sequence used in the sampling process. Unfortunately, in attempting to evaluate the error power spectrum or its Fourier transform (the auto-correlation function), we encounter an infinite summation which cannot be expressed in a closed form. Therefore it is very difficult to predict the variation of the spectrum theoretically.

The output \( g(t) \) even though it may be statistically identical for a number of different distributions, can never be exactly identical because of the phase distortion.
Fig. (2a) - Normalised Cross Correlation Function for (NORMALISED DELAY) the sample distribution (1 2 6 5 4 8 1 7)
Fig. (25) - Normalised Cross Correlation Function for the sample distribution.

\[ \frac{\sin(\tau - \tau_0)}{\tau - \tau_0} \]
Fig. (26) - Normalised Cross Correlation Function for the sample distribution

\[ \sin(\tau - \tau_0) \over (\tau - \tau_0) \]
Fig. (27) - Normalised Cross Correlation Function for the sample distribution (1 8 1 8 1 8 1 8)
introduced by the recovery scheme in transforming the non-uniform sequences into uniform ones. To make this point clear, consider a sample distribution. By interchanging the order of the numbers in the set specifying it, we can obtain a whole family of distributions which involve different sampling processes being performed on the signal. The theoretical expressions (3.2.18), (3.3.19) and (3.4.20) contain no provisions for the order of the numbers, and so therefore, they would predict identical results for all these different distributions. This is not surprising in view of the fact that the above mentioned expressions represent average quantities and as such are expected to be independent of the order of the numbers specifying the distributions. This point also can be confirmed by the computer and experimental results, see Fig. 28 - 31 where curves for the signal to error ratio are shown for two families of different distributions, specified by sets containing the same numbers, namely:

3 1 6 7 8 2 5 4
5 2 4 8 7 6 1 3
and 4 4 3 3 6 5 8 1
4 3 4 3 5 6 1 8

The outputs $g(t)$ resulting from such distributions cannot be exactly identical because the order of the numbers to a large extent determines the amount of phase distortion suffered by each frequency component of the original signal in the recovery scheme. Therefore, the frequency spectrum of $g(t)$ varies according to the sample distributions. To show this point, experimental frequency analysis of the error function for a number of distributions were conducted.
Fig. (28) - Computer Simulation results for two sample distributions specified by the same numbers.
Fig. (29) - Experimental results for the two sample distributions of Fig. (28).
Fig. (30) - Computer Simulation results for two sample distributions specified by the same numbers.
Fig (31) - Experimental results for the two sample distributions in Fig. (30).
During these experiments, the delay was set at the optimum value $\gamma_0$. Figs. 32-35 show these error spectra normalised with respect to the constant power spectrum of the signal; the sharp rise occurring near the cut-off frequency is a result of the sharp decrease of the signal power spectrum and the relative rise in the high frequency content of the error function. Sample distributions specified by sets containing the same numbers but arranged in different orders result in error functions having different shape power spectra, as can be seen in Figs. 32-35. The area under these curves however, should be identical since they represent the minimum value of the mean square error. This appears to be the case in these figures (note the logarithmic frequency scale).

6.5 Approximate Relationships between the Maximum Value of the S/E Ratio, the Optimum Delay and a given Distribution.

In the last section, it was mentioned that different sets of numbers specifying different sample distributions may give rise to identical results if the numbers comprising the sets are identical, but arranged in different orders. Can this suggest any simple relationships between the distributions of the numbers, the value of the optimum delay, and the maximum value of the S/E Ratio?

To answer this question, we should note that the optimum delay $\gamma_0$ which really gives the effective delay introduced by the recovery scheme might well be related to the average value of the numbers comprising the set, since these numbers do in fact define the amount of time shifts suffered by the sample points. Intuitively speaking, we
Fig. (32) - Power Spectrum of the Error Function for the sample distribution

(1 2 6 5 4 8 1 7)
Fig. (33) - Power Spectrum of the Error Function for the sample distribution

(6 7 7 8 5 3 2 2)
Fig. (34) - Error Power Spectra for the sample distributions:

- - - - - (1 3 4 5 8 3 6 4)
- - - - - (4 3 4 3 4 6 1 8)
- - - - - (4 4 3 3 6 5 8 1)
Fig. (35) - Error Power Spectra for the sample distributions:

- - - - - (3 1 6 7 8 2 5 4)
- - - - - (5 2 4 8 7 6 1 3)
should also expect that the S/E ratio decreases as the sample distribution departs more and more from the uniform case.

Indeed, it was found, and can be seen on all the graphs showing curves plotted against the delay, that the optimum values appear to occur at a delay value which is very approximately given by the average value of the numbers comprising the sets multiplied by \( \frac{f}{8} \).

As far as the magnitude of the optimum values is concerned, the variance of these numbers, given by \( (X_1 - \bar{X})^2 \) where \( \bar{X} \) is the average value of the numbers, was considered to be the best parameter to be used as a measure of departure of a sample distribution from the uniform case. A number of sample distributions specified by sets of numbers having different means and variances were considered and the minimum values of the mean square error were evaluated for each of them. The percentage of the error power, relative to that of the signal, was plotted against a variance axis to see whether any relationship exists. The results are shown plotted in Fig. 16, where the curve appears to be a straight line, passing through the origin which obviously corresponds to the uniform sampling case. The case was proved further, by considering a number of different distributions having different means and equal variances. The results are shown in Table 1, which gives almost identical results for all these different distributions.

Unfortunately, these relationships were found to be difficult if not impossible to prove analytically. They
Fig. (36) - Relationship between the minimum Error/Signal ratio and the sample distribution.
should therefore be treated, at this stage, as approximations only. However, they appear to be very useful, since by using them the optimum values of the delay and S/E ratio can be found instantly and with good accuracy without having to refer to the exact theoretical expressions. This allows a quick estimate of the errors involved in the recovery scheme under a given distribution.

6.6 Summary of Results.
1) Results obtained by the three methods were found to be in good agreement. The theoretical results were verified and thus confirmed.

2) Different sample distributions obtained by interchanging the order of numbers specifying a given distribution, give rise to different outputs $g(t)$, which are nevertheless equally correlated to the original signal $f(t)$.

3) The optimum value of the delay $\gamma_0$ is very approximately given by the average value of the numbers specifying a given distribution multiplied by $\frac{p}{8}$. The minimum value of the mean square error is found proportional to the variance of the numbers.

<table>
<thead>
<tr>
<th>Sample Distribution</th>
<th>Mean Value</th>
<th>Variance</th>
<th>Max. Value of $S/E$ ratio (DB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5 3 1 1 1 3 1</td>
<td>2</td>
<td>2</td>
<td>10.08</td>
</tr>
<tr>
<td>2 5 5 1 2 4 2 3</td>
<td>3</td>
<td>2</td>
<td>9.98</td>
</tr>
<tr>
<td>1 7 1 2 1 1 2 1</td>
<td>2</td>
<td>3.75</td>
<td>7.84</td>
</tr>
<tr>
<td>7 3, 5 7 6 5 6 1</td>
<td>5</td>
<td>3.75</td>
<td>7.68</td>
</tr>
<tr>
<td>5 1 6 4 1 3 6 6</td>
<td>4</td>
<td>4</td>
<td>7.09</td>
</tr>
<tr>
<td>8 8 7 3 6 3 8 5</td>
<td>6</td>
<td>4</td>
<td>7.09</td>
</tr>
</tbody>
</table>
CHAPTER 7

A Simple Method of Channel Coding with an Exact Reconstruction Scheme.

In the preceding chapters, emphasis was placed on the problem of non-uniform sampling and the estimate of the amount of distortion resulting from the use of an approximate recovery scheme.

The minimum value of the mean square error varies in relation to the sampling distributions and as we found, it is almost proportional to the variance of the set of numbers specifying them. In a dynamic situation where any of these distributions might be used, the approximate recovery scheme would produce outputs varying considerably in their degree of correlation with the original signal. In dealing with speech signals however, subjective tests indicated that the quality and intelligibility of these signals differed little in relation to the different sample distributions. The reason of course, is the relatively high degree of self-correlation and the large amount of redundancy characterising the speech.

In looking at the approximate recovery scheme, we find that the error arises from the time-shifting the non-uniform sequence is subjected to in order to make it uniform prior to the filtering process. It is therefore possible that the reverse process, if carried out at the transmitter, before the channel multiplexing, might well eliminate the error completely. This is the essence of the following method of channel coding that can be used in conjunction with the approximate recovery scheme and results in exact reproduction of the original signal.
Each one of the signal sources is first sampled uniformly and the sample values are stored on the capacitor of a sample and hold circuit. The outputs of the capacitors are then sampled non-uniformly, in a manner determined by the multiplexing sequence, and the data is transmitted. At the receiver, the prior knowledge of the multiplexing sequence is essential for the decoding process. The demultiplexed non-uniform data belonging to each channel is essentially uniform data in that it was originally obtained by regular sampling at the signal source, but was transmitted non-uniformly for security reasons. This data will be restored to its original regular form by the action of the sample and hold circuit included in the approximate recovery scheme which reverses the process undergone at the transmitter. The output $g(t)$, is then simply a delayed version of the original signal. This method is basically a possible realisation of method(1), in Chapter One, which involves uniform sampling of the signal plus buffer storage. The sample and hold circuit provided the necessary storage for this application. Non-uniform sampling can be indirectly employed in this method for the purpose of achieving the form of channel multiplexing necessary to increase the system security.

In comparing this method to the other two methods which employ direct non-uniform sampling of signal sources, it can be said that this method is much superior to the other two, as far as this application is concerned. It has the advantages of simplicity, flexibility and perfect recovery. The cost of achieving this is small involving simply the addition of a sample and hold circuit prior to the multiplexing process, at the transmitter.
Conclusions.

The security of the normal pulse modulation system can be greatly increased if the sequence of channel multiplexing is made variable in a manner known only to the sender and the receiver. A simple method of achieving this was presented at the beginning of this study.

It was found that the realisation of this method requires non-uniform sampling prior to the multiplexing process. Either directly applied to the signal sources or indirectly to intermediary signals, derived from these by regular sampling and capacitor holding (see Chapter 7).

Non-uniform sampling applied directly to the signal sources poses certain problems as signal reconstruction at the receiver and the accuracy of the sampling process at the transmitter are concerned. These were dealt with in detail in Chapter (2). In contrast, it was found that indirect non-uniform sampling together with a relatively simple recovery scheme is capable of achieving the increased security requirement and at the same time combines simplicity and flexibility with exact recovery. Qualities which cannot be achieved by the exact and complex scheme needed in the direct case. Therefore as far as this application is concerned the choice is obvious.

Nevertheless, it was worthwhile investigating the problem of non-uniform sampling and its effects on band limited signals. Low pass filtering which is all that is required to reconstruct a band limited signal from its sample values taken regularly at the Nyquist rate, is
invariably insufficient if the sampling process was non-uniform. In this case, more sophisticated and complex filtering is required. However, because of the simplicity of low pass filtering it would be interesting to investigate the effects this would have on a non-uniformly sampled signal and whether it is possible to incorporate modifications in order to reduce some of the adverse effects. This was carried out, and though the investigation considered a particular form of non-uniform sampling only it is possible to make the following general points.

1) A necessary and generally sufficient condition for the exact recovery of a signal from its non-uniformly sampled data is that these data must occur at an average rate equal to or above the Nyquist rate. During the investigation and in what follows, it is assumed that this condition is fulfilled.

2) The use of a direct low pass filter to interpolate between samples which have been obtained by non-uniform sampling, would in all cases result in erroneous outputs differing from the original signal. The magnitude of the error involved depends largely on the sampling sequence and it is a function of the deviation of the sample distribution from the uniform case. This may well be a linear function of the variance of the sample distribution. It is also useful to mention that in all cases, the interpolation would not reproduce the correct values at the sampling instants.

3) The situation can be somewhat improved if the non-uniform data sequence is converted into a uniform one
immediately prior to the low pass filtering process as was suggested in the approximate recovery scheme. Although this results in the information contained in the relative time displacements being lost, it has the advantage of reproducing the original sample values, albeit at different instants to the sampling instants. Further, if the original signal is fairly correlated, then this scheme is capable of yielding an output highly correlated to the signal. Experiments using speech signals showed that the output of the approximate recovery scheme was far superior to the output obtained by direct low pass filtering in terms of speech quality and intelligibility. The superiority is expected to be higher for signals which are highly redundant, like speech.

4) Formulae (3.19), (3.23) & (3.43) derived in Chapter 3 for the mean square error, the cross correlation function and the signal/error ratio may be extended to situations where the non-uniform sampling process is not of the 'recurrent' type and has no periodicity at all. In this more general case \( M \) in the formulae assumes the value \( \infty \), and we need to specify an infinite number of the quantities \( \Delta_j \)'s which define the deviation of the sampling process from the uniform case. It would however, be more difficult to evaluate the resulting infinite summation and its limit as \( M \) tends to infinity.

5) The optimum value of the delay at which the maximum S/E ratio occurs, appears to be equal to the average time shift required to convert the non-uniform sequence into a uniform one and the maximum value of the S/E ratio appears
to be proportional to the variance of the time shifts. If these results can be extended to the general case, it would be possible to calculate the mean square error and other quantities without using equations (3.2.18), (3.3.19) & (3.4.20), thus avoiding the inconvenience of an infinite sum. This is a reasonable thing to do but care must be taken to treat results obtained in this manner as estimates only which may not be accurate, particularly if the sample distribution is uneven, having large gaps separating aggregates of sample points.

Returning to the security aspect, we note that this is a function of the number of possible sequences at one's disposal in the multiplexing process. It is also a function of the number of channels actually in use. To obtain the full security potential of this system requires the use of all channels simultaneously, transmitting noise if necessary. The security also increases if the sequence period is increased, i.e., by increasing M. Apart from the corresponding increase in the possible sequences resulting from this, there is the added but significant advantage that the periodic information relating to each channel, which can easily be obtained by a third party, wishing to decipher the system, would be occurring at too low a rate to be of any use. This is particularly important if the system is used to transmit speech signals, in view of their high redundancy.

One of the proposed applications of this system is to provide a secure communication link between mobile units such as police cars and a stationary base. For such an application it is desirable to have a system which allows
random access to channels. The design of such a system and its capability to avoid deciphering attempts are related intimately and can provide the basis for further and separate investigation.

The above investigation demonstrated both the possibility and the feasibility of a secure pulse modulation system and presented a simple method of achieving the increased security. In addition, the investigation dealt in some detail with the problem of non-uniform sampling, particularly with the problem of reconstruction. It is hoped that this contribution will be of some value in the development of a secure, practical system, and more generally, in assessing the practical significance this form of sampling may have in a given application.
If $f(t)$ is a band limited function, then using the sampling theorem

$$f(t) = \sum_{k=-\infty}^{\infty} f(kT) \sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)} \quad \cdots \quad (A.1)$$

provided

$$T \leq \frac{\pi}{\omega_0} \quad \cdots \quad (A.2)$$

If the Fourier transform of $f(t)$ is $F(\omega)$, then the transform of $f(t-a)$ equals $e^{-ja\omega} F(\omega)$, where $a$ is an arbitrary constant. Therefore $f(t-a)$ is band limited with the same bound $\omega_0$ as $f(t)$ and so $f(t-a)$ can be expressed as

$$f(t-a) = \sum_{k=-\infty}^{\infty} f(kT-a) \sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)} \quad \cdots \quad (A.3)$$

which is another form of the sampling theorem. Replacing $t-a$ by $t$, (A.3) becomes

$$f(t) = \sum_{k=-\infty}^{\infty} f(kT-a) \sin \frac{\omega_0(t+a-kT)}{\omega_0(t+a-kT)} \quad \cdots \quad (A.4)$$

As an example, $\sin \frac{\omega_0 T}{\omega_0 t}$ is in fact a band limited function with an upper cut-off radian frequency $\omega_0$, therefore if $f(t) = \sin \omega_0 t$ we obtain

$$\sin \frac{\omega_0 t}{\omega_0 T} = \sum_{k=-\infty}^{\infty} \sin \frac{\omega_0 (kT-a)}{\omega_0 (kT-a)} \sin \frac{\omega_0 (t+a-kT)}{\omega_0 (t+a-kT)} \quad \cdots \quad (A.5)$$

valid for all $t$.

Let $a=t_1$, $t+a=t_2$ so that $t=t_2-t_1$, we obtain the following identity:
\[ \sum_{k=-\infty}^{\infty} \frac{\sin \omega_0 (t_1 - kT)}{\omega_0 (t_1 - kT)} \frac{\sin \omega_0 (t_2 - kT)}{\omega_0 (t_2 - kT)} = \sin \frac{\omega_0 (t_1 - t_2)}{\omega_0 (t_1 - t_2)} \quad \cdots \quad (A.6) \]

valid for all \( t_1 \) and \( t_2 \).

If \( t_1 = t_2 = t \), then as a special case of (A.6)
\[ \sum_{k=-\infty}^{\infty} \frac{\sin^2 \omega_0 (t - kT)}{\omega_0^2 (t - kT)^2} = 1 \quad \cdots \quad \text{valid for all } t. \]

(A.6) is a very valuable identity and is used extensively in evaluating infinite series usually arising in sampling problems. It follows directly from the sampling theorem and therefore only holds true provided condition (A.2) is satisfied, i.e. \( T \leq \frac{\pi}{\omega_0} \).
APPENDIX B

The Mean and Variance of $g(t)$

Intuitively speaking, both the mean value and the variance of $g(t)$, the output obtained from the recovery scheme, are expected to be the same as those of $f(t)$ if this is assumed to be a stationary process. Nevertheless a more sophisticated and rigorous proof is needed and is given below.

Once again using the notations introduced in section 3.1. we have

$$g(t) = \sum_{k=\infty}^{\infty} f(kT + \Delta_k) \sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)} \ldots \ldots (B.1)$$

Taking the ensemble average of both sides of (B.1), the mean value is given by

$$\langle g(t) \rangle = \langle f(kT + \Delta_k) \rangle \sum_{k=\infty}^{\infty} \frac{\sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)}}{\omega_0(t-kT)} \ldots \ldots (B.2)$$

therefore

$$\langle g(t) \rangle = 0 \ldots \ldots \ldots (B.3)$$

since

$$\langle f(kT + \Delta_k) \rangle = \langle f(t) \rangle = 0$$

Because the mean value of $g(t)$ equals zero, the variance of $g(t)$ therefore equals the mean square value of $g(t)$, which is given by

$$\langle g^2(t) \rangle = \sum_{k=\infty}^{\infty} \sum_{n=\infty}^{\infty} f(kT + \Delta_k) f(nT + \Delta_n) \sin \frac{\omega_0(t-kT)}{\omega_0(t-kT)} \cdot \frac{\sin \frac{\omega_0(t-nT)}{\omega_0(t-nT)}}{\omega_0(t-nT)} \ldots \ldots (B.4)$$

The ensemble average needs only be performed on the product of the sample values since they are the only...
random variables. Therefore, since \( f(t) \) is a low pass process,

\[
\langle f(kT + \Delta k) f(nT + \Delta n) \rangle = \sigma^2 \frac{\sin (\omega_0(k-n)T + \Delta k - \Delta n)}{\omega_0(k-n)T + \Delta k - \Delta n} \quad \ldots \text{B.5}
\]

and

\[
\langle g^2(t) \rangle = \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{\sin \omega_0((k-n)T + \Delta k - \Delta n)}{\omega_0((k-n)T + \Delta k - \Delta n)} \cdot \frac{\sin \omega_0(t-kT)}{\omega_0(t-kT)} \cdot \frac{\sin \omega_0(t-nT)}{\omega_0(t-nT)} \quad \ldots \text{B.6}
\]

Replacing \( k-n \) by \( m \), so that the infinite summation over \( k \) is now replaced by one over \( m \) we obtain:

\[
\langle g^2(t) \rangle = \sigma^2 \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{\sin \omega_0(mT + \Delta m + n - \Delta n)}{\omega_0(mT + \Delta m + n - \Delta n)} \cdot \frac{\sin \omega_0(t-(m+n)T)}{\omega_0(t-(m+n)T)} \cdot \frac{\sin \omega_0(t-nT)}{\omega_0(t-nT)} \quad \ldots \text{B.7}
\]

The summation over \( n \) can now be split into two summations, an infinite summation over \( k \) and a finite summation of \( M \) terms over \( r \). Thus replacing \( n \) by \( Mk+r \) and making use of the group periodicity equation (3.2.7), we obtain

\[
\langle g^2(t) \rangle = \sigma^2 \sum_{m=-\infty}^{\infty} \sum_{r=1}^{M} \sum_{k=-\infty}^{\infty} \frac{\sin \omega_0(mT + \Delta r + m - \Delta r)}{\omega_0(mT + \Delta r + m - \Delta r)} \cdot \frac{\sin \omega_0(t-(m+r)T)}{\omega_0(t-(m+r)T)} \cdot \frac{\sin \omega_0(t-(Mk+r)T)}{\omega_0(t-(Mk+r)T)} \quad \ldots \text{B.8}
\]

Neither infinite summation can be evaluated with the
aid of identity A6 (see Appendix A). This is because the 
m summation includes a quantity $\Delta r_{m}$ which varies with the 
summing index $m$ and in the case of the $k$ summation, the 
effective sampling period is $MT$, thus violating condition 
A2.

To overcome this difficulty we make use of the 
stationarity property of $g(t)$, namely

$\langle g^{2}(t) \rangle = \langle g^{2}(t-a) \rangle \quad ... \quad B.9$

where $a$ is an arbitrary constant.

Thus, putting $a=jT$, where $j=1,2,..,M$, replacing $t$ in (B.8) 
by $t-jT$, adding $M$ such terms and rearranging, we obtain

$$M \left\langle g^{2}(t) \right\rangle = \mathcal{C}^{2} \sum_{m=-\infty}^{\infty} \sum_{r=1}^{M} \frac{\sin \omega_{0}(mt+\Delta_{m+r}-\Delta r)}{\omega_{0}(mt+\Delta_{m+r}-\Delta r)} \cdot$$

$$\cdot \sum_{j=1}^{M} \sum_{k=0}^{\infty} \frac{\sin \omega_{0}(t-mT-rT-(Mk+j)T)}{\omega_{0}(t-mT-rT-(Mk+j)T)} \cdot \sin \frac{\omega_{0}(t-rT-(Mk+j)T)}{\omega_{0}(t-rT-(Mk+j)T)} \quad ... \quad B.10$$

Now, replacing $Mk+j$ by $n$, so that the two summations 
over $k$ and $j$ are lumped into one summation over $n$, (B.10) 
becomes

$$M \left\langle g^{2}(t) \right\rangle = \mathcal{C}^{2} \sum_{m=-\infty}^{\infty} \sum_{r=1}^{M} \frac{\sin \omega_{0}(mt+\Delta_{m+r}-\Delta r)}{\omega_{0}(mt+\Delta_{m+r}-\Delta r)} \cdot$$

$$\cdot \sum_{n=0}^{\infty} \frac{\sin \omega_{0}(t-mT-rT-nT)}{\omega_{0}(t-mT-rT-nT)} \cdot \frac{\sin \omega_{0}(t-rT-nT)}{\omega_{0}(t-rT-nT)} \quad ... \quad B.11$$

The identity A6 can now be used to evaluate the summation 
over $n$ and so by letting $t_{1}=t-mT-rT$, $t_{2}=t-rT$, (B.11) becomes:
\[ M \left\langle g^2(t) \right\rangle = \sigma^2 \sum_{m=-\infty}^{\infty} \sum_{r=1}^{M} \frac{\sin \omega_0(mT+\Delta m+r-\Delta r)}{\omega_0(mT+\Delta m+r-\Delta r)} \cdot \frac{\sin \omega_0(mT)}{\omega_0(mT)} \ldots \text{B.12} \]

Since however,
\[ \frac{\sin \omega_0(mT)}{\omega_0(mT)} = \begin{cases} 1 & m=0 \\ 0 & m \neq 0 \end{cases} \]

the summation over \( m \) is simply the value of B.12, when \( m=0 \), therefore the mean square value of \( g(t) \) is given by

\[ \left\langle g^2(t) \right\rangle = \sigma^2 \sum_{r=1}^{M} \frac{\sin \omega_0(\Delta r - \Delta r)}{\omega_0(\Delta r - \Delta r)} = \sigma^2 \ldots \text{B.13} \]

The variance of \( g(t) \) is thus \( \sigma^2 \), being the same as that of \( f(t) \). The recovery scheme has therefore preserved the two basic statistical properties of the signal, namely the mean value and the variance.
PART II
CHAPTER ONE.

Introduction.

One of the fundamental problems encountered in the theory of continuous stochastic processes is that concerning the time instants at which a random function \( f(t) \), crosses a certain threshold level \( h \). The randomness of \( f(t) \) means that the crossing instants are random variables, which can only be expressed in probabilistic terms. In such studies we are concerned with finding the probability distributions of the number of crossings, in a given interval of time \( T \), and/or the distribution of intervals between successive crossings.

This problem finds application in fields such as statistical communication theory, oceanography and control engineering. In the first field, the problem may arise in the detection of radar or sonar signals where the information contained in the zero crossing instants of the received waveform can be used to detect the presence of signals in noise. The important parameter here is the interval between successive zero crossings, which is related to the instantaneous frequency of the waveform. The distribution of this interval is strongly influenced by the signal to noise ratio of the received waveform, and can thus be used to detect the presence or absence of targets. Further, since both false alarm and detection probabilities are intimately related to the distribution of threshold crossings, the knowledge of this distribution can aid a great deal in the development of detection methods aimed at improving the probability of target
detection and minimising false alarm errors.

In Oceanography, the random function \( f(t) \) may represent the height of sea surface above a fixed level. Here it is of interest to find the statistical distributions of the points at which this height exceeds a given level.

In Control Engineering, the problem may arise when a controlling signal \( s(t) \), initially negative, triggers a certain action or process on reaching the zero level for the first time. If the signal is corrupted by noise (additive), the initiating instant is no longer deterministic but becomes a random variable, whose probability distribution is required in estimating the effects of noise on the controlling process. This distribution is equivalent to that of the noise waveform first intersecting the level \( -s(t) \).

The problem also arises whenever a data transmission system codes the information to be transmitted, as a frequency shift of a carrier. For example, in the transmission of binary data over telephone lines, by frequency modulation, where upper and lower frequencies are used to represent the two binary states. The receiver computes the number of zero level crossings by the received waveform, during a time interval equal to the pulse width, and compares the obtained number with a predetermined threshold. Based on this comparison, a decision is made on whether the upper or lower frequency was transmitted. The presence of noise causes the number of crossings to become a random variable whose probability distribution, corresponding to the lower frequency, overlaps that corresponding to the upper
frequency. This gives rise to wrong decisions at the receiver. In order to determine the error rate of such a system, it is necessary to know the probability distribution of the zero level crossings of the combined signal plus noise in both cases.

The zero crossing problem is a special case of the threshold crossing problem, in which the threshold level, h, assumes the zero value, which is normally taken to be the mean value of the function f(t). It is easier to deal with this problem than the more general one. Even so, the zero crossing problem remains one of the most challenging problems of the theory of noise today. S.O. Rice, in his well-known paper (1944) introduced it and discussed some of the mathematical difficulties arising in the course of its solution. He obtained a related probability distribution, which for a certain class of random functions, can be used as an approximation to the distribution of intervals between successive crossings, providing these are sufficiently small. Since then, many mathematicians and scientists have investigated this problem, and their attempts at finding satisfactory solutions or approximations are widely reported in the extensive literature on the subject. Most of these attempts, however, succeeded in obtaining approximations which are valid only over a limited range of interval lengths and for a limited class of random functions, which were generally assumed to be Gaussian. Exact solutions at this stage seem to be prohibitive, and approximations valid over the entire range of interval lengths and for all classes of random functions are, to the best knowledge of the author, non-existent. The difficulty
appears to be mainly due to mathematical and computational difficulties arising in the course of finding a solution which are problematic at this stage.

In view of the importance of the zero crossing problem it was thought worthwhile to conduct a further investigation in an attempt to find a satisfactory solution. The investigation would survey the attempts and approaches advanced by those who have worked in this field, which have succeeded to a large extent in clarifying the theoretical aspects of the problem and the difficulties arising in its solution. This survey then provides the basis for formulating new ideas and devising new techniques aimed at finding a satisfactory solution or improving the already existing ones.

After surveying some of the contributions reported in the literature on the subject in Chapter two, an approach based on the multivariate normal integral is discussed in detail in Chapter three. The theory of Markov processes is used in Chapter four to construct a Markov model which provides a framework for obtaining further approximations as well as the possibility of exact solutions for a certain class of Gaussian random functions. Chapter five deals with a practical system which was used to obtain direct measurements of the distribution of zero crossing intervals. In Chapter six, we present some results for random functions having different power spectra and discuss in detail their implications, particularly the relationship between the signal power spectra and its zero crossing distribution. The study is finally concluded in Chapter seven.
CHAPTER TWO.

Literature Survey.

2.1 Introduction.

In view of its importance and the fact that its solution is required for many Engineering Problems, the zero-crossing problem has merited the attention of many engineers, mathematicians and scientists, S.O. Rice; J.A. McFadden, and M.S. Longuet-Higgins are well-known for their significant contributions on the subject. Other useful contributions have been made by R.L. Stratonovich V.I. Tikhanov, and P.I. Kuznestov; A.J. Rainal; D. Slepian and A.J. Siegert.

In this chapter, a survey of the most significant of these contributions will be presented and their various approaches are discussed and analysed. Because certain of these contributions are directly related to, and form the basis of the material reserved for later chapters, only a brief mention will be made here. Wherever possible, such things as proofs of formula and other lengthy pieces of analysis which the reader can find in the original material are omitted. The aim of this chapter is not to provide simply a summary of these contributions, but rather to present a critical review of the ideas and approaches contained in them and the extent to which they have succeeded in providing a solution to the zero-crossing problem.

2.2 Expected Number of Zeros per Unit Time.

In a random process, the instants at which zero
crossings occur are random. In a stationary process, the
probability of a crossing in an infinitesimal interval, $\Delta t$
is independent of the actual value of time $t$. So. Rice in
his well-known paper \(^{(1)}\) evaluated this probability for a
stationary Gaussian Process and showed it given by:

$$P(0, \Delta t) = \frac{\Delta t}{\pi} \left[ \frac{-\gamma''(0)}{\gamma(0)} \right]^{\frac{1}{2}} \ldots \ldots \ (2.2.1)$$

where $\gamma(\tau)$ is the auto-correlation function of the
random process and

$$\gamma''(\tau) = \frac{d^2 \gamma(\tau)}{d \tau^2}$$

The number of crossings per unit time is obtained by
integration over 1 unit. This gives:

$$N_0 = \frac{1}{\pi} \left[ \frac{-\gamma''(0)}{\gamma(0)} \right]^{\frac{1}{2}} \ldots \ldots \ (2.2.2)$$

On average these crossings are equally divided between
upward (those with negative slopes) and downward (those
with negative slopes) zeros.

For processes whose power spectra are of the form

$$W(f) = \frac{1}{f^2 + a^2} \ldots \ldots \ (2.2.3)$$

The second derivative $\gamma''(\tau)$ evaluated at zero is $\infty$
and this leads to $N_0 = \infty$. These processes form the class
of first-order Markovian processes whose axis crossing
behaviour is discussed later in the chapter.

2.3 The statistical distribution of Zeros.

For the class of random processes whose power spectra
not of the form (2.2.3), the average zero crossing rate is given by expression (2.2.2). This however, is not sufficient to completely determine the statistical behaviour of these crossings. In order to do that one needs, in addition, to know the following probability distributions.

a) $P_n(\zeta)$; the probability density that the interval between a given zero crossing and the $(n+1)$th succeeding one has a length that lies in the range $\zeta, \zeta + d\zeta$.

b) $P(n,\zeta)$; the probability that a given time interval of length $\zeta$ contains exactly $n$ zero crossings.

The evaluation of these distributions is generally very difficult and to the best of the author's knowledge, no-one has yet given a satisfactory solution. $P_0(\zeta)$ which is the probability density of the intervals between successive zeros, should be the easiest distribution to determine in view of the fact that correlation between successive intervals need not necessarily be specified, (in the case of $P_n(\zeta)$; for $n > 0$ it is invariably necessary), yet no satisfactory solution for this case has been found. Rice (1) derived an expression for a probability distribution $Q(\zeta)$ which is related to $P_0(\zeta)$ and could serve as a good approximation in the region of small $\zeta$. The distribution $Q(\zeta)$ is the probability density of a Gaussian random function $f(t)$ passing through zero in the interval $\zeta, \zeta + d\zeta$ with a negative slope when it is known that it has passed through zero at $\zeta = 0$ with a positive slope. It is given by

$$Q(\zeta) = \frac{1}{2\pi} \left[ \frac{\psi_0}{-\psi_0} \right]^{-\frac{1}{2}} \frac{M_{25}}{H} \left( \psi_0^2 - \psi^2 \zeta^2 \right)^{3/2} \left( 1 + H \cot^{-1}(-H) \right)$$

... (2.3.1)
where $M_{22}$ and $M_{23}$ are the co-factors of $M_{22}' (= -\psi^0_0)$ and $M_{23}' (= -\psi^0_\zeta)$ in the correlation matrix:

$$M = \begin{bmatrix}
\tau_0 & 0 & \psi_\zeta & \psi_\zeta \\
0 & -\psi^0_0 & -\psi^0_\zeta & -\psi^0_\zeta \\
\psi^0_\zeta & -\psi^0_\zeta & -\psi^0_0 & 0 \\
-\psi^0_\zeta & -\psi^0_\zeta & 0 & \tau_0
\end{bmatrix} \ldots (2.3.2)$$

and $H = M_{23} \left[ M_{22}^2 - M_{23}^2 \right]^{-\frac{1}{2}} \ldots (2.3.3)$

We choose $0 < \text{Cot}^{-1}(-H) \leq \frac{\pi}{2}$, the value $\frac{\pi}{2}$ being taken at $\zeta = 0$ and the value $\frac{\pi}{2}$ being approached as $\zeta \to 0$. The arguments of the correlation functions are written as subscripts.

Another related distribution can be obtained by removing the restrictions on the slope of $f(t)$ when it goes through zero in $\zeta$, $\zeta + d\zeta$: positive as well as negative slopes are considered and this leads to the distribution $U(\zeta)$ which is the probability density of $f(t)$ passing through zero in $\zeta$, $\zeta + d\zeta$, given that $f(t)$ has passed through zero at $\zeta = 0$ with a positive slope. $U(\zeta)$ is given by:

$$U(\zeta) = \frac{1}{2\pi} \left[ \frac{\psi_0}{-\psi^0_0} \right]^{\frac{3}{2}} \left( M_{22}^2 - M_{23}^2 \right)^{\frac{1}{2}} \left( \psi^2_0 - \psi^2_\zeta \right)^{-\frac{3}{2}} \cdot (1 + H \tan^{-1}(H)) \ldots (2.3.4)$$

For small values of $\zeta$ both of these distributions approximate well to the distribution $P_0(\zeta)$, since a zero crossing occurring only a short time after the one at $\zeta = 0$ is most likely to be the next crossing.
Further, in view of the fact that the first zero crossing following an upward crossing must be downward, it is logical to expect the distribution \( Q(\tau) \) to be a better approximation than \( U(\tau) \), since in deriving the latter no consideration of the slope of the zero at \( \tau, \tau + \Delta \tau \) was made. This may also be proved by considering the following series relationships:

\[
U(\tau) = \sum_{n=0}^{\infty} P_n(\tau) = P_0(\tau) + P_1(\tau) + P_2(\tau) + \ldots \\
Q(\tau) = \sum_{n=0}^{\infty} P_{2n}(\tau) = P_0(\tau) + P_2(\tau) + P_4(\tau) + \ldots
\]

\[\text{...(2.3.5)}\]

\[\text{...(2.3.6)}\]

The interpretation is relatively simple. Given an upward zero at \( \tau = 0 \), then if there is a downward zero at \( \tau, \tau + \Delta \tau \), then it must be the first, the third, or in general, the \((2n+1)\)th zero. Hence series \( (2.3.6) \). Similarly, using the same argument and removing the restriction on the slope of the crossing at \( \tau, \tau + \Delta \tau \) we arrive at the series \( (2.3.5) \).

By close examination of \( (2.3.5) \) & \( (2.3.6) \) it is clear that

\[
U(\tau) \geq Q(\tau) \geq P_0(\tau), \text{ for all } \tau \\
\text{...(2.3.7)}
\]

The equality sign only strictly applies when \( \tau = 0 \). Clearly this means that both \( U(\tau) \) and \( Q(\tau) \) form upper limits on \( P_0(\tau) \), and that \( Q(\tau) \) is the better approximation of the two. As \( \tau \) becomes larger and larger, the behaviour of \( f(t) \) is influenced less and less by the fact that it has a zero crossing at \( \tau = 0 \). Hence both \( U(\tau) \) and \( Q(\tau) \) are expected to approach the probability densities \( P(0, \Delta t) \) and \( P(0, \Delta t) \) where \( P(0, \Delta t) \) is given by \( (2.2.1) \) and:

\[
P(0, \Delta t) = P_+(0, \Delta t) = \frac{1}{2} P(0, \Delta t) \ldots \ (2.3.8)
\]

These results are true because as \( \tau \) approaches \( \infty \) the matrix \( \mathbb{H} \) approaches a diagonal matrix and both \( M_{23} \) and \( \mathbb{H} \) approach zero with the quantities \( \frac{M_{23}}{\mathbb{H}} \) and \( M_{22} \) approaching \( -\psi_0^2 \psi_0' \), hence:
\[ \lim_{\tau \to \infty} Q(\tau) = \frac{1}{2\pi} \left[ \frac{-\psi_0}{\psi_0} \right]^{1/2} \quad \ldots (2.3.9a) \]

and

\[ \lim_{\tau \to \infty} U(\tau) = \frac{1}{\pi} \left[ \frac{-\psi_0}{\psi_0} \right]^{1/2} \quad \ldots (2.3.9b) \]

Rice evaluated both \( U(\tau) \) and \( Q(\tau) \) for an ideal low pass spectrum cutting off at \( f_b \) Hz. Fig.(1) shows both quantities plotted against the normalised time \( \phi = 2\pi f_b \tau \) over the range \( \phi = 0-9 \). Beyond this range \( Q(\phi) \) oscillates with an ever decreasing amplitude about the constant level 0.0919, whereas \( U(\phi) \) does the same about the level 0.1838. These being the values of the limits (2.3.9a) and (2.3.9b) respectively. The figure shows that over the range (0-3.5) both curves are almost identical, which by considering series (2.3.5) & (2.3.6) clearly means that the probability distributions \( P_n(\phi) \) where \( n > 0 \) must have negligible values compared to that of \( P_0(\phi) \). Over this range both \( U(\phi) \) and \( Q(\phi) \) are considered to be very close to \( P_0(\phi) \). Beyond \( \phi = 3.5 \), \( Q(\phi) \) continues to be the better approximation of the two, but ceases to be reliable when \( P_2(\phi) \), the second term in series (2.3.6), starts to assume a significant value comparable to that of \( P_0(\phi) \).

For large and intermediate values of \( \tau \) Rice pointed out that it seems especially difficult to obtain an expression for \( P_0(\tau) \). One method suggested by Gouldsmit is to amend the conditions leading to (2.3.1) by adding conditions that \( f(t) \) be positive at equally spaced points in the interval \( \tau \). However, this leads to multiple integrals which are hard to evaluate. It was found that for one point between 0 and \( \tau \), the integral is of the form:
A = U(\Phi)
B = Q(\Phi)
\circ = Experimental Points

Fig. (1) - (Taken from Rice's Paper)

Distribution of Intervals between Zeros - Low Pass Filter.
Another method of approach is to use the method of 'in and exclusion' of zeros, between 0 and \(\tau\), which is the subject of the next section.

2.4 The In and Exclusion Method

Consider the ensemble of curves \(f(t)\), having an upward zero at \(\tau = 0\). Then in theory at least it is possible to compute the following probability distribution functions.

i) \(U(\tau)\) as defined above.

ii) \(U_2(r, \tau)\) \(dr\) \(d\tau\); the joint conditional probability of \(f(t)\) having zero crossings in \(r, r+dr\) and \(\tau, \tau+d\tau\).

iii) \(U_3(s, \tau, \tau)\) \(dr\) \(ds\) \(d\tau\); the joint conditional probability of \(f(t)\) having zero crossings in \(s, s+ds; \tau, \tau+d\tau\) and \(\tau+\tau+d\tau\) etc. The method of inclusion and exclusion leads to the following equation for \(P_0(\tau)\)

\[
P_0(\tau) = U(\tau) - \int_0^\tau U_2(r, \tau) \, dr + \int_0^\tau U_3(s, \tau, \tau) \, ds \, dr - \int_0^\tau \int_0^\tau \int_0^\tau U_4(t, s, \tau, \tau) \, dt \, ds \, dr + \ldots \quad \ldots(2.4.1)
\]

Another way of writing it is:

\[
P_0(\tau) = U(\tau) - \int_0^\tau U_2(r, \tau) \, dr + \int_0^\tau U_3(s, \tau, \tau) \, ds - \int_0^\tau \int_0^\tau \int_0^s U_4(t, s, \tau, \tau) \, dt \, ds + \ldots \quad \ldots(2.4.1a)
\]

By restricting the slopes of \(f(t)\) at the crossing points...
\( t, s, p, \tau \) etc, to negative values, so that only negative values are considered; similar equations to (2.4.1) & (2.4.1a) are obtained:

\[
P_0(\tau) = Q(\tau) - \int_0^\tau Q_2(r, t) \, dr + \frac{1}{2} \int_0^\tau \int_0^\tau Q_3(s, r, \tau) \, dr \, ds - \frac{1}{3} \int_0^\tau \int_0^\tau \int_0^\tau Q_4(t, s, r, \tau) \, dr \, ds \, dt + \ldots \quad \ldots (2.4.2)
\]

and

\[
P_0(\tau) = Q(\tau) - \int_0^\tau Q_2(r, \tau) \, dr + \int_0^\tau dr \int_0^\tau Q_3(s, r, \tau) \, ds \ldots \quad (2.4.2)
\]

Here \( Q, Q_2 Q_3 \) etc are defined similarly to \( U, U_2 U_3 \) etc, with the rider that the zero crossings at \( t, s, p, \tau \) etc are all downward. Rice gives equation (2.4.1) but does not derive it. It was thought useful to do so. A possible method of derivation is given in Appendix (1).

The in and exclusion method equations lead to difficult integrals. In order to appreciate the mathematical problems involved, the following points are worth mentioning:

1) In deriving expression (2.3.1) for \( Q(\tau) \) Rice started from the expression

\[
Q(\tau) \propto \int \int d\eta_1 \int_0^\infty d\eta_2 \, P(0, \eta_1; 0, \eta_2) \ldots \quad (2.4.3)
\]

where \( P(\lambda, \eta_1; \lambda_2, \eta_2) \) is the joint probability density function for the four Gaussian random variables.

\[
f(0) = \lambda_1 = 0
\]

\[
f(\tau) = \lambda_2 = 0
\]

\[
f(0) = \eta_1
\]

\[
f(\tau) = \eta_2
\]

\[
\ldots \ldots \quad (2.4.4)
\]
2) It follows that the integral in the second term of the inclusion equation is itself a triple integral whose integrand is a six-dimensional Gaussian probability distribution. The multiplicity of integrals as well as their integrands increases as we consider higher and higher terms.

3) As will be made clear in a later chapter, the evaluation of integrals of multi-dimensional Gaussian probability distribution over one quadrant of the n-dimensional space seems only possible for \( n \leq 3 \). This means that at least for the time being we cannot express in a closed form, more than the first two terms of these series. This would have been satisfactory if these series converge quickly so that higher terms need not be considered. Unfortunately, this is not the case and although series (2.4.2) is more rapidly convergent than the corresponding series (2.4.1), for the reasons mentioned in the previous section, both series with only two terms considered fail to give reliable results for large \( \tau \). As will become clear, this failure manifests itself by yielding negative probabilities.

2.5 Series Expansions for \( P_n(\tau) \).

In order to see how rapidly (2.4.1) & (2.4.2) converge it is useful to consider some further series expansions involving the distributions \( P_n(\tau) \). Using these series, each term of the in and exclusion series (2.4.1) and (2.4.2) may then be expressed in terms of \( P_n(\tau) \). As shown below this will be useful in estimating the order of magnitude of the error arising as a result of truncating the in and exclusion series. The truncation error can then be used to
compare the rapidity of convergence of (2.4.1) with (2.4.2) and also to estimate the range of interval lengths over which this error is negligible.

M.S. Longuet-Higgins (2) showed that in addition to the two series (2.3.5) and (2.3.6), it is possible to obtain further series relationships expressing higher order probability distributions of the type \( U_2, U_3 \ldots \) etc, and \( Q_2, Q_3 \ldots \) etc in terms of \( P_n(\zeta) \). He showed that:

\[
\int_0^{t_n} \int_0^{t_{n-1}} \cdots \int_0^{t_2} Q_n(t_1, t_2, \ldots, t_n) \, dt_1 \cdots dt_{n-1} = \sum_{r=0}^{\infty} \left( \frac{n-1+r}{r} \right) P_{2n+2r-2}(\zeta) \quad \cdots \quad (2.5.1)
\]

and

\[
\int_0^{t_n} \int_0^{t_{n-1}} \cdots \int_0^{t_2} U_n(t_1, t_2, \ldots, t_n) \, dt_1 \cdots dt_{n-1} = \sum_{r=0}^{\infty} \left( \frac{n-1+r}{r} \right) P_{n+r-1}(\zeta) \quad \cdots \quad (2.5.2)
\]

where \( t_n > t_{n-1} > \cdots > 0 \) and \( \binom{P}{q} \) denotes the coefficient of \( x^q \) in the binomial expansion \((1+x)^P\). \( U_n \) and \( Q_n \), where \( n > 1 \) are the same as those contained in the in and exclusion series. Letting \( n = 1 \) in both (2.5.1) & (2.5.2) gives the series (2.3.5) and (2.3.6) respectively.

Footnote *

Longuet-Higgins actually used different notations to ours. Under his notations the probability distribution \( Q_n \) is equivalent to \( \frac{W(+, -, \ldots, -)}{W(+)} \) and \( U_n \) is the same as \( \frac{W(+, \pm, \pm \ldots \pm)}{W(+)} \) where the number of signs in the argument of \( W \) is \( (n+1) \) in both cases.
Denoting the L.H.S. of equations (2.5.1) & (2.5.2) by $X_n$ and $Y_n$ respectively we have:

\[ X_n = \sum_{r=0}^{\infty} \frac{(n-1+r)}{r} P_{2n+2r-2} \quad \ldots \quad (2.5.3) \]

\[ Y_n = \sum_{r=0}^{\infty} \frac{(n-1+r)}{r} P_{n+r-1} \quad \ldots \quad (2.5.4) \]

Using these notations the in and exclusion equations (2.4.2) & (2.4.1) may be written as:

\[ P_0(\tau) = X_1 - X_2 + X_3 - X_4 \quad \ldots \quad (2.5.5) \]

and

\[ P_0(\tau') = Y_1 - Y_2 + Y_3 - Y_4 \quad \ldots \quad (2.5.6) \]

By considering the first few terms of (2.5.5) we can obtain successive approximations to the distribution $P_0(\tau)$.

Denoting the truncated series of $N$ terms by $P_0^{(N)}(\tau)$ and using the relation (2.5.3) we obtain the following successive approximations:

1) $P_0^{(1)}(\tau) = X_1 = P_0(\tau) + (P_2(\tau) + P_4(\tau) \ldots \ldots) \quad (2.5.7a)$
2) $P_0^{(2)}(\tau) = X_1 - X_2 = P_0(\tau) - (P_4(\tau) + 2P_6(\tau) \ldots \ldots) \quad (2.5.7b)$
3) $P_0^{(3)}(\tau) = X_1 - X_2 + X_3 = P_0(\tau) + (P_6(\tau) + 3P_8(\tau) \ldots \ldots) \quad (2.5.7c)$

In general:

\[ P_0^{(N)}(\tau) = X_1 - X_2 \ldots + (-1)^{N-1}X_N = P_0(\tau) + (-1)^{N-1}(P_{2N}(\tau) + NP_{2N+2}(\tau) \ldots \ldots) \quad (2.5.8) \]

The above list of approximations shows that the order of magnitude of the error in the truncated series becomes that of the probability distribution $P_{2N}(\tau)$. This distribution has an average value which is $(2N+1)$ times that of the distribution $P_0(\tau)$. Although the shape of the distribution $P_{2N}(\tau)$ varies with respect to $N$, it can generally be said
that this distribution is negligible over the range of interval lengths which are reasonably small compared to its mean value. This is true in the case of the low pass spectrum considered by Rice and shown in Fig (1), which shows that the distributions \( P_n(\phi) \) where \( n > 0 \) are negligible over the range \( (\phi = 0-3.5) \). As \( \phi \) reaches the value 3.5, \( P_1(\phi) \) starts to become significant and this causes \( U(\phi) \) to depart from \( Q(\phi) \). The latter continues to be the better approximation until \( P_2(\phi) \) starts to assume an appreciable value at which point \( Q(\phi) \) ceases to be reliable. It is worth mentioning here that \( Q(\phi) \) is the same as \( P_0(\tau) \) and that \( P_2(\phi) \) represents the order of magnitude of the truncation error involved.

Increasing \( N \) corresponds to increasing the number of terms in the truncated series. This causes \( P_0(\tau) \) to become a truer representation of \( P_0(\tau) \) extending over a wider range. This range is approximately given by that over which \( P_{2N}(\tau) \) has a negligible value compared to \( P_0(\tau) \).

Next by considering the first \( N \) terms of series (2.5.6) and using the relation (2.5.4), we obtain the following expression for the truncated sum \( S_N \)

\[
S_N = P_0(\tau) + (-1)^N \left[ P_N(\tau) + N P_{N+1}(\tau) \ldots \right] \quad (2.5.9)
\]

This shows that the order of magnitude of the error involved in truncating the series is that of the probability distribution \( P_N(\tau) \). Comparing this error with that obtained using the truncated series (2.5.8), it is clear that twice as many terms would be needed if series (2.5.6)
was used instead of (2.5.5) for the same order of truncation error magnitude. This means that the in and exclusion equation (2.4.2) is twice more rapidly convergent than the equation (2.4.1). In both cases however, many terms are required in order to obtain a satisfactory approximation which extends over a wide range of interval lengths.

The successive approximations (2.5.7a,b,c) can easily be seen to alternate between upper and lower limits on the true distribution \( P_0(\tau) \) and get progressively closer to it. Indeed, if \( S_N \) denotes the sum of the first \( N \) terms in any inclusion and exclusion equation, then we obtain the following inequalities:

\[
\begin{align*}
S_1 &> S_3 \cdots > S_k > P_0(\tau) > S_{k-1} \cdots > S_2 \quad \cdots (2.5.10)
\end{align*}
\]

where \( k = N \) or \( N - 1 \) depending on whether \( N \) is odd or even.

As \( N \) tends to \( \infty \), \( S_k \) approaches \( P_0(\tau) \) from above, whereas \( S_{k-1} \) approaches \( P_0(\tau) \) from below. Since \( P_0(\tau) \) must tend to zero as \( \tau \) gets larger, it follows that all the lower limits \( S_2, S_4 \ldots S_{k-1} \) must have negative values over that range.

Finally, by manipulating expressions (2.5.3) and (2.5.4) and using the binomial coefficients properties we can obtain series expressions for the distributions \( P_n(\tau) \) in terms of the functions \( X_n \) and \( Y_n \). These may be thought of as the inverse series of (2.5.3) and (2.5.4) and represent the generalisations of the in and exclusion equations (2.4.1) and (2.4.2). The expressions are:

\[
P_{2n}(\tau) = \sum_{r=0}^{\infty} (-1)^r \binom{n+r}{r} X_{n+1+r} \quad \cdots (2.5.11)
\]

and

\[
P_n(\tau) = \sum_{r=0}^{\infty} (-1)^r \binom{n+r}{r} Y_{n+1+r} \quad \cdots (2.5.12)
\]
which for \( n=0 \) give the series (2.5.5) & (2.5.6) respectively.

2.6 The Asymptotic behaviour of \( P_n(\tau) \).

As previously pointed out, the problem of determining \( P_n(\tau) \) is generally a very difficult one. But some asymptotic information is available and is presented here. Before doing so it is necessary to differentiate between two classes of random functions.

1) The regular case which includes functions whose auto-correlation functions can be expressed in Power series of the form:

\[
\Psi(\tau) = \psi(0) + \frac{\psi''(0)}{2} \tau^2 + \frac{\psi^{(4)}(0)}{4} \tau^4 + \cdots \tag{2.6.1}
\]

i.e. the function and its derivatives are all continuous at the origin.

2) A singular class which includes functions whose auto-correlation functions can be expressed in the form

\[
\Psi(\tau) = \psi(0) + \frac{\psi''(0)}{2} \tau^2 + \frac{\psi^{(4)}(0)}{3} |\tau|^3 + \cdots \tag{2.6.2}
\]

In other words, the third derivative of \( \Psi(\tau) \) possesses a finite discontinuity at the origin. Such a case occurs whenever the power spectrum of \( f(t) \) is of the order \((\text{frequency})^{-4}\) at high frequencies. This case is not to be mistaken for the Markovian process with power spectrum given by (2.2.3). The existence of \( \psi''(0) \) in (2.6.2) ensures a finite average number of crossings per unit time.

In future work, whenever reference is made to the 'regular case' and the 'singular case' this should be understood in the above context.

2.6.1 Assymptotic behaviour at small interval lengths.

Longuet-Higgins (2) showed that if \( \Psi(\tau) \) is regular
at the origin then if \( t \) is small, \( P_n(t) \) is of the order of \( t^{m-2} \), where
\[
m = \frac{1}{2} (n+2) (n+3)
\]
in other words
\[
P_n(t) = O(t^{m-2})
\]
As \( m \) increases \( P_n(t) \) becomes proportional to a rapidly increasing power of \( t \), for example
\[
P_0(t) \propto t
\]
\[
P_1(t) \propto t^4
\]
\[
P_2(t) \propto t^8
\]
\[
P_3(t) \propto t^{13}
\]
... etc

2.6.1.1 The Singular Case.

In this case all the distributions \( P_n(t) \) approach non-zero values as \( t \) tends to zero. Because of this the truncated in and exclusion series fails to even give the values of \( P_0(t) \) or \( P_1(t) \) at the origin. However, the truncated series yields the following upper and lower limits which were given by Longuet-Higgins\(^2\).
\[
1.147 \alpha < P_0(0) < 1.218 \alpha \quad \ldots (2.6.6)
\]
\[
0.193 \alpha < P_1(0) < 0.218 \alpha \quad \ldots (2.6.7)
\]
where
\[
\alpha = \frac{\psi''_+(0)}{-6 \psi''(0)} \quad \ldots (2.6.8)
\]
and \( \psi''_+(0) \) is the third derivative of the auto-correlation function evaluated at \( \tilde{t} = 0^+ \).

The above results show that in the regular case the distribution \( P_n(t) \) have zero values at the origin. Further the relationship (2.6.5) seem to justify the assumption
that if \( n > 0 \), then \( P_n(\tau) \) are negligible compared to \( P_0(\tau) \) over the range of small intervals. The order of magnitude of \( P_n(\tau) \) diminishes rapidly as \( n \) increases. Hence it is possible for even the first term of the in and exclusion series to represent closely the distribution \( P_0(\tau) \) over this range. In contrast to this we find that in the singular case all the distributions \( P_n(\tau) \) have non zero values at the origin, and this makes the truncated in and exclusion series unable to represent \( P_0(\tau) \) even at the origin.

Recently, Longuet-Higgins, in a private correspondence,* pointed out that E.Wong proved that \( P_0(0) \) in the singular case is exactly given by

\[
P_0(0) = 1.15625 \times \ldots (2.6.9)
\]

and that Wong has also evaluated \( P_n(0) \) when \( n > 0 \). However, it is not clear whether the proof has yet been published.

2.7 Behaviour of \( P_0^{(N)}(\tau) \) at large \( \tau \).

As defined above, \( P_0^{(N)}(\tau) \) denotes the partial sum of the first \( N \) terms of the in and exclusion series. As \( N \) increases, these partial sums alternately represent upper and lower limits on the true distribution \( P_0(\tau) \) as shown in the relation (2.5.12). In the regular case the first two partial sums \( P_0^{(1)}(\tau) \) and \( P_0^{(2)}(\tau) \) represent close approximations in the vicinity of \( \tau = 0 \), but are not so useful in the singular case. Here we are concerned about the order of magnitude of errors when \( \tau \) is large. The following results apply for both cases.

1) \( P_0^{(n)}(\tau) = X_1 \)

As explained earlier, \( X_1 \) is the same as \( Q(\tau) \). As \( \tau \) gets large \( Q(\tau) \) tends to \( P-(0, \Delta t) \) which is the prob-

* According to Higgins this proof is in SIAM journal 1966.
ability density of \( f(t) \) having a downward zero crossing in the infinitesimal interval \( \Delta t \). The distribution \( P_0(\zeta) \) must however tend to zero in order that its integral shall converge to unity. Hence the error in the approximation \( P_0^{(1)}(\zeta) \) when \( \zeta \) is large is given by

\[
E_1 = \frac{1}{2\pi} \left[ -\frac{\int_0^\infty g_0(z) dz}{\sqrt{g_0(0)}} \right]^2
\]

which is the value of \( P_0(0, \Delta t) \).

ii) \( P_0^{(2)}(\zeta) = X_1 - X_2 \)

In order to evaluate the error \( E_2 \) the above is written (see Section 2.5) in the form:

\[
P_0^{(2)}(\zeta) = Q(\zeta) - \int_0^\zeta Q_2(r, \zeta) \, dr \quad \text{(2.7.2)}
\]

As \( \zeta \to \infty \), we have

\[
Q(\zeta) \to P_-(0, \Delta t) \quad \text{and} \quad Q_2(r, \zeta) \to P_2(0, \Delta t), \text{ hence}
\]

\[
P_0^{(2)}(\zeta) \to P_-(0, \Delta t) - P_2(0, \Delta t) \cdot \zeta \quad \text{(2.7.3)}
\]

and

\[
E_2 \to - P_2(0, \Delta t) \cdot \zeta \quad \text{(2.7.4)}
\]

Since the second term is the dominant one. The negative limit is expected in view of the fact that \( P_0^{(2)}(\zeta) \) is a lower limit. This means that although this is a better approximation than \( P_0^{(1)}(\zeta) \) over small and possibly intermediate values of \( \zeta \), it nevertheless fails radically at large \( \zeta \). This failure manifests itself in yielding negative probabilities. The linear behaviour of \( P_0^{(2)}(\zeta) \) at large values of \( \zeta \) means that the probability integral \( \int_0^{\zeta} P_0^{(2)}(\zeta) \, d\zeta \), is divergent and negative thus violating the two basic laws of probability theory.

2.8 Approximations over the range of large intervals.

The results of the previous section show that partial
sums $P_0(\tau)$ and $P_0(\tau)$ are not satisfactory approximations to the distribution $P_0(\tau)$ when large or even intermediate intervals are considered. Even if it is possible to evaluate more terms (say, by computers) there will be some value of beyond which $P_0(\tau)$ cannot be considered reliable, though admittedly this value of $\tau$ could be made as large as one requires by evaluating more terms. This proposition, which is sound in theory, is nevertheless unrealistic in practice in view of the complexity of those terms, which will take present day computers a very long time to evaluate. The need for other methods which might be used in conjunction with $P_0(\tau)$ and $P_0(\tau)$ is obvious. A discussion of these methods follows, but after dealing first with the expected behaviour of $P_0(\tau)$ at large interval lengths since this is useful for the purpose of the discussion.

2.8.1 Behaviour of $P_0(\tau)$ at large $\tau$.

It is generally assumed that for large values of $\tau$, $P_0(\tau)$ decreases exponentially. Rice (3) gives an informal argument in support of this assumption and points out that no exact and useable formulae describing this asymptotic behaviour exists although progress in this direction has been made by Kuznestov, Stratonovich and Tikhanov (4). Their results give the cumulative distribution $F(T)$ defined by

$$F(\tau) = 1 - \int_0^\tau P_0(U) \, dU$$

\tag{2.8.1}

in terms of an exponential whose argument consists of an infinite series, the $n^{th}$ term of which is a multiple
integral of order $n-1$. When $\tau$ becomes large, each term approaches the form $a_n \tau + b_n$ where $a_n$ and $b_n$ are independent of $\tau$. Here, a similar argument to that of Rice, but which is expressed in more mathematical terms is presented in order to illustrate this exponential behaviour.

Define $q(\tau)$ as the conditional probability density (in $\tau, \tau + \Delta\tau$ when it is known that $f(t)$ has an upward crossing) of $f(t)$ having a downward zero crossing at $\tau=0$ and no other crossings in between. Then by the law of conditional probabilities we have,

$$P_0(\tau) = q(\tau) \cdot \left\{ 1 - \int_0^\tau P_0(u) \, du \right\} \quad \ldots(2.8.2)$$

As $\tau$ gets larger and larger, the event at $\tau, \tau + \Delta\tau$ becomes increasingly less influenced by the event at $\tau=0$. For finite memory processes, if $\tau$ is sufficiently long compared to the effective range ($\tau_e$ say) of the autocorrelation function $\psi(\tau)$ then the event at $\tau, \tau + \Delta\tau$ becomes dependent only on the most recent past with duration of order $\tau_e$. This means that if $\tau$ is greater than $\tau_e$, $q(\tau)$ will approach a constant value $a$. The solution of (2.8.2) where $q(\tau)$ is constant, is an exponential of the form $A e^{-a\tau}$.

The above argument, useful as an estimate for the expected behaviour at large $\tau$ leaves two important questions unanswered. Namely, those concerning the parameters of the exponential decay. In general, neither $a$ nor $A$ are known exactly. Further, the value of $\tau$ which defines the start of the range over which this behaviour is valid depends on the choice of $\tau_e$ and the manner in which it is chosen. For some random functions, e.g. a function with an ideal low pass power spectrum, $\tau_e$ must be chosen very large since such a function has an infinite memory. In such a
case, the exponential behaviour is only approximately valid over the range of very large values of τ.

Next two methods of working out an overall approximation to the distribution \( P_0(\tau) \) are described. These methods combine the approximations provided by \( P_0(\tau) \) and \( P_0(\tau) \) for small values of \( \tau \) with an assumed exponential behaviour over large values of \( \tau \), resulting in an overall approximation.

2.8.2. The Graphical Extrapolation Method.

This method was first suggested by Rice (3) who used it in conjunction with the first term approximation \( P_0(\tau) \). The method consists of suitably extrapolating \( P_0(\tau) \) over intermediate and large values of \( \tau \) to produce a probability curve which integrates to unity and whose mean value coincides with that of \( P_0(\tau) \). The mean value of \( P_0(\tau) \) is simply given by

\[ \beta = \frac{1}{N_0} \quad \text{(2.8.3)} \]

where \( N_0 \) is given by (2.2.2). The method is as follows: replacing \( \tau \) by the normalised time \( U = \frac{\tau}{\beta} \) equation (2.8.1) can be written as:

\[ F(U) = 1 - \int_0^U P_0(V) \, dV \quad \text{(2.8.4)} \]

Define

\[ G(U) = 1 - \int_0^U F(V) \, dV \quad \text{(2.8.5)} \]

Then it can be easily shown that:

\[ \lim_{U \to \infty} \int_0^U F(V) \, dV = U = \frac{\tau}{\beta} = \frac{1}{\beta} = \bar{\tau} \quad \text{(2.8.6)} \]

where the bars denote mean values. (2.8.6) follows from integrating \( \int_0^U F(V) \, dV \) by parts, which leads to:
\[
\left[ F(V) \cdot \dot{V} \right]_0^U + \int_0^U P_0(V) \cdot \dot{V} \, dU \quad \text{...(2.8.7)}
\]

As the limit of integration tends to \(\infty\), the first term tends to zero, and the second term tends to the mean value of \(P_0(\zeta)\) which after normalisation becomes unity. Further examination of relations (2.8.4) and (2.8.5) reveals the following results.

When \(U = 0\), both \(F(0)\) and \(G(0)\) are unity.

When \(U \to \infty\), \(F(\infty)\) and \(G(\infty)\) both tend to zero.

Plotting \(F(U)\) against \(G(U)\) should result in a curve starting at the point \((1,1)\), see Fig.(2), corresponding to \(U=0\) and finishing at point \((0,0)\) corresponding to \(U \to \infty\).

Rice's extrapolation method consists of plotting \(F(U)\) and \(G(U)\) where both functions are calculated using the \((1)\) approximation \(P_0(U)\) as a substitute for \(P_0(U)\) in (2.8.4). This resulted in a curve similar to \(A\) in Fig.(2). Because \(P_0(U)\) ceases to be a good representation of \(P_0(U)\) after some value of \(U\); this curve will not pass through the origin. Using this curve as a guide, Rice chose a point from which he extrapolated the curve in a 'reasonable' manner so as to pass through the origin. The extrapolation is shown as a dotted line on Fig.(2).

In order to obtain \(F(U)\) as a function of \(U\) we need to compute \(U\) at selected points on the estimated curve. This is done by numerical integration using the easily proven formula

\[
U = \int_G^1 \frac{dG}{F(G)} \quad \text{...(2.8.8)}
\]

where \(F(G)\) is the estimated curve. Once \(F(U)\) is obtained as a function of \(U\), the distribution \(P_0(U)\) may be obtained by numerical differentiation.
Fig. (2) - Graphical method for estimating the distribution $P_0(U)$ at large values of $U$ - dotted line represents estimated $(F, G)$ curve.
Two properties of the \((G,F)\) curve are worth mentioning. First if \(dF/dG=a\) in the small region around the origin, then (2.8.8) gives:

\[
F = C_3 e^{-au}
\]

...(2.8.9)

where \(c\) is independent of \(U\). Hence the slope of the \((G,F)\) curve at the origin reveals some information about the behaviour of \(F\) at large values of \(U\). Secondly, by using \(P_0(U)\) instead of \(P_0(U)\) we can obtain similar relationships to (2.8.4) & (2.8.5), but with \(F(U)\) and \(G(U)\) replacing \(F(U)\) & \(G(U)\). Subtracting these from (2.8.4) & (2.8.5) we obtain:

\[
F(U) - F(U) = \int_0^U \left[ P_0(V) - P_0(V) \right] dV \quad \text{...(2.8.10)}
\]

\[
G(U) - G(U) = \int_0^U \left[ F(V) - F(V) \right] dV \quad \text{...(2.8.11)}
\]

Since \(P_0(U) \geq P_0(U)\) (see 2.5.12), the quantity \(F(U) - F(U)\) starts at zero and monotonically increases as \(U\) runs from 0 to \(\infty\). The same is true of \(G(U) - G(U)\). Therefore the point \((G,F)\) tends to lie above and to the left of the corresponding (i.e. same value of \(U\)) point, \(G, F\). The distance \((F-F)^2 + (G-G)^2\) between the points is zero at \(U=0\) and never decreases as \(U=\infty\).

One of the things to be decided in drawing the \((G,F)\) curve is where it should start to diverge appreciably from the \((G,F)\) curve. Rice chose the region at which \(P_0(U)\) begins to approach its constant value \(\mathcal{P}(0, \Delta t)\), and he pointed out that it is helpful to draw the curves on both linear and log-log graph paper, since the object of the whole exercise is to find a suitable estimate of the distribution behaviour at large values of \(U\).
resulting distribution curve will integrate to unity, thus satisfying one of the basic probability laws as well as giving the true value of the mean interval between zero crossings.

This method can be extended by using the approximation \( P_0(U) \) to obtain a second guiding curve referred to as \( B \) in Fig. (2). This represents a better approximation over small values of \( U \). Further, there is the added advantage of having an extra curve, thus enabling one to make a more accurate extrapolation.

Since \( P_0(U) \geq P_0(U) \), a similar argumentation to that used above, the point \((G(2), F(2))\) on curve \( B \) tends to lie above and to the left of the corresponding (i.e., same value of \( U \)) point \((G, F)\). Curves \( A \) & \( B \) coincide over the range of small \( U \) and they start to diverge when \( P_0(U) \) ceases to be a good representation of \( P_0(U) \). Our estimated curve should start to diverge from \( B \) not very long after this happens. Hence the main advantage of having the curve \( B \), lies in the information it gives regarding the range of applicability of \( A \) and the region at which the estimated curve should start to diverge appreciably from both \( A \) and \( B \).

2.8.3 Analytical method.

This method, advanced by B.R. Levin and Ya.A. Fomin

\[ \omega(\tau) = ce^{-a\tau} \quad \cdots (2.8.12) \]
These two curves intersect at a point $\tau_n$. The three unknown parameters of the approximating function (2.8.12), namely $C, a$ and $\tau_n$ are determined by the following system of three simultaneous equations:

i) \[ \omega(\tau_n) = Q(\tau_n) \quad \text{(2.8.13a)} \]
Since at the point of intersection $\tau_n$, the values of $Q(\tau)$ and $\omega(\tau)$ must be the same.

ii) The total area below the approximating curves must be equal to unity, hence:
\[ \int_0^{\tau_n} Q(\tau) \, d\tau + \int_{\tau_n}^{\infty} \omega(\tau) \, d\tau = 1 \quad \text{(2.8.13b)} \]

iii) The mean value of the crossing interval, calculated on the basis of the approximating curve must be equal to the mean crossing interval $\beta$ as determined by (2.8.3), hence
\[ \int_0^{\tau_n} Q(\tau) \cdot \tau \, d\tau + \int_{\tau_n}^{\infty} \omega(\tau) \cdot \tau \, d\tau = \beta \quad \text{(2.8.13c)} \]

Substituting expression (2.8.12) for $\omega(\tau)$ in the system (2.8.13), we obtain:
\[ Q(\tau_n) = C e^{-a \tau_n} \quad \text{(2.8.14a)} \]
\[ \int_0^{\tau_n} Q(\tau) \, d\tau = 1 - \frac{C}{a} e^{-a \tau_n} \quad \text{(2.8.14b)} \]
\[ \int_0^{\tau_n} Q(\tau) \cdot \tau \, d\tau = \beta - \frac{1}{a} \tau_n + \frac{C}{a} e^{-a \tau_n} \quad \text{(2.8.14c)} \]

Hence we can obtain the following equation in the unknown $\tau_n$:
\[ Q(\tau_n) \cdot \int_0^{\tau_n} Q(\tau) \, d\tau + \tau_n \cdot Q(\tau_n) \cdot \frac{F(\tau_n) + F(\tau_n)}{F(\tau_n)} = \beta Q(\tau_n) \quad \text{(2.8.15)} \]

where $F(\tau_n) = 1 - \int_0^{\tau_n} Q(\tau) \, d\tau \quad \text{(2.8.16)}$

The roots of (2.8.15) may either be determined by numerical
or graphical means. Once this is done \( C \) and \( a \) are then evaluated using (2.8.14a) \&(2.8.14b) and hence the overall approximate curve for the distribution \( P_0(\mathcal{C}) \).

It is very likely that (2.8.15) can have more than one root over the entire range of \( \mathcal{C} \) which means that a number of possible curves for the distribution \( P_0(\mathcal{C}) \) can be obtained, all satisfying the area and mean value requirements.

2.8.4 Comparison between the Graphical and Analytical Method.

Comparing the above method with that suggested by Rice, there is little to choose between them. Both methods attempt to construct an overall solution approximately valid over the entire range of \( \mathcal{C} \). In both methods \( Q(\mathcal{C}) \) provides the initial section of the curves whereas an assumed exponential behaviour provides the second section. Where they seem to differ is in the determination of the point or the region at which the sections merge. Rice takes that region to be the one at which \( Q(\mathcal{C}) \) starts to approach its constant level, whereas Levin and Fomin obtain this as a solution of an equation. This equation however may have several roots resulting in several possible approximations. There is no mention of how to overcome this problem in their paper, but it is possible to suggest a guideline which can help in selecting an approximation. This is, to select the root at which the exponential behaviour can justifiably be assumed and which provides a smooth transition between the two sections of the curve.

2.9 Approximate methods utilising Integral equations.

In this section, development of integral equations
whose solutions provide further approximations for the distribution $P_n(\tau)$, is dealt with. The correlation between successive intervals is of prime importance in deriving these equations and so will be discussed in some detail. The main contribution in this field has been made by McFadden, J.A., though significant contributions were made also by D.S. Palmer (6), J. Rice, & F. Beer (7).

J.A. McFadden attempted to relate the zero crossing distributions to the properties of $x(t)$, which is derived from the random function $f(t)$ by infinite clipping. $x(t)$ is thus given by

$$x(t) = 1, \text{ when } f(t) \geq 0$$
$$x(t) = -1, \text{ when } f(t) < 0$$

...(2.9.1)

under infinite clipping the zero crossings remain invariant though all other information contained in $f(t)$ is lost. If $f(t)$ is Gaussian, its normalised auto-correlation function $\rho(\tau)$ is related to that of $x(t), r(\tau)$ by the following relationship:

$$r(\tau) = \frac{2}{\pi} \sin^{-1} \rho(\tau)$$

...(2.9.2)

This relationship implies that if $\rho(\tau)$ is regular at $\tau=0$, i.e. it can be expanded in a power series containing even powers of $\tau$ only, its corresponding $r(\tau)$ when expanded in a power series, would contain only odd powers of $|\tau|$ and the constant term which is unity. Further, whereas $\rho(\tau)$ is continuous and equal to zero, the first derivative of $r(\tau)$ is discontinuous at the origin and it is necessary to distinguish between $r'(0^+)$ and $r'(0^-)$, both of which have non-zero values.

2.9.1 Preliminary results.

The following are some of the results obtained by
McFadden in his earlier work\(^8\). They are listed without proofs as these can be found in the original work; the results are useful for the ensuing discussion.

i) The expected number of crossings/unit time \(N_0\) is given by:
\[
N_0 = \frac{1}{2}r'(0+)
\]  
\[\text{(2.9.3)}\]

ii) \(P_0(\tau)\) is zero over a finite range \(0 < \tau < T\) if, and only if \(r(\tau)\) is linear in \(|\tau|\) over that range and vice-versa.

iii) If the assumption holds true that given a random function \(f(t)\), there exists a quantity \(T_1\), such that the probability of more than two zeros occurring in the interval \((t, t+T_1)\) is negligibly small, then in the range \(0 < \tau < T_1\) the expression
\[
P_{MC}(\tau) = \frac{r''(\tau)}{-2r'(0+)} = \frac{r''(\tau)}{4N_0}
\]  
\[\text{(2.9.4)}\]
is a very close approximation to the true distribution \(P_0(\tau)\). In terms of the normalised auto-correlation function \(R(\tau)\) of \(f(t)\), (2.9.4) may be written as
\[
P_{MC}(\tau) = \frac{1}{2 \pi N_0} \left[ \frac{R(\tau)}{1 \cdot R^2(\tau)} + \frac{R(\tau)}{1 - R^2(\tau)} \right] \frac{1}{\delta^2}
\]  
\[\text{(2.9.5)}\]

The above results show that the distribution \(P_0(\tau)\) is more readily related to the statistical properties of \(x(t)\) than to those of \(f(t)\). If \(f(t)\) is not Gaussian, the arc sine law (2.9.2) does not apply and in general the relation between \(r(\tau)\) and \(R(\tau)\) is not known. However if \(r(\tau)\) is known but no other information about \(f(t)\) is available, then the average number of zero crossings is given by (2.9.3) and the results listed in (ii) and (iii) apply provided the appropriate assumptions are valid.

2.9.2 Relationships between \(P_n(\tau)\) and \(P(n,\tau)\).

Earlier in the chapter we defined \(P(n,\tau)\) as being
the probability of \( f(t) \) having exactly \( n \) zero crossings in a given interval of length \( T \). Intuitively, we can expect these distributions to be related in some manner to the distributions \( P_n(T) \). McFadden provided the relationships which are listed below.

\[
P^0(0,T) = N_0 P_0(T) \quad \ldots (2.9.6)
\]

\[
P^0(1,T) = N_0 [P_1(T) - 2P_0(T)] \quad \ldots (2.9.7)
\]

and when \( n \geq 2 \)

\[
P^0(n,T) = N_0 [P_n(T) - 2P_{n-1}(T) + P_{n-2}(T)] \quad \ldots (2.9.8)
\]

where

\[
P^0(n,T) = \frac{\partial^2}{\partial T^2} P_n(T) \quad \ldots (2.9.9)
\]

Considering a record of the function \( x(t) \) of length \( T \) and taking an ensemble of such records, then if the number of zero crossings is even, \( x(t) \) and \( x(t+T) \) have the same sign. Whereas if the number is odd, \( x(t) \) and \( x(t+T) \) have opposite signs. The ensemble average \( \langle x(t) \cdot x(t+T) \rangle \) is thus given by:

\[
\langle x(t) \cdot x(t+T) \rangle = \sum_{n=0}^{\infty} P(2n,T) \cdot 1 - \sum_{n=0}^{\infty} P(2n+1,T) \quad \ldots (2.9.10)
\]

By definition, the ensemble average is the auto-correlation function \( r(T) \), hence

\[
r(T) = \sum_{n=0}^{\infty} (-1)^n P(n,T) \quad \ldots (2.9.11)
\]

Using the relations (2.9.6-2.9.8), the above series can be written in terms of \( P_n(T) \), namely

\[
\frac{r^2(T)}{4N_0} = \sum_{n=0}^{\infty} (-1)^n P_n(T) \quad \ldots (2.9.12)
\]

If \( T \) is small such that \( P_n(T) \) for \( n > 0 \) is negligible compared to \( P_0(T) \), we have

\[
P_0(T) = \frac{r^2(T)}{4N_0} \quad \ldots (2.9.13)
\]

which is the same as (2.9.4) obtained earlier. This explains
the assumption which had to be made before, to justify (2.9.4), (see section 2.9.1), (2.9.13), when used in conjunction with the series (2.3.5) and (2.3.6), obtained earlier for \(U(t)\) and \(Q(t)\), yields the following relationships

\[
\frac{1}{2}(U(t) + \frac{\dot{r}(\zeta)}{4N_0}) = \sum_{n=0}^{\infty} p_{2n}(\zeta) = Q(\zeta) \quad \ldots(2.9.14)
\]

\[
\frac{1}{2}(U(t) - \frac{\dot{r}(\zeta)}{4N_0}) = \sum_{n=0}^{\infty} p_{2n+1}(\zeta) = P_{\text{ODD}}(\zeta) \quad \ldots(2.9.15)
\]

where \(P_{\text{ODD}}(\zeta)\) is also given by

\[
P_{\text{ODD}}(\zeta) = U(\zeta) - Q(\zeta) \quad \ldots(2.9.16)
\]

For small interval lengths (2.9.14) provides an approximation for \(P_0(\zeta)\) whereas (2.9.15) provides one for \(P_1(\zeta)\).

### 2.9.3 Derivation of Integral Equations based on Statistically Independent Intervals.

The series formulae provide means of determining the probability distributions \(P_n(\zeta)\) for all \(n\) and hence \(P(n, \zeta)\) if one can express \(P_n(\zeta)\) for all \(n\) in terms of the distribution \(P_0(\zeta)\). For this purpose J.A. McFadden postulated that successive interval lengths between zero crossings are independent. This postulate enabled him to relate \(P_n(\zeta)\) to \(P_0(\zeta)\) by a simple formulae, since under this assumption the distribution \(P_n(\zeta)\) is simply the \(n\)th fold self-convolution of \(P_0(\zeta)\). Or alternatively, using the Laplace Transform, this means:

\[
P_n(s) = \left[ P_0(s) \right]^{n+1} \quad n \geq 0 \quad \ldots(2.9.17)
\]

where

\[
P_n(s) = \mathcal{L}(P_n(\zeta)) = \int_0^\infty P_n(\zeta)e^{-s\zeta} \, d\zeta \quad \ldots(2.9.18)
\]

Taking the Laplace transform of both sides of series (2.3.5) yields:

\[
U(s) = \sum_{n=0}^{\infty} \left[ P_0(s) \right]^{n+1} \quad \ldots(2.9.19)
\]

and therefore
\[ U(s) = \frac{P_0(s)}{1-P_0(s)} \quad \cdots (2.9.20) \]
similarly, using (2.9.12) we arrive at
\[ \frac{r^*(s)}{4N_0} = \frac{P_0(s)}{1+P_0(s)} \quad \cdots (2.9.21) \]

Further, taking the Laplace Transforms of (2.9.14 & 15) we obtain:
\[ \frac{1}{2} \left[ U(s) + \frac{r^*(s)}{4N_0} \right] = Q(s) = \frac{P_0(s)}{1-P_0^2(s)} \quad \cdots (2.9.22) \]
and
\[ \frac{1}{2} \left[ U(s) - \frac{r^*(s)}{4N_0} \right] = P_{ODD}(s) = P_0(s)Q(s) \quad \cdots (2.9.23) \]
Together, (2.9.22) & (2.9.23) yield:
\[ Q(s) = P_0(s) / \left[ 1 - \frac{P_{ODD}(s)P_0(s)}{Q(s)} \right] \quad \cdots (2.9.24) \]

Because the functions \( r(\tau), Q(\tau) \) and \( U(\tau) \) are complicated functions, the solutions of (2.9.20), (2.9.21) & (2.9.24) cannot be easily obtained. Instead it is simpler to solve the corresponding integral equations obtained by taking the inverse Laplace Transform of these equations. These are:
\[ P_0(\tau) = U(\tau) - U(\tau) * P_0(\tau) \quad \cdots (2.9.25) \]
\[ P_0(\tau) = \frac{r^*(\tau)}{4N_0} + \frac{r^*(\tau)}{4N_0} * P_0(\tau) \quad \cdots (2.9.26) \]
\[ P_0(\tau) = Q(\tau) - P_{ODD}(\tau) * P_0(\tau) \quad \cdots (2.9.27) \]
where \((*)\) denotes the convolution integral.
\[ Y_1(\tau) * Y_2(\tau) = \int_0^{\tau} Y_1(1) Y_2(\tau - 1) d\tau \quad \cdots (2.9.28) \]

In general the solutions obtained from (2.9.25) and (2.9.26-27) are not identical if the independence assumption does not hold true. McFadden obtained (2.9.25 & 2.9.26) by postulating independence between successive interval lengths. However in obtaining (2.9.27) he postulated what he termed 'quasi-independence'. This assumption, which using Laplace Transforms can be stated as
\[ P_{2n+2}(s) = P_{2n+1}(s) \cdot P_0(s) \quad \ldots (2.9.29) \]

implies that a given interval is independent of the sum of the \((2n+2)\) intervals immediately preceding it. The relation \((2.9.29)\) is automatically satisfied if \((2.9.17)\) is valid. The converse is not necessarily true since it may be possible to have an interval correlated to the one immediately preceding it, but which is independent of the sum of the two, four, or in general, the \((2n+2)\) preceding intervals.

It can therefore be said that the integral equations \((2.9.27)\) applies to a wider range of processes in which correlation between successive intervals is permitted so long as \((2.9.29)\) is satisfied. Further when applying the integral equations \((2.9.25-27)\) to processes whose zero crossing intervals are correlated such that neither \((2.9.17)\) nor \((2.9.29)\) is strictly valid; we would expect that \((2.9.27)\) gives a better approximation than either \((2.9.25)\) or \((2.9.26)\).

Results obtained by McFadden seem to confirm this.

McFadden solved \((2.9.27)\) numerically for Gaussian noise in anumber of cases stated below corresponding to different power spectra. The normalised auto-correlation function was obtained by the Weiner-Khintchine theorem, then \( U(\tau) \) and \( \frac{2\pi^2(\tau)}{4N_0} \) were obtained. The convolution integral was approximated by the trapezoidal rule, using a suitable mesh width of \(\Delta\tau\). Starting with \(P_0(0)\) he solved successively for \(P_0(\Delta\tau), P_0(2\Delta\tau)\) \ldots etc.

McFadden obtained solutions for the following spectra:

i) \[ W_0(\tau) = \frac{(2 \pi \tau)^{2m}}{(1+4\pi^2 \tau^2)^n} \quad \ldots (2.9.30) \]

where \(m=0, n=2\). In this case

\[ \rho (\tau) = (1 + |\tau|) e^{-|\tau|} \quad \ldots (2.9.31) \]

ii) \(m=0, n=3\) correspondingly

\[ \rho (\tau) = (1 + |\tau| + \frac{1}{2} \tau^2) e^{-|\tau|} \ldots (2.9.32) \]
iii) m=2, n=4 and correspondingly
\[ \rho(\tau) = (1 + |\tau| - 2\tau^2 + \frac{3}{2} |\tau|^3) e^{-|\tau|} \] \hspace{1cm} \text{(2.9.33)}

iv) 7-pole Butterworth Spectrum, i.e.
\[ W(f) = \frac{1}{1 + (f/f_0)^{14}} \] \hspace{1cm} \text{(2.9.34)}

Then if \( 2\pi f_0 = 1 \)
\[ \rho(\tau) = \sin \frac{\pi}{14} \left[ e^{-|\tau|} + 2 \sum_{n=1}^{3} e^{-|\tau|} \cos(\pi n/7) \cos(\frac{\pi n}{7} - |\tau| \sin \frac{\pi n}{7}) \right] \] \hspace{1cm} \text{(2.9.35)}

v) Ideal low-pass Spectrum, where
\[ \rho(\tau) = \frac{\sin 2\pi f_0 \tau}{2\pi f_0 \tau} \] \hspace{1cm} \text{(2.9.36)}

vi) Ideal band-pass Spectrum given by
\[ \omega(f) = \begin{cases} 0 & 0 < f < f_0 \\ \text{Const.} & f_0 < f < 2f_0 \\ 0 & f > 2f_0 \end{cases} \] \hspace{1cm} \text{(2.9.37)}

Hence:
\[ \rho(\tau) = \sin \frac{4\pi f_0 \tau}{4\pi f_0 \tau} - \sin \frac{2\pi f_0 \tau}{4\pi f_0 \tau} \] \hspace{1cm} \text{(2.9.38)}

Cases (i) & (iii) belong to the class where \( \omega(f) \) behaves like \( f^{-4} \) as \( f \to \infty \). This was referred to earlier as the singular case for which \( \beta(\tau) \) is discontinuous at the origin, \( U(0), Q(0) \) and consequently \( P_0(0) \) do not vanish at \( \tau=0 \). For all other cases, these quantities do vanish at \( \tau=0 \); though they do not all belong to the class where \( \beta(\tau) \) is regular at the origin. Cases (ii) & (iv) involve processes whose \( \rho(\tau) \) when expanded in power series of \( \tau \) do contain odd powers of \( |\tau| \), though not terms in \( |\tau| \) or \( |\tau|^3 \).

According to McFadden, the solution of (2.9.27) obtained in the cases (ii) & (iv) showed a close agreement
with experimental results previously obtained by Favreau, Low and Pfeffer\((10)\). In cases (i) & (iii) the agreement is also good except around \( \zeta = 0 \). In the remaining cases the solution dips negative and therefore cannot be accepted for large \( \zeta \). He evidently concluded that the assumption of quasi-independence may be justified when spectra with gradual high frequency cut-off are involved; cases (i-iv). However this assumption is not valid when spectra with sharp cut-off frequencies are involved (cases v & vi). In any case McFadden pointed out that the assumption is not good for extremely large intervals and therefore should not be used for the calculation of moments.

2.9.4 Markov Chain of Intervals\((9)\)

McFadden also derived expressions for the variance of zero crossing intervals and the correlation co-efficient between two successive intervals under the assumption that successive intervals form a Markov chain in the wide sense or that:

\[
k_{ij} = k^{|i-j|} \quad \ldots (2.9.39)
\]

where \( k_{ij} \) is the correlation co-efficient between the \( i^{th} \) & \( j^{th} \) crossing intervals and \( k \) is that between two successive intervals. Denoting the variance of intervals by \( \sigma_C^2 \), we have

\[
k_{ij} = \left( \langle \zeta_i \zeta_j \rangle - \frac{4}{N_0^2} \right) / \sigma_C^2; (i \neq j) \quad \ldots (2.9.40)
\]

McFadden derived the following results

\[
\sigma_C^2 = \left[ \frac{2A}{N_0} \frac{(1+2B)}{N_0^2} \right]^\frac{1}{2} \quad \ldots (2.9.41)
\]

and

\[
\frac{1-k}{1+k} = \left( \frac{2A}{N_0} \right)^{\frac{1}{2}} \left( \frac{1+2B}{N_0} \right)^{\frac{1}{2}} \quad \ldots (2.9.42a)
\]

where

\[
A = \int_0^\infty r(\zeta) d\zeta \quad \ldots (2.9.43a)
\]
\[ B = \int_0^\infty [U(\tau) - N_0] \, d\tau \quad \cdots (2.9.43b) \]

and evaluated \( \sigma_\tau \) and \( k \) for the above cases, (except case iv). These are tabulated in Table 1.

<table>
<thead>
<tr>
<th>Case</th>
<th>Mean Value</th>
<th>( \sigma_\tau )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.1416</td>
<td>3.22</td>
<td>0.058</td>
</tr>
<tr>
<td>2</td>
<td>5.4414</td>
<td>4.68</td>
<td>0.005</td>
</tr>
<tr>
<td>3</td>
<td>1.4050</td>
<td>0.775</td>
<td>0.486</td>
</tr>
<tr>
<td>5</td>
<td>5.4414</td>
<td>3.74</td>
<td>0.009</td>
</tr>
<tr>
<td>6</td>
<td>2.0567</td>
<td>0.565</td>
<td>0.442</td>
</tr>
</tbody>
</table>

As might be expected, the ratio \( \sigma \langle \tau \rangle \) is smallest in the two cases 3 & 6 where the bandwidth is relatively narrow. In these cases the correlation between successive intervals appears to be strong. The ideal low pass spectrum (case 5) is particularly significant since the numerical solution of (2.9.25) which is based on the assumption of independent intervals was stated by McFadden to be far from correct. Yet \( k \) in this case is nearly zero. Evidently the assumption that successive intervals form a Markov Chain is not true in this case. On the other hand, for a low pass spectrum such as that in case 2, where the cut-off is gradual, the correlation co-efficient \( k \) is nearly zero. Further, according to McFadden the solutions of the integral equations (2.9.25-27) were in close agreement with each other as well as with the experimental results of Low, Pfeffer and Favreau. Evidently the assumption of independence between successive intervals is not far from correct in this case.

2.9.5 Zero Crossings as a Renewal Process.

J. Rice and F. Beer\(^{(7)}\) considered the level crossing
problem of which the zero crossing is a special case. In their attempt to find an approximation to \( P_0(\tau) \), they postulated that the events represented by level crossing points, form a renewal process. That is to approximate the conditional probability of a downward crossing of level at \( \tau \), \( \tau + d\tau \) given several prior downward crossings in the interval \( 0 \rightarrow \tau \) and an upward crossing at \( \tau = 0 \) by the conditional probability of a downward crossing at \( \tau', \tau + d\tau \) given only the last known prior downward crossing. This postulate simplifies the evaluation of each term of the in and exclusion series (2.4.2). These terms, which are multiple integrals with an ever increasing multiplicity become convolution integrals and this facilitates the use of Laplace transform methods that lead finally to an integral equation which is identical to (2.9.27). The renewal assumption as stated above is thus equivalent to McFadden's assumption of quasi independence between successive intervals. The equivalence is demonstrated with the aid of the following illustration which shows an interval of length \( \tau \) over which several crossing points (not necessarily successive) are indicated. It is assumed however that the zeros at \( t_0 \) and \( \tau \) and those at \( t_2 \) and \( t_1 \) are successive.

\[
\begin{array}{cccccccc}
  & + & - & + & - & + & - & \\
  t=0 & t_3 & t_2 & t_1 & t_0 & \tau
\end{array}
\]

According to the renewal assumption, the knowledge that a downward zero crossing at \( t_1 \) is sufficient to determine the probability of a downward zero crossing at \( \tau, \tau + d\tau \) and that the history of \( f(t) \) prior to \( t_0 \) is of no relevance as far as this event is concerned. It follows therefore that the interval length \( (\tau - t_1) \) is independent of the interval length \( (t_1 - t_2) \). Similarly, it is true to say that the length of the present interval \( (\tau - t_0) \) is independent.
of the interval \((t_0 - t_2)\). This latter interval which is the interval between two upward crossings can be the sum of two, four or in general \((2n+2)\) successive zero crossing intervals depending on the number of crossings contained in \((t_0 - t_2)\). Hence under the renewal assumption the present interval \((\zeta - t_0)\) is independent of the sum of the \((2n+2)\) interval immediately preceding it. This is equivalent to the quasi independence assumption of McFadden.

The renewal assumption becomes more valid when crossings of high threshold levels are concerned. Due to the rarity of such events and the relatively long intervals between successive ones, it is logical to expect that a downward crossing is almost entirely dependent on the last known downward crossing.

2.9.6 The Advantages of the Integral Equations Approach.

The solutions of the integral equations developed in this section provide further overall approximations to the distribution \(P_0(\zeta)\). Of all the integral equations developed so far, (2.9.27) provides the best approximation. When compared to methods discussed previously this approach has the following advantages.

i) No assumptions regarding initial or final behaviour are necessary.

ii) Because of (i) it is possible to obtain better approximations for the singular case.

iii) Less computations are involved in the solution of (2.9.27) than in methods stated previously.

These advantages are weighed against the errors resulting from making wrong assumptions regarding the
dependence between successive intervals. Thus in the ideal low-pass spectrum; the integral equation fails to give reliable results at large $\tau$ simply because the assumption of quasi-independence is not valid.

2.10 Relationship between $P_0(\tau)$ and the probability integral of Multivariate Distributions.

We define $V(\tau)$ as the probability that the process $f(t)$ is positive over an arbitrary time interval of length $\tau$. By symmetry $V(\tau)$ is equal to $Z(\tau)$, the probability that $f(t)$ is negative over the period $\tau$. Hence by adding them together, we obtain the probability that $f(t)$ has no zero crossings in the interval $\tau$, i.e.

$$V(\tau) + Z(\tau) = P(0, \tau) \quad \cdots (2.10.1)$$

This together with (2.9.6) results in

$$P_0(\tau) = \frac{2}{M_0} \cdot \frac{d^2}{d\tau^2} \left[ V(\tau) \right] \quad \cdots (2.10.2)$$

which is a similar relation to the one Longuet-Higgins derived from first principles and without using (2.9.6). The relation (2.10.2) furnishes us with another method for determining the distribution of the zero crossing intervals $P_0(\tau)$. This problem is now reduced to the evaluation of $V(\tau)$, which having been obtained enables $P_0(\tau)$ to be evaluated, by twice differentiating the outcome.

2.10.1 The Evaluation of $V(\tau)$.

One possible method of evaluating $V(\tau)$ would be to take $n$ points (normally equally spaced) over the interval with $t_1$ and $t_n$ forming the beginning and end points of the interval. Defining the quantity $V_n(t_1, t_2, \ldots, t_n)$ by:

$$V_n(t_1, t_2, \ldots, t_n) = \text{Prob} \left[ f(t_1) > 0; f(t_2) > 0; \ldots; f(t_n) > 0 \right] \quad \cdots (2.10.3)$$
and assuming it is possible to evaluate $V_n$, then by letting $n \to \infty$ in such a manner that the intervals between the points tend to zero, where $\zeta$ is kept fixed, it is reasonable to expect (if $f(t)$ is continuous) that

$$\lim_{n \to \infty} V_n(t_1, \ldots, t_n) = V(\zeta) \quad ...(2.10.4)$$

Because the samples at $t_1, \ldots, t_n$ are normally correlated, the limiting value (2.10.4) may well be approached by taking a finite number of samples, depending on the length $\zeta$ and the degree of correlation.

Using the definition of $V_n$, we can write immediately the following expression:

$$V_n(t_1, \ldots, t_n) = \int_0^\infty \ldots \int_0^\infty P_n(\xi_1, \ldots, \xi_n) \, d\xi_1 \ldots d\xi_n \quad ...(2.10.5)$$

where $\xi_i$ denotes $f(t_i)$ and $P_n(\xi_1, \ldots, \xi_n)$ is the joint probability density of the $n$ variates $\xi_i$.

Equation (2.10.5) relates $V_n$ and hence $P_0(\zeta)$ to the integral, in the $n$th dimensional space of the joint probability density function which is often referred to in the literature as the multivariate probability integral. This is, in itself, an extensive subject which is discussed in detail in Chapter three. Suffice it to mention here that this is an important area which has the possibility theoretically at least of providing further methods of approach to the problem under discussion.

2.11 The First-Order Markov Process.

It was mentioned earlier that if $f(t)$ has the following spectrum:

$$W(f) = \frac{1}{\alpha e^2 + \psi f^2} \quad ...(2.11.1)$$
then the second derivative of the auto-correlation function evaluated at the origin is infinity. The average distance, between successive zeros as calculated by using formula (2.2.3) is therefore zero. For this class of random function the distribution $P_0(\tau)$ does not exist in the usual sense. Instead, some other related distributions are defined. S. Rice listed several results which were obtained earlier by A.J.F. Seigert. McFadden and Slepian have also made significant contributions on this aspect of the zero crossing problem. Here some of these results are listed, in order to illustrate the main differences between this class of function and the general case.

S. Rice, in an attempt to shed some light on why formula (2.2.2) breaks down for this class of function, considered the power spectrum to consist of two bands of noise. One band is confined to relatively low frequencies whose power spectrum is denoted by $W_1(f)$, whereas the other band is very narrow and centered at the relatively high frequency $f_2$. The complete power spectrum is then:

$$W(f) = W_1(f) + A^2 \delta (f-f_2) \quad \ldots (2.11.2)$$

where the $\delta$-function is used to represent the very narrow band.

Now since

$$\psi(\tau) = 2 \int_0^\infty W(f) \cos 2\pi f \tau \ df \quad \ldots (2.11.3)$$

therefore

$$\psi(0) = 8 \pi^2 \int_0^\infty W(f) f^2 df \quad \ldots (2.11.4)$$

In the argument that ensued, Rice concluded that if $A \to 0$ and $f_2 \to \infty$ in such a way that $Af_2 \to \infty$, the integral (2.11.3) for $\psi(\tau)$ converges, whereas the integral (2.11.4) for $\psi(0)$ diverges. The process $f(t)$ behaves something
like a continuous function which has no derivative. Under these conditions, it is possible for \( f(t) \) which has just crossed a particular threshold level, say the zero level, to cross it again and again. This gives rise to clustering of axis crossings. It is not meaningful therefore to speak of intervals between consecutive zeros, but rather to define probabilities of clusters at a particular threshold level and the probability \( W(I, T) \) is the probability that \( f(t) \) reaches the value zero for the first time between \( \zeta, \zeta + d\zeta \), when it is known that \( f(t) = I \) at \( \zeta = 0 \).

For Gaussian processes, Siegert \(^{42} \) obtained the probability \( W(I, \zeta) \) as a solution of an integral equation. It is given by

\[
W(I, \zeta) \, d\zeta = \left( \frac{2}{\pi} \right)^{1/2} I \exp(-I^2 \zeta^2/2) \left(-d\zeta \right) \tag{2.11.5}
\]

where

\[
\zeta_1 = e^{-\zeta} (1 - e^{-2\zeta})^{-1/2}, \quad d\zeta_1 = e^{2\zeta} \zeta_1^3 d\zeta \tag{2.11.6}
\]

In addition, he derived the following results:

i) The probability that \( f(t) \) will go through zero for the first time in \( \zeta, \zeta + d\zeta \) given that \( f(0) > 0 \) is

\[
Q_0(\zeta) \, d\zeta = \left( \frac{2}{\pi} \right) e^{-\zeta} (1 - e^{-2\zeta})^{-1/2} d\zeta \tag{2.11.7}
\]

By symmetry, the condition \( f(0) > 0 \) may be replaced by \( f(0) \neq 0 \). This is also the probability that the interval between an instant selected at random and the first zero crossing of \( f(t) \) following it; have a length \( \zeta \) lying between \( \zeta, \zeta + d\zeta \).

ii) The chance that \( x \) exceeds \( \zeta \) is given by

\[
\int_0^\infty Q_0(1) \, dl = \frac{2}{\pi} \sin^{-1}(e^{-\zeta}) \tag{2.11.8}
\]

where the arc lies between 0 and \( \pi/2 \).

D. Slepian \(^{13} \) obtained (2.11.8) by using a much easier method than that used by Siegert. Equation (2.11.8) also
gives the probability of no zero crossings occurring in an arbitrary interval of length $\tau$.

i.e. $P_0(0, \tau) = \frac{2}{\pi} \sin^{-1} e^{-\tau}$ \hspace{1cm} (2.11.11)

Let $\tau = \Delta$ and $N(\Delta)\Delta$ be the probability that one or more zero crossings occur in the interval $\Delta$, therefore using (2.11.11) we obtain.

$$N(\Delta)\Delta = 1 - \frac{2}{\pi} \sin^{-1} e^{-\Delta} \hspace{1cm} (2.11.12)$$

If $\Delta$ is small, $N(\Delta)$ is approximately given by

$$N(\Delta) \approx \frac{2\sqrt{2}}{\pi} \Delta^{-\frac{1}{2}} + O(\Delta^{\frac{3}{2}}) \hspace{1cm} (2.11.13)$$

As $\Delta \to 0, N(\Delta) \to \infty$. $N(\Delta)$ is the analogue of $P(0, \Delta t)$ defined earlier for the general Gaussian Process. McFadden derived the above result. In addition he defined the two quantities $P(\Delta, \tau)$ and $U(\Delta, \tau)$ for the Markovian process, in analogy to $P_0(\tau)$ and $U(\tau)$ which apply to the general process. $P(\Delta, \tau)$ is the probability density of $f(t)$ crossing the zero level for the first time in $(t+\tau, t+\Delta+\tau)$ when it is known that $f(t)$ had crossed that level once or more in $(t, t+\Delta)$. McFadden derived the following expression for $P(\Delta, \tau)$

$$P(\Delta, \tau) = \frac{e^{-\tau}(1-e^{-2\tau})^{-\frac{1}{2}} - e^{-\tau}(1-e^{-2(\tau+\Delta)})^{-\frac{1}{2}}}{\frac{\pi}{2} \sin^{-1}(e^{-\Delta})} \hspace{1cm} (2.11.14)$$

He also gave the following results for the mean and mean square values of this distribution.

$$E(\tau) = \frac{\pi}{2\sqrt{2}} \Delta^{\frac{1}{2}} + O(\Delta) \hspace{1cm} (2.11.15)$$

$$E(\tau^2) = \frac{\pi}{\sqrt{2}} \log 2 \Delta^{\frac{3}{2}} + O(\Delta^{3/2}) \hspace{1cm} (2.11.16)$$

Hence

$$E(\tau) N(\Delta) = 1 + O(\Delta^{\frac{3}{2}}) \hspace{1cm} (2.11.17)$$

which is analogous to $\beta N_0 = 1$ in the case of the general process. (2.11.15) & (2.11.16) together, give the variance
to mean value ratio:
\[
\frac{E(\zeta^2)}{E(\zeta)} = 2 \log 2 + O(\Delta^{3/2}) \quad \ldots (2.11.18)
\]

which is finite as \(\Delta \to 0\). Hence although both \(E(\zeta)\) and \(E(\zeta^2)\) vanish as \(\Delta \to 0\), their ratio remains finite. \(U_{\Delta \zeta}(\zeta)\) is the probability density of \(f(t)\), having one or more zero crossings in \((t+\zeta, t+\zeta+\delta)\) when it is known that \(f(t)\) had crossed that level once or more in \((t, t+\Delta)\).

McFadden referred to \(U_{\Delta \zeta}\) as the probability of recurrence and derived the following approximation:
\[
U_{\Delta \zeta}(\zeta) \approx N(\delta) \left(1 - e^{-2\zeta}\right)^{-\frac{\zeta}{2}} \quad \ldots (2.11.19)
\]

Hence for small but finite values of \(\zeta\), \(U_{\Delta \zeta}(0)\) is infinite but \(U_{\Delta \zeta}(\zeta)\) decreases monotonically as \(\zeta\) increases approaching the value \(N(\delta)\). This limit is correct since \(N(\delta)\) is the probability of one or more crossings in an interval of length \(\delta\) and the initial condition has a vanishing influence as \(\zeta \to \infty\).

To conclude, \(N_0, P_0(\zeta)\) and \(U(\zeta)\) do not exist for a stationary Gaussian, Markov process. Instead, analogous quantities are defined for which asymptotic formulae are given.

2.12 Conclusion of survey.

The aim of this survey, which is by no means exhaustive was to review some basic ideas and approaches advanced by several people who have contributed in this field. Further, to present them, in a framework where they could be discussed and assessed individually and in relation to each other The material in this chapter forms the basis of the present research and is directly related to the work
continued in succeeding chapters. Thus it is hoped that this survey will help the reader to appreciate and place in context, the material that follows.
CHAPTER THREE.

The Multivariate Normal Integral.

3.1. Introduction.

In Chapter two, the relationship between \( P_0(\xi) \) and the multivariate integral \( V_n(t_1 \ldots t_n) \) was developed (see section 10), and the following expression for \( V_n \) was given

\[
V_n = \int_0^\infty \cdots \int_0^\infty P_n(\xi_1 \ldots \xi_n) \, d\xi_1 \cdots d\xi_n \quad \cdots (2.10.5)
\]

If \( f(t) \) has a Gaussian amplitude probability distribution with zero mean and unity variance, the multidimensional density \( P(\xi_1 \ldots \xi_n) \) is given by the following expression:

\[
P_n(\xi_1 \ldots \xi_n) = \frac{1}{(2\pi)^{n/2}} \Delta_n^{1/2} \exp\left(-\frac{1}{2} \sum_{i,j=1}^n M_{ij} \xi_i \xi_j\right) \quad \cdots (3.1.1)
\]

where \( M_{ij} \) is the \( ij \)th element of the matrix \([M]\), the inverse of the correlation matrix \([ \rho_{ij} ]\) and \( \Delta_n \) is the determinant of \([ \rho_{ij} ]\). Substituting expression (3.1.1) for \( P_n(\xi_1 \ldots \xi_n) \) in (2.10.5), we obtain:

\[
V_n = \int_0^\infty \cdots \int_0^\infty \frac{1}{(2\pi)^{n/2}} \Delta_n^{1/2} \exp\left[-\frac{1}{2} \sum_{i,j=1}^n M_{ij} \xi_i \xi_j\right] \cdot d\xi_1 \cdots d\xi_n \quad \cdots (3.1.2)
\]

This expression is usually referred to, in the literature, as the 'multivariate normal integral'. This chapter is mainly devoted to this integral and methods of its evaluation. The chapter also discusses further approximations to the distribution \( P_0(\xi) \) based on this integral.

3.2 Methods of evaluating the multivariate normal integral.

One method of evaluating the integral (3.1.2) is to
express the quadratic form $\sum_{i,j}^{n} M_{ij} \xi_i \xi_j$ as the sum of squares by means of a suitable linear transformation of variables. For this purpose, we write the form in matrix notation.

$$\sum_{i,j}^{n} M_{ij} \xi_i \xi_j = \xi^T \mathbf{M} \xi \quad \ldots (3.2.1)$$

where $\xi$ is the column vector $(\xi_1 \ldots \xi_n)$ and the prime denotes its transpose. By introducing the linear transformation

$$\xi = \mathbf{C} \eta \quad \ldots (3.2.2)$$

where $\mathbf{C}$ is an $n \times n$ square matrix satisfying the matrix relation

$$\mathbf{C}^T \mathbf{C} = \mathbf{I} \quad \ldots (3.2.3)$$

and $\mathbf{I}$ is the Unit Matrix, the integral (3.1.2) becomes:

$$V_n = \frac{1}{(2\pi)^{n/2} \Delta_n^{1/2}} \int_V \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \xi_i^2 \right) J(Y) \, dY_1 \ldots dY_n \quad \ldots (3.2.4)$$

Here $V$ denotes the new domain of integration bound by the hyperplanes

$$\xi_i = \sum_{j=1}^{n} C_{ij} \eta_j \geq 0; \ i = 1, 2 \ldots n \quad \ldots (3.2.5)$$

and $J(Y)$ is the Jacobian of the transformation, which is given by $\text{Det} \, \mathbf{C}$.

Now it can be easily shown that: (using 3.2.3)

$$\text{Det} \, \mathbf{C} = (\text{Det} \, \mathbf{M})^{-\frac{1}{2}} = \Delta_n^{\frac{1}{2}} \quad \ldots (3.2.6)$$

Therefore

$$V_n = \frac{1}{(2\pi)^{n/2} \Delta_n^{1/2}} \int_V \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \eta_i^2 \right) \, dY_1 \ldots dY_n \quad \ldots (3.2.7)$$

A second transformation which is given by

$$Y_i = r \left( \prod_{k=1}^{i-1} \sin \phi_k \right) \cos \phi_i; 1 \leq i \leq n-2$$

$$Y_{n-1} = r \left( \prod_{k=1}^{n-2} \sin \phi_k \right) \cos \theta \quad \ldots (3.2.8)$$

$$Y_n = r \left( \prod_{k=1}^{n-2} \sin \phi_k \right) \sin \theta$$
transforms the integral to:

$$V_n = (2\pi)^{-n/2} \int_0^\infty dr r^{n-1} \exp(-\frac{1}{2}r^2) \int_{S} \frac{n-2}{k=1} (\sin \phi_k)^{n-1-k} \cdot d\Theta \cdots d\phi_1 \cdots d\phi_{n-2} \quad \cdots(3.2.9)$$

where the region of integration $S$ will be discussed later.

The transformation (3.2.8) corresponds to changing the Cartesian coordinate system $Y_i, i=1, \ldots, n$, to a spherical system $(r, \Theta, \Phi_1, \ldots, \Phi_{n-2})$. Denoting the $(n-1)^{th}$ integral over $(\Theta, \Phi_1 \ldots \Phi_{n-2})$ by $S_n$ and integrating over $r$ yields for $V_n$ the following expression:

$$V_n = \frac{1}{2} (\pi)^{-n/2} \Gamma(n/2) \cdot S_n \quad \cdots(3.2.10)$$

where $\Gamma(n/2)$ is the Gamma function defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}dt \quad \cdots(3.2.11)$$

which for positive integer values of $x$ is simply factorial $(x-1)$

3.3. The geometric Interpretation of the multivariate integral.

In the previous section an expression for $V_n$ is obtained by means of two transformations. The first was a linear transformation whereas the second was effected by changing from Cartesian to Spherical co-ordinates. In the course of these transformations the region of integration has accordingly been mapped into new domains determined by the actual transformations. In this section we give a geometrical interpretation of the integral (3.1.2) and its consequent development, in order to assist understanding of this integral and the problems arising in its evaluation.

The integral $V_n$ written in the form (3.1.2) represents the integral of the joint density $P_n(\xi_1 \ldots \xi_n)$ over the region in the $n^{th}$ dimensional space bounded by the hyperplanes:
This region defines one of the \((2^n)\) subspaces into which the \(n\)th dimensional space is divided, by the \(n\) Cartesian co-ordinate planes. The linear transformation \(\xi = C\ Y\) whose main advantage is to simplify the quadratic form (3.2.1), also transforms the region of interaction (3.3.1) into the region given by (11)

\[
\sum_{j=1}^{n} C_{ij} Y_j \geq 0 ; i=1,2... n \quad ...(3.3.2)
\]

This defines a solid angle \(V\) in \(n\) dimensions bounded by the hyperplanes

\[
\sum_{j=1}^{n} C_{ij} Y_j = 0 ; i=1,2...n \quad ...(3.3.2a)
\]

The interior angles between the hyperplanes \(\xi_i = 0\) are right angles. In contrast those existing between the hyperplanes (3.3.2a) are not necessarily so. The transformation \(\xi = C\ Y\) in simplifying the quadratic form does this at the expense of increasing the complexity of the region of integration.

The change to spherical co-ordination allows radial integration over \(r\) to be performed and the reduction of the integral \(V_n\) to an \((n-1)\)th order integral over the angles \((\phi, \phi_1... \phi_{n-2})\). This integral (denoted by \(S_n\)) is the \((n-1)\) dimensional content of the intersection of the region \(V\) with the surface of the unit hypersphere having centre at the origin. For \(n=2\), this intersection is the arc on the unit circle subtending angle \(\Theta_{12}\) contained by the two straight lines (see Fig.3a), whose equations are:

\[
C_{11} Y_1 + C_{12} Y_2 = 0 \quad ...(3.3.3)
\]

For \(n=3\), \(S_3\) is the area of the surface on the unit sphere
Fig. (3) - Geometric interpretation of $S_n$.

(a) $S_2$; Arc on Unit Circle.

(b) $S_3$; Spherical triangle on a Unit Sphere, dotted lines represent the intersections of the planes (3.3.4) with surface of the sphere.
which subtends the solid angle contained by the three planes (see Fig. 3b), given by

\[
\begin{align*}
C_{11} Y_1 + C_{12} Y_2 + C_{13} Y_3 &= 0 \\
C_{21} Y_1 + C_{22} Y_2 + C_{23} Y_3 &= 0 \\
C_{31} Y_1 + C_{32} Y_2 + C_{33} Y_3 &= 0
\end{align*}
\] (3.3.4)

Using vector algebra it can be shown that:

\[
S_2 = \Theta_{12} = \cos^{-1}(-\rho_{12}) \quad \cdots (3.3.5)
\]

As for $S_3$ this is the area of spherical triangle drawn on a unit sphere. This area is given by:

\[
S_3 = \Theta_{23} + \Theta_{31} + \Theta_{12} - \Pi \quad \cdots (3.3.6)
\]

where the $\Theta$s are the interior angles between the three planes (3.3.4).

For general values of $n$, $S_n$ is obviously a function of the $\frac{n(n-1)}{2}$ interior angles between the $n$ bounding hyperplanes (3.3.2a). Again by using vector algebra, it can be shown that the interior angle $\Theta_{ij}$ between the $i^{th}$ and $j^{th}$ hyperplane is given by:

\[
\cos \Theta_{ij} = -\sum_{k=1}^{n} \frac{C_{ik} C_{jk}}{(\sum_{k=1}^{n} C_{ik}^2) \left(\sum_{k=1}^{n} C_{jk}^2\right)^{\frac{1}{2}}} \quad \cdots (3.3.7)
\]

where $0 \leq \Theta_{ij} \leq \Pi$.

For $n > 3$, the region $S_n$ becomes the content of what is known in geometry as the 'spherical simplex' which is the generalisation of the spherical triangle (corresponding to $n=3$). Geometers have studied the problem of expressing the content of the spherical simplex in terms of the angles between its bounding surfaces. However, it would appear that for $n > 3$, closed-form expressions do not yet exist. Series expansions, upper and lower bounds can be written for $S_n$ either directly, using the geometrical properties of the spherical simplex, or indirectly, using the probability
theory to write these for $V_n$ first and then $S_n$, since these are related by (3.2.10). D. Slepian (13) obtained several upper and lower limits for $V(\mathcal{C})$ (the limit of $V_n$ as $n$ tends to $\infty$). He also listed many references dealing with the geometry of the $n^{th}$ dimensional simplex (ref. 33-35). He pointed out that the geometrical picture suggests a large number of upper and lower bounds but that unfortunately none has been found, which in the limit, yield useful results for the probability $V(\mathcal{C})$ or $P_0(\mathcal{C})$.

The quantity $S_n$ is referred to sometimes as the Schlaffli (15) function after Schaffli who first introduced the spherical simplex, in 1858.

3.4 Series Expansions for the multivariate integral.

The following probability expression:

$$P_2(\xi_1,\xi_2) = \frac{1}{2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2(1-\rho^2)} \cdot \left(\xi_1^2 - 2\rho\xi_1\xi_2 + \xi_2^2\right)\right] \quad \ldots (3.4.1)$$

represents the joint probability density of two Gaussian variables $\xi_1, \xi_2$ of zero mean, unity variance and correlation coefficient $\rho$. Next, if the polynomial $h_n(\xi)$ of degree $n$ is defined by

$$h_n(\xi) = e^{-\xi^2/2} \cdot (-1)^n \cdot \frac{d^n}{d\xi^n} \left(e^{-\xi^2/2}\right) \quad \ldots (3.4.2)$$

it can be shown that

$$P_2(\xi_1,\xi_2) = P_1(\xi_1) \cdot P_1(\xi_2) \sum_{n=0}^{\infty} \frac{\rho^n}{n!} \cdot h_n(\xi_1)h_n(\xi_2) \quad \ldots (3.4.3)$$

where

$$P_1(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} \quad \ldots (3.4.4)$$

The polynomial $h_n(\xi)$ is the Hermite polynomial of degree $n$. 
n which is orthogonal over the interval $-\infty$ to $\infty$ with respect to the Gaussian distribution (3.4.4). This means that

$$
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h_n(\xi) h_m(\xi) e^{-\frac{\xi^2}{2}} \, d\xi = n! \delta_{mn} \ldots (3.4.5)
$$

where $\delta_{mn}$ is the Kronecker delta.

The relation (3.4.3) is normally referred to as 'Mehler's formula', which applies to Gaussian distributions. The formula is very useful from a computational and mathematical point of view since it converts the somewhat complicated form (3.4.1) into a series whose terms consist of products of functions of single variables. The use of orthogonal polynomials has resulted in separating the variables and rendered the function $P_2(\xi_1, \xi_2)$ readily integrable over $\xi_1$ and $\xi_2$. This is a significant result as far as the evaluation of the multivariate integral is concerned and is therefore worthwhile investigating the possibility of extending it for higher order probability densities. The first step is to investigate the possibility of extending Mehler's result to include third and higher order probability densities. This was carried out (16) and it was found that $P_3(\xi_1, \xi_2, \xi_3)$ is given by:

$$
P_3(\xi_1, \xi_2, \xi_3) = \sum_{l=1}^{3} P_1(\xi_l) \sum_{i, j, k, \epsilon=0}^{\infty} \frac{\rho_{12}^{i} \rho_{13}^{j} \rho_{23}^{k}}{i! j! k!} \cdot h_r(\xi_1) h_s(\xi_2) h_t(\xi_3) \ldots (3.4.6)
$$

where

$$
\begin{align*}
 r &= i + j \\
 s &= i + k \\
 t &= j + k
\end{align*}
$$

A general expression for the density $P_n(\xi_1 \ldots \xi_n)$ was
also derived. Before writing this general expression it is useful to make the following points:

i) The order of the multiple series in the general expression is given by \( \frac{1}{2}(n^2-n) \), which is half the number of the non-diagonal elements in the correlation matrix \( \rho_{ij} \). This is also the number of the interior angles \( \Theta_{ij} \), defined in the last section.

ii) In writing the general expression, the letter (i) with subscripts 1, 2, ..., \( \frac{1}{2}(n^2-n) \) is used for the summing indices. The letter (r) subscripted 1, 2, ..., n is used to indicate the order of the associated Hermite polynomials. Bearing this in mind, the general expression can be written thus:

\[
P_n(\xi_1 \cdots \xi_n) = \prod_{l=1}^{n} P_l(\xi_l) \sum_{i_1 \cdots i_n=0}^{\infty} \frac{(\rho_{12})^{i_1} \cdots (\rho_{1n})^{i_{n-1}} (\rho_{23})^{i_n} \cdots (\rho_{n-1,n})^{i_N}}{(i_1)! \cdots (i_{n-1})! (i_n)! \cdots (i_N)!}
\]

\[
\cdot h_{r_1}(\xi_1) \cdot h_{r_2}(\xi_2) \cdots h_{r_s}(\xi_s) \cdots h_{r_n}(\xi_n)
\]

\[\ldots(5.4.8)\]

where \( n = \frac{1}{2}(n^2-n) \)

and \( r_s (s=1,2, \ldots n) \) is the sum of the elements in the \( s^{th} \) row of the following symmetric matrix:

\[
\begin{bmatrix}
0 & i_1 & i_2 & \cdots & i_{n-1} \\
0 & i_n & & & i_{2n-3} \\
0 & i_{2n-2} & & & i_{3n-6} \\
& & \ddots & \ddots & \ddots \\
& & & & i_N \\
& & & & 0
\end{bmatrix}
\]
This matrix is constructed in order to relate the subscripts \( r_s; s=1,2,.. n \) to the summing indices \( i_1...i_N \). To construct this matrix we note that the diagonal elements are of no significance and so can be written as zeros. Starting at row one, and working down the rows, we write the indices \( i_1...i_N \) in ascending order; making sure to start each row at the element immediately to the right of the principal diagonal. Because of symmetry, the elements to the left of the diagonal can now be inserted and hence the whole of the matrix (2.4.9) is obtained. The power to which the correlation coefficient \( \rho_{ij} \) is raised in the general expression (2.4.8) can be found from this matrix. It is simply the \((ij)\)th element of the matrix (2.4.9). This matrix is therefore very useful when writing the general expression (2.4.8).

The general expression (2.4.8) expresses \( P_n(\xi_1...\xi_n) \) as a multiple series and makes it more readily integrable than the form (3.1.1). The multivariate integral \( V_n \) is now transformed into a multiple series whose terms are products of one dimensional integrals. A typical such integral is of the form:

\[
I_{r_s} = \int h_{r_s} (\xi_s) P_1(\xi_s) \, d\xi_s \quad \cdots (3.4.10)
\]

Hence, by using (3.4.2) we obtain:

\[
I_{r_s} = \lim_{s \to 0} h_{r_s-1} (\xi_s) P_1(\xi_s) \quad \cdots (3.4.11)
\]

From the properties of the Hermite Polynomials we find that:

\[
I_{r_s} = 0 \quad ; \quad r_s \text{ even}
\]
\[
I_{rs} = \frac{(-1)^2 (r_s^{-1})!}{\sqrt{2\pi} \left( \frac{r_s^{-1}}{2} \right)! \left( \frac{r_s^{-1}}{2} \right)} \quad ; \quad r_s \text{ odd} \quad \ldots(3.4.12)
\]

Hence \( V_n \) can be expressed as a multiple series of order \( \frac{1}{2} n(n-1) \) whose terms are products of quantities of the form \((3.4.12)\). For \( n=2 \) the above approach leads to:

\[
V_2 = \frac{1}{4} + \frac{1}{2\Pi} \left( \rho + \frac{\rho^3}{3!} + \frac{9\rho^5}{5!} \right) \quad \ldots(3.4.13)
\]

which is identical to

\[
V_2 = \frac{1}{4} + \frac{1}{2\Pi} \sin^{-1}(\rho) \quad \ldots(3.4.14)
\]

For \( n=3 \) it was found that the only values which the summing indices \( i_1, i_2, i_3 \) can assume in order to obtain non-vanishing terms are the following:

\[
\begin{align*}
&i_1 = i_2 = i_3 = 0 \quad \ldots(3.4.15) \\
i_1 \text{ odd}; \quad i_2 = i_3 = 0 \\
i_2 \text{ odd}; \quad i_1 = i_3 = 0 \\
i_3 \text{ odd}; \quad i_2 = i_1 = 0
\end{align*}
\]

This reduces the series to the sum of a constant term and three single series of the form similar to that in \((3.4.13)\) involving \( \rho_{12}, \rho_{13}, \) and \( \rho_{23} \), namely:

\[
V_3 = \frac{1}{8} + \frac{1}{4\Pi} \left[ \sin^{-1}(\rho_{12}) + \sin^{-1}(\rho_{13}) + \sin^{-1}(\rho_{23}) \right] \quad \ldots(3.4.17)
\]

\((3.4.14)\) and \((3.4.16)\) may also be written as

\[
\begin{align*}
V_2 &= \frac{1}{2\Pi} \cos^{-1}(-\rho) \quad \ldots(3.4.17) \\
V_3 &= \frac{1}{4\Pi} \left( \cos^{-1}(-\rho_{12}) + \cos^{-1}(-\rho_{13}) + \cos^{-1}(-\rho_{23}) \right)
\end{align*}
\]

which are the same as obtained previously (see \((3.3.5)\)& \((3.3.6)\) in conjunction with \((3.3.7)\)).

For \( n > 3 \) however, it does not seem possible to reduce the multiple series into a summation of simple series as in \((n=3)\). Consequently, \( V_n \) cannot be expressed in a closed
form. The above approach which allowed the integration
to be carried out readily, could only yield series solutions
for $V_n$ in the general case. For general $V_n$ it was found that

$$V_n = \frac{1}{2^n} + \frac{1}{2^{n-1}n!} \sum_{1 \leq i < j}^n \sin^{-1}(\rho_{ij}) + \frac{1}{2^{n-2}n!^2} \sum_{1 \leq i < j < k}^n (\rho_{ij} \rho_{kl} + \rho_{ik} \rho_{jl} + \rho_{il} \rho_{jk}) \cdots (3.4.18)$$

+ terms containing higher order products of
correlation coefficients.

e.g. for $n=4$

$$V_4 = \frac{1}{16} + \frac{1}{8} \sum_{1 \leq i < j}^4 \sin^{-1}(\rho_{ij}) + \frac{1}{4n!^2} (\rho_{12} \rho_{34} + \rho_{13} \rho_{24} \cdots + \rho_{14} \rho_{23}) \cdots (3.4.19)$$

An attempt was made to evaluate the limit of $V_n$ as $n \to \infty$
using the multiple series solution. As stated earlier, this
limit (if it exists) is equivalent to the probability
density $V(\zeta)$ which is related to $P_0(\zeta)$ by (2.10.2). Un-
fortunately, in taking the limits of the first few terms
in the multiple series it was found that some terms tended
to zero whereas others tended to infinity. It thus became
clear that meaningful results for the limit were not
possible using the series solution.

The rapidity of the convergence of the multiple series
depends primarily on the values of the correlation co-
efficients $\rho_{ij}$. If these are relatively small, then only
the first few terms (say those indicated in (3.4.18)) need
be evaluated to obtain a fairly accurate estimate for $V_n$.
If, on the other hand, these are relatively high, then a
larger number of terms needs to be evaluated for reasonable
accuracy. In either case, the multipleseries is useful in providing an estimate of the multivariate integral $V_n$. In addition the series is very useful in the calculation of product moments of multidimensional Gaussian distributions\(^{(16)}\).

Finally it should be noted that the results contained in this section were obtained independently by the author. It was afterwards found that similar approaches and results were obtained earlier by Aitken, Moran & Kendall\(^{(17,19)}\).

3.5 McFadden's Urn Model Approximation.

The mathematical difficulties arising in the two methods of evaluating the multivariate integral discussed above make it necessary to seek approximate results. As stated in section(3.4), the first terms of the multiple series can provide an approximation to the integral $V_n$ whose accuracy is largely dependent on the values of the correlation coefficients.

In this section, another method leading to approximate results is discussed. The method is based on a generalisation of the urn scheme first introduced by Polya\(^{(20)}\). In this method an urn contains initially equal numbers of black and red balls and the probability that the first $n$ draws are all of the same colour is interpreted as a function of the correlation coefficient between successive outcomes which is assumed equal throughout. The generalised model, due to McFadden\(^{(21)}\), allows for correlation between pairs of outcomes which may differ from pair to pair. The probability of successive draws being of the same colour was then considered and compared to some results already
available for the multivariate normal integral. The comparison suggest a possible approximation which, according to McFadden appears to be useful only in the Polya case where the correlation coefficients are all equal.

Even then it is considered useful to discuss these models briefly (referring the reader to the original contribution for details) in order to provide some more insight into this problem.

3.5.1 Polya's Urn Scheme.

Initially, the urn contains $a$ red balls, and $b$ black balls. Successive drawings are performed, with replacements and with the provision that $\Delta$ extra balls are added after each draw, all of the same colour as the ball most recently drawn. $\Delta$ may be negative but it must obey the inequality:

$$a + (n-1) \Delta \geq 0 \quad \ldots(3.5.1)$$

where $n$ is the total number of draws, in order that neither colour may be overdrawn. The probability of drawing $n$ black balls in the first $n$ trials is given by

$$P_n = \frac{(a/\Delta)_n}{(2a/\Delta)_n} \quad \ldots(3.5.2)$$

where

$$(\alpha)_n = \alpha(\alpha+1) \ldots(\alpha+n-1), \quad \text{and} \quad (\alpha)_0 = 1 \quad \ldots(3.5.3)$$

Let $x_i = +1$ if the $i^{th}$ draw is black and let $x_i = -1$ if the $i^{th}$ draw is red. The correlation coefficient between $x_i$ and $x_j (i \neq j)$, is given by

$$r_{ij} = \frac{\Delta}{2a + \Delta} \quad \ldots(3.5.4)$$

which is equal for all $i$ and $j$.

3.5.2 The generalised Urn Scheme.

Initially, the urn contains $a$ black balls and $b$ red
balls. Instead of adding all balls after each draw, this scheme makes use of a matrix of elements $\Delta_{ij}$. One ball is drawn and replaced, and $\Delta_{12}$ balls are added of the colour drawn. Again one ball is drawn and replaced, then $\Delta_{13}$ are added of the first colour drawn and $\Delta_{23}$ of the second colour drawn. After the $(k-1)^{th}$ draw (and replacement), $\Delta_{1k}$ are added of the first colour drawn, $\Delta_{2k}$ of the second, and $\Delta_{k-1,k}$ of the $(k-1)^{th}$ colour drawn.

Let

$$D_{mk} = \sum_{i=m+1}^{k} \Delta_{mi}; \quad m=1,2,\ldots,k-1 \quad \ldots(3.5.5)$$

$D_{mk}$ is then the total number of balls which have been added, up to the time immediately preceding the $k^{th}$ draw, as a result, and of the colour of the $m^{th}$ draw.

To prevent overdrawing of either colour, the inequality:

$$a + \sum_{m=1}^{k-1} D_{mk} \geq 0 \quad \ldots(3.5.6)$$

must be satisfied. Again let $x_i = +1$ if the $i^{th}$ draw is black, and $x_i = -1$ if the $i^{th}$ draw is red. McFadden showed that:

$$r_{in} = E(x_i x_n) = \frac{\sum_{j=1}^{n-1} r_{ij} D_{jn}}{2a + \sum_{j=1}^{n-1} D_{jn}}; \quad i=1,2,\ldots,n-1 \quad \ldots(3.5.7)$$

where $r_{ii} = 1$ for all $i$. Notice that if all correlation coefficients are known for the first $(n-1)$ draws, then equation $(3.5.7)$ gives the remaining coefficients necessary to correlate the $n^{th}$ draw.
Defining a new variable $G_{kn}$ as

$$G_{kn} = \frac{D_{kn}}{2a + \sum_{j=1}^{n-1} D_{jn}} \quad k=1,2...n-1 \quad \ldots(3.5.8)$$

$(3.5.7)$ may be written as

$$r_{in} = \sum_{j=1}^{n-1} r_{ij} G_{jn} \quad i=1,2...n-1 \quad \ldots(3.5.9)$$

For the ratio $P_n/P_{n-1}$ McFadden derived the following relationship:

$$-1 = \sum_{j=1}^{n-1} G_{jn} - \frac{P_n}{P_{n-1}} \quad \ldots(3.5.10)$$

$(3.5.9)$ and $(3.5.10)$ constitute $n$ equations in the $n$ unknowns $G_1n, G_2n ... G_{n-1}, n$ and $2P_n/P_{n-1}$. Solving this system of equations for the last quantity yields:

$$\frac{P_n}{P_{n-1}} = \frac{1 + r_{1n}}{r_{12} + r_{1n}} \frac{r_{12} + r_{2n}}{1 + r_{2n}} \cdots \frac{r_{1,n-1} + r_{1n}}{1 + r_{1,n-1}} \cdots \frac{r_{1,n-1} + r_{2,n-1}}{1 + r_{2,n-1}} \cdots \frac{1 + r_{2,n-1}}{1 + r_{2,n-1}} \cdots \frac{1 + r_{n-1,n}}{1 + r_{n-1,n}} \quad , \quad n=2,3... \quad (3.5.11)$$

The correlation coefficients may now be regarded as continuous rather than discrete. The coefficients may assume any values between (-1) and (+1) which do not violate
(3.5.6), i.e. which do not lead to negative probabilities. By comparison with equation (3.5.10) it follows that the inequality (3.5.6) may be rewritten as
\[ P_k / P_{k-1} \geq 0 \quad k=2,3,...,n \quad \text{...(3.5.12)} \]
Finally since \( P_n \) is given by
\[ P_n = \frac{P_2}{P_1} \cdot \frac{P_3}{P_2} \cdots \frac{P_{n-1}}{P_{n-2}} \cdot \frac{P_n}{P_{n-2}} \cdot P_1 \quad \text{...(3.5.14)} \]
and \( P_1 = \frac{1}{2} \).
Then by using (3.5.11) we have
\[
\begin{bmatrix}
1 + r_{1n} & r_{12} + r_{13} & \cdots & r_{1,n-1} + r_{1n} \\
r_{12} + r_{23} & 1 + r_{23} & \cdots & r_{2,n-1} + r_{23} \\
\vdots & \vdots & \ddots & \vdots \\
r_{1,n-1} + r_{n-1,n} & r_{2,n-1} + r_{n-1,n} & \cdots & 1 + r_{n-1,n}
\end{bmatrix}
\]
\[ 2^{-n} (1+r_1^2) r_1 r_2 \cdots r_n \quad \text{...(3.5.14)} \]

3.5.3 Comparison with the multivariate normal integral.

As explained earlier, this integral represents the probability that \( n \) mutually correlated Gaussian variates with mean zero and correlation matrix \( \rho_{ij} \) are simultaneously positive (see eq.(3.1.2)). By means of the transformation:
\[ x_i = +1 \text{ if } \xi_i > 0 \\
-1 \text{ if } \xi_i < 0 \quad i=1,2,...,n \quad \text{...(3.5.15)} \]
the above problem reduces to the probability that the \( n \) discrete random variables \( x_1, x_2, \ldots, x_n \) are simultaneously
equal to +1. The mean values of these variates are zero and their correlation matrix $[r_{ij}]$ is related to the matrix $[\rho_{ij}]$. For Gaussian processes the correlation coefficient $r_{ij}$ is related to the corresponding $\rho_{ij}$ by the well-known arc sine law.

$$r_{ij} = \frac{2}{\pi} \sin^{-1} \rho_{ij} \quad \ldots(3.5.16)$$

the transformation (3.5.15) means that we can immediately apply the results of the urn schemes to the problem in the discrete variates $x_i$. Then, by using (3.5.16), the multivariate integral may be expressed as a function of the correlation coefficients $\rho_{ij}$. This is demonstrated below.

For $n=2$ and $n=3$ respectively, (3.5.14) yields for $P_2$ and $P_3$

$$P_2 = \frac{1}{8}(1+r_{12}) \quad \ldots(3.5.17)$$

$$P_3 = \frac{1}{8}(1+r_{12}+r_{13}+r_{23})$$

This together with (3.5.16) gives

$$P_2 = \frac{1}{8}(1+\sin^{-1} \rho_{12}) \quad \ldots(3.5.18a)$$

$$P_3 = \frac{1}{8}(1+\sin^{-1} \rho_{12}+\sin^{-1} \rho_{13}+\sin^{-1} \rho_{23}) \quad \ldots(3.5.18b)$$

Relations (3.5.18a & b) are exactly the same expressions as those obtained earlier for the bivariate and trivariate normal integral using the geometrical approach or the multiple series method.

This suggests that the urn model can provide some results for the multivariate case. McFadden however, stated that for $n \geq 3$ and arbitrary $\rho_{ij}$ the agreement between the power series obtained by expanding (3.5.14) and that obtained by Aitken, Kendall and Moran (see end of last section) does not extend beyond the second order terms. McFadden
compared the 'urn' approximation with several known results for the multivariate integral. The following points emerged:

i) The closed form (3.5.14) is exact when \( n=2,3 \) for arbitrary \( \rho_{ij} \). It is also exact for all \( n \) when \( \rho_{ij} \) are all equal to 0, 1 or \( \frac{1}{2} \).

ii) The approximation is generally good if the correlation coefficients \( \rho_{ij} \) are all equal and small. For a given value of this equal correlation coefficient, the approximation grows steadily worse as \( n \) increases.

iii) In the general case of unequal \( \rho_{ij} \), the results are much less satisfactory. For \( n>3 \), (3.5.14) is not symmetric with respect to interchange of indices. Different results are possible if the outcome of the \( n^{th} \) draw was represented by the discrete variable \( x_2 \) say instead of \( x_n \).

Finally, McFadden derived an expression for \( P_2 \) in terms of the product moments of the variates \( x_1 \ldots x_n \) which are defined by

\[
E(x_1, x_2, \ldots, x_n) = \sum_{y_1 y_2 \ldots y_n} y_1 y_2 \ldots y_n P(x_1 = y_1, x_2 = y_2, \ldots, x_n = y_n)
\]

where the sums are taken over all the combinations of \( y_1 = \pm 1, y_2 = \pm 1 \ldots y_n = \pm 1 \). If the symmetry of the distribution is such that moments of odd order are zero, then \( P_n \) can be expressed in the form:

\[
P_n = 2^{-n} \left[ 1 + \sum_{1 \leq i < j} E(x_i x_j) + \sum_{1 < k < j} E(x_i x_j x_k) \ldots \right]
\]

ending with a product moment of order \( n \), if \( n \) is even or with a moment of order \( (n-1) \) if \( n \) is odd.

Now it should become clear why \( P_2 \) and \( P_3 \) could be
obtained exactly from the urn model, whereas for higher $P_n$ the model fails to be exact. The bivariate and trivariate integrals do not involve product moments higher than the second, since (3.5.20) terminates at the second product moment. On the other hand, higher order integrals involve higher product moments than the second. In applying the result of the urn model to the multivariate integral, it is assumed that there is a one to one correspondence between the correlation matrix of the urn model and that of the multivariate integral. This means that the two distributions are identical as far as their second order product moments are concerned. However, this does not mean that their corresponding higher product moments are equivalent. Hence $P_2$ and $P_3$ are given exactly whereas higher $P_n$ can at best be given approximately.

Exact expressions for higher product moments of infinitely clipped Gaussian processes are again difficult to evaluate. McFadden (22) obtained some reliable results for the fourth product moment $E(x_1x_2x_3x_4)$ but in so doing he found it necessary to consider non-Gaussian processes.

3.6 Approximations for $P_0(\tau)$ based on the Schlaffli Function $S_n$

This section deals with a new set of approximations to the probability distribution $P_0(\tau)$ based on the Schlaffli function $S_n$ (see section 3.2). Although it was mentioned there that no closed form expressions exist for $S_n(n > 3)$, it is possible, through the use of a special reduction formula, to obtain approximations for $P_0(\tau)$ based on $S_4$ & $S_5$. Before explaining this approach, a discussion on the reduction formula is necessary and is
given first.

The multi-dimensional probability density \( P_n(\xi_1...\xi_n) \) may be written as the inverse Fourier transform of its characteristic function:

\[
P_n(\xi_1...\xi_n) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\mathbf{U}^T \mathbf{\xi} - \frac{1}{2} \mathbf{U}^T \mathbf{\Sigma} \mathbf{U}\right) dU_1 \cdots dU_n \tag{3.6.1}
\]

where \( \mathbf{U} \) & \( \mathbf{\xi} \) are column vectors, and \( \mathbf{\Sigma} \) is the correlation matrix \( \mathbf{\Sigma} = \sqrt{-1} \).

Differentiating (3.6.1) with respect to the element \( \rho_{ij} \) of the matrix \( [\rho] \) we obtain

\[
\frac{\partial P_n}{\partial \rho_{ij}} = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} U_i U_j \exp\left(-\mathbf{U}^T \mathbf{\xi} - \frac{1}{2} \mathbf{U}^T \mathbf{\Sigma} \mathbf{U}\right) dU_1 \cdots dU_n \tag{3.6.2}
\]

Now differentiating \( P_n \) twice: first with respect to \( \xi_i \) and then \( \xi_j \) we obtain

\[
\frac{\partial^2 P_n}{\partial \xi_i \partial \xi_j} = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} U_i U_j \exp\left(-\mathbf{U}^T \mathbf{\xi} - \frac{1}{2} \mathbf{U}^T \mathbf{\Sigma} \mathbf{U}\right) dU_1 \cdots dU_n \tag{3.6.3}
\]

Since the right hand sides of (3.6.2) & (3.6.3) are identical, we obtain the following system of partial, differential equations

\[
\frac{\partial^2 P_n}{\partial \xi_i \partial \xi_j} = \frac{\partial^2 P_n}{\partial \xi_j \partial \xi_i} \quad \text{for all } i > j \tag{3.6.4}
\]

Applying (3.6.4) with \( i=1, j=2 \) to the multivariate integral \( V_n \) we obtain

\[
\frac{\partial V_n}{\partial \rho_{12}} = \int_0^\infty \int_0^\infty \frac{\partial^2}{\partial \xi_1 \partial \xi_2} \left( P_n(\xi_1...\xi_n) \right) d\xi_1 d\xi_2 \cdots d\xi_n \tag{3.6.5}
\]
Hence it can be shown that

\[
\frac{\partial V_n}{\partial \rho_{12}} = \int_0^\infty \int_0^\infty P_n(0,0, \xi_1, ... \xi_n) d\xi_1 ... d\xi_n \quad \text{(3.6.6)}
\]

By expressing \( P_n(\xi_1, ..., \xi_n) \) as

\[
P_n(\xi_1, ..., \xi_n) = P_2(\xi_1, \xi_2) \cdot P(\xi_3, ..., \xi_n/\xi_1, \xi_2) \quad \text{(3.6.7)}
\]

where \( P(\xi_3, ..., \xi_n/\xi_1, \xi_2) \) is the conditional probability density of the variates \( \xi_3, ..., \xi_n \) given \( \xi_1, \xi_2 \), and evaluating (3.6.7) at \( \xi_1 = \xi_2 = 0 \), we find

\[
P_n(0,0, \xi_3, ..., \xi_n) = \frac{1}{2\pi \sqrt{1-\rho_{12}^2}} P_{n-2}(\xi_3, ..., \xi_n; \rho^{(12)}) \quad \text{(3.6.8)}
\]

where \( P_{n-2}(\xi_3, ..., \xi_n; \rho^{(12)}) \) is the \((n-2)\)th dimensional normal probability density of the variates \( \xi_3, ..., \xi_n \) with a modified correlation matrix \( \rho^{(12)} \). \( \rho^{(12)} \) is discussed below. This together with (3.6.6) gives

\[
\frac{\partial V_n}{\partial \rho_{12}} = \frac{1}{2\pi \sqrt{1-\rho_{12}^2}} \cdot V_{n-2}(\rho^{(12)}) \quad \text{(3.6.9)}
\]

The above result can be generalised for arbitrary indices and we obtain

\[
\frac{\partial V_n}{\partial \rho_{i,j}} = \frac{1}{2\pi \sqrt{1-\rho_{i,j}^2}} \cdot V_{n-2}(\rho^{(i,j)}) \quad \text{(3.6.10)}
\]

As for the modified correlation matrix \( [\rho^{(i,j)}] \), this is derived from \([M] \), the inverse of the correlation matrix \([\rho] \) by first deleting the \( i \)th & \( j \)th rows and columns of \([M][\rho^{(i,j)}] \) is then the inverse of the resulting \((n-2)x(n-2)\) matrix. D. Slepian (13) points out that the elements of the matrix \( [\rho^{(i,j)}] \) are what is normally referred to, in statistics, as **Partial correlation**.
coefficients. Slepian expressed these elements in terms of the complete correlation coefficients $\rho_{ij}$, and showed that the $(rs)^{th}$ element of the modified matrix $[\rho_{ij}]$ is given by $(r,s \neq i,j)$

$$
\rho_{rs} = \begin{vmatrix}
\rho_{rs} & \rho_{ri} & \rho_{rj} \\
\rho_{is} & 1 & \rho_{ik} \\
\rho_{js} & \rho_{ij} & 1
\end{vmatrix}
$$

...(3.6.11)

Using the relation between $S_n$ & $V_n$, an alternative way of writing (3.6.10) is:

$$
\frac{\partial S_n}{\partial \rho_{ij}} = \frac{1}{(n-2)\sqrt{1-\rho_{ij}^2}} \cdot S_{n-2}(\rho_{ij})
$$

...(3.6.12)

Relation (3.6.12) is the reduction formula connecting the simplex $S_n$ to the $(n-2)$ dimensional simplex corresponding to the $(n-2)\times(n-2)$ matrix $[\rho_{ij}]$.

In terms of the interior angles $\theta_{ij}$ between the $i^{th}$ & $j^{th}$ hyperplanes bounding the simplex $S_n$, (3.6.12) may be written as:

$$
\frac{\partial S_n}{\partial \theta_{ij}} = \frac{1}{n-2} \cdot S_{n-2}(\rho_{ij}); \; n \geq 4
$$

...(3.6.13)

the quantity (3.6.11) can now be interpreted as the minus cosine of the $(rs)^{th}$ angle of the simplex $S_{n-2}(\rho_{ij})$.

The reduction formula (3.6.13) was first proved by Schlaflii (1858) who dealt with simplexes and obtained many relationships connecting their interior angles. In statistics these relationships take the form of algebraic identities connecting partial correlation coefficients.

R.L. Plackett (23) started from the system of partial
differential equations (3.6.4) and proceeded to obtain the reduction formula (3.6.10). Plackett however, did not relate the matrix $[\rho^{(ij)}]$ to the correlation matrix $[\rho_{ij}]$. M.S. Longuet-Higgins later gave a similar formula to (3.6.11) relating the interior angles of the simplex $S_{n-2} [\rho^{(ij)}]$ to the elements of the correlation matrix. D. Slepian in obtaining the reduction formula previously obtained by Flackett, identified the cosines of these interior angles as the partial correlation coefficients which are well known quantities in statistics.

The relation (3.6.13) is very useful since it reduces the number of consecutive integrations required in evaluating $S_n$ to $\frac{1}{2}n$ or $\frac{1}{2}(n-1)$ according to whether $n$ is even or odd. For example, $S_2$ involves essentially one integration since:

$$\theta_{12} = \cos^{-1}(-\rho_{12}) = \int_{-\rho_{12}}^{1} \frac{dx}{(1-x^2)^{\frac{1}{2}}} \quad \ldots (3.6.14)$$

Also (3.6.13) allows the evaluation of the partial derivatives of $S_n$ for $n$ as high as 5. The significance of this is demonstrated below in discussing further approximations to the distribution $P_0(\zeta)$.

We recall that $P_0(\zeta)$ is related to $V(\zeta)$ by:

$$P_0(\zeta) = \frac{d^2V(\zeta)}{d\zeta^2} \quad \ldots (3.6.15)$$

where

$$V(\zeta) = \lim_{n \to \infty} V_n(t_1, t_2 \ldots t_n) \quad \text{and} \quad t_n - t_1 = \zeta$$

It is not possible at this stage to obtain a meaningful result for this limit. However, by taking $n$ to be finite, successive approximations to $V(\zeta)$ and hence $P_0(\zeta)$ may be obtained. Longuet-Higgins obtained the following two sets of approximations.
\[ P_{\text{on}}(\zeta) = -\frac{2}{N_0} \frac{\partial^2}{\partial t_i^2} \cdot V_n(t_1 \ldots t_n) \ldots (3.6.15a) \]

and

\[ P_{\text{on}}(\zeta) = \frac{2}{N_0} \frac{\partial^2}{\partial t_i^2} \cdot V_n(t_1 \ldots t_n) \ldots (3.6.15b) \]

In the first set, \( P_{\text{on}}(\zeta) \), \( V_n(t_1 \ldots t_n) \) is being differentiated partially with respect to \( t_1 \) and \( t_n \) only, \( t_2 \ldots t_{n-1} \) being kept constant. In the second, each of \( t_1 \ldots t_n \) is considered to be a function of \( \zeta \), say \( t_i = \frac{i\zeta}{n-1} \) corresponding to equispaced points. Differentiation with respect to \( \zeta \) then involves all the points.

3.6.1 The approximation \( P_{\text{on}}(\zeta) \).

From (3.2.10) and (3.6.15a) we obtain:

\[ P_{\text{on}}(\zeta) = -\frac{\sqrt{(n/2)}}{N_0 \sqrt{n/2}} \frac{\partial^2 S_n}{\partial t_1 \partial t_n} \ldots (3.6.16) \]

The first interesting case is \( n=3 \), here we obtain:

\[ P_{03}(\zeta) = -\frac{1}{2N_0} \frac{\partial^2}{\partial t_1 \partial t_3} (\Theta_{12} + \Theta_{13} + \Theta_{23} - \Pi) \ldots (3.6.17) \]

Since only \( \Theta_{13} \) depends on both \( t_1 \) and \( t_3 \) and because:

\[ \Theta_{13} = \cos^{-1}(-\rho_{13}) = \cos^{-1}(-\rho(\zeta)) \ldots (3.6.18) \]

it follows that

\[ P_{03}(\zeta) = \frac{1}{2N_0} \frac{\partial^2}{\partial t_1^2} \cos^{-1}(-\rho(\zeta)) \ldots (3.6.19) \]

Hence

\[ P_{03}(\zeta) = \frac{1}{2N_0} \left[ \rho'(\zeta) \left[ 1 - \rho^2(\zeta) \right] + \rho(\zeta) \rho'(\zeta) \left[ 1 - \rho^2(\zeta) \right]^{3/2} \right] \ldots (3.6.20) \]

For \( n > 3 \) we recall that \( S_n \) is a function of the angles \( \Theta_{ij} \), therefore:

\[ \frac{\partial S_n}{\partial t_i} = \sum_{i > j} \frac{\partial S_n}{\partial \Theta_{ij}} \cdot \frac{\partial \Theta_{ij}}{\partial t_i} \ldots (3.6.21) \]

Since \( \Theta_{ij} \) is a function of \( \rho_{ij} \) and consequently a function of \( (t_i - t_j) \) only, (3.6.21) reduces to:
\[ \frac{\partial S_n}{\partial t_1} = \sum_{j=2}^{n} \frac{\partial S_n}{\partial \theta_{ij}} \cdot \frac{\partial \theta_{ij}}{\partial t_1} \quad \ldots (3.6.22) \]

(3.6.22) after using the reduction formula (3.6.13), becomes:

\[ \frac{\partial S_n}{\partial t_1} = \frac{1}{n-2} \sum_{j=2}^{n} S_{n-2}(\theta^{(ij)}) \cdot \frac{\partial \theta_{ij}}{\partial t_1} \quad \ldots (3.6.23) \]

Differentiating with respect to \( t_n \) yields

\[ \frac{\partial^2 S_n}{\partial t_1 \partial t_n} = \frac{1}{n-2} S_{n-2}(\theta_{1n}) \cdot \frac{\partial^2 \theta_{1n}}{\partial t_1 \partial t_n} + \frac{1}{n-2} \sum_{j=2}^{n} \frac{\partial S_{n-2}}{\partial t_n} \cdot \frac{\partial \theta_{ij}}{\partial t_1} \quad \ldots (3.6.24) \]

where \( S_{n-2}(\theta^{(ij)}) \) is abbreviated as \( S_{n-2}^{(ij)} \).

It is clear from the case \( n=3 \), that \( \frac{\partial^2 \theta_{1n}}{\partial t_1 \partial t_n} \) is proportional to \( P_{03}(t_n-t_1) \). Bearing this in mind, (3.6.16) together with (3.6.24) gives:

\[ P_{0n}(\gamma) = \frac{\sqrt{(n-1)}}{n-1} S_{n-2}(\theta_{1n}) \cdot P_{03}(\gamma) - \frac{\sqrt{(n-2)-1}}{2(n-2)-1} \sum_{j=2}^{n} \frac{\partial S_{n-2}^{(ij)}}{\partial t_n} \cdot \frac{\partial \theta_{ij}}{\partial t_1} \quad ; \ n>3 \quad \ldots (3.6.25) \]

The fact that we can express \( S_n \) as far as \( n=3 \), means that we can obtain successive approximations for \( P_{0n}(\gamma) \) as far as \( n=5 \). This is made possible by the reduction formula.

Higgins showed that \( P_{03}(\gamma) \) has both the correct value and gradient at the origin. He also showed that as \( \gamma \to 0 \), \( P_{0n}(\gamma) \) approaches \( P_{03}(\gamma) \). For \( n=4 \) (3.6.25) gives:

\[ P_{04}(\gamma) = \frac{\sqrt{3}}{3} P_{03}(\gamma) - \frac{1}{2 \sqrt{n-1}} \left[ \frac{\partial \alpha}{\partial t_4} \cdot \frac{\partial \theta_{12}}{\partial t_1} + \frac{\partial \beta}{\partial t_4} \cdot \frac{\partial \theta_{13}}{\partial t_1} + \frac{\partial \gamma}{\partial t_4} \cdot \frac{\partial \theta_{14}}{\partial t_1} \right] \quad \ldots (3.6.26) \]

where

\[ \alpha = \Theta^{(12)}_{34} = S_2^{(12)} = \cos^{-1} \rho_{34}^{(12)} \quad \ldots \ a \]

\[ \beta = \Theta^{(13)}_{24} = S_2^{(13)} = \cos^{-1} \rho_{24}^{(13)} \quad \ldots \ b \]
\[ \delta = \theta_{23}^{(24)} = \theta_{2}^{(14)} = \cos^{-1} \rho_{23}^{(14)} \quad \ldots \quad (3.6.27) \]

and \( \rho_{rs} \) are the partial correlation coefficients given by (3.6.11). Similarly for \( n=5 \) we obtain:

\[ P_{05}(\tau) = \frac{s_3^{(15)}}{2\pi} \cdot P_{03}(\tau) - \frac{1}{4\pi^2 n_0} \left[ \frac{\partial s_3^{(12)}}{\partial t_5} \cdot \frac{\partial \theta_{12}}{\partial t_4} + \frac{\partial s_3^{(13)}}{\partial t_5} \cdot \frac{\partial \theta_{13}}{\partial t_4} \right. \]

\[ \left. + \frac{\partial s_3^{(14)}}{\partial t_5} \cdot \frac{\partial \theta_{14}}{\partial t_4} + \frac{\partial s_3^{(15)}}{\partial t_5} \cdot \frac{\partial \theta_{15}}{\partial t_4} \right] \]

\[ \ldots (3.6.28) \]

where:

\[ s_{3}^{(12)} = \theta_{23}^{(12)} + \theta_{25}^{(12)} + \theta_{45}^{(12)} \quad \ldots \quad a \]

\[ s_{3}^{(13)} = \theta_{23}^{(13)} + \theta_{25}^{(13)} + \theta_{45}^{(13)} \quad \ldots \quad b \]

\[ s_{3}^{(14)} = \theta_{23}^{(14)} + \theta_{24}^{(14)} + \theta_{35}^{(14)} \quad \ldots \quad c \]

\[ s_{3}^{(15)} = \theta_{23}^{(15)} + \theta_{24}^{(15)} + \theta_{34}^{(15)} \quad \ldots \quad d \]

6.2 The Approximation \( P^*_{on}(\tau) \).

Using (3.6.15b) we obtain a second set of approximations

\[ P_{on}(\tau) = \frac{\Gamma(n)}{\pi^{4+n_0}} \frac{d^2 s_n}{d\tau^2} \quad \ldots \quad (3.6.30) \]

For \( n=3 \) all angles contribute to the derivative & we obtain

\[ P^*_{03}(\tau) = P_{03}(\tau) + \frac{1}{4} P_{03}(\frac{\tau}{2}) \quad \ldots \quad (3.6.31) \]

Higgins showed that the gradient of \( P^*_{03}(\tau) \) at the origin is not equal to the true gradient but exceeds it by 25 per cent. For \( n > 3 \) Higgins obtained:

\[ \frac{d s_n}{d\tau} = \frac{1}{n-2} \sum_{i \neq j} s_{n-2}^{(ij)} - \rho_{ij} \frac{\rho_{ij}}{(1-\rho_{ij}^2)^{\frac{1}{2}}} \quad \ldots \quad (3.6.32) \]

To obtain \( \frac{d^2 s_n}{d\tau^2} \) Higgins pointed out that this may be obtained by computing (3.6.32) at regular intervals of \( \tau \), and performing the differentiation numerically. Again since \( s_{n-2}^{(ij)} \)
is of degree $n-2$, this may be done up to, and including, $n=5$.

Higgins evaluated both sets of approximations for the case of an ideal low pass spectrum. He found that $P_{03}, P_{04}$ and $P_{05}$ coincide, for small $\zeta$ and have the true gradient of $P_0(\zeta)$. On the other hand all these approximations result in negative probabilities at some values of $\zeta$. $P_{03}^*(\zeta), P_{04}^*(\zeta)$ and $P_{05}^*(\zeta)$ are generally worse than the other set for small $\zeta$ (none of them gives the true gradient), but appear to behave better for large values of $\zeta$. We have evaluated the approximations $P_{03}(\zeta), P_{04}(\zeta)$ and $P_{05}(\zeta)$ for various power spectra. The results are presented and discussed in Chapter six.

### 3.7 Final Note on the Multivariate Normal Integral.

If $E_i$ denotes the event that $f(t_i)$ is non-negative then it can easily be shown that:

$$P(E_1 + E_2 + \cdots + E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j}^{n} P(E_i E_j) + \cdots + (-1)^{n-1} \cdot P(E_1 E_2 \cdots E_n) \quad \cdots (3.7.1)$$

where $P(E_1 + E_2 + \cdots + E_n)$ is the probability that at least one of the events $E_i$ occurs and $P(E_i E_j \cdots)$ is the probability of the simultaneous occurrence of the events $E_i E_j \cdots$.

This later probability is also referred to as the probability of the intersection of the events $E_i E_j \cdots$ etc.

If $\bar{E}_i$ denotes the compliment of the event $E_i$ then

$$P(E_1 + E_2 + \cdots + E_n) = 1 - P(\bar{E}_1 \bar{E}_2 \cdots \bar{E}_n) \quad \cdots (3.7.2)$$

Now if the random signal $f(t)$ is symmetrical about the zero level, then by symmetry:

$$P(\bar{E}_1 \bar{E}_2 \cdots \bar{E}_n) = P(E_1 E_2 \cdots E_n) \quad \cdots (3.7.3)$$
Hence for n odd we have:

\[
P(E_1E_2 \cdots E_n) = \frac{1}{n} \left[ 1 - \sum_{i=1}^{n} P(E_i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} P(E_iE_j) \cdots \text{etc} \right]
\]

the last term on the right hand side being the sum of the joint probabilities of the E's taken n-1 at a time. For n even we obtain an identity.

Now \(P(E_1E_2 \cdots E_n)\) is the multivariate integral \(V_n\).

F.N. David (24) in a note he wrote about this integral, used the above approach to derive the recurrence formula (3.7.4). He was able to obtain \(V_3\) from the knowledge of \(V_1\) & \(V_2\). He correctly pointed out that \(V_5\) may be obtained from the knowledge of \(V_4, V_3, V_2 \cdots \text{etc.}\)

However, because for n even, there is no corresponding recurrence formula, the step from \(n=3\) to \(n=4\) cannot be made. What is needed therefore he stated, is some means by which \(V_n\) for \(n=\text{even}\) can be evaluated. Hence by using (3.7.4) \(V_n\) for n odd follows.

At the time David wrote this note, 1953, there was the series solutions by Moran and Kendal, discussed in section 3. David pointed out that the series were slow in convergence and as such were of little use in evaluating \(V_n\) for n even. Since then, there have been some attempts to evaluate the quadrivariate (n=4) normal integral, but most of these attempts were directed to obtain numerical results which are not very helpful as far as approximations to \(P_0(\gamma)\) are concerned. It is more helpful to use Schlafli's reduction formula (see last section), since this permits evaluation of approximations to \(P_0(\gamma)\) based on the accurate value of the quadrivariate integral (\(V_4\)) and that of (\(V_5\)).
It seems that further progress towards solving this problem is intimately bound to a thorough understanding of the behaviour of the Schlaffli function \((S_n)\). This was the view expressed by Longuet-Higgins in a private correspondence with the author. Higgins acknowledged the difficulty of this problem, but noted that its solution is necessary for the evaluation of the multivariate integral.
CHAPTER FOUR.

The Markovian Model.

4.1. Introduction.

In Chapter 2, section 9, approximations for the distribution of zero crossing intervals were expressed as the solutions of integral equations based on the assumption of independence and the 'quasi independence' between successive intervals. The latter assumption was shown to be equivalent to treating the zero crossing events as forming a renewal process. Both amount to the same thing; that the event of a downward crossing at \( t \), \( t + dt \), given several downward crossings in the prior times \( t_0, t_1, t_2, \ldots \) etc., depends only on the most recent event at \( t_0 \). This is also equivalent to the assumption that the events represented by downward crossings form a Markov Process.

In this chapter it will be shown that this integral equation is only the first of a series of such equations which can be derived using a general Markovian model. A generalised integral equation will be derived whose solution provides the basis for obtaining exact results for the zero crossing distributions \( P_n(\tau) \) for a class of random functions. The equation also suggests a new series of approximations for the distribution \( P_0(\tau) \).

4.2 Introduction to the theory of Markov Processes.

A stochastic process \( x(t) \) can be completely described in the statistical sense only if for any sequence of times \( t_1 \cdots t_n \), the process at those times is known. This (the joint probability density function \( P(x_1 x_2 \ldots x_n) \) of the values)
function is hereafter referred to as the (Pdf) and must satisfy the following probability laws:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(x_1 x_2 \cdots x_n) \, dx_1 \cdots dx_n = 1 \quad \cdots (4.2.1)$$

$$\int_{-\infty}^{\infty} P(x_1 x_2 \cdots x_n) = P(x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_n) \quad (4.2.2)$$

The process \( x(t) \) is stationary if the Pdf's are independent of the time origin, i.e. if \( P(x_1 \cdots x_n) \) is a function of the time intervals \( (t_1 - t_i) \) rather than the actual instants \( t_i \).

The collection of the Pdf's needed to describe the stochastic process \( x(t) \) is much too broad to be obtained by measurements of \( x(t) \), even if many realisations of the process were available. Instead, a hypothetical model is set up which accounts for what is known about the process \( x(t) \) from which theoretical results for the statistical behaviour of the process can be derived.

One such model associates the process \( x(t) \) with a system whose condition or state at any time can be specified by a finite number of easily measured or estimated variables. These variables, referred to as the state variables are random, and their joint probability distributions at any one time provide the statistical description of the state of the system. If the process \( x(t) \) can be related to these state variables, its joint probability density functions \( P(x_1, x_2 \cdots x_n) \) can be derived from those of the state variables.

To describe a system statistically, a relatively small number of variables are needed. These are called the macroscopic variables. A complete description of a
system however, requires a large number of additional variables to account for the microscopic details of the state of the system. These are called the microscopic variables. Macroscopic variables define the average state or motion of the system, whereas the microscopic ones describe the rapid fluctuations about the average state.

If the number of the macroscopic variables are sufficient to provide a complete statistical description of the system, then the state of these variables at any one time, say \( t_0 \), is sufficient to determine their joint probability density at all times \( t \) in the future. In other words the state of the macroscopic variables at time \( t \) is independent of the past history of the system which has brought it to the state at \( t_0 \). This state of affairs is obtained if the state variables, denoted collectively by the vector \( \mathbf{X} \) satisfies a first order differential equation of the form:

\[
\frac{d\mathbf{X}}{dt} = g(\mathbf{X}) \quad \cdots \quad (4.2.3)
\]

The solutions of this vector equation give the average state of the system at any time \( t \), given its initial stage at time \( t_0 \). The effect of the microscopic variables will be to cause the components of the vector \( \mathbf{X} \) to fluctuate randomly about the solutions of (4.2.3). These fluctuations have much shorter periods compared to the relaxation times of the vector \( \mathbf{X} \) and cannot be specified precisely due to the nature of the microscopic variations. It is possible however, to make some statistical assumptions about the variations. Hence, we can, at most, give the probability distributions of the vector \( \mathbf{X} \). Yet, because of the rapid
and chaotic variations of the microscopic variables, it is logical to expect that, given the initial values of the component of the vector \( X \) at time \( t_0 \); their probability distributions at all future times \( t \) will be independent of the precise past history of the system that brought it to the state \( X(t_0) \). If the above is true, the system will be Markovian and the collection of the state variables \( X \) forms a vector Markov process. In order that \( X \) is Markovian however, it must contain sufficient numbers of components providing a complete description of the system from the macroscopic viewpoint and the vector \( X \) must satisfy a first order differential equation of the form (4.2.3). If any of the macroscopic variables were omitted from \( X \), the state of the vector \( X \) at time \( t \) will become dependent on its history prior to the time \( t_0 \) through the omitted variables, and the vector \( X \) ceases to be Markovian.

If the vector \( X \) is Markovian, the conditional Pdf of its component values at time \( t \), given their values at the past times \( t_0,t_1,t_2\ldots t_n \) depends only on their values at the most recent time, namely

\[
P \left[ X(t)/X(t_0), X(t_1)\ldots X(t_n) \right] = P \left[ X(t)/X(t_0) \right]
\]

where

\[
t_0 > t_1 > t_2 \ldots > t_n
\]

and \( P \left[ X(t)/X(t_0) \right] \) is called the 'transitional probability density' of the Markov process. The number of components constituting the vector \( X \) defines the order or the dimension of the vector Markov process.

The transitional probability density satisfies the so-called Smoluchowski equation:

\[
P(X(t)/X(t_0)) = \int P(X(t)/X(t_1)) P(X(t_1)/X(t_0)) \, dx_1 \ldots (4.2.5)
\]
where the integration is performed over the entire space of the vector $\mathbf{X}$.

4.2.1 Linear Systems.

We next consider a linear Markovian system whose state can be specified by the vector $\mathbf{X}$ having $n$ components. A linear system is one in which $g(\mathbf{X})$ is related linearly to the vector $\mathbf{X}$, or in other words, the first order differential equation (4.2.3) is of the form:

$$\frac{d\mathbf{X}}{dt} = A \mathbf{X} + B \quad \ldots \text{(4.2.6)}$$

where $A$ & $B$ are matrices, independent of the vector $\mathbf{X}$, but may be functions of time.

C.W. Helstrom (25) showed that, if a system is linear and Markovian, whose state can be represented by a vector $\mathbf{X}$ obeying a first order differential equation of the form (4.2.6), but with $A$ & $B$ both constant, then the steady state fluctuations of each component of the vector $\mathbf{X}$ will have a spectral density $\phi(\omega)$ which is a rational function of the frequency $\omega$. This means that $\phi(\omega)$ is of the form:

$$\phi(\omega) = \frac{N(\omega^2)}{D(\omega^2)} \quad \ldots \text{(4.2.7)}$$

where $N(\omega^2)$ and $D(\omega^2)$ are both polynomials in $\omega^2$. Conversely, according to Helstrom, if a stochastic process $x(t)$ has a rational spectral density of the form (4.2.7), then $x(t)$ can be represented as one component of a vector Markov process $\mathbf{X}(t)$ which obeys the first order differential equation (4.2.6) and whose dimension is equal to the degree of the polynomial $D(\omega^2)$.

This is an important result since it means that any
random function \( f(t) \) with a rational spectral density can be associated with a vector Markov process \( \mathbf{X}(t) \). The statistical behaviour of the random function \( f(t) \) including that describing its zero crossing distribution can be derived from the statistics of the vector process \( \mathbf{X}(t) \).

A new approach to the zero crossings distribution is thus possible using this method. This will be discussed later, after dealing first with other remaining components which constitute the vector process \( \mathbf{X}(t) \).

4.2.2 Components constituting the vector process \( \mathbf{X}(t) \).

It was stated above that a random function \( x(t) \), \( x(t) \) is used here in place of \( f(t) \) with a rational spectral density can be represented as one component of a vector Markov process \( \mathbf{X}(t) \) whose dimension is given by the degree of the polynomial \( D(\omega^2) \). The remaining \((n-1)\) components of this process will now be given and discussed.

First, (4.2.7) is written in the form:

\[
\phi(\omega) = \frac{F(j\omega)F(-j\omega)}{G(j\omega)G(-j\omega)} \quad \ldots (4.2.8)
\]

where \( F(j\omega) \) and \( G(j\omega) \) are polynomials assumed to have degrees \( m \), and \( n \) respectively and whose zeros are in the left half of the \( s \) plane.

In order that the function \( f(t) \) has a finite variance, the degree of the numerator must be less than the denominator, i.e. \( m < n \). Also, since a differentiation of the random function \( x(t) \) in the time domain, corresponds to a multiplication of its frequency spectrum \( F(j\omega)/G(j\omega) \) by a factor of \( j\omega \) and consequently raises the degree of the polynomial \( F(j\omega) \) by one, it follows that the maximum number of times one can differentiate \( x(t) \) and still obtain
derivatives with finite variances is given by \((n-m-1)\).

Next, the components of the vector process \(X(t)\) are given. Helstrom showed that these components are related to each other by the following differential equations:

\[
\frac{dX_k}{dt} = X_{k+1} + C_k y(t) ; \ 1 \leq k \leq n \quad (4.2.9a)
\]

\[
\frac{dX_n}{dt} + \sum_{r=0}^{n-1} g_r X_{n-r} = C_n y(t) \quad (4.2.9b)
\]

\[
X_1 = x(t) \quad (4.2.9c)
\]

where \(y(t)\) is a white noise process, and the coefficients \(g_r\) are those of the polynomial \(G(j\omega)\), namely:

\[
G(j\omega) = (j\omega)^n + g_0(j\omega)^{n-1} + \ldots + g_{n-1} \ldots (4.2.10)
\]

The coefficients \(C_k\) are obtained from the contour integration

\[
C_k = \frac{1}{2\pi j} \int_C (j\omega)^{k-1} \frac{F(j\omega)}{G(j\omega)} \, dj\omega \quad (4.2.11)
\]

where the contour \(C\) encloses the zeros of \(G(j\omega)\).

It can easily be shown that the above relationships lead to the following results:

\[
C_k = 0; \ 1 \leq k \leq n-m-1 \quad (4.2.12)
\]

and

\[
C_{n-m} = f_0 \quad (4.2.13)
\]

where \(f_0\) is the leading coefficient of \(F(j\omega)\), hence equations \((4.2.9a)\) and \((4.2.9c)\) may be written as:

\[
X_1(t) = x(t) \quad (4.2.14a)
\]

\[
X_k(t) = \frac{d^{k-1} x(t)}{dt^{k-1}}, \ 1 \leq k \leq n-m-1 \quad (4.2.14b)
\]

The rest of the coefficients \(C_k\) depend on the other coefficients of \(F(j\omega)\) and \(G(j\omega)\).

Equations \((4.2.14a)\) &\((4.2.14b)\) show that the first \((n-m)\) components of the vector Markov process \(X(t)\) contain
the random function \( f(t) \) and its available \((n-m-1)\) derivatives.

Finally, if \( y(t) \) is a Gaussian white noise process, then the \( n \) components form a Gaussian Markov process and it can be shown that the transitional Pdf \( P(x(t)/x(t_0)) \) satisfies a Fokker-Planck type differential equation and its adjoint.

4.3 Derivation of the generalised Integral equation.

We shall now use the Markovian model representation of a random function \( x(t) \) discussed above, to derive the generalised integral equation whose solution can be related to the crossing interval distribution \( P_0(\mathcal{C}) \) and which has the possibility of yielding exact results for the zero crossing problem. The method of derivation can be considered as the generalisation of the method used by A. Seigert\(^{(26)}\), which dealt with a two-dimensional Markov process.

The random function \( x(t) \) is assumed to be stationary with a rational spectral density of the form \((4.2.7)\). Using the above theory, the function \( x(t) \) is considered as one of the components forming the \( n \)-dimensional Markov process \( \mathcal{X}(t) \). The remaining \((n-1)\) components will be referred to as \( \dot{x}(t), \ddot{x}(t) \ldots x^{(n-1)}(t) \), bearing in mind that only the first \((n-m-1)\) components of these represent actual derivatives of the function \( x(t) \). The transitional Pdf of the process is defined by:

\[
P_n(x(t)) = \frac{\mathcal{X}/x(t_0) = x_0}{dy \, dy \ldots dy}^{(n-1)} \quad \text{(Conditional Probability)}
\]

if \( x(t_0) = y_0, \dot{x}(t_0) = \dot{y}_0, \ldots, x = y_0, \dot{x} = \dot{y}_0, \ldots, x^{(n-1)} = y^{(n-1)} \) and \( dy \).

Further, we define \( q_k(\dot{y}, \dot{y} \ldots \dot{y}, t_0; y_0, y_0 \ldots y_0, t_0) dt \) as the conditional probability that if \( x(t_0) = 0, \dot{x}(t_0) = \dot{y}_0 \ldots \dot{y}_0, \dot{y}_0 \ldots \dot{y}_0, t_0 \), then the random process \( x(t) \) will go through...
zero in the interval \( t, t + dt \) with \( \dot{y} < x(t) < \dot{y} + dy, \ldots \) and will have \( k \) other zero crossings in the interval \((t-t_0)\).

Since the vector process \( \mathbf{x}(t) \) is Markovian we have the following recursion formula for \( q_k \):

\[
q_{k+1}(\dot{y}, \ldots, \dot{y}, t; \dot{y}_0, \ldots, \dot{y}_0, t_0) = \int_{t_0}^{t} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} q_0(z, \ldots, z, t; \dot{z}, \ldots, \dot{z}, t_1) \, dz \, dz \ldots dz
\]

if we define the integral operator \( \Delta \) which operates on the function \( F \) in the manner given by

\[
\Delta F = \int_{t_0}^{t} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} q_0(z, \ldots, z; t; \dot{z}, \ldots, \dot{z}, t_1) \, dz \, dz \ldots dz \]

we can write (4.3.1) in the form:

\[
q_{k+1} = \Delta q_k
\]

and

\[
q_k = \Delta^k q_0 = \Delta q_k
\]

Let \( P(\dot{y}, \ldots, \dot{y}, t; \dot{y}_0, \ldots, \dot{y}_0, t_0) \) be the conditional probability that \( x(t) \) goes through zero between \( t \) and \( t + dt \) with \( \dot{y} < x(t) < \dot{y} + dy, \ldots \dot{y} < x(t) < \dot{y} + dy \), given that \( x(t_0) = 0 \), etc; and without regard to other zero crossings.

It is clear that

\[
P = \sum_{k=0}^{\infty} q_k
\]

where \( P \) and \( q_k \) are written without their arguments for brevity. (4.3.5) together with (4.3.4) yields

\[
P = \sum_{k=0}^{\infty} \Delta^k q_0 = \frac{1}{1-\Delta} q_0
\]
Hence \( q_0 = P - \Delta P \) \ ...(4.37)

Using the definition of the operator \( \Delta \) given by (4.3.2),
we obtain the following equation for \( q_0 \)

\[
q_0(y', y' \ldots y', t/y_0, \ldots y_0, t_0) = P(y', y' \ldots y', t/y_0, \ldots y_0, t_0)
\]

\[
- \int_{t_0}^{t} dt_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} q_0(z', z' \ldots z', t_1/y_0, \ldots y_0, t_0) P(y', y' \ldots y, t/
\]

\[
z', z', \ldots z', t_1) \ dz' \ dz' \ldots \ dz \ldots \ (4.3.8)
\]

This is the generalised integral equation for \( q_0 \), the probability that \( x(t) \) crosses the zero level for the first time at \( t, t + dt \) with \( y' < x(t) < y' + dy' \ldots \) etc., given that \( x(t) \) has crossed this level at \( t^* \) with \( y' < x(t^*) < y' + dy' \ldots \) etc. It will be shown \( q_0 \) can be related to \( F_0(\gamma) \), the distribution of the intervals between successive zero crossings. At this stage it must be pointed out that for \( q_0 \) and more generally \( q_k \) to exist, the random function \( x(t) \) must be differentiable at least once. This means that \( n \), the degree of polynomial \( D(\omega^2) \) must be greater than one. For \( n=1 \), which corresponds to a one-dimensional Markov process, \( x(t) \) does not have a continuous derivative, and once \( x(t) \) has reached the level \( h \) say, it may cross it again and again during an arbitrary, infinitesimal interval. Hence, the rate of level crossings are infinite. In this case we have the following integral equation:

\[
f(h, t/y_0) = \phi(h, t/y_0) - \int_{0}^{t} q(t_1, h/y_0) \phi(h, t-t_1/h) \ dt_1 \ldots \ (4.3.9)
\]

where \( f(h, t/y_0) \) is the conditional probability that given \( x(0)=y_0 \) which is greater than \( h \), the function \( x(t) \) has not crossed the level \( h \) in the interval \( (0, t) \). \( q(t_1, h/y_0) dt_1 \) is
the probability that \( x(t) \) first crosses the level \( h \) in \( t_1, t_1 + dt \), given that \( x(0) = y_0 \) and finally \( \phi(h, t/y_0) \) is the conditional probability that \( x(t) \) has an amplitude greater than \( h \) at time \( t \), given the same initial condition. \( h \) is assumed positive throughout.

The first passage time probability, i.e. the probability density that \( x(t) \), starting from \( y_0 \) at \( t=0 \), crosses the level \( h \) for the first time at \( t, t+dt \) is given by:

\[
q_0(h, t/y_0) = -\frac{\partial \Phi}{\partial t} (h, t/y_0) \quad ...(4.3.10)
\]

The derivation of (4.3.9) and its solution are discussed in Appendix (2).

For vector processes whose dimension is greater than 1, it can be shown that the transition Pdf \( P_n \) is related to the joint density

\[
P(\vec{y}, \vec{y}, ..., \vec{y}, t/\vec{y}_0, \vec{y}_0 ..., \vec{y}_0, t_0) \quad by
\]

\[
P(\vec{y}, \vec{y}, ..., \vec{y}, t/\vec{y}_0, \vec{y}_0 ..., \vec{y}_0, t_0) = P_n(0, \vec{y}, \vec{y}, ..., \vec{y}/0, \vec{y}_0, \vec{y}_0 ..., \vec{y}_0, t_0) \quad \| \vec{y} \|
\]

...(4.3.11)

Here \( \vec{y} \) represents the slope of \( x(t) \) at time \( t \). The method of deriving (4.3.9) is similar to that used by Rice(1) in obtaining the relation between the probability of a random function being in the range \((0-dy)\) at a given instant of time \( t \), and the probability that that random function goes through zero in the interval \( t, t+dt \).

For stationary processes, all of the probability functions discussed above are functions of \((t-t_0)\) and not the actual values of \( t \) & \( t_0 \).

4.4 The relationship between the zero crossing distribution
... $P_0(\tau)$ and the solution of the generalised integral equation.

Let $P(\tilde{y}_0, \tilde{y}_0, \ldots, \tilde{y}_0, t_0)$ be the joint probability that if the random function $x(t)$ has a zero value at time $t_0$, then the remaining $(n-1)$ components of the vector process $X(t)$ have values at time $t_0$ given by:

$$\dot{y}_0 < \dot{x}(t_0) < \dot{y}_0 + d\dot{y}_0 \quad \ldots \quad (n-1)(n-1)$$

Further let $P_0(0, t_0, t_0)$ be the probability density that if $x(t_0)$ is zero, then $x(t)$ goes through zero for the first time in $t, t+dt$. We have then the following relationship:

$$P_0(0, t_0, t_0) = \int_{-\infty}^{0} P(\tilde{y}_0, \ldots, \tilde{y}_0, \tilde{y}_0, t_0) \cdot q_0(\tilde{y}_0, \tilde{y}_0, \ldots, t_0) \, dq_{0}(\tilde{y}_0, \ldots, \tilde{y}_0, \tilde{y}_0, t_0)$$

where $q_0$ is the solution of the generalised integral equation (4.3.8).

But $P_0(0, t_0, t_0)$ is the same as $P_0(\tau)$, if $\tau = t - t_0$; hence (4.4.1) provides a relationship between $P_0(\tau)$ and the probability distribution $q_0$. It follows therefore, that for random functions with rational spectral densities, the Markovian model can provide exact solutions for the probability distribution of the zero crossing intervals. This is achieved by solving first for the conditional probability density $q_0$ using the integral equation (4.3.8), and then using the relationship (4.4.1) to obtain the distribution $P_0(\tau)$.

This is the second method which has the possibility of yielding exact results for the distribution $P_0(\tau)$. The other method is the 'in and exclusion' method which expresses $P_0(\tau)$ as an infinite series whose terms are integrals of ever-increasing multiplicity (see equations 2.4.2). Comp-
-aring these two exact methods, it is clear that the Markovian model is far superior to the in and exclusion method since it is more direct and involves far less computational work. The number of integrals involved in the Markovian model is finite and proportional to \( n \), the degree of the polynomial \( D(\omega^2) \); whereas the number of integrals in the in and exclusion equation (2.4.2) is infinite on the other hand, the in and exclusion method applies to a wider class of random functions.

4.5 Approximations based on the integral equation.

Although in theory, the Markovian model has the potential of yielding exact results for \( P_0(C) \), in practice mathematical difficulties again limit our ability to obtain exact results, for random functions with arbitrary rational spectral densities. Apart from the case corresponding to \( n=1 \), when the integral equation is of the form (4.3.9), which can be solved using the Laplace transform method, there are no such exact methods for solving the integral equation (4.3.8) for arbitrary \( n \). Numerical methods may be considered, but here again practical considerations limit our ability to solve (4.3.8) when \( n \) has an arbitrary value.

To obtain exact results for \( P_0(C) \), the value of \( n \) must be equal to the degree of the polynomial \( D(\omega^2) \). If \( n \) is chosen lower than this value in order to facilitate a numerical solution for (4.3.8), the associated vector process \( \bar{X}(t) \) ceases to be Markovian, since this results in the reduction of the number of macroscopic state variables. The implications of reducing the dimension of the vector process are that (4.3.8) can at best lead to app-
proximate results for the distribution $P_0(t)$. We shall now discuss some of these approximations and begin by applying (4.3.8) with $n=1$ to random functions having finite mean rates of zero crossings. Examples of such functions are those with the following spectral densities:

1. $\phi(\omega) = \frac{(\omega^2)^m}{(1+\omega^2)^n}$ \hspace{1cm} ...(4.5.1)
   
   where $n > 1$ and $n-m > 1$.

   Also those with Butterworth spectra

2. $\phi(\omega) = \frac{1}{1+\omega^{2n}}$ \hspace{1cm} ...(4.5.2)

   The ideal low pass filter case may be obtained from (4.5.2) by letting $n \to \infty$.

   Letting $n=1$ in equation (4.3.8), we obtain:

   $$q_0(t/t_0) = P(t/t_0) - \int_{t_0}^{t} dt_1 q_0(t_1/t_0) P(t/t_1) \hspace{1cm} ...(4.5.3)$$

   where $q_0(t/t_0)dt$ is the probability that $x(t)$ first crosses the zero level at $t,t+dt$, given that $x(t_0) = 0$, and $P(t/t_0)dt$ is the probability that $x(t)$ crosses the zero level at $t,t+dt$, given $x(t_0)=0$. It is clear from these definitions that $q_0(t/t_0) = P_0(t)$ and $P(t/t_0) = U(t)$.

   Hence (4.4.4) can be written as:

   $$P_0(t) = U(t) - \int_0^t P_0(y) U(t-y)dy \hspace{1cm} ...(4.5.4)$$

   which is the integral equation obtained by McFadden assuming independence between successive intervals.

   Since a random function with a finite rate of zero crossings must have a continuous first derivative, we
expect the slope of \( x(t) \) at the zero crossing points to have some influence on the distribution of these zeros.

We can therefore improve the approximation given by (4.5.4) which does not take into account the slope of \( x(t) \), by noting that for the zero crossing at \( t, t+dt \) to be the first zero crossing after the one at \( t_0 \), the slope of \( x(t) \) must be of opposite sign to that of \( x(t_0) \). Assuming the slope at \( t_0 \) is positive corresponding to an upward zero crossing (4.5.4) becomes:

\[
q_0(t/t_0) = P_{-1+}(t/t_0) - \int_{t_0}^{t} dt_1 q_0(t_1/t_0)P_{-1-}(t/t_1)
\]

where \( P_{-1+} \) and \( P_{-1+} \) are defined similarly to \( P \) with the added constraints on the signs of the slope of \( x(t) \) at the crossing points \( t_0, t_1 \) and \( t \) as indicated in the subscripts.

In the notations used in Chapter One, this is the same as

\[
P_0(\tau) = Q(\tau) - \int_{0}^{\tau} P_0(y)P_{\text{odd}}(\tau-y) \, dy \quad \ldots (4.5.6)
\]

which is the integral equation obtained by McFadden and Rice and Beer assuming quasi independence between successive intervals. The results obtained by McFadden may now be explained in the light of the above theory.

McFadden considered three processes with power spectra of the form (4.5.1) with \( m \) and \( n \) having the following values

(i) \( m = 0, n = 2 \)
(ii) \( m = 0, n = 3 \)
(iii) \( m = 2, n = 4 \)

For the Markovian model to provide exact solutions for the distribution \( P_0(\tau) \), the dimensions of the associated vector Markov processes should be two in case
(i), three in case (ii); and four in case (iii). Therefore, the integral equation (4.5.4) which is based on a one-dimensional Markovian representation can at most provide approximations. These approximations, according to the above theory should have their best agreement with $P_0(\tau)$ in case (i) which deteriorates as $n$ increases. The results of McFadden seem to confirm this point. The remaining cases which McFadden considered are:

(iv) $\phi(\omega) = \frac{1}{1 + \omega^2 n}$, $n = 7$

(v) Ideal low pass spectrum which may be expressed in the form:

$$\phi(\omega) = \lim_{n \to \infty} \frac{1}{1 + \omega^2 n} \quad \text{(4.5.7)}$$

(vi) Ideal octave band-pass whose $\phi(\omega)$ may be expressed in the form:

$$\phi(\omega) = \lim_{n \to \infty} \frac{1}{[1 + (\omega^2)^{2n}][1 + \frac{1}{\omega^2 n}]} \quad \text{(4.5.8)}$$

Cases (v) and (vi) both correspond to sharp cut-off frequencies. A vector Markovian model must have a very large number of components, if it is to represent random functions with spectra such as these adequately. This means that these functions have infinite memories, and a vector representation with finite number of components can never account for all effects of memory. This is why in such cases the solution of an integral equation of the form (4.5.4) fails to provide reliable approximations for the distribution $P_0(\tau)$.

In all these six cases, the assumption of independence or quasi independence can not strictly be true, since
according to the above theory, the solutions of (4.5.4) and (4.5.6) can at best provide approximations for the distribution $P_0(\tau)$. This was the contention of Longuett-Higgins (2) also; who by considering the initial behaviour of the distributions $P_n(\tau)$ arrived at the same conclusion. The quasi independence assumption is however justifiable when considering spectra with gradual cut-off frequencies such as cases (i - iii) and to some extent (iv). In all these cases the associated Markov model is of a relatively low dimension making it possible for (4.5.6) to yield meaningful results. It is clear that McFadden's integral equation is the first in a series of such relations which can be obtained from the generalised integral equation (4.3.6). In the next section we discuss the second equation in this series, which corresponds to $n=2$. This takes into account the actual value of the slope of $f(t)$ at the zero crossing points and leads to the development of a higher approximation to the distribution $P_0(\tau)$.

4.6 Approximations based on the Second Order Markovian Model.

Letting $n=2$ in (4.3.6) we have

$$q_0(\dot{y}, t/\dot{y}_0, t_0) = P(\dot{y}, t/\dot{y}_0, t_0) - \int_{t_0}^{t} dt_1 \int_{-\infty}^{\infty} q_0(\dot{z}, t_1/\dot{y}_0, t_0).$$

(4.6.1)

If both sides are multiplied by $P(\dot{y}_0/t_0)d\dot{y}_0$, the probability that the slope at $t_0$ lies between $\dot{y}_0$ and $\dot{y}_0+d\dot{y}_0$ and then integrating over all positive values of $\dot{y}_0$ and finally divide both sides by the probability that the slope at $t_0$ is positive (irrespective of its actual value), (4.6.1) becomes:
\[ P_0(\dot{y}, t/t_0) = Q(\dot{y}, t/t_0) - \int_{t_0}^{t} dt_1 \int_{-\infty}^{0} dz \frac{d}{dz} P_0(z, t_1/t_0) P(y, t/z, t_1) \] ...

The integration over the slope \( z \) runs from \(-\infty \) to 0 since the slope at \( t_1 \) must be negative, i.e.
\[ P_0(\dot{z}, t_1/t_0) = 0 \quad z > 0 \] ...

\( P_0(\dot{y}, t/t_0) \) \( dy \) \( dt \) is the probability that the random function \( x(t) \) first crosses the zero level at \( t, t+dt \) with a slope between \( \dot{y} \) and \( \dot{y}+dy \) given that it has crossed the zero level at \( t_0 \) with a positive slope. \( Q(\dot{y}, t/t_0) dy dt \) is the probability that a downward crossing with a slope between \( \dot{y} \) and \( \dot{y}+dy \) occurs at \( t, t+dt \) given the same conditions. \( P(\dot{y}, t/\dot{z}, t_1) \) is the same as defined previously.

The distribution \( P_0(\gamma) \) is thus related to the solution \( P(\gamma, t/t_0) \) of (4.6.2) by
\[ P(\gamma, t/t_0) = \int_{-\infty}^{0} P_0(\dot{y}, t/t_0) dy \] ...

Hence by solving (4.6.2) and using (4.6.4), the distribution \( P_0(\gamma) \) can be obtained. It is expected that this method would yield exact results for case (i) considered by McFadden and improved approximations for the other cases. A pre-requisite for the solution of (4.6.2) is the evaluation of both \( Q(\dot{y}, t/t_0) \) and \( P(\dot{y}, t/\dot{z}, t_1) \) as functions of time and slope. This is discussed in the following section.

4.7 Derivation of expressions for \( Q(\dot{y}, t/t_0) \) and \( P(\dot{y}, t/\dot{y}_0, t_0) \) for a Gaussian process.

Rice(1) showed that for a Gaussian process \( x(t) \), the joint probability density that \( x(t) \) goes through zero with a slope between \( \dot{y}_0 \) and \( \dot{y}_0+dy_0 \) at \( t_0, t_0+dt_0 \) and through zero with a slope between \( \dot{y} \) and \( \dot{y}+dy \) at \( t, t+dt \) is given by
\[
\text{dt}_0 \text{ dt}_1 \text{ d}\hat{\gamma}_0 \text{ d} \gamma \mid \hat{\gamma}_0 \mid \gamma \left[ \frac{\text{M}}{4\pi} \right]^{\frac{1}{2}} \exp \left[ -\frac{1}{2|\text{M}|} \left( M_{22} \gamma_0^2 + 2M_{23} \hat{\gamma}_0 \gamma + M_{33} \gamma^2 \right) \right] \ldots (4.7.1)
\]

where \( M, M_{22}, M_{33}, M_{23} \) are explained in Section (2.3).

The probability that \( x(t) \) goes through zero with a slope \( \hat{\gamma}_0, \gamma_0 + d\gamma_0 \) at \( t_0, t_0 + dt_0 \) can be shown to be given by:

\[
d\hat{\gamma}_0 \text{ dt}_0 \left| \hat{\gamma}_0 \right| P(0, \hat{\gamma}_0) \ldots (4.7.2)
\]

where \( P(0, \hat{\gamma}_0) \) is the two-dimensional Gaussian density of the two variates \( x(t_0) \) and \( \dot{x}(t_0) \). For the function \( x(t) \) which is assumed to have zero mean, unity variance and a normalised auto-correlation function \( \rho(\tau) \), \( P(y, \dot{\gamma}_0) \) is given by:

\[
P(y, \dot{\gamma}_0) = \frac{1}{2\pi\left[-\rho_0^2\right]^{\frac{1}{2}}} \exp \left[ \frac{y^2}{2\rho^2} - \frac{\dot{\gamma}_0^2}{2\rho_0^2} \right] \ldots (4.7.3)
\]

and so:

\[
P(0, \dot{\gamma}_0) = \frac{1}{2\pi\left[-\rho_0^2\right]^{\frac{1}{2}}} \exp \left[ \frac{-\dot{\gamma}_0^2}{2\rho_0^2} \right] \ldots (4.7.4)
\]

where

\[
\rho_0 = \left[ \frac{\partial^2}{\partial \tau^2} \rho(\tau) \right]_{\tau_0}
\]

The conditional probability density \( P(\dot{y}, t/\dot{\gamma}_0, t_0) \) is therefore given by (4.7.1) divided by (4.7.2), i.e.

\[
P(\dot{y}, t/\dot{\gamma}_0, t_0) = \frac{1}{2\pi\left[-\rho_0^2\right]^{\frac{1}{2}}} \left| \dot{\gamma}_0 \right| \exp \left[ -\frac{1}{2|\text{M}|} \left( M_{22} \gamma_0^2 + 2M_{23} \dot{\gamma}_0 \gamma + M_{33} \gamma^2 \right) \right] \ldots (4.7.5)
\]

In deriving (4.6.2) from (4.6.1) it is clear that:

\[
Q(\dot{y}, t/t_0) = \frac{\int_0^{\infty} P(\dot{y}_0/t_0) \ P(\dot{y}, t/\dot{\gamma}_0, t_0) \ d\dot{\gamma}_0}{\int_0^{\infty} P(\dot{y}_0/t_0) \ d\dot{\gamma}_0} \ldots (4.7.6)
\]
Now it can easily be shown that:

\[ P(\dot{y}_0/t_0) = -\frac{1}{2\rho_0} |\dot{y}_0| \exp \left[ -\frac{\dot{y}_0^2}{2\rho_0} \right] \quad \ldots (4.7.7) \]

and hence \( \int_0^\infty P(\dot{y}_0/t_0) \, d\dot{y}_0 = \frac{1}{2} \) therefore,

\[ Q(t,\dot{y}/t_0) = \frac{1}{2\pi|\mathbb{M}|^2 (\rho_0')^2} |\dot{y}| \int_0^\infty \dot{y}_0 \exp \left[ -\frac{1}{2|\mathbb{M}|} (M_{22} \dot{y}_0^2 + 2M_{23} \dot{y}_0 \dot{y}) \right] \, d\dot{y}_0 \quad \ldots (4.7.8) \]

Let

\[ I = \int_0^\infty \exp \left[ -\frac{1}{2|\mathbb{M}|} (M_{22} \dot{y}_0^2 + 2M_{23} \dot{y}_0 \dot{y}) \right] \, d\dot{y}_0 \quad \ldots (4.7.9) \]

Then

\[ \frac{dI}{d\dot{y}} = \int_0^\infty -\frac{M_{23}}{|\mathbb{M}|} \dot{y}_0 \exp \left[ -\frac{1}{2|\mathbb{M}|} (M_{22} \dot{y}_0^2 + 2M_{23} \dot{y}_0 \dot{y}) \right] \, d\dot{y}_0 \quad \ldots (4.7.10) \]

In terms of \( I \) (4.7.8) becomes:

\[ Q(t,\dot{y}/t_0) = \frac{1}{2\pi|\mathbb{M}|^2 (\rho_0')^2} |\dot{y}| \exp \left( \frac{M_{23}}{2|\mathbb{M}|} \dot{y}^2 \right) \cdot \frac{1}{2|\mathbb{M}|} \cdot \frac{dI}{d\dot{y}} \quad \ldots (4.7.11) \]

To evaluate \( I \), we first express the quantity inside the exponent as a complete square and then make the substitution

\[ x = \frac{1}{2|\mathbb{M}|} \left( \sqrt{M_{22}} \dot{y}_0 + \frac{M_{23}}{\sqrt{M_{22}}} \dot{y} \right) \]

we thus obtain

\[ I = \frac{2|\mathbb{M}|}{\sqrt{M_{22}}} \exp \left( \frac{M_{23}^2}{2|\mathbb{M}| M_{22}} \dot{y}^2 \right) \int_a^\infty \exp(-x^2) \, dx \quad \ldots (4.7.12) \]

where \( a = \frac{M_{23}}{\sqrt{2|M| M_{22}}} \quad \ldots (4.7.13) \)

and the integral \( \int_a^\infty \exp(-x^2) \, dx \) is related to the Error Function
Differentiating $I$, with respect to $\dot{y}$ and substitute in (4.7.11) we obtain the following expression for $Q(t,\dot{y}/t_0)$

$$Q(t,\dot{y}/t_0) = \frac{1}{2\pi(-\dot{y}/t_0)^{1/2}} \left[ \frac{|M|^{1/2}}{M_{22}} \dot{y} \exp \left( \frac{-M_{33} \dot{y}^2}{2|M|} \right) - \frac{M_{23}}{M_{22}} \frac{\sqrt{\Pi/2}}{3/2} \right] \cdot \dot{y} |\dot{y}| \exp \left( a^2 - \frac{M_{33}}{2|M|} \dot{y}^2 \right) \cdot (1 - \text{erf}(a\dot{y})) \quad \cdots (4.7.15)$$

where care must be taken regarding the sign of the slope $\dot{y}$ when evaluating this expression. Thus if the slope is negative, then $\dot{y}|\dot{y}| = -\dot{y}^2$ (negative) whereas if the slope is positive then $\dot{y}|\dot{y}| = \dot{y}^2$. This is the reason why it is written as $\dot{y}|\dot{y}|$ instead of $\dot{y}^2$.

If we integrate this expression over all negative values of $y$, we should obtain the distribution $Q(\tau)$, i.e.

$$Q(\tau) = \int_{-\infty}^{0} Q(t,\dot{y}/t_0) d\dot{y} \quad \cdots (4.7.16)$$

This was done for the purpose of checking the above expression for $Q(t,\dot{y}/t_0)$. The result was the expression obtained by Rice for $Q(\tau)$.

The dependence of both $Q(t,\dot{y}/t_0)$ and $P(t,\dot{y}/t_0\dot{y}_0)$ on the time is implicit through the quantities $M_{13}$, $M_{23}$ and $M_{22}$ which are functions of the auto correlation function and its derivatives.

**4.8 Asymptotic behaviour of $Q(t,\dot{y}/t_0)$ and $P(t,\dot{y}/t_0\dot{y}_0)$**

Before discussing numerical methods for solving the second order integral equation, it is useful to give some results concerning the asymptotic behaviour of the dist-
ributions $Q(t, \dot{y}/t_0)$ and $P(t, \dot{y}/t_0, \ddot{y}_0)$. The results follow, by closely investigating the expressions (4.7.15), (4.7.5) for these quantities.

i) $t - t_0$ fixed.

$$Q(t, \dot{y}/t_0) \to 0 \text{ as } \dot{y} \to 0 \text{ or } \infty.$$  
and

$$P(t, \dot{y}/\ddot{y}_0, t_0) \to 0 \text{ as } \dot{y} \to 0 \text{ or } \infty.$$  

ii) $t \to t_0$ is finite ($\neq 0$)

$$P(t, \dot{y}/\ddot{y}_0, t_0) \to \delta(\dot{y} - \ddot{y}_0)$$  
and

$$Q(t, \dot{y}/t_0) = 0 \text{ if } \dot{y} < 0$$  
$$= 2P(\dot{y}/t_0), \dot{y} > 0$$

This last result is obtained by substituting $\delta(\dot{y} - \ddot{y}_0)$ for $P(t, \dot{y}/\ddot{y}_0, t_0)$ in (4.7.6). It is expected, since the probability that the slope lies between $\dot{y}$ and $\dot{y} + d\dot{y}$ at the point $t$ (which is asymptotically the same as the initial point $t_0$) is the same as the probability that the positive slope at the initial point $t_0$, does indeed lie between $\dot{y}$ and $\dot{y} + d\dot{y}$.

iii) $t - t_0 \to \infty$

In this case

$$Q(t, \dot{y}/t_0) \to P(t, \dot{y}/\ddot{y}_0, t_0) \to \frac{1}{2} \left| \frac{1}{-2\lambda_0^2} \dot{y}^2 \right| \exp\left(\frac{1}{-2\lambda_0^2} \dot{y}^2\right)$$  

...(4.8.1)

(4.8.1) is the same as $P(t, \dot{y})$ which is the probability density that there is a zero crossing with a slope $\dot{y}, \dot{y} + d\dot{y}$ in the time interval $t, t + dt$. This is again expected, since the zero crossing at $t_0$ has little or no influence on the event at $t$, which is separated from $t_0$ by a very long time.

4.9 The solution of the 2nd Order Integral Equation.

In this section, the solution of the integral equation
(4.6.2) is discussed. Analytical solution is not possible and numerical methods have to be considered. Equation (4.6.2) simply expresses the fact that the class of random functions which cross the zero level with slope \( \dot{y}, \dot{y} + d\dot{y} \) at time \( t, t + dt \), given that they have crossed the level at \( t_0 \) with a positive slope, may be classified according to the first time they cross the zero level downwards in the interval \( (t_0 - t) \). We can thus have two categories:

i) Those which cross the zero level for the first time at \( t, t + dt \) with slopes \( \dot{y}, \dot{y} + d\dot{y} \) represented by the term \( P_0(t, y/t_0) \).

ii) Those which cross the zero level for the first time at \( t_1, t_1 + dt \) with slopes between \( \dot{z}, \dot{z} + d\dot{z} \), followed by another downward crossing at \( t, t + dt \) with slopes between \( \dot{y}, \dot{y} + d\dot{y} \). These are represented by the term \( \int_{t_0}^{t} dt_1 \int_{-\infty}^{\infty} dz P_0(t_1, \dot{z}/t_0) P(t, y/t_1, \dot{z}) \).

For stationary random functions, (4.6.2) can be written in the form:

\[
P_0(\dot{y}, \tau) = Q(\dot{y}, \tau) - \int_{-\infty}^{\tau} dV \int_{-\infty}^{\infty} dz P_0(\dot{z}, V) P(\dot{y}, \dot{z}, \tau - V) \]

where \( \tau = t - t_0, V = t_1 - t_0 \) and the conditional probabilities \( P_0, Q, P \) are written in the functional forms indicated, to show their dependence on the appropriate variables.

Writing

\[
G(\dot{y}, V) = \int_{-\infty}^{\infty} dz P_0(\dot{z}, V) P(\dot{y}, \dot{z}, \tau - V) \quad \text{... (4.9.2)}
\]

and

\[
F(\dot{y}, \tau) = \int_{-\infty}^{\tau} dV G(\dot{y}, V) \quad \text{... (4.9.3)}
\]

we have

\[
P_0(\dot{y}, \tau) = Q(\dot{y}, \tau) - F(\dot{y}, \tau) \quad \text{... (4.9.4)}
\]

The numerical solution of (4.9.1) consists of finding \( P_0(\dot{y}, \tau) \) for selected values of the slope \( \dot{y} \), at the discrete time intervals \( 0, \Delta \tau, 2\Delta \tau, \ldots \) etc. In order to
obtain a new set of $P_0(\dot{y}, \zeta)$ corresponding to the slopes at the discrete time $\zeta + \Delta \zeta$, requires the knowledge of the sets at all the past times $0, \Delta \zeta, \ldots, \zeta$. Starting from the initial point $\zeta = 0$, we can successively obtain $P_0(\dot{y}, \Delta \zeta)$, $P_0(\dot{y}, 2\Delta \zeta)$, etc., as a function of the slope $\dot{y}$.

The integrals (4.9.2) and (4.9.3) are evaluated using numerical integration. The formulae used to perform the integration are discussed in Appendix 3. This appendix also contains in more details the procedure used to solve (4.9.1).

Since the accuracy of numerical integration depends largely on the number of ordinates or points at which the integrand is given, it follows that the integral (4.9.3) will not be sufficiently accurate if $\zeta$ was less than $4\Delta \zeta$ as this means that the number of available ordinates are less than 4. This causes $P_0(\dot{y}, \zeta)$ to be inaccurate for small values of $\zeta$. Further, since $P_0(\dot{y}, \zeta)$ depends on its past values, it follows that these inaccuracies are transmitted forward causing errors in the overall distributions.

To avoid this, the distribution $Q(\dot{y}, \zeta)$ was used to provide the initial part of $P_0(\dot{y}, \zeta)$. This is valid since these two distributions are almost identical for small values of $\zeta$, provided the random function $x(t)$ does not have the singularity mentioned in Chapter two. This singularity occurs whenever the power spectrum of $x(t)$ varies as $f^{-4}$ at high frequencies.

Having solved for the distribution $P_0(\dot{y}, \zeta)$, the distribution of the zero crossing intervals $P_0(\zeta)$ may be obtained by numerical means, by integrating $P_0(\dot{y}, \zeta)$ over all the values of the slope $\dot{y}$.

Since random functions having the singularity mentioned above can be truly represented by a two-dimensional
Markov process, it follows that the second order Markovian equation (4.9.1) should lead to exact results for the distribution \( P_0(\tau) \). Unfortunately, the fact that \( Q(\dot{y}, \tau) \) and \( P_0(\dot{y}, \tau) \) are not the same even for very small values of \( \tau \) means that it is not possible for \( Q(\dot{y}, \tau) \) to provide the initial part of the solution \( P_0(\dot{y}, \tau) \). It is also difficult to obtain the initial, non-zero values of the distribution \( P_0(\dot{y}, 0) \). If these values were assumed to be the same as those provided by \( Q(\dot{y}, 0) \) and we proceeded to solve for \( P_0(\dot{y}, \tau) \), it would be doubtful whether the accuracy of the overall result for \( P_0(\tau) \) would be better than that obtained by solving McFadden's first order integral equation.

Further integral equations leading to higher approximations for \( P_0(\tau) \) can be obtained by letting \( n=3, 4 \) etc in the generalised integral equation. However, the increased complexity of the probability distributions and the computational problems involved, make them impractical at this stage.

4.10 An Expression for \( P_n(\tau) \) based on the Markovian Model

Suppose it is possible to solve for the conditional \((n-1)\) \((n-1)\) probability density \( q_0(\dot{y}, \ddots, \dot{y}, t/\dot{y}_0, \ddots, \dot{y}_0, t_0) \) using \((4.3.8)\), then by successive applications of equation \((4.3.1)\) we obtain the conditional probability densities \( q_1, q_2, \ldots, q_n \) etc.

The probability density \( P_n(\tau) \) is then related to \( q_n \) by

\[
P_n(\tau) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} q_n(\dot{y}, \ddots, \dot{y}, t/\dot{y}_0, \ddots, \dot{y}_0, t_0) \, \frac{dy}{dy_0} \ldots \frac{dy}{dy_{(n-1)}} \frac{dy}{dy_0} \ldots \frac{dy}{dy_{(n-1)}}
\]

\((4.10.1)\) is a generalisation of \((4.4.1)\) which if used in conjunction with \((4.3.6)\) and \((4.3.1)\) can, in theory at least
yield exact results for the distribution $P_n(\tau)$, provided the power spectrum is a rational function of $\omega$. Here, as in case of $P_0(\tau)$, mathematical and computational difficulties limit the dimension of the Markov Model. For example, if the model is of dimension one, the relevant equations become:

$$q_n(t/t_0) = \int_0^t dt_1 q_0(t_1/t_0) q_{n-1}(t/t_1) \quad \ldots (4.10.2)$$

$$q_0(t/t_0) = P(t/t_0) - \int_0^t dt_1 q_0(t_1/t_0) P(t/t_1) \quad \ldots (4.10.3)$$

and

$$P_n(\tau) = q_n(t/t_0).$$

Hence in order to obtain an approximation for $P_n(\tau)$ based on the simplest Markovian model we need to perform $(n)$ integrations and this number increases sharply as the dimension of the model increases. It seems therefore that the computations involved in obtaining approximations for $P_n(\tau)$ are formidable.

### 4.11 Summary

In this chapter, a random function with a rational power spectrum is represented as one component of a vector Markov process and a generalised integral equation is derived. This integral equation is shown to form the basis for obtaining exact solutions for the distribution $P_0(\tau)$ and $P_n(\tau)$. Because of mathematical and computational limitations it was necessary to consider Markovian models of low dimensionality and hence be content with approximated results. McFadden's integral equation was identified as the first in a series of approximations based on the generalised...
equation. The second one in the series was considered and its solution was discussed.
CHAPTER FIVE.

Experimental Measurement of the Distribution $P_0(\zeta)$.

5.1 Introduction.

In the preceding chapters, attention was focused on a discussion of analytical solutions to the zero crossing problem. It became clear in the course of discussion that exact solutions are difficult to obtain in view of the mathematical and computational problems involved. Approximate methods however were presented but these are mainly applicable in the case of Gaussian Processes. It follows that for non-Gaussian processes and those Gaussian ones for which the zero crossing distribution is required to be of a better accuracy than that offered by the theoretical approximations, one is compelled to turn to practical measurements of the distribution. These measurements in addition provide the results needed to compare the various approximations, test their validity and that of the assumptions made in deriving them, thus providing further insight into this complex problem. For these purposes it is thought useful to devote this chapter to discuss some of the methods which could be used to obtain practical measurements of the distribution $P_0(\zeta)$. First a brief outline of the basic system of measurements is given and possible methods of its implementation are presented. Later, the method used to measure $P_0(\zeta)$ involving use of the departmental computer is presented.

5.2 Basic system of Axis Crossings Histogram Measurement.

Fig. (4) shows the basic system used to measure the probability histogram of the zero crossing intervals of
Fig. (4) - Basic System for measuring the histogram of zero crossing intervals.
random signals. The system can be used to measure the distribution of intervals at levels other than the zero level by simply subtracting the value of that level from the function $f(t)$ prior to the infinite clipping operation.

The random function generator produces a random function $f(t)$ having the desired statistics and power spectrum. Gaussian functions are normally generated by passing Gaussian white noise through a linear filter with the appropriate frequency response. The infinite clipper transforms $f(t)$ into a binary waveform whose state changes at the zero crossing instants. The binary waveform is next applied to a differentiating network which produces a sequence of positive and negative spikes corresponding to upward and downward crossings respectively. The length of time intervals between successive spikes is measured using a timing circuit which produces an output proportional to this length. The intervals are consequently sorted into classes according to their length. The sorting of intervals is carried out by the multichannel interval analyser which upon receiving an interval uses its length to determine a certain memory location. A count is placed at this location indicating that an interval of a certain length has occurred. In this manner the number of intervals whose lengths lie in the ranges covered by each class are automatically counted. At the end of a run the memory of the analyser contains the frequency distribution of the intervals which is proportional to $P_0(\tau)$ and can be made to approach it by processing a sufficiently long record of $f(t)$ and using a small size class interval.

5.3 System Implementation.
Two possible methods of implementing the above system are now presented. The first is due to Favreau, Low & Pfeffer (10) and the second of J. Rainal (27).

In the first method, an analogue computer was used to simulate the linear filter required to shape the input white noise in order to produce the random function $f(t)$ with the desired power spectrum. The spectrum of the input noise used was flat in the range 0-40HZ. The filter cut-off frequency was designed to be much lower (about 2HZ), in order that frequencies above 40HZ contained in the input noise would have very little effect on the spectrum of the function $f(t)$. $f(t)$ is next infinitely clipped and the binary waveform is differentiated to produce the positive and negative spike sequence. The positive spikes are used to trigger a ramp generator whose output voltage actuates a matrix of high speed relays. The relay contacts may be represented schematically as a switch (see Fig 5) which advances from contact to contact at present values of the ramp voltage. As the switch advances, its arm remains in contact with the $r$th channel during the time interval $(t_{r-1} - t_r)$ where $t_r$ is the time taken by the switch to reach the $(r+1)$th channel, after a positive spike has occurred. The next occurring negative spike signifying the next zero crossing is applied to the $r$th channel, registering a count if it occurs within this time interval. It also resets the ramp generator and relays, ready for the next positive spike to begin the next measurement. The total number of negative spikes counted by the $r$th channel represents the number of intervals between successive zeros, whose lengths lie between the specified values of $t_{r-1}$ and $t_r$. The counters attached to these channels thus contain the frequency distribution of these intervals.
Fig. (5) - Practical System used by Fawreau, Low and Pfeffer to measure zero crossing intervals

Histogram.
In this system, the ramp generator performs the timing function and channel selection. The high-speed relay matrix is one possible realisation of the multichannel interval analyser.

J. Rainal's method is essentially digital in nature. The system initiates commands at intervals nearly one second long. Following each 'initiate' command an electronic gate is opened at the beginning of a zero-crossing interval and closed at the end of it. During the time the gate is open an electronic counter counts the number of clock pulses produced by a high-frequency pulse generator. The count stops at the instant the gate closes and the number is punched on a paper tape. This number is proportional to the interval length and hence can be taken as a measure of it. At the end of a run, the paper tape contains the frequency distribution of the interval lengths. By processing this tape, the frequency distribution and hence $P_0(\tau)$ can be obtained.

In the first method, only half of the intervals, i.e. those initiated by the positive spikes, are analysed. It is possible however to analyse all intervals with additional equipment, and thus double the amount of data obtainable in a given time. This is important since in this method a 2Hz cut-off frequency means a relatively low number of zero crossings. Consequently, in order to obtain sufficiently accurate results, especially when the distribution $P_0(\tau)$ is expected to have very low values, may require a run time of several hours.

The ramp voltage must be very nearly linear in order to represent the interval length. The range of linearity must extend over all the values covered by the relay.
matrix. Hence one of the disadvantages of this method is the strict linearity requirement it places on the ramp generator. Since it is generally difficult to obtain a linear ramp voltage extending over a very wide range, this method is not very suitable for measuring \( P_0(\tau) \) when \( \tau \) is very large. The second method overcomes this problem by using a counter circuit to count the number of high frequency pulses occurring during the interval between successive zeros. Although this method can not determine the precise length of this interval, due to its digital nature, its accuracy improves when long intervals are concerned. Therefore it is more suitable for long interval measurements than the first method. It is however less efficient since only one measurement is performed during the interval between 'initiate' commands, which contains several zero crossings.

5.4 Alternative Method of Probability Histogram Measurement.

An alternative method which we used and which proved to be fast, efficient and reasonably accurate is described shortly. This method made use of the digital computer facilities available in the department which permitted the filtering, zero detection and interval classification to be carried out, mostly by means of a special software programme. Before describing the method it is useful to outline briefly the main features of the departmental computer.

5.4.1 The departmental computer.

The computer contains the usual facilities, namely the store, the processor, the console teletype and the tape reader. The processor is supervised by a special programme
called the 'Executive' which allows control to be exercised on the programme under run.

In addition to these basic facilities there are Analog to digital (AD) and digital to analog (DA) converters. The AD converter allows continuous waveforms to be fed into the computer to be processed and analysed and the DA converter permits the display of information onto the oscilloscope or the X-Y platter. These extra facilities make the computer suitable for conducting real time measurements and analyses of waveforms. The size of the accessible registers in the processor is 16 bits except the one which acts as a programme pointer which is only 13 bits long. The last bit is used to indicate the signs of the numbers held in the registers. The AD and DA converters deal however with words of 10 bits long. This covers 1024 quantisation levels separated by 5 millivolts intervals. In order to prevent overloading, the maximum input voltage to the AD converter must not exceed 5 volts and the maximum input number to the DA must be less than 1024.

5.4.2 Outline of the Measurement System.

The system used in measuring the distribution $P_0(\tau)$ is outlined in block diagram in Fig (6). The output from the white noise generator is fed into the computer, via the AD converter, where it is first filtered digitally. Digital filtering is accomplished by setting up the appropriate difference equation (see below). The filtered waveform is then analysed and its zero crossings are detected. The intervals between successive zeros are measured and classified in categories according to their length. The numbers of intervals in each category or class are
Fig. (6) - System Outline of the Computer Method used to measure the Histogram of the Zero Crossing Intervals.
registered in locations in the computer store. Using the DA converter the frequency distribution of these intervals can be displayed on the oscilloscope, or alternatively a punched tape containing this information may be produced. Next the blocks constituting this system are discussed in detail.

5.4.3 The white noise Generator.

The output noise is derived from a temperature limited thermionic diode noise source whose noise current is proportional to the square root of the dc anode current. In the noise generator used, the anode current is held constant by a means of feedback controlling the filament current. Feedback is also used to stabilise the gain of the amplifiers. Both of these feedbacks resulted in giving a reasonably stationary noise output which is essential for the measurements of $P_0(\gamma)$.

The generator also has a wide noise spectrum which is flat within $\pm 1$db in the range 20HZ-100HZ and $\pm 1.5$db in the range 10-20 MHz. The amplitude probability distribution is very nearly Gaussian in the range covered by $\pm 3\sigma$, where $\sigma$ is the standard deviation.

5.4.4 Digital Filtering.

The frequency shaping of the white noise in order to generate the random function $f(t)$ with the desired power spectrum was carried out using digital filtering. Digital filters are used in sampled data systems to perform equivalent functions to those performed by analog filters in continuous systems. They can therefore be used in computer simulation of analog filters. The physical real-
isation of both types is however different. Linear continuous filter theory is based on the mathematics of linear differential equations whereas linear digital filter theory is based on that of linear difference equations. An \( m \)th order linear difference equation may be written as

\[
Y(nT) = \sum_{i=0}^{m} a_i X \left[ (n-i)T \right] - \sum_{i=1}^{m} b_i Y \left[ (n-i)T \right] \quad \text{...}(5.4.1)
\]

where \( X(nT) \) and \( Y(nT) \) are the input and output sample sequences respectively, \( T \) being the sampling interval. The form \((5.4.1)\) emphasises the iterative nature of the difference equation; given the \( m \) previous samples of the output \( Y \) and the \( r+1 \) most recent values of the input \( X \), the new output may be computed from \((5.4.1)\). Equation \((5.4.1)\) represents what is known as a 'Recursive Digital Filter'. Such a filter involves feedback loops as can be seen from Fig.\((7)\) which gives a possible schematic representation of \((5.4.1)\) consisting of unit delays, adders and multipliers. Non recursive digital filters have no feedback, consequently the previous output samples have no influence on the new sample value. This situation corresponds to \( m=0 \) in \((5.4.1)\) which would represent a zero order difference equation.

Real time digital filtering consists of performing the iteration of \((5.4.1)\) for each arrival of a new input sample. The frequency response of the filter is determined by the coefficients \( a_i \) and \( b_i \) whose evaluation forms the filter design problem. The accuracy of the filtering process depends on the quantisation error in the (A-D) convertor and also the error arising, due to the finite length of the computer word, from quantising the coefficients \( a_i, b_i \) and the intermediate results of various multiplications and
Fig. (7) - Pictorial representation of the $n$th order difference equation.
additions required to compute (5.4.1).

The design procedure used to evaluate digital filters with various frequency spectra will not be discussed in this chapter. It is sufficient to point out here that the digital filters we set up on the computer for the purpose of probability histogram measurements can be represented in the form (5.4.1) and their measured frequency spectra were found to be very close to their analog counterparts. However, a brief description of design procedures used to arrive at the difference equations is given in Appendix 4, since the filtering process represents an integral part of the measurement method.

5.4.5 Zero Detection and Histogram Measurements.

Fig(6) shows a flow chart of the computer programme which was used to perform the functions of the blocks encircled by the dotted lines in the block diagram of Fig.(6). The white analog noise is sampled and the sample values are converted into numbers by the (A-D) convertor which is operated in the 'clocked' mode. This means that acquisition of a new input sample is only achieved successfully if a clock pulse occurs. This makes it possible to eliminate variations in the programme cycle time and to regularise the sampling operation which is necessary if the frequency response of the digital filter is to be stationary. In order to conduct real time analyses, the clock (sampling) interval must be greater than the maximum value of the programme cycle time.

The programme sets up the appropriate difference equation which is used to produce a new filtered sample value \( Y(nT) \) every time a new input sample \( X(nT) \) is
Input Sequence, 
\[ x_n = x(nT) \]

\[ Y(nT) = \sum_{n=0}^{n} x_n + \sum_{n=0}^{n-1} b_i Y_i \]

\[ Z = Y_n \cdot Y_{n-1} \]

is \( Z \leq 0 \) ?

YES

\[ I(N) = 0; N = 1, 2 \ldots M \]

NO

\( N = 0 \)

\[ Y(nT) = \sum \ldots + \sum \ldots \]

\[ Z = Y_n Y_{n-1} \]

is \( Z \leq 0 \) ?

YES

\[ N = N+1 \]

NO

is \( Y_n = 0 \) ?

YES

\[ |Y_n| \geq |Y_{n-1}| \]

NO

\( N = N+1 \)

YES

\[ I(N) = I(N) + 1 \]

PUNCH TAPE

Fig. (8) - Flow Chart of the Computer Programme used to implement the System of Fig. (6).
acquired from the A-D convertor. The sequence \( Y(nT) \) represents the sample values of the random function \( f(t) \) whose zero crossing distribution histogram is required. In order that this sequence sufficiently approximates to the continuous waveform \( f(t) \), the sampling interval must be several times smaller than the period of the highest significant frequency present in the spectrum of \( f(t) \). \( T \) was taken to be at least 10 times smaller than the period of the 3 db cut-off frequency. Even then, only a few of the sample sequence \( Y(nT) \) would actually have zero values. These samples provide a small fraction of the total number of zero crossings of \( f(t) \), since it would be more likely for these crossings to occur between the samples rather than at the sample points themselves.

In order to detect zero crossings which occur between samples, the following procedure was used. This consists of evaluating the product of the two successive samples \( Y(nT) \) and \( Y(nT-\tau) \). If this product is negative, then a zero crossing must have occurred somewhere in the interval separating the two samples. If, on the other hand, the product is positive, then no zero crossing can have occurred. If the product is zero, then one of the two samples, or both, must be zero. Using this procedure, we can detect almost all the zero crossings of \( f(t) \).

If the product is positive, the programme asks for a new input sample \( X(nT) \) and the above procedure of digital filtering and zero detection is carried out. This is repeated again and again until a zero crossing is detected by obtaining a zero or a negative product. At this point the programme sets up an area of locations in the
computer store which can be denoted by the array \( I(N) \). The relative addresses of the locations within the array are made to coincide with the relative class sizes of interval length. This makes it possible to automatically classify a given interval if its length is decided. The size of this length, which will be in multiples of the sampling interval \( \tau \), is simply used to construct the address of the location assigned to count the events, represented by the occurrence of intervals having this length. Using this address, the location is accessed and the present count inside it is updated by adding 1. After the end of a run, the array will contain the frequency distribution of the zero crossing intervals.

Initially the array locations are all set to zero. When the first zero crossing is detected using the procedure described above, a counter denoted by \( N \) is also set to zero. The programme now asks for a new input sample \( x(n\tau) \) and the filtering and zero detection procedures are applied again. If the product \( y(n\tau) \cdot y(n\tau - \tau) \) is positive, indicating that no zero crossing has occurred, the counter \( N \) is updated by adding 1 and the whole procedure is repeated again, and again until this product becomes zero or negative. At this point, the value of \( N \) which is clearly proportional to the length of the interval between successive zero crossings is used to construct the address of the location within the array \( I(N) \) at which this event is registered, by updating the count already present in that location by 1, i.e., \( I(N) = I(N) + 1 \). Having done this, the counter is reset to zero, \( (N=0) \) and a new input sample \( x \) is acquired ready for the next measurement.

After a zero crossing (not the initial one) is detected, the programme may or may not update the value \( N \) by 1,
(see the flow chart) prior to using it to determine the address of the location $I(N)$. This depends on whether the zero crossing is assumed to occur nearer to the present sample $Y(n\cdot T)$ than to the previous one $Y(n\cdot T - T)$ or vice-versa. If the sampling rate is sufficiently high, so that linear interpolation between the two sample values may be assumed, the zero crossing based on this interpolation occurs nearer to the sample with the smaller absolute value.

The programme tests for the relative size of the absolute values of the samples $Y(n\cdot T)$ and $Y(n\cdot T - T)$ and decides to update $N$ by 1 if $Y(n\cdot T)$ is smaller than $Y(n\cdot T - T)$ otherwise, it leaves $N$ as it is. Updating $N$ is equivalent to the decision that a zero crossing has occurred at $(n\cdot T)$, whereas leaving $N$ as it is, is to decide that a zero crossing has occurred at $(n-1)\cdot T$. In this latter case $N$ must be reset to 1 instead of zero before the next measurement is performed, since the minimum value the next interval can have is $T$.

The programme also tests for whether $Y(n\cdot T)$ is actually zero and updates the present value of $N$ prior to using it in constructing the address of the location $I(N)$.

Real time interval analyses are achieved if the sampling interval $T$ is chosen sufficiently long to allow analog to digital conversion, digital filtering, zero detection, interval measurement and classification to be performed. Yet $T$ must be sufficiently short in order to achieve a small class interval which is necessary for accurate measurement of the distribution $P_0(\mathcal{D})$. It is possible to satisfy both requirements simultaneously by designing the filters to have relatively low cut-off frequencies. It was found that
the programme cycle time which is mostly consumed by the digital filtering process, varied between 50 and 180 microseconds. The first figure corresponds to a second order difference equation—whereas the second corresponds to a seventh order equation describing a 7-pole Butterworth spectrum. This permitted the use of sampling rates in the range 5-10KHz which consequently resulted in choosing the filters' cut-off frequencies in the range of 250-500 Hz. This gave a ratio of sampling frequency to the 3-db cut-off frequency of at least 10.

Low cut-off frequencies lead to low rates of zero crossings. In our case however, it was possible to obtain, over one million zero crossing intervals in one hour measurement time. This is sufficient to minimise the effects of statistical variance of the number of events counted in most of the classes and would permit reasonable accuracy in the measured P0(\bar{T}) over a wide range of interval lengths.

5.5 Accuracy of the Computer Method.

The method discussed in the last section has many advantages over the other two methods described earlier. Provided suitable digital computer facilities are available, the implementation of this method is straightforward and can mostly be achieved by a software programme. The method is efficient since it considers and analyses every zero crossing interval of the sampled function f(t). It is also reasonably fast since, as we found, it is possible to analyse a very large number (nearly one million) intervals in one hour. If digital filtering is replaced by analog filtering prior to the analog to digital
conversion, the programme cycle time will be considerably reduced. This permits the use of high sampling rates and would result in increasing the speed and accuracy of the method. This is particularly useful if the distribution $P_0(\tau)$ for random functions with complex power spectra is to be measured, especially when digital simulation of these spectra may involve high order difference equations whose evaluation consumes considerable computing time.

The method however suffers from some inaccuracies which are inherent in sampled data systems. Some of these inaccuracies arise in the digital filtering and were mentioned in Section (5.4.4). These are mainly due to the finite length of the computer word and the quantisation errors of the analog to digital conversion. These inaccuracies caused the power spectrum of the filtered noise to somewhat deviate from its desired shape.

Other inaccuracies arise from the fact that the filtered noise whose $P_0(\tau)$ is to be measured, is only available in the form of quantised samples at regular intervals. This fact makes it extremely difficult to detect the exact instants at which zero crossings occur. The procedure for detecting zero crossings, explained in the last section, can only determine the exact sampling interval over which a zero crossing has occurred, but is unable to decide with certainty the exact location of that crossing. This location is approximated to be that of either one of the two samples defining the sampling interval, depending on the relative size of their absolute values. This leads to errors in estimating the length of the intervals between zero crossings. Errors also arise if two zero crossings occur over one sampling
interval. In this case the procedure would fail to detect either zero crossing, since the product \( Y(nT) \cdot Y(nT-T) \) would be positive.

Most of these errors can be minimised by making the sampling rate sufficiently high compared to the high frequency cut off of the function spectrum. If the power spectrum has a gradual cut-off, the sampling rate needs to be particularly high since in this case, short intervals between successive zero crossings are very likely to occur.

The quantisation of the time scale in steps of \( T \) and the inability to exactly measure the length of the zero crossing intervals gives rise to another significant inaccuracy. This results from the overlap between neighbouring histogram classes and is discussed below.

5.5.1 Histogram Class Overlap.

The length of intervals as decided by the computer method is quantised in steps of \( T \). Consequently the classes into which the intervals are sorted, are separated by a class interval of length \( T \). If we were dealing with a continuous time scale, then class \( r \), say, would contain all the intervals measured by an analog device, whose length lies between \( rT \) and \( (r+1)T \), (see Fig.9). In the sampled data system set up by the computer method, it can be demonstrated that intervals whose exact length lies in the range \( (rT-(r+1)T) \) are not necessarily classified in Class \( r \), and that it is possible for a given interval to be classified in either one of two neighbouring classes, depending on the positions of the zero crossings relative to those of the sampling points. This is shown below. Fig(\#) shows two successive zero crossings separated by an interval of length 1 which is
Fig. (9) - Non-Overlapping System of Interval Classification.

Fig. (10) - Effect of Sample Point distribution on Interval Classification.
equal to \( rT + x \). The distribution of the sample points is such that the zero crossing at \( B \) is situated mid-way between the two sampling points. If the sample points were shifted to the left so that the sample point \( C \) lies anywhere in the shaded area, then using the procedure of the computer method, the zero crossing at \( A \) would be decided to occur at the sampling point \( C \) since it would be nearer to \( C \) than to \( C' \). Similarly the zero crossing at \( B \) would be assumed to occur at the sampling point \( D \). This means that the length of the interval would be taken as \((r+1)T\) and consequently is classified in class \((r+1)\). If on the other hand, the sampling points are shifted in either direction so that the sample point \( C \) lies outside the shaded area, the length of the interval would be taken as \( rT \) which is the length of either \( CD \) or \( C'D \). This would result in the interval being classified in class \( r \).

The above argument clearly shows that it is possible for an interval of the same length to be classified in either one of two neighbouring classes, depending on the relative displacement of the zero crossing instants in relation to the sample points. In order to estimate the effects of this class overlap on the frequency distribution of the intervals, it is necessary to determine the probability that an interval of a given length \( l=rT+x \) is classified in either neighbouring class. For this purpose it is assumed that the relative displacement of the sampling points in relation to the zero crossing instants is a random variable with a uniform probability density given by

\[
P(d) = \begin{cases} 
\frac{1}{T} & \text{if } 0 \leq d \leq T \\
0 & \text{elsewhere} 
\end{cases} \quad (5.5.1)
\]
where \( d \) is the relative displacement.

Let \( P(r/l) \) be the probability that given an interval of length \( l = rT + x \), it would be classified in class \( r \), then we have:

\[
P(r/l) = \frac{T-x}{T} \quad ; \quad 0 \leq x \leq T \quad \ldots (5.5.2)
\]

and hence

\[
P(r+1/l) = \frac{x}{T} \quad \ldots (5.5.3)
\]

where \( P(r+1/l) \) is the probability that the interval \( l \) is classified in class \( (r+1) \).

Equations (5.5.2) & (5.5.3) quantify what is intuitively expected, namely that as \( x \) increases, it would be more likely to have a sample point inside the shaded area, resulting in the interval being classified in class \( (r+1) \).

The range of interval lengths which could be classified in class \( r \) may be deduced from (5.5.2) & (5.5.3). This range is given by:

\[
(r-1)T \leq l \leq (r+1)T
\]

where the upper limit is obtained by letting \( x=T \) in equation (5.5.2). The lower limit is obtained by letting \( r \) assume its previous value \((r-1)\), and \( x=0 \) in equation (5.5.3). It is therefore possible for an interval whose length lies in the range \( (r-1)T - (r+1)T \) to be classified in class \( r \). The probability, \( P(r) \), that a given interval whose length \( X \) lies in this range, is classified in \( r \) is given by:

\[
P(r) = \frac{X - (r-1)T}{T} \quad (r-1)T \leq X \leq rT \quad \ldots (5.5.4)
\]

\[
P(r) = \frac{(r+1)T - X}{T} \quad rT \leq X \leq (r+1)T \quad \ldots (5.5.5)
\]
(5.5.4) & (5.5.5) follows immediately from (5.5.2) and in Fig.(11) 
(5.5.3). F(r) is sketched as a function of X for successive 
values of r. The sketch clearly shows the class overlap 
between neighbouring classes. The sketch also shows that 
intervals whose length lies in the range \((rT-\frac{T}{2})\rightarrow(rT+\frac{T}{2})\) are 
more likely to be classified in class \(r\) than in either 
neighbouring class. This approximates to a possible non-
overlapping situation in which class \(r\) would only contain 
intervals of length in that range.

5.5.2 Effect of class overlap on the Computed Relative Frequency.

The effects of class overlap on the computed relative 
frequencies of intervals are now considered. It is shown that 
these effects cancel out if the number of events counted 
in each class are expected to be equal. This corresponds to 
a uniform distribution of intervals. The effects of overlapping 
also cancel out if this distribution is linear or piecewise 
linear in the range \((r-1)T\rightarrow(r+1)T\) (for all \(r\)).

In the case of uniform distribution, the number of 
intervals in the range \((rT\rightarrow rT+\frac{T}{2})\) classified in class 
\((r+1)\), because of overlapping would be the same as the 
number of intervals in the range \((rT+\frac{T}{2}\rightarrow(r+1)T)\) class-
ified in class \(r\). Similarly, the number of intervals in 
the range \((rT-\frac{T}{2}\rightarrow rT)\) classified in \((r-1)\), is equal to 
that of intervals in the range \(((r-1)T\rightarrow rT-\frac{T}{2})\) classified 
in \((r)\). The equivalence is due to the symmetry of the 
probabilities \(P(r)\) (see Fig.11) and the equal number of 
intervals expected to be classified in each class.

Because of this, the system of classification is eff-
fectively the same as the non-overlapping system shown 
superimposed in dotted lines in Fig.(11). Under this 
system, class \(r\) contains intervals whose length lies in
Fig. (11) - Sketch of the probability $P(r)$, for successive values of $r$. 
in the range \( rT - \frac{T}{2} \rightarrow rT + \frac{T}{2} \). The two systems are equivalent since the effects of overlapping between adjacent classes tend to cancel each other out resulting in zero net effect. In general, the distribution of interval lengths is non-uniform. In this case the effects of class overlap do not cancel out since it is possible for a class, say \( r \), to have a net gain of events from neighbouring classes, if intervals whose lengths lies in the range \((r-1)T \rightarrow (rT - \frac{T}{2})\) and \((rT + \frac{T}{2}) \rightarrow (r+1)T\) have higher probabilities of occurrence then intervals in the range \((rT - \frac{T}{2}) \rightarrow rT + \frac{T}{2}\). This results in distorting the frequency distribution and must be accounted for. However if the distribution of interval lengths is linear or alternatively it can be approximated over the ranges \((r-1)T \rightarrow (r+1)T\) (for all values of \( r \)) by segments which are piecewise linear, then in general the effects of class overlap tend to cancel each other out. This is demonstrated using Fig. (17).

The figure shows a typical probability distribution of zero crossing intervals which is approximated by a linear curve over the range \((r-1)T \rightarrow (r+1)T\). Because of the symmetry of the probability \( P(r) \), we need only to consider the effects of overlapping resulting from the unequal number of intervals expected due to the linear rise of the interval distribution curve.

The shaded area represents the difference between the number of intervals expected to have lengths in the range \((rT + \frac{T}{2}) \rightarrow (r+1)T\) and that of the intervals whose length is expected to lie in the range \( rT \rightarrow rT + \frac{T}{2} \). Similarly \( B \) represents the difference between the number of intervals lying in the range \((rT - \frac{T}{2} \rightarrow rT)\) and that of those lying in the range \((r-1)T \rightarrow (rT - \frac{T}{2})\).
Fig. (12) - Effects of class overlap tends to cancel if $P_0(T)$ is piecewise linear range.
Hence the overlap between the classes \((r-1)\) and \((r)\), results in class \(r\) incurring a net loss in the number of events recorded in it by an amount proportional to the area \(B\). The overlap between class \((r)\) and \((r+1)\) on the other hand, results in a net gain in the number of events recorded in class \((r)\) whose amount is proportional to the area \(A\). If the slope of the interval distribution curve is constant over the range \((r-1)T \rightarrow (r+1)T\), then the two areas \(A\) and \(B\) are equal. Hence, the net gain in the number of events acquired by class \((r)\) from the neighbouring class \((r+1)\) is cancelled by the net loss incurred by \((r)\) to \((r-1)\). The overall gain is thus zero and the system is equivalent to the non-overlapping system mentioned above.

It is possible to approximate any probability curve by a piecewise linear curve whose slope remains reasonably constant over a short interval if that interval is chosen sufficiently small; then the effects of class overlap can be very nearly eliminated and the accuracy of the computer method can be made comparable to that resulting from analog measurements of the zero-crossing intervals.

The effects of class overlap cannot be eliminated completely. The effects are noticeable at the points where the distribution of intervals is expected to have maximum and minimum values. In the case of a maximum, the slope changes sign from positive to negative. If the maximum value occurs at a point which lies in the range covered by class \(r\), then this class incurs losses to both neighbouring classes \((r-1)\&(r+1)\). This results in a computed frequency which is somewhat lower than its true value. Conversely in the case of a minimum, the computed frequency would be somewhat higher than its true value. These effects
were noted in some of the frequency distributions measured by the computer method and were consequently minimised by further increasing the sampling rate. The effects were corrected by adding the estimated losses to the classes covering these values.

Experimental results were obtained for a number of random functions having various frequency spectra. These results are presented and discussed in Chapter Six.
CHAPTER SIX.

Review and Discussion of Results.

6.1 Introduction.

In this chapter, results are presented for the probability distribution $P_0(\tau)$ obtained by various methods. In particular, emphasis is laid on theoretical results based on the solution of the second order Markovian Integral Equation, and those based on the multivariate integral method which involves the Schlaffli Function $S_n$ and its derivatives. Results using these two methods, as well as experimental results using the computer method described in the previous chapter, were obtained for a number of random functions with different power spectra. These results are presented first, and the agreement between the theoretical distributions and their experimentally measured counterparts are then discussed. The effects of the variation of the power spectra on the theoretical results and their deviation from the experimental distributions is then studied in detail. The study reveals some interesting results about the correlation of successive intervals and the behaviour of the distribution $P_0(\tau)$ at large interval lengths.

In order to investigate the effects of variations in the power spectrum on the zero crossing distribution, random signals with different spectra were considered for which both theoretical and experimental results were obtained. The spectra considered can be grouped into three main categories:

i) Power Spectra which can be expressed in the form

$$
\phi_{1n}(\omega) = \frac{1}{(1+\omega^2)^n}
$$

(...6.1.1)
ii) The Butterworth class which may be expressed in the form

\[ \Phi_{2n}(\omega) = \frac{1}{1 + \omega^{2n}} \]  \hspace{1cm} (6.1.2)

iii) Power spectra displaying infinitely sharp frequency cut-off characteristics, examples of which are the ideal low pass and band pass spectra, given by

\[ \Phi_3(\omega) = 1 \; ; \; \omega \leq 1 \]  \hspace{1cm} (6.1.3)
\[ = 0 \; ; \; \omega > 1 \]

and

\[ \Phi_4(\omega) = 1 \; ; \; 1 \leq \omega \leq 2 \]  \hspace{1cm} (6.1.4)
\[ = 0 \; ; \; \text{elsewhere} \]

where \( \omega \) represents the radian frequency which has been normalised with respect to some reference frequency \( \omega_0 \).

In Group (i), \( \omega_0 \) is the frequency at which \( \Phi_{4n}(\omega) = \frac{1}{2^n} \), whereas in group (ii), \( \omega_0 \) is the 6db cut-off frequency, since \( \Phi_{2n}(\omega) = \frac{1}{2} \). In group (iii), \( \omega_0 \) is the upper cut-off frequency in the ideal low pass case, but it represents the lower cut-off frequency in the band pass case.

Both (6.1.3) & (6.1.4) may be expressed in the following form

\[ \Phi_3(\omega) = \lim_{n \to \infty} \left[ \frac{1}{(1 + \omega^{2n})} \right] = \lim_{n \to \infty} \left[ \Phi_{2n}(\omega) \right] \]  \hspace{1cm} (6.1.5)

\[ \Phi_4(\omega) = \lim_{n \to \infty} \left[ \frac{1}{(1 + (\omega^2/2)^{2n})(1 + \frac{1}{\omega^{2n}})} \right] \]  \hspace{1cm} (6.1.6)

Hence the ideal low pass spectrum can be considered as the limiting form of the Butterworth spectrum.

Random functions with group (i) spectra may be obtained by passing white noise through a cascade of \( n 'RC' \) networks all having the same time constant. Varying the number of
cascaded networks n results in random signals, having different bandwidths and cut-off frequencies. As n increases, higher frequencies are increasingly attenuated and consequently the bandwidth become narrower, reaching the zero value as n gets infinitely large. This results in the random function degenerating to a constant dc value.

Group (ii) spectra have in general sharper cut off frequencies than their corresponding (same value of n) group (i) counterparts. This may be seen from Figs. (13) & (14) in which spectra of both groups are plotted for several values of n. Increasing n in this case results in sharper frequency cut-off with bandwidths which become more and more rectangular. As n approaches infinity, the spectrum tends to the ideal low pass spectrum.

Group (iii) spectra have an interesting property which is not present in the other groups, namely band-limitedness which implies that frequencies outside a certain range are completely absent from the power spectrum. It will be seen later that this property has an important influence on the correlation between successive intervals.

These three groups provide a reasonably wide variety of spectra, whose bandwidths and frequency cut-off rates can be made variable. In general group (i) spectra have gradual frequency cut-off rates, whereas (iii) have infinitely sharp frequency cut-off rates. In between, group (ii) have bandwidths with moderate to sharp frequency cut-offs but which are not band limited.

Various probability distributions concerning the zero crossing distribution, for example $Q(\gamma), Q(\gamma, \zeta)$, are more directly related to the auto correlation function and its derivatives than to the corresponding power
Fig. (13) - Examples of group (i) spectra.

Fig. (14) Examples of group (ii) spectra.
spectrum. Likewise \( P_0(\tau) \) is expected to be more directly related to the auto-correlation function than the power spectrum. Hence it would be useful to consider the auto-correlation functions of the three groups of spectra given above, in order to gain further insight into the effects spectrum variation has on the distribution \( P_0(\tau) \).


The auto-correlation function is related to the spectrum by the Fourier integral

\[
\psi(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\omega) e^{-j\omega \tau} d\omega \quad \ldots (6.1.7)
\]

Using (6.1.7) and evaluating the integral by means of contour integration in the complex plane, the normalised auto-correlation functions \( \rho(\tau) \) corresponding to the three groups of power spectra can be shown to be respectively given by the following expressions:

i) \( \rho_{1n}(\tau) = \sum_{r=0}^{n-1} \frac{(n-1)!}{r! (n-1-r)!} \left( \frac{2^r}{2n} \right) \frac{|\tau|^r}{r!} e^{-|\tau|} \quad \ldots (6.1.8) \)

ii) \( \rho_{2n}(\tau) = \sin \frac{\pi}{2n} \left[ e^{-|\tau|/2} \sum_{r=1}^{n-1/2} e^{-|\tau|} \cos \left( \frac{\pi n}{n} \right) \right] \cos \left( \frac{\pi n}{n} \sin \left( \frac{\pi n}{n} \right) \right) \quad n \text{ odd} \quad \ldots (6.1.9a) \)

and

\( \rho_{2n}(\tau) = \sin \frac{\pi}{2n} \left[ \sum_{r=1}^{n-1/2} \cos \left( \frac{\pi n}{2n} \right) \cos \left( \frac{\pi n}{2n} - |\tau| \sin \frac{\pi n}{2n} \right) \right] \quad n \text{ even} \quad \ldots (6.1.9b) \)

iii) \( \rho_3(\tau) = \frac{\sin \tau}{\tau} \quad \ldots (6.1.10) \)

and

\( \rho_4(\tau) = \frac{\sin 2\tau - \sin \tau}{\tau} \quad \ldots (6.1.11) \)
If \( n \) assumes the value 1, then both expressions for \( \rho_{1,n}(\tau) \) and \( \rho_{2,n}(\tau) \) yields \( e^{-|\tau|} \). This is expected since in this case both expressions yield the normalised auto-correlation function of the first-order Markov Process, which has an infinite zero crossing rate. For values of \( n \) other than zero and 1, \( \rho_{1,n}(\tau) \) and \( \rho_{2,n}(\tau) \) display different characteristics which are explained below.

Upon close examination of expression (6.1.8), it can be seen that it is made up of the product of two quantities. These are the exponential \( e^{-|\tau|} \) and a polynomial in \( |\tau| \) whose coefficients are all positive.

The polynomial is therefore a continuously increasing function of \( \tau \). Further examination however reveals that this polynomial never exceeds the exponential \( e^{1\tau} \), since the coefficient of \( |\tau| (\tau > 1) \), can be shown to be smaller or at most equal to \( \frac{1}{\tau!} \) which is the corresponding coefficient in the series expansion of \( e^{1\tau} \). Hence as \( \tau \) increases the polynomial gets increasingly smaller than \( e^{1\tau} \). It follows therefore that the auto-correlation function \( \rho_{1,n}(\tau) \) is a continuously decaying function of \( \tau \) which has a unity value at \( \tau = 0 \), and decays continuously towards zero as \( \tau \) increases.

Increasing \( n \) tends to increase the value of the coefficients of the polynomial and hence increases its overall value. This results in slowing the rate at which \( \rho_{1,n}(\tau) \) decays towards zero. As \( n \) approaches infinity it can easily be shown that each coefficient of the polynomial tends to \( \frac{1}{\tau!} \), resulting in the limit

\[
\lim_{n \to \infty} \rho_{1,n}(\tau) = 1 \quad \ldots (6.1.2)
\]

which is the auto-correlation function of a dc signal.
Both expressions for the auto-correlation function $\rho_{2,n}(\tau)$ consist of series of exponentially damped cosine terms. In addition, there is an exponential term $e^{-|\tau|}$ when $n$ is odd. The frequency of the $r^{th}$ cosine term is given by $\frac{1}{2\pi} \sin \left( \frac{\pi r}{n} \right)$ for odd $n$ and $\frac{1}{2\pi} \sin \left( \frac{\pi r}{2n} \right)$ for even $n$. The exponential damping of the $r^{th}$ term is respectively given by $e^{-|\tau|} \cos \left( \frac{\pi r}{n} \right)$ and $e^{-|\tau|} \cos \left( \frac{\pi r}{2n} \right)$. In both cases the damping is greatest where $r=1$, and progressively decreases as $r$ increases. The cosine term with the highest frequency which is $\frac{1}{2\pi} \sin \left( \frac{n-1}{2n} \cdot \pi \right)$ in both cases, is the least damped.

Consequently, the auto-correlation function $\rho_{2,n}(\tau)$ is expected to be largely influenced by this term over intermediate and large values of $\tau$. Over this range, $\rho_{2,n}(\tau)$ oscillates about the zero value with an ever decreasing amplitude and with frequency approximately equal to the highest frequency. This was found to be the case when considering low values of $n$. As $n$ tends to infinity, the highest frequency tends to $\frac{1}{2\pi}$ which is the frequency at which $\frac{\sin \tau}{\tau}$ oscillates about the zero value, over large and intermediate values of $\tau$. This is not surprising since (6.1.9) should approach $\frac{\sin \tau}{\tau}$ as $n$ tends to infinity.

The main difference between the two types of auto-correlation $\rho_{1,n}(\tau)$ & $\rho_{2,n}(\tau)$ lies in the manner in which both of these functions which start at unity at $\tau=0$, decay towards the zero value as $\tau$ increases. $\rho_{1,n}(\tau)$ is monotonic and never assumes negative values. In contrast $\rho_{2,n}(\tau)$ continuously oscillates about the zero level with an ever decreasing amplitude before finally reaching it at infinity. $\rho_{2,n}(\tau)$ thus assumes negative values, as well as positive values. Increasing $n$ in both cases tends to decrease the rate of decay and hence increase the memory of the ran-
dom function. This is clear in the first case. In the second case, increasing \( n \) results in progressively reducing the damping of the oscillations. As a result, they become more pronounced and persist for longer time.

Power spectra with infinitely sharp frequency cut-off which constitute group (iii), also have auto-correlation functions which oscillate about the zero value but the rate at which the oscillations decay is smaller than that encountered in group (ii) spectra.

6.3 Review of Theoretical and Experimental Results.

In this section, theoretical results based on the second order Markovian Equation and the Schaffli Function obtained for the three groups of power spectra are presented and the agreement between them and the experimental distributions is discussed. The following cases from each group were considered.

**Group (i)**

1) \( n=2 \); \( \phi_{1,2}(\omega) = \frac{1}{(1+\omega^2)^2} \) \hspace{1cm} (6.3.1)

2) \( n=3 \); \( \phi_{1,3}(\omega) = \frac{1}{(1+\omega^2)^3} \) \hspace{1cm} (6.3.2)

3) \( n=5 \); \( \phi_{1,5}(\omega) = \frac{1}{(1+\omega^2)^5} \) \hspace{1cm} (6.3.3)

**Group (ii)**

4) \( n=3 \); \( \phi_{2,3}(\omega) = \frac{1}{1+\omega^6} \) \hspace{1cm} (6.3.4)

5) \( n=4 \); \( \phi_{2,4}(\omega) = \frac{1}{1+\omega^8} \) \hspace{1cm} (6.3.5)

6) \( n=5 \); \( \phi_{2,5}(\omega) = \frac{1}{1+\omega^{10}} \) \hspace{1cm} (6.3.6)
7) \( n=6 \); \( \phi_{2,6}(\omega) = \frac{1}{1+\omega^{12}} \) \( \ldots (6.3.7) \)

8) \( n=7 \); \( \phi_{2,7}(\omega) = \frac{1}{1+\omega^{14}} \) \( \ldots (6.3.8) \)

Group (iii)

9) Ideal low pass spectrum; \( \phi_2 = 1 \), \( 0 \leq \omega \leq 1 \)
\[ = 0 \], \( \omega > 1 \) \( \ldots (6.3.9) \)

10) Ideal band pass spectrum \( \phi_4 = 1 \); \( 1 \leq \omega \leq 2 \)
\[ = 0 \] elsewhere. \( \ldots (6.3.10) \)

Future reference to these cases is achieved by using either the power spectrum \( \phi(\omega) \) or the corresponding autocorrelation function \( \rho(\tau) \), with the appropriate subscripts, (In order to facilitate neater expressions, the comma between subscripts will henceforth be omitted). Also in order to avoid frequent repetition of the full title of each approximation, the abbreviation 'MK2' is used to denote the second order Markovian approximation. The set of three approximations \( P_{0n}(\tau), (n=3,4,5) \) based on the Schaffli Function \( S_n \) and described in Chapter Three, will be denoted collectively by the abbreviation 'P0N', with 'P03', 'P04' and 'P05' corresponding to \( P_{03}(\tau), P_{04}(\tau) \) and \( P_{05}(\tau) \) respectively. Occasionally, 'MK1' is used to denote McFadden's approximation based on "quasi independence" of successive intervals.

6.3.1 Experimental Results.

The probability distribution of intervals between successive zero crossings was measured for a number of cases using the digital computer method discussed in Chapter 5. These cases included \( \phi_{12}, \phi_{13} \) from group (i) and \( \phi_{23}, \phi_{24} \)
\( \phi_{25}, \phi_{26}, \text{and } \phi_{27} \) from group (ii). For the ideal low pass spectrum, it was decided to use the experimental results due to Blötekjaer (28). These were found reasonably close to the results corresponding to the spectrum \( \phi_{27} \) (see Fig. 43), which may be considered as an approximate physical realisation of the idealised spectrum.

In obtaining the experimental distributions, over one million intervals were analysed in each case. This number makes it possible to obtain reasonably accurate results over a wide range of interval lengths. Higher sampling rates were used in the cases \( \phi_{12}, \phi_{13}, \phi_{23} \) and \( \phi_{24} \) than in the other cases, since these have a significantly high frequency content. The ratio of the sampling frequency to \( \omega_0 \) (see section 6.1) was chosen to be 20, whereas in the cases \( \phi_{25}, \phi_{26}, \text{and } \phi_{27} \) a ratio of 10 was found sufficient. The experimental measured distributions are shown plotted in Figs. (15 - 23) for the cases mentioned above as well as the case \( \phi_{15} \) for which experimental results were borrowed from the paper of Low, Favreau and Pfeffer (10). The experimental curves are those denoted by 'EXP'. Also plotted in these figures are the distributions \( Q(\gamma) \) and the approximations MK2 corresponding to these cases, (except case \( \phi_{12} \), for which MK1 is plotted instead of MK2). In order to show the variations of the various distributions more accurately over the range of large interval lengths this part of the distributions is plotted again using an expanded probability scale. Also, photographs of the oscilloscope displays, showing these distributions are presented in Figs. (24 - 30). The higher sampling rate used in the cases \( \phi_{12}, \phi_{13}, \phi_{23}, \text{and } \phi_{24} \) can be seen from these photographs.
The interval length $\tau$ which forms the abscissae of the graphs plotted in these and subsequent figures, is expressed in normalised units. The normalisation is effected by using the following relationship

$$\tau = \omega_0 \tau_0$$

where $\tau$ is the normalised time, $\tau_0$ is the time in seconds and $\omega_0$ is the reference radian frequency defined for the three groups of power spectra in section 6.1, p. 201.
Fig. (16) - Zero crossing distribution for the spectrum $\Phi_x(\omega) = 1/(1 + \omega^2)^2$
Fig. (16) - Zero Crossings distribution for the spectrum $\phi(\omega) = 1/(1 + \omega^2)^3$. 
Fig. (17) - Zero crossings distribution for the spectrum $Q_0(\omega) = \frac{1}{(1 + \omega^2)^5}$
The figure shows probability density distribution curves for the spectrum $Q(\omega) = 1/1+\omega^6$. The curves are labeled as $\Pi^2$ and show the distribution over interval length $\xi$. The expanded view highlights the behavior at higher $\omega$ values.
Fig. (19) - Zero crossing distributions Curves for the spectrum $\phi_4(\omega) = 1/1 + \omega^8$
Fig. (20) - Zero Crossing Distribution Curves for the spectrum $f_{\omega}(\omega) = 1/1 + \omega^{10}$
Fig. (21) - Zero Crossings Distribution Curves for the spectrum $\phi_\omega(\omega) = 1/1 + \omega^{12}$
Fig. (22) - Zero crossings distribution for the spectrum $\phi(\omega) = 1/1 + \omega^4$. 
Fig. (23) - Zero crossing distribution curves for the low-pass spectrum.
Fig (24) - Oscillogram of $P_0 (\tau)$ for the spectrum $\phi_{12} = 1/(1 + \omega^2)^2$.

Fig. (25) - Oscillogram of $P_0 (\tau)$ for the spectrum $\phi_{13} = 1/(1 + \omega^2)^3$. 
FIG (26) - Oscillogram of $P_0(\tau)$ for the spectrum

$\phi_{23} = \frac{1}{1 + \omega^6}$.

FIG (27) - Oscillogram of $P_0(\tau)$ for the spectrum

$\phi_{24} = \frac{1}{1 + \omega^8}$.
Fig. (28) - Oscillogram of $P_0(\zeta)$ for the spectrum

$$\phi_{25} = 1/1 + \omega^{10}.$$
Fig. (30) - Oscillogram of $P_0(\tau)$ for the spectrum

$\phi_2 = 1/1 + \omega^{14}$. 
A brief account of the experimental results is given below. Case \( \Phi_{12} \) represents a 'singular case' for which the distribution \( P_0(\tau) \) has a non-zero value at the origin, which also happens to be the maximum value of this distribution. The measured curve for this case fails to give accurate results at \( \tau = 0 \) and its immediate vicinity. This is due to the fact that discrete methods do not accurately measure the very small intervals which are highly probable in this case. Over the rest of the range, the experimental curve should be reasonably accurate.

In all other cases, the experimentally measured distributions start from zero at the origin, and increasing as \( \tau \) increases, reach a maximum value at some value of \( \tau \). For \( \Phi_{13} \) this value is approximately 2, whereas for group (ii) spectra, it varies between 3 & 4. The gradient of the distribution at the origin and in the immediate vicinity is higher for \( \Phi_{13} \) than for \( \Phi_{23} \). Further within group (ii) spectra, it can be seen from the graphs that this gradient decreases as \( n \) increases. The reason for this lies in the continuous attenuation suffered by high frequencies contained in the power spectra as a result of increasing \( n \).

6.3.2 Theoretical Results based on the second order Markovian Equation.

Approximations for the distribution \( P_0(\tau) \) were obtained for cases (2-10) as listed in section (6.3). The approximations are represented by the curves 'MK2' in Figs. (6-23(i)-(ii)). As explained earlier, this method cannot be solved accurately for case \( \Phi_{12} \). Instead the approximation 'MK1' based on McFadden's Integral equation is plotted in Fig. (15).
Fig. (31) - Comparison of the distribution curves $Q(\tau)$ and $Wx2$ of the ideal low-pass spectrum with the corresponding ones for the spectrum:

$$Q(\omega) = \frac{1}{1 + \omega^2L}$$
The corresponding distribution \(Q(\tau)\) which gives the probability distribution of a downward zero crossing at \(t+\tau; t + \tau + d\tau\), given an upward zero crossing at time \(t\) is superimposed on each one of these graphs. Since, in general \(Q(\tau)\) is very close to \(P_0(\tau)\) over the range of small values of \(\tau\), it follows that this distribution will be useful in testing and checking the accuracy of the experimental method used to measure the distribution \(P_0(\tau)\). If this method yields reasonably close results to \(Q(\tau)\) over this range, then it can be assumed that this method would be reasonably accurate throughout the range. \(Q(\tau)\) is also useful when results for the different groups of spectra are compared.

By studying the graphs shown in Figs. (6,25,31), the following points regarding the agreement of the approximation 'MK2' with the experimental curves 'EXP' can be made: i) The best agreement between 'MK2' and the experimental results corresponds to power spectra of low order, i.e. those whose denominators consist of low degree polynomials, e.g. \(\phi_{13}\) and \(\phi_{23}\). In \(\phi_{13}\) the agreement is almost perfect in the entire range over which the integral equation was solved. For \(\phi_{23}\), the agreement, although good, is nevertheless less close over the range of large interval lengths, than that obtained in the case of \(\phi_{13}\). Over this range, 'MK2' tends to be slightly lower than 'EXP' and in fact dips negative at about \(\tau = 23\).

ii) For the same value of \(n\), the agreement between 'MK2' and 'EXP' is better if the spectrum belongs to group (i) than that obtained if it belongs to group (ii). This can be seen by comparing the agreement obtained in the cases \(\phi_{13}\) and \(\phi_{15}\) with that resulting in the cases \(\phi_{23}\) and \(\phi_{25}\).
Further, whereas 'MK2' decays monotonically towards zero as \( \tau \) increases for Group (i) spectra; it invariably dips negative for Group (ii) spectra. The value at which this happens is large \( \tau \) (approximately 23) for low order spectra and approximately 12.5 for high order spectra (see Fig. 31). iii) Allowing for the observations made above, the approximation 'MK2' in general, agrees reasonably well with the experimental results for both groups (i) and (ii) spectra, provided that \( n \) which represents the order of these spectra is reasonably low.

iv) In contrast, MK2 does not agree very well when the infinitely sharp cut-off spectra of group (iii) are considered. For the ideal low pass spectrum, see Fig. 23 where agreement with 'EXP.' extends over a limited range (zero to \( \gamma \)). The curve MK2 dips negative at approximately \( \tau = 12 \) and from then on oscillates about the zero value. Fig. (31) also shows the distribution \( Q(\tau) \) and the approximation MK2 for the case \( \Phi_{2,12} \) superimposed on the corresponding graphs of the ideal low-pass filter. \( \Phi_{2,12} \) corresponds to a 12th order Butterworth spectrum which has a very sharp cut-off and as such can be considered as a very close physical realisation of an ideal low-pass spectrum. Although the corresponding graphs for \( Q(\tau) \) and 'MK2' agree reasonably well with each other, appreciable differences still exist. This indicates that the two spectra which are nearly identical when viewed in the frequency domains are not so identical when their zero crossing distributions are concerned.

v) For the band pass spectra, MK2 dips negative at \( \tau = 5.5 \) and from then on continues to oscillate about the zero value. The distribution appears to be very narrow where it is estimated that about 90 per cent of the area under the curve is concentrated in the range (zero to three). Over this range 'MK2' may be considered as
a close approximation to $P_0(\tau)$ since it agrees closely with $Q(\tau)$. Beyond the value $\tau = 4$, the distribution $P_0(\tau)$ is expected to have a relatively small value differing appreciably from those suggested by 'MK2'.

iv) The distribution $Q(\tau)$, after reaching its maximum value which normally co-incides with the maxima of both 'MK2' and 'EXP', decreases towards its steady state value $P(0, \Delta t)$. In group (i), $Q(\tau)$ has a negative slope over the entire range following the point at which the maximum value occurs. In group (ii) and (iii) the slope may have positive as well as negative values. Consequently $Q(\tau)$ decays monotonically towards $P(0, \Delta t)$ for group (i), whereas for group (ii) & (iii), the distribution oscillates about the steady state value with a continuously decaying amplitude. Figs. (18 - 22) show that these oscillations are weak for low order spectra, becoming progressively more pronounced as $n$ increases. The oscillations are generally infinitely more marked for group (iii) spectra which possess/sharp cut-off frequencies. The oscillations displayed by the distribution $Q(\tau)$ appear to have frequencies similar to those of the oscillations experienced by the corresponding auto-correlation function. In group (ii) the frequency is approximately that of the cosine term with the highest frequency in the expressions for $\rho_{2n}(\tau)$. In group (iii) the frequencies are nearly $\frac{1}{2\pi}$ and $\frac{1}{\pi}$ which are the frequencies at which $\rho_3(\tau)$ and $\rho_4(\tau)$ oscillate about the zero level.

6.3.3 Theoretical results based on the Schaffli Function.

Further approximations for the distribution $P_0(\tau)$ were evaluated using the multivariate method, based on the
Schaffli function $S_n$ which is described at the end of Chapter Four. The results are plotted for the cases (2-10) in Figs. (33-40). The curves are denoted by 'P03', 'P04' and 'P05'. Also superimposed on these graphs are the 'EXP' curves in order to test their agreement with the experimental results. The graphs reveal the following points.

i) In the range of small $\zeta$, the three curves $P03$, $P04$ and $P05$ (denoted collectively by $PON$) coincide with each other as well as the 'EXP' curves. The coincidence of the $PON$ curves implies that the multivariate integrals $V_3$, $V_4$ & $V_5$ must be very close to the probability distribution $V(\zeta)$. Consequently the $PON$ curves are very close to the actual distribution $P_0(\zeta)$ over this range. As $\zeta$ increases, the three curves depart from each other and 'P05' must be considered as the best approximation.

ii) For most of the cases considered 'P05' does not provide as close an approximation to $P_0(\zeta)$ as that obtained from 'MK2'. This can be assessed by comparing the agreement of both approximations with the experimental results in each of the cases.

iii) In group (i) spectra, the agreement appears to be better in case $\phi_{15}$ than in case $\phi_{13}$. For both of these cases 'P05' yields lower values than those suggested by the experimental curves over intermediate and large values of $\zeta$. However, for these spectra, 'P05' does not yield negative values.

iv) In contrast 'P05' dips negative for all group (ii) spectra. This is not very noticeable for low values of $n$. As $n$ increases, the approximation 'P05' oscillates more and more noticeably about the zero value, but the agreement of 'P05' and the experimental curves appears to improve slightly in the range over which $P_0$ is positive. The oscillations appear also in both 'P03' and 'P04'.
but at an earlier stage.

v) In group (iii) spectra, the agreement of approximation P05 appears to be comparable to that obtained from 'MK2'. In the ideal band pass spectrum the agreement might even be closer over the range (0-Ω).

vi) The oscillations in the approximations 'FON' seem to be once again related to those of the corresponding auto-correlation functions.

6.4 General Discussion of Results.

It is clear from the various results presented in the last section, that the shape of the power spectrum has a significant influence on the probability distribution of the zero crossing intervals as determined by experiment as well as on the theoretical approximations for this distribution. In this section we present a detailed discussion on the effects, the variation of power spectrum has on these results. In the course of discussion, some interesting results on the relationship between the power spectrum and the correlation of successive zero crossings emerge. The discussion deals with the following topics:

i) Effects of spectrum variations on P0(τ).

ii) Relationship between the power spectrum and the correlation of successive intervals.

iii) Effects of spectrum variations on the approximation MK2.

iv) Effects of spectrum variations on the approximation FON.

v) Behaviour of the distribution P0(τ) at large interval lengths.

vi) Comparison between various approximations.
Figure 33 - Graphs of the approximations 'PON' for the spectrum $\phi_n(\omega) = \sqrt{1 + \omega^2}$
Fig. 3h - Zero crossings distribution curves for the spectrum

\[ \Phi_{15}(\omega) = \frac{1}{1 + \omega^2}^5 \]
Fig. (36) - Zero crossings distribution curves for the spectrum $\Phi_2(\omega) = 1/(1 + \omega^2)$.
Fig. (37) - Zero crossings distribution curves for the spectrum \( \phi_2(\omega) = 1/(1+\omega^10) \)
Fig. (33) - Zero crossings distribution curves for the spectrum $\Phi_{R_0}(\omega) = 1/1 + \omega^{12}$
Fig. (40) - Zero crossings distribution curves for the ideal low-pass spectrum.
Fig. (M) - Zero crossings distribution curves for the ideal band pass spectrum.
In dealing with topic (i), the experimental distributions will be used, since these represent the most accurate results for $P_0(\tau)$. Using these distributions in conjunction with the corresponding $Q(\tau)$ and $P_{ODD}(\tau)$, some interesting results regarding the correlation of successive interval lengths are obtained. These results are then used to explain the differences in the agreement with the experimental results of both theoretical approximations 'MK2' and 'PON'.

The behaviour of the distribution $P_0(\tau)$, at large interval lengths is then discussed, in order to establish whether the exponential assumption is valid in all cases. Finally a comparison between the various proposed approximations, based on the discussion of these topics is given.

6.4.1 Effects of Spectrum Variations on $P_0(\tau)$.

In order to study the effects of power spectrum variations on the distribution $P_0(\tau)$, the experimental distributions obtained for various spectra are plotted. Fig. (4.2) shows the distributions corresponding to the spectra $\phi_{12}$, $\phi_{13}$ and $\phi_{15}$ which belong to group (i), super imposed on each other, whereas Fig (4.3) superimposes the experimental curves for $\phi_{23}$, $\phi_{24}$, $\phi_{25}$, $\phi_{26}$ and $\phi_{27}$ as well as that for the ideal low pass filter due to Blöteker.

Considering group (i) spectra first, we note that as $n$ varies between 0 and $\infty$, the spectrum $\phi_{4n}(\omega)$ varies between that of white noise ($n=0$) and that of a constant dc signal ($n=\infty$). The probability distribution of intervals between successive zero crossing of white noise consists of a delta function at the origin, whereas that of dc signal consists of a line which coincides with the $\tau$-axis. When
Graph - $P_0(\gamma)$ for examples of the spectrum $\phi_2(\omega) = 1/(1 + \omega^2)^n$, for $\nu = 2, 5$. The graph shows the distribution $P_0(\gamma)$ as a function of the interval length $\gamma$. The spectrum is given by the equation $\phi_2(\omega) = 1/(1 + \omega^2)^n$. The graph includes curves for different values of $n$, with $n = 2$ and $n = 5$ specifically highlighted.
Fig. (43) - Graphs of the distribution $P_0 (\tau)$ for examples of the spectrum $\phi_{2n}(\omega) = 1/1 + \omega^{2n}$
n=1, \( \Phi_{11}(\omega) \) corresponds to a first order Markov process which has an infinite rate of zero crossings. The distribution \( P_0(\tau) \) in this case also consists of a delta function at the origin. For \( n > 1 \), the distribution \( P_0(\tau) \) is therefore expected to lie between the limits corresponding to \( n=0,1 \) on one hand and that corresponding to \( n=\infty \) on the other. The graphs corresponding to \( n=2,3 \) and 5 confirm this.

As \( n \) varies between 1 and \( \infty \), the 6db cut-off frequency which is given by \( \frac{1}{2\pi} \sqrt{\frac{2}{\pi} (\omega - 1)} \), progressively moves towards the zero frequency. As more and more higher frequencies are attenuated, the zero crossings of the random function becomes less rapid. Hence the mean value of the distribution \( P_0(\tau) \) increases as \( n \), the order of the spectrum, is increased.

The behaviour of \( P_0(\tau) \) at the origin and its immediate vicinity is governed by the high frequency content of the spectrum. Consequently in this range, the distribution \( P_0(\tau) \) must have higher values for spectra which corresponds to small values of \( n \). Fig. (42) shows that over the range (0-3), \( P_0(\tau) \) is larger for \( \Phi_{12} \) than \( \Phi_{13} \), which in turn, is higher than \( P_0(\tau) \) for \( \Phi_{15} \). Conversely over the range of large \( \tau \), which is largely influenced by the low frequency content of the spectrum, the distribution \( P_0(\tau) \) is expected to have the largest values corresponding to higher values of \( n \).

Apart from the case \( \Phi_{12} \), the distribution \( P_0(\tau) \) starts from zero and achieves a maximum value at some value of \( \tau \). This value of \( \tau \) progressively increases as \( n \) increases, since this results in shifting the frequency cut-off towards the zero frequency. The frequency shift also results in increasing the value of \( P_0(\tau) \) at large interval lengths which leads to the widening of the range of \( \tau \) over which \( P_0(\tau) \)
has significant values.

Hence in order that the area under the curve $P_0(\bar{c})$ shall be unity, this increase over the large range is compensated for by the decrease in the values of $P_0(\bar{c})$ over the range of small $\bar{c}$. Thus as $n$ increases, the maximum value of the distribution $P_0(\bar{c})$ continuously decreases as well as shifting to the left.

The overall effect of increasing $n$ is therefore a corresponding increase in the mean value and standard deviation of the distribution $P_0(\bar{c})$. The effect on the mean value can be seen from Table 2, whereas the effect on the standard deviation may be assessed from the graphs.

**Table 2.**

<table>
<thead>
<tr>
<th>$n$</th>
<th>Mean value of $P_0(\bar{c})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.14</td>
</tr>
<tr>
<td>3</td>
<td>5.44</td>
</tr>
<tr>
<td>4</td>
<td>7.023</td>
</tr>
<tr>
<td>5</td>
<td>8.311</td>
</tr>
</tbody>
</table>

Next the experimental distributions for cases involving group (ii) spectra and the ideal low pass spectrum are considered. These are plotted in Fig. (43). The part of these distributions over the range of large intervals is plotted again in Fig. (43) using an expanded probability scale in order to compare the distributions more accurately. As mentioned earlier, the agreement between the experimental curve for $\phi_27$ and that for the ideal low pass spectrum due to Blotekjaer(28) appears to be quite close particularly over large values of $\bar{c}$.

As $n$ increases, frequencies higher than $\frac{1}{2\pi}$, which is the 6db cut-off frequency of all group (ii) spectra, are increasingly attenuated, whereas those frequencies just
below $\frac{1}{2\pi f}$ are enhanced relatively. This results in sharpening the frequency cut-off of the spectrum, but at the same time leaves the cut-off frequency unchanged. Hence as $n$ increases, the mean value of the distribution $P_0(\gamma)$ is expected to increase by a relatively small amount. Table 3 gives a list of the mean values

<table>
<thead>
<tr>
<th>$n$</th>
<th>Mean value of $P_0(\gamma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4.44</td>
</tr>
<tr>
<td>4</td>
<td>4.88</td>
</tr>
<tr>
<td>5</td>
<td>5.09</td>
</tr>
<tr>
<td>6</td>
<td>5.20</td>
</tr>
<tr>
<td>7</td>
<td>5.27</td>
</tr>
<tr>
<td>$\infty$</td>
<td>5.45</td>
</tr>
</tbody>
</table>

Here again by using similar argument to that used in dealing with group (i), we can explain why near the origin, the greater gradient of $P_0(\gamma)$ corresponds to the case having the lower value of $n$. Similarly it can be explained why the larger value of $P_0(\gamma)$ over the intermediate range corresponds to the case with the higher value of $n$. However, when comparing group (ii) with group (i) spectra the following points can be made:

i) For group (ii) spectra, it can be seen from Fig(43) that over the range of large $\gamma$, the graphs appear to be much closer to each other than those of group (i). The reason lies in the cut-off frequency of group (ii) spectra which remains constant and is considerably sharper than group (i) spectra. This means that the low frequency content of group (ii) spectra changes only slightly as $n$ increases. Consequently the part of the distribution $P_0(\gamma)$ over large interval lengths which is governed by the low frequency content
varies only slightly as \( n \) increases. On the other hand the gradual frequency cut-off of group (i) spectra causes the low frequency content to change more appreciably as \( n \) increases, which results in larger differences between the distributions \( P_0(\gamma) \) corresponding to the different values of \( n \) over this range.

ii) The close agreement between the curves for various group (ii) spectra at large values of \( \gamma \) and the appreciable difference between these curves at small values of \( \gamma \), causes the curves with the lower gradients at the origin to have the higher maximum values. This is necessary in order that all the curves shall have a unity area. This means that all the curves corresponding to higher values of \( n \) will achieve higher maximum values. In group (i) spectra, the opposite situation was obtained.

iii) As \( n \) varies between 0 and \( \infty \) in group (ii) spectra the distribution \( P_0(\gamma) \) varies between the delta function at the origin corresponding to \( n=0,1 \) and the distribution \( P_0(\gamma) \) for the ideal low pass spectrum. In contrast, the distribution \( P_0(\gamma) \) for group (i) varies between a delta function at the origin and a line coinciding with \( \gamma \) axis. Further, since the bandwidth and cut-off frequency of group (ii) spectra remains approximately constant as \( n \) increases, whereas both of these quantities vary considerably for group (i) spectra, it follows that the shape of the distribution \( P_0(\gamma) \) varies less dramatically in the first case, as \( n \) varies, than in the latter. This explains the small differences existing between the experimental distributions for various group (ii) spectra and the considerably larger ones that exist between those corresponding to various group (i) spectra.
6.4.2 Relationship between the power spectrum and the Correlation of Successive Intervals.

The close agreement between the distributions \( Q(\tau) \) and \( P_0(\tau) \) over the range of small \( \tau \) which can be seen from the graphs shown in Figs. (16-23), implies that over this range the distributions \( P_{2N}(\tau) \), \((N \geq 1)\) must have negligible values compared to \( P_0(\tau) \), since \( Q(\tau) \) is related to \( P_{2N}(\tau) \) by the following series:

\[
Q(\tau) = P_0(\tau) + P_2(\tau) + P_4(\tau) + \cdots (6.4.1)
\]

The range over which close agreement between \( Q(\tau) \) and \( P_0(\tau) \) is obtained also varies from one case to another, but in general seems to increase as the order of the spectrum represented by the value of \( n \) increases.

A similar series to (6.4.1) which was given before is the following

\[
P_{ODD}(\tau) = P_1(\tau) + P_3(\tau) + P_5(\tau) + \cdots (6.4.2)
\]

where \( P_{ODD}(\tau) \) is also given by

\[
P_{ODD}(\tau) = U(\tau) - Q(\tau) \quad \cdots (6.4.3)
\]

If the distributions \( Q(\tau) \) and \( P_{ODD}(\tau) \) are used in conjunction with the distribution \( P_0(\tau) \), the initial parts of the distributions \( P_1(\tau) \) and \( P_2(\tau) \) can be obtained.

Since \( P_N(\tau) \) represents the probability distribution of the sum of \((N+1)\) successive intervals, it would be possible, using \( P_0(\tau) \) and the initial parts of \( P_1(\tau) \) and \( P_2(\tau) \) to obtain some results relating to the dependence between successive intervals.

The graphs shown in Figs. (16-23) reveal that in general the agreement between \( P_0(\tau) \) and \( Q(\tau) \) extends over a wider range for group (ii) spectra than for group (i) and that for both groups the range of agreement increases as \( n \) increases. Further, whereas \( Q(\tau) \) for group (i) decays monotonically towards its steady state value after reaching
its maximum value, in group (ii), this distribution oscillates about the steady state value. This difference in the behaviour of \( Q(\tau) \) in the two groups is expected to have some effect on the correlation of successive intervals.

In order to study the effects of the power spectrum on the behaviour of \( P_0(\tau) \), and the correlation between successive intervals, two spectra, representative of their groups are chosen. These are the spectrum \( \Phi_{\text{L}} \) and the ideal low pass spectrum which can also be considered as the limiting spectrum of group (ii). For each one of these spectra, the following distributions are plotted in Figs. (44) & (45).

1) \( P_0(\tau) \) as measured experimentally.
2) \( Q(\tau) \)
3) \( P_{\text{ODD}}(\tau) \)
4) \( Q(\tau) - P_0(\tau) \)

The initial part of the last curve provides a close approximation for the distribution \( P_2(\tau) \) whereas that of curve 3 provides another for \( P_1(\tau) \).

The oscillations displayed by \( Q(\tau) \) in the ideal low pass spectrum case is now explained with the aid of the auto-correlation function \( \frac{\sin \tau}{\tau} \). This function which starts from unity at \( \tau = 0 \) changes sign at \( \tau = \pi, 2\pi, 3\pi \ldots \text{etc} \) and achieves alternately maximum positive and negative values at \( \tau = 0, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots \text{etc} \). Positive auto-correlation means that on average, the signs of the random function \( f(t) \) at two instants of time, separated by \( \tau \), are more likely to be the same, than opposite. The converse is true if the auto-correlation is negative. Hence if an upward zero crossing occurs at the earlier instant, then using this argument, the probability of a downward crossing at an instant \( \tau \) later, would be greater than the steady state value \( P_-(0, \Delta t) \) if \( \tau \) lies in the ranges \((\pi-2\pi);(3\pi-4\pi)\ldots \text{etc}\). Conversely
Fig. (44) - Graphs of the distribution $\mathbb{P}_n(\tau)$ for the Ideal low-pass spectra.
Fig. (45) - Graphs of some distributions for the spectrum $\Phi_{13}(\omega) = 1/(1 + \omega^2)^3$
this probability would be smaller than $P_\tau(0, \Delta t)$ if $\tau$ lies in the ranges $(2\pi \to 3\pi); (4\pi \to 5\pi); \ldots$ etc. Hence $Q(\tau)$ is expected to oscillate about the steady state value with frequency which is nearly the same as that of $\frac{\sin \tau}{\tau}$. Using the same argument, explains the oscillations in $P_{\text{ODD}}(\tau)$.

In the case $\Phi_{13}$ and more generally group (i) spectra, the auto correlation function is always positive but continuously decaying. Here after a certain time has elapsed following the occurrence of an upward zero crossing, the probability of a downward crossing becomes increasingly less influenced by the initial event. As this influence progressively weakens, the probability of a downward zero crossing approaches $P_\tau(0, \Delta t)$. Hence $Q(\tau)$ decays continuously towards this value.

The band-limitedness of the ideal low pass spectrum means that frequencies higher than a certain frequency are infinitely attenuated. In our case this frequency is taken to be $\frac{1}{2\pi}$. The zero crossings of this frequency alone occurs regularly at intervals of $\pi$. This means that the maximum number of zero crossings contained in a time interval of length $N\pi$ is given by $(N+1)$. By adding to this frequency the other lower frequencies which constitute the ideal low pass spectrum, the rate of zero crossings is reduced and the average value of the interval length between successive zero crossings is raised from $\pi$ to 5.441. Although intervals of length less than $\pi$ are now possible, since $P_0(\tau)$ is finite over this range it would nevertheless be extremely unlikely that more than two zero crossings occur over this range. More generally over an interval of length less than $N\pi$, it would be very unlikely that more than $(N+1)$ zero crossings occur. If this was likely it would mean that zero
crossings can occur at a rate higher than that of the highest frequency \( \frac{1}{2\pi} \), which is unlikely.

The above argument leads to the result that the probability distribution \( P_N(\tau), (N > 0) \) has very small values over the range of interval lengths given by \((0-N\pi)\). The curves for \( P_{ODD}(\tau) \) and \((Q(\tau) - P_0(\tau)) \) appear to confirm this. \( P_{ODD}(\tau) \) which is initially the same as \( P_1(\tau) \) does not start to assume significant values till after \( \tau = \pi \). It is not till \( \tau \) is nearly equal to \( 2\pi \), that \( P_1(\tau) \) gets larger than \( P_0(\tau) \), which indicates that over this range \((0-2\pi)\), it is more likely to have two zero crossings instead of three. The other curve which is initially the same as \( P_2(\tau) \) does not rise significantly till \( \tau \) is nearly \( 2\pi \) which further confirms the above result. Using this result it can be assumed that \( P_{ODD}(\tau) \) provides a close approximation to \( P_1(\tau) \) over the range \( 0-3\pi \), whereas \((Q(\tau) - P_0(\tau)) \) provides another close approximation for the distribution \( P_2(\tau) \) over the range \( 0-4\pi \).

From the graphs in Fig. 44 it would appear that the maxima of distribution \( Q(\tau) \) occur as a result of one of the distributions \( P_{2N}(\tau) \) achieving its maximum value. This is true of the two maxima at \( \tau = 4 \) and \( 11 \) which coincide with the maximum values of \( P_0(\tau) \) and \( P_2(\tau) \) respectively. Similarly those for \( P_{ODD}(\tau) \) which is expected to have maxima coinciding with those of the distribution \( P_{2NN}(\tau) \). Further the maxima of \( Q(\tau) \) coincide with the minima of \( P_{ODD}(\tau) \) and vice-versa. This is not surprising since these two distributions express the probabilities of two mutually exclusive events.

The significance of the point made in the last paragraph lies in the following: that given an interval of
length $\tau$ between an upward zero crossing at $t_0$ and another zero crossing at $t_0 + \tau$, it would be possible to decide with some certainty, which one of the succeeding zero crossings is most likely to occur at $t_0 + \tau$. For example, if $\tau$ is in the range $(2\pi - 3\pi)$, then the most likely zero crossing is the second following the one at $t_0$. More generally if $\tau$ is in the range $(N+1)\pi - (N+2)\pi$, the most likely zero crossing after the one at $t_0$ would be the $(N+1)^{st}$ succeeding crossing.

The gradual frequency cut-off of the spectrum $\Phi_1^\omega(\omega)$ on the other hand contains higher frequencies than $\frac{1}{2\pi}$. This makes it possible to have more zero crossings in a given interval of time than the maximum number obtained in the ideal low pass spectrum. Here, although $P_N(\tau), (N \gg 1)$ is very small near the origin, indicating that very small intervals are not likely to contain more than two zero crossings, it is difficult to estimate the values of at which $P_N(\tau)$ starts to rise significantly. Consequently, it is not clear how far the distributions $P_{ODD}(\tau)$ and $(Q(\tau) - P_0(\tau))$ can closely represent $P_1(\tau)$ and $P_2(\tau)$. Judging by the range over which $Q(\tau)$ is in close agreement with $P_0(\tau)$, which is relatively small, (see Fig 45), it is expected that these two curves would represent closely $P_1(\tau)$ and $P_2(\tau)$ over a comparable range. This means that the distributions $P_N(\tau)$ start to rise significantly at a relatively much lower value of $\tau$ than those obtained in the case of the ideal low pass spectrum. Consequently, it would be quite probable that a short interval of time contains many zero crossings.

The distributions $P_N(\tau)$ in this case overlap considerably in contrast to the ideal low pass spectrum in
which each one of the distributions \( P_N(\tau) \) appears to have a range of dominance where it is larger than any other distribution. In the \( \Phi_{15} \) case, given an interval \( \tau \) between an upward zero crossing at \( t_0 \) and a succeeding zero at \( t_0 + \tau \), it would not be possible to decide which succeeding zero is most likely to correspond to the one at \( t_0 + \tau \) unless \( \tau \) is very small. This is because the considerable overlap between the distributions \( P_N(\tau) \) makes it comparatively likely that this zero crossing corresponds to one of many possible succeeding crossings.

The above discussion leads to interesting results relating to the correlation between successive interval lengths. For the ideal low pass spectrum, the fact that \( P_N(\tau), (N > 1) \) have negligible values over the range \((0 - N\pi)\) leads to the following qualitative results.

i) Given an interval of length \( \zeta_0 \) which is less than \( \pi \), it would be very unlikely that this interval would be followed by an interval of length \( \tau \), such that \( \zeta_0 + \tau < \pi \).

In other words, the conditional probability density \( P_1(\tau | \zeta_0 < \pi) \) is negligibly small if \( \tau \) lies in the range \( (\pi - \zeta_0) \).

ii) Given two successive intervals whose sum \( \zeta_2 \) is less than \( 2\pi \), it would be very unlikely that they would be followed by an interval \( \zeta \) whose length is less than \( (2\pi - \zeta_2) \), i.e. the conditional probability density \( P_2(\tau | \zeta_2 < 2\pi) \) is very small if \( \tau \) lies in the range \( (2\pi - \zeta_2) \).

If, on the other hand \( \zeta_2 \) is greater than \( 2\pi \), then it is quite probable that the succeeding interval may have a relatively small length in the range \((0 - \pi)\).

iii) In general if \( \zeta_N \) represents the sum of \( N \) successive zero crossing intervals, which is less than \( N\pi \), then the conditional probability density \( P_N(\tau | \zeta_N < N\pi) \) that the
next interval length $\tau$ lies in the range $(N\pi - \tau_N)$ is very small. On the other hand, if $\tau_N$ is greater than $N\pi$, then it is quite likely that the following interval length may have relatively small values in the range $(0 - \pi)$.

Point (i) leads to the conclusion that a very short interval is not likely to be followed by a similarly short one, i.e. short intervals tend to repel each other. Points (ii) and (iii) imply that the length of a given interval is largely dependent on the sum of the preceding intervals. This clearly contradicts the assumption of "quasi independence" and renders it invalid in this case.

On the other hand, since the distribution $P_N(\tau)$ overlap considerably more in the case of $\phi_{13}$ spectrum which permits more zero crossings in a given interval, it follows that less stringent requirements are placed on the lengths of successive intervals. Here it is quite probable that a short interval may follow another similarly short one. Further, whereas correlation between successive intervals is not discounted completely, due to the attenuation of the higher frequencies present in this spectrum. It is nevertheless logical to assume that the length of an interval is only slightly influenced by the sum of the two, or more generally, $(2N+2)$ proceeding intervals. Hence, in this case McFadden's assumption of quasi independence is not far from correct.

These results may be generalized for other spectra within the two groups (i) and (ii). For $\phi_{12}$ which corresponds to the singular case, the zero crossings occur even more rapidly than in $\phi_{13}$ due to the smaller attenuation of high frequencies. Because of this, the correlation between successive intervals is expected to be weak. This is con-
firmed by McFadden's results obtained for this case, and based on the assumption of total independence between intervals which were reported to be in good agreement with the 'NK1' based on the quasi independence assumption. For the general spectrum $\phi_1 n$. The correlation between successive intervals is expected to get stronger as $n$ increases since this results in more attenuation of high frequencies which causes less rapid zero crossings and leads to a reduction in the overlap between the distributions $P_N(\tau)$. This reduction can be deduced from the fact that agreement between $Q(\tau)$ and $P_0(\tau)$ extends over a wider range for $\phi_{15}$ than both $\phi_{13}$ and $\phi_{12}$.

For Group (ii) spectra as $n$ decreases, we notice that the agreement between $Q(\tau)$ and $P_0(\tau)$ extends over a range which progressively narrows. Hence the distributions $P_N(\tau)$ starts to rise significantly at lower values of $\tau$ than those obtained in the ideal low pass spectrum. This results in increasing the overlap between these distributions and leads to weaker correlation between successive intervals. The distributions overlap is however, less considerable than in Group (i) spectra, since in Group (ii) spectra high frequencies are considerably more attenuated. Hence for low order Group (ii) spectra McFadden's assumption of quasi-independence may be used, but it is not generally as valid as in Group (i) spectra.

The above results may be summarised as follows. In general, correlation between successive intervals appears to be strongly related to the frequency cut-off characteristics of the power spectrum. Gradual frequency cut-off involves relatively less attenuation of high frequencies. This results in more rapid zero crossings and also permits
the occurrence of many zeros over a relatively short time. This leads to relatively weak correlation between successive interval lengths. On the other hand, sharp frequency cut-off tends to set a maximum limit on the number of zero crossings likely to occur over a given interval. This inevitably leads to strong correlation between successive zero crossing intervals.

6.4.3 The effect of spectrum variations on the MK2 Approximation

The theory presented in Chapter 4, which dealt with the representation of random functions in terms of vector Markov processes predicted that approximations based on the second order integral equation is expected to be in good agreement with the true distribution $P_0(\mathcal{V})$ if the order of the power spectrum is not much higher than 2. Moreover, this agreement is expected to deteriorate as the order of the spectrum increases. These predictions are confirmed by the results obtained for the various spectra constituting groups (i), (ii) and (iii). In particular, we notice that the agreement is almost perfect for $\Phi_1$ spectra whereas in the ideal low-pass filter case, appreciable differences between MK2 and the experimental curve exist. In this section, an attempt is made to explain these differences making use of some of the results developed in Section 6.4.2, regarding the correlation between successive intervals.

Referring to Fig. (16 & 23) we note that in addition to the curves $Q(\mathcal{V}), MK2$ and EXP the corresponding curve $MK1$ are plotted. These curves were taken from McFadden's paper (9). The graphs show that the two curves $MK1 & MK2$ are close to each other in both cases, with $MK2$ being, as expected, closer to the experimental curve as it represents a higher approximation. The close agreement between $MK1$ and $MK2$ suggests that in the case
of \( \Phi_{12} \), McFadden's assumption of quasi-independence is very nearly true, since 'MK1', based on this assumption is very close to \( P_0(\tau) \) as obtained experimentally. By including the actual value of the slope at the crossing points, which is what 'MK2' involves, can only lead to a slight improvement. On the other hand, in the ideal low pass spectrum case, the correlation between successive intervals is so strong that 'MK1' based on McFadden's assumption fails to provide a close approximation to \( P_0(\tau) \). Including the actual slope value at the crossing points here, is still so far from accounting for the correlation between successive intervals, that 'MK2' can only result in a marginal improvement. In order to achieve that, a much higher Markovian model is required.

To explain the differences between 'MK2' and 'EXP' in the ideal low pass spectrum, we shall use McFadden's first order integral equation since its solution 'MK1' is close to 'MK2'. This equation is written below:

\[
P_0(\tau) = Q(\tau) - \int_0^\tau P_0(y) P_{\text{ODD}}(\tau-y) \, dy \quad \ldots (6.4.4)
\]

Since \( P_{\text{ODD}}(\tau) \) is nearly zero over the range \((0-\pi)\), and relatively small over the range \((\pi-2\pi)\) as can be seen from Fig(44), the convolution integral over the range \((0-2\pi)\) is negligible compared to \( Q(\tau) \), hence \( P_0(\tau) \) and \( Q(\tau) \) are very close over this range. In order to make equation (6.4.4) exact, \( P_{\text{ODD}}(\tau) \) must be replaced by the probability \( X(-/+,0,-) \) which is the conditional probability that a downward crossing occurs at \( \tau, \tau+d \tau \), given that an upward crossing has occurred at the origin, and that the first zero crossing after this has occurred at \( y \). Ignoring the history prior to \( y \) as (6.4.4) does, leads to erroneous estimates.
of the probability of a downward crossing at \( \zeta \), \( \zeta + d\zeta \).

Thus if \( \zeta \) lies in the range approximately given by \((2\pi - 3\pi)\), then the most likely number of zero crossings over this interval is three. However, the minimum number required in order that an upward zero at the origin is followed by two downward ones at \( y \) and \( \zeta \) is four, hence the probability of a downward crossing at \( \zeta \), based on the knowledge of the history prior to \( y \) is relatively small. On the other hand if \( y \) is small so that it is quite probable that \((\zeta - y)\) contains three crossings, then by ignoring the history prior to \( y \), the probability of a downward crossing at \( \zeta \) will be estimated at a higher value than that based on the knowledge of the prior history to \( y \). This means that the convolution integral would be estimated higher than the value which would have been obtained if \( P_{\text{odd}}(\zeta - y) \) was replaced by the true density \( X(-/+, 0, -) \). This causes the approximation 'MK1' to be lower than the experimental curve over the range \((2\pi - 3\pi)\).

The converse situation arises if \( \zeta \) is approximately in the range \((3\pi - 4\pi)\). There, the convolution integral has a lower value than the exact value obtained by replacing \( P_{\text{odd}}(\zeta - y) \) by \( X(-/+, 0, -) \). Consequently 'MK1' is higher than the experimental curve. In general 'MK1' tends to be higher than 'EXP' in the range over which \( Q(\zeta) \) is higher than \( P_{\text{odd}}(\zeta) \) and lower in the range over which \( P_{\text{odd}}(\zeta) \) is higher than \( Q(\zeta) \). As \( \zeta \) increases, it is possible for 'MK1' in the range over which it is lower than 'EXP' to dip negative, giving rise to negative probabilities. The effects of accounting for the actual value of the slope at the crossing points is not sufficient to describe and thus substitute for the knowledge of the history prior to \( y \). In particular, the
knowledge of the slope at the crossing point \( y \) is not sufficient to replace the knowledge that an upward zero crossing has occurred at the origin, and the influence this has on the event of a downward crossing at \( \tau \). Hence it is not surprising that 'MK2' could not improve much on the approximation 'NK1'.

An approximation which is an improvement on both 'MK1' and 'MK2' is provided by the solution of the following integral equation which takes account of the upward zero crossing at the origin, as well as the downward one at \( y \), when estimating the probability of a downward crossing at \( \tau \). This is given by:

\[
P_0(\tau) = Q(\tau) - \int_0^{\tau} P_0(y) P(\tau/y,0) \, dy \quad (6.4.5)
\]

where \( P(\tau/y,0) \) is the conditional probability density that a downward crossing occurs at \( \tau \), \( \tau + \Delta \tau \), given that an upward zero crossing and a downward crossing have occurred at \( t=0 \) and \( y \) respectively. This equation is discussed in Section (6.4.6).

Accounting for the slope value at the zero crossing points is expected however to yield 'MK2' approximations which are appreciably closer than 'MK1' to the distribution \( P_0(\tau) \), for spectra which fall between \( \phi_{13} \) and the ideal low pass spectrum. Such spectra are provided by group (ii) spectra of low order. In these cases, the assumption of quasi independence is less valid than in \( \phi_{13}(\omega) \), consequently 'MK1' is not expected to be as close to \( P_0(\tau) \) as that of \( \phi_{13}(\omega) \). On the other hand, the low order of these spectra makes it possible to obtain reasonably close approximations for \( P_0(\tau) \) using the second order Markovian model.
6.4.4 Effect of Spectrum Variation on the Multivariate Approximation.

The agreement of the approximations 'FON' with the true distribution \( P_0(t) \) depends to a large extent on the closeness of the multivariate integrals \( V_3, V_4 \) and \( V_5 \) to the probability \( V(\zeta) \) that the function \( f(t) \) is positive throughout the interval \( \zeta \). For small values of \( \zeta \), the agreement is very close since for \( f(t) \) to be positive over three, four or five sample points distributed over the interval \( \zeta \), almost certainly requires that \( f(t) \) is positive throughout the interval \( \zeta \). This explains why \( P_03 \), \( P_04 \) and \( P_05 \) agree closely with one another over the range of small \( \zeta \).

As \( \zeta \) increases, \( V_3, V_4 \) and \( V_5 \) become increasingly larger than \( V(\zeta) \). This is because \( V_N \) contains in addition to \( V(\zeta) \), the probability that \( f(t) \) is negative in between the sample points, which increases as \( \zeta \) increases. Therefore \( V_5 \) represents the closest approximation to \( V(\zeta) \) since having more sample points tends to reduce the probability that \( f(t) \) is negative in between.

Because of the close relationship that exists between the Schlaffli Function \( S_N \) and the auto-correlation function, the approximations 'FON' are largely influenced by the power spectrum. Hence it was not surprising that different types of results were obtained corresponding to various spectra.

For group (i) spectra, the multivariate integral \( V_N \) starts from the value \( \frac{1}{2} \) at \( \zeta=0 \) and decreases monotonically to its steady state value \( (\frac{1}{2})^N \) as \( \zeta \) approaches \( \infty \). For group (ii) and group (iii) spectra, \( V_N \) starts from the same initial value. However, because of the oscillations
displayed by the auto-correlation functions over intermediate and large values of $\gamma$, the integral $V_N$ oscillates about the value $\langle \gamma \rangle^N$ over this range. The monotonic decay of $V_N$ in group (i) and the oscillations in groups (ii) and (iii) tend to produce approximations which are positive throughout the entire range in the first case and approximations which are partly negative in the second case.

The probability that $f(t)$ is positive at the sample points but negative in between can be reduced by increasing the memory of the function $f(t)$. This corresponds to increasing the order of the power spectrum in both group (i) and group (ii). In group (i) increasing the order leads to less rapid zero crossings. Consequently the probability that $f(t)$ goes negative between the sample points which involves at least two zero crossings over any interval between successive sample points, progressively diminishes. This leads to $V_N$ becoming closer to $\sqrt{\langle \gamma \rangle}$ which in turn leads to closer agreement between the approximations 'PON' and the probability distribution $P_0(\gamma)$. The graphs in Figs. (33) & (34) appear to confirm this point, since the agreement between 'PON' and 'EXP' seem to be closer in the case $\phi_{15}$ than that in $\phi_{13}$.

In group (ii) spectra, increasing the order of the spectrum tends to reduce the rate of zero crossings by a marginal amount. This is because the cut-off frequency remains practically constant. Here, although increasing the order of the spectrum tends to increase the effective memory of the function $f(t)$, the probability that $f(t)$ is negative between the sample points is only reduced slightly. Consequently the agreement between 'PON' and the experimental results improves only marginally. In any case
this agreement is only obtained over a limited range since the 'FOH' curves invariably dip negative and hence result in negative probabilities.

From the above discussion it seems that the multivariate method is very suitable to use for group (i) spectra with relatively high order. This is fortunate in view of the fact that 'HK2' becomes increasingly less close to $P_0(\tau)$ as the order increases. Hence, for group (i) spectra, these two methods can be used to provide close approximations.

6.4.5 The behaviour of $P_0(\tau)$ at large interval lengths.

In Chapter Two, a brief discussion of the expected behaviour of $P_0(\tau)$ at large interval lengths is presented. There it was suggested that this behaviour could be exponential. The validity of this suggestion can now be assessed using the results obtained for the various power spectra.

Figs. (46) & (47) show several experimental distributions plotted on log-linear graph paper over the range $(\tau = 8 \to 20)$. Fig. (46) shows the curves corresponding to the spectra $\phi_{12}$, $\phi_{13}$ and $\phi_{15}$ of group (i) whereas Fig. (47) shows the curves corresponding to $\phi_{23}$, $\phi_{24}$ and $\phi_{27}$ spectra of group (ii).

The experimental points for the three spectra of group (i) seem to fit closely a straight line relationship with very slight scatter. For group (ii) spectra, the experimental points for $\phi_{23}$ and $\phi_{24}$ seem to be fairly close to a straight line, but the scatter of the points about the lines are somewhat larger than in group (i). For $\phi_{27}$ spectrum this scatter is appreciable with the points oscillating about the straight line.

Since a linear relationship on a log-scale corresponds
Fig(46) - Graphs of $P_\alpha (\tau)$ for interval length $\tau$. Examples of $\phi_\alpha (\omega)$ spectra.
Fig. (47) - Graphs of $P_0(\nu)$ for examples of $\phi_{2n}(\omega)$ spectra.
to an exponential one over a linear scale, the above results
show that in general, low order spectra lead to interval
distribution \( P_0(\tau) \) which varies almost exponentially at
large values of \( \tau \). Also for group (i) spectra, the exponential
assumption seems to be more valid and starts earlier than
in the corresponding group (ii) spectra. Further, for a
relatively high order group (ii) spectra, the exponential
assumption does not seem to be valid over the range (8-20).
It is not clear whether this assumption may be valid over
extremely large values of \( \tau \), but if it is it would have
very little significance as far as approximations based on
this assumption are concerned.

These results are not surprising in view of the fact
that the exponential assumption is valid in the cases
where the influence of the event of the upward zero cross-
ing at \( t_0 \), on the downward crossing at \( t_0+\tau, t_0+\tau+\delta \tau \),
progressively diminishes as \( \tau \) increases. Such cases are
obtained if \( f(t) \) has a group (i) type spectrum, since this
corresponds to a continuously decaying auto-correlation
function. In contrast the auto-correlation functions of
corresponding group (ii) spectra oscillate about the zero
value with continuously decaying amplitude as \( \tau \) increases.
Consequently, the distribution \( P_0(\tau) \) tends to oscillate
about a curve which decays exponentially at large values
of \( \tau \).

For low order spectra of both groups the value of \( \tau \) at
which the exponential or near exponential behaviour starts
progressively increases as \( n \) is increased. This is attr-
bributed to the increase of the effective memory time of
\( f(t) \) resulting from increasing the order of its power
spectrum.
6.4.6 Comparisons of Various Approximations.

This Chapter is concluded by presenting a brief comparison between the various approximations proposed for the distribution $P_0(\tau)$. The comparison is done in the form of the following points:

i) In general, the approximations 'MK2' based on the second order Markovian equation is very close to the true distribution $P_0(\tau)$, if the function $f(t)$ has a rational power spectrum of relatively low order. The approximation progressively deteriorates as the order of the spectrum increases which is in accordance with the theory on which this approximation is based.

ii) The multivariate method yields approximations 'PON' which gets closer to $P_0(\tau)$ as the memory of the function $f(t)$ is increased. These approximations are therefore expected to yield very close results for group (i) spectra of relatively high order. For group (ii) and group (iii) spectra, the negative probabilities which invariably occur using this method, cause these approximations to be of limited use.

iii) The approximations explained in Chapter Two (Section 2) which aim at constructing an overall solution whose initial part is provided by $Q(\tau)$ and whose second part is made up of an exponential curve, has limited use generally. For group (i) spectra in which the exponential assumption is approximately valid, the agreement between $Q(\tau)$ and $P_0(\tau)$ extends over a relatively small range. Hence over the intermediate range which forms the range of values of $\tau$ between the point at which $Q(\tau)$ ceases to be adequately representative of $P_0(\tau)$, and the point at which the exponential behaviour can be assumed to start, neither curve is cap-
able of yielding reliable results for $P_0(\tau)$. On the other hand for group (ii) and group (iii) spectra, $Q(\tau)$ represents closely the distribution $P_0(\tau)$ over a relatively large range, but unfortunately the exponential behaviour cannot be assumed to start until $\tau$ reaches a very large value.

This approximation is useful however for low order group (i) spectra, since $Q(\tau)$ agrees with $P_0(\tau)$ over a wider range than that in group (i) and the exponential behaviour can be assumed to start at a smaller value of $\tau$ than that in group (iii) or high order group (ii) spectra.

iv) The relatively wide range over which $Q(\tau)$ agrees closely with $P_0(\tau)$ if $f(t)$ has a high order group (ii) or group (iii) spectrum suggests that the sum of the first few terms of the In and Exclusion equation (2.4.2) may well provide a very close approximation to $P_0(\tau)$ over a relatively wide range. The first of the approximations denoted in Chapter Two by $P_0^{(1)}$ which is the same as $Q(\tau)$, is close to $P_0(\tau)$ in the range over which $P_2(\tau)$ is negligible. The second approximation $P_0^{(2)}(\tau)$ is close to $P_0(\tau)$ in the range over which $P_4(\tau)$ is negligible. From the discussion of section (6.4.2) these ranges for the ideal low pass spectrum are approximately $2\pi$ and $4\pi$ respectively. Hence the approximation $P_0^{(2)}(\tau)$ is expected to be very close to $P_0(\tau)$ in the range $(0-4\pi)$. Longuet-Higgins evaluated $P_0(\tau)$ for this case and his results showed that over the range $(\tau=0-14)$, this approximation agrees closely with the experimental distribution. Beyond $\tau=14$, this approximation goes negative, and from then on it varies approximately as $(-\tau)$. Longuet-Higgins also evaluated $P_0(\tau)$ for the band pass spectrum and stated that in this case, this approximation is even closer to $P_0(\tau)$ than that in the ideal low pass case.
v) Finally, the integral equation (6.4.5) which is given by:

\[ P_0(\tau) = Q(\tau) - \int_0^\tau P_0(y) \cdot P(\tau/y,0) \, dy \quad \ldots (6.4.5) \]

is discussed for the ideal low pass spectrum.

In formulating this equation, the upward zero crossing at the origin and its influence on the event at \( \tau \) is accounted for. It is true that the probability \( P(\tau/y,0) \) is not exactly the same as the probability \( X(-\tau,0,\tau) \). However they are very close if \( y \) is less than \( 2\pi \), since it is unlikely that more zero crossings can occur over the interval \( (y < 2\pi) \) in addition to those already at 0 and \( y \). Further, these two probabilities are reasonably close if \( y \) lies in the range \((2\pi - 3\pi)\). Therefore it is expected that this integral equation would yield a reasonably close approximation to \( P_0(\tau) \) at least in the range \((0-3\pi)\).

Next this equation is compared to the sum of the first two terms of the In and Exclusion equation \( P_0(\tau) \) which is written in the form:

\[ P_0^{(2)}(\tau) = Q(\tau) - \int_0^\tau Q_2(y,\tau) \, dy \quad \ldots (6.4.6) \]

where \( Q_2(y,\tau) \) is the probability density that downward zero crossings occur at \( y, y+dy \) and \( \tau, \tau + d\tau \) given that an upward crossing has occurred at 0.

In terms of \( Q_2(y,\tau) \) (6.4.5) may be written as:

\[ P_0(\tau) = Q(\tau) - \int_0^\tau \frac{P_0(y)}{Q(y)} \cdot Q_2(y,\tau) \, dy \quad \ldots (6.4.7) \]

Therefore the solution of both equations coincide over the range in which \( Q(\tau) \) and \( P_0(\tau) \) are very close together.
As \( \tau \) increases, the solution of the integral equation (6.4.5) lies above the approximation \( P_0(\tau) \), since \( P_0(\tau) \) is smaller than \( Q(y) \). Since \( P_0(\tau) \) is a lower bound on the distribution \( P_0(\tau) \) it follows that the solution of the integral equation (6.4.5) may well be a closer approximation than \( P_0(\tau) \). As \( \tau \) continues to increase, \( P_0(\tau) \) crosses the zero value and from then on continues to increase in a negative direction. In contrast, the solution of the integral equation tends to zero as \( \tau \) gets very large, but as it does that it may oscillate about the zero value. Hence the solution of the integral equation (6.4.5) is a better approximation than \( P_0(\tau) \) over this range. Longuet-Higgins (2) obtained an expression for the probability distribution \( Q_2(y, \tau) \), which is equivalent to \( W^{(+)} \) in his notations and used it to evaluate \( P_0(\tau) \). Using this expression the integral equation (6.4.7) which is the same as (6.4.5) may be solved, but the computations involved are expected to be more complex than those involved in evaluating \( P_0(\tau) \).
CHAPTER SEVEN.

Conclusions.

A comprehensive and detailed study of the statistical distribution of zero crossings of random processes was carried out. The study revealed the theoretical complexity of this problem and the computational difficulties involved in obtaining a satisfactory solution.

The infinite 'In and Exclusion' series (2.5.11) provides exact solutions for the distribution $P_0(T)$ and more generally $P_n(T)$. The convergence of these series is however slow, making it necessary to evaluate several terms in order to obtain satisfactory approximations which extend over a wide range of interval lengths. Nevertheless with present-day computing techniques it does not seem practical to evaluate more than the first two or three terms of these series.

Another method of approach relates the distribution $P_0(T)$ to the multivariate integral $V_n$. This method was discussed in detail, for Gaussian processes for which $V_n$ assumes its easiest form. Even then it is not possible at this stage, to obtain a closed form expression for this integral which could be used to obtain an analytical solution for the distribution $P_0(T)$. It was however possible using the $n^{th}$ dimensional analog of Mehl's Formula (see Chapter 3, section 3.4) to obtain an expression in the form of a multiple series for the multivariate normal integral. This series which again converges slowly was found to be of little use in obtaining a satisfactory solution for $P_0(T)$.

Through the use of a special reduction formula, approximations to the distribution $P_0(T)$ were found using
the multivariate integral approach. These were found useful for some examples of random processes.

Perhaps one of the most significant methods is the one introduced in Chapter Four. In this method, a model is constructed in which the random function $f(t)$, assumed to have a rational power spectrum, is represented as one component of a vector Markov Process. Using this model it is possible to derive a multi-dimensional integral equation whose solution is related to the probability distribution $P_0(\gamma)$. McFadden's integral equations, based on the assumption of independence and quasi-independence were identified as special cases of this equation, in which the dimension of the associated Markov process is unity. For the class of random processes with rational power spectra, this method, in theory at least, is capable of yielding exact solutions to the distributions $P_n(\gamma)$. Further to obtain such solutions, we need only deal with finite order probability distributions as well as finite order integrals. This is in contrast to the 'In and Exclusion' series which involves integrals and distributions of infinite multiplicity.

Difficulties, both mathematical and computational prevent us at this stage, from obtaining the solution of the general order integral equation. Instead, we can obtain approximations for the distribution $P_0(\gamma)$ by considering equations of low order. The second order equation was considered and its solution was numerically evaluated for a number of random functions having various power spectra. For spectra with not too sharp frequency cut-off characteristics, the agreement between the approximations based on the second order equation and the experimental results was found to be generally close. For sharp cut-off
spectra, the approximations resulted in negative probabilities at large values of $\zeta$.

Next the effects of power spectrum variation on the distribution $P_n(\zeta)$ and consequently on the correlation between successive intervals was investigated. The relatively small overlap that exists between $P_n(\zeta)$ for spectra with sharp frequency cut-off causes the variance of the number of zero crossings over a given period of time, to be small. On the other hand, the considerable overlap between $P_n(\zeta)$ for spectra with gradual frequency cut-off causes the variance to be relatively large. As explained in Chapter Six, this results in strong correlation between successive intervals in the first case and weak correlation in the second.

The assumed exponential behaviour of $P_0(\zeta)$ at large interval length is justified if the auto-correlation function $\rho(\zeta)$ decays monotonically towards its zero value as $\zeta$ increases. The region at which this behaviour may be assumed is given by the value of $\zeta$ at which $\rho(\zeta)$ is negligibly small. On the other hand, if $\rho(\zeta)$ is of the oscillatory type, e.g. that corresponding to a low-pass spectrum, then strictly speaking the exponential assumption is not valid. In this case the assumption maybe justified at very large values of $\zeta$.

It is felt that further progress with this problem lies in two directions. Firstly in the approach based on the multivariate integral and the associated Schaffli function $S_n$. Secondly in the construction of a model which adequately represents the random process, yet at the same time is sufficiently simple to permit mathematical manipulation. Such a model must account for the correlation of successive intervals in order to yield satisfactory solutions. It is
It is hoped that the discussion on this aspect presented in Chapter Six, will be of use in achieving this model.

It is further hoped that the material included in this study will be useful in any further research related to or conducted on this problem.
APPENDIX ONE.

Derivation of the In and Exclusion Equation.

The In and Exclusion Equations (2.4.1) & (2.4.2) are derived below. The probability distributions \( U(\tau), U_2(\tau, \tau) \), \( U_3(s, \tau, \tau) \) ... etc and \( Q(\tau), Q_2(\tau, \tau), Q_3(s, \tau, \tau) \) ... etc have the same definition as those given in Chapter Two, Section (2.4). In addition, we define the following conditional probability distributions, given that an upward zero crossing has occurred at \( t=0 \).

i) \( U_{20}(r, \tau) d\tau d\tau \); the joint conditional probability that \( f(t) \) has its next zero crossing at \( r, r+dr \) and another zero crossing at \( \tau, \tau+d\tau \).

ii) \( U_{30}(s, \tau, \tau) d\tau d\tau d\tau \); the joint conditional probability that \( f(t) \) has its next zero crossing at \( s+ds \) and two more crossings at \( r, r+dr \) and \( \tau, \tau+d\tau \).

We can also define \( Q_{20}(r, \tau), Q_{30}(s, \tau, \tau) \) ... etc similarly to \( U_{20} \) and \( U_{30} \) etc, except that the zero crossings at \( \tau, r \) ... etc are limited to downward crossings only.

Next, Equation (2.4.1) is derived.

From the definitions of \( U(\tau) \) and \( U_{20}(r, \tau) \), the following equation can be immediately written

\[
U(\tau) = P_0(\tau) + \int_0^\tau U_{20}(r, \tau) \, dr \quad \ldots (A.1.1)
\]

Equation (A.1.1) expresses the fact that the event of a zero crossing at \( \tau, \tau+d\tau \), given an upward zero crossing at \( t=0 \), can be realised in two mutually exclusive ways. It is either the first crossing after the initial one; or alternatively any other zero crossing but the first. The first probability occurs with probability \( P_0(\tau) \). The second consists of all the realisations which cross the zero
level for the first time after \( t=0 \), at \( r, r+\text{d}r \), where \( r \) lies in the range \((0-\tau)\). This has the probability given by the second term in (A.1.1).

The same argument, if used to relate the probability distributions \( U_2(r,\tau) \) and \( U_{30}(s,r,\tau) \) yields the following equation:

\[
U_2(r,\tau) = U_{20}(r,\tau) + \int_0^r U_{30}(s,r,\tau) \, ds \quad \cdots (A.1.2)
\]

where the sum of the probabilities on the right-hand-side exhausts that represented by the probability on the left-hand-side. In general:

\[
U_n(t_1,t_2,\ldots,t_n) = U_{n0}(t_1,t_2,\ldots,t_n) + \int_0^{t_1} U_{n+1,0}(t,t_1,t_2,\ldots,t_n) \, dt \quad \cdots (A.1.3)
\]

Equation (A.1.3) represents a recurrent relationship which relates the \( n \)th order conditional probability density \( U_{n0} \) to the \((n+1)\)st density \( U_{n+1,0} \). Using (A.1.2) we can eliminate the density \( U_{20}(r,\tau) \) in equation (A.1.1), which becomes after rearranging:

\[
P_0(\tau) = U(\tau) - \int_0^{\tau} U_2(r,\tau) \, dr + \int_0^{\tau} \int_0^r U_{30}(s,r,\tau) \, ds \, dr \quad \cdots (A.1.4)
\]

Now, by successfully eliminating the density \( U_{n0} \), using the recurrent relation (A.1.3) which involves the higher density \( U_{n+1,0} \), we obtain for \( P_0(\tau) \) the following infinite series as \( n \) approaches \( \infty \).

\[
P_0(\tau) = U(\tau) - \int_0^{\tau} U_2(r,\tau) \, dr + \int_0^{\tau} \int_0^r U_{30}(s,r,\tau) \, ds \, dr - \cdots
\]
\[ \int_0^r \int_0^s \int_0^r U_4(t,s,r,s) \,dt \,ds \,dr + \ldots \] 

...(A.1.5)

which is the equation termed by S. Rice as the 'In and Exclusion' equation.

If it is desired that all limits of integration should run between 0 and \( \tau \), then in order to maintain the value of each term, it must be multiplied by a factor \( \frac{1}{N!} \) where \( N \) represents the corresponding number of zero crossings between the initial one at \( t=0 \) and the final one at \( t=\tau \). This factor represents the number of possible ways in which these zeros may be interchanged.

Finally, the above method used to derive the In and Exclusion equation involving the probability distributions \( U(\tau), U_2(r, \tau) \ldots \) etc can be exactly used to define the corresponding equation involving the distributions \( Q(\tau), Q_2(r, \tau) \ldots \) etc.
Derivation of the Probability $V(\tau)$ for a first order Markov Process.

In this appendix, the probability that a first order Markov process $x(t)$ remains positive over an interval of length $\tau$ denoted by $V(\tau)$, is derived from the solution of an integral equation. This equation is that numbered (4.3.9) in Chapter 4, Section (4.3.) and written again below:

\[
\Phi(h, t/y_0) = \phi(h, t/y_0) - \int_0^t q(t_1, h/y_0) \phi(h, t-t_1/h)dt_1 \ldots (A.2.1)
\]

where $\Phi(h, t/y_0)$ is the conditional probability that given $X(0)=y_0$ which is greater than $h$, the function $x(t)$ has not crossed the level $h$ at all in the interval $(0, t)$.

$q(t_1, h/y_0)dt$ is the probability that $x(t)$ first crosses the level $h$ in $t_1, t_1+dt$, given that $x(0)=y_0$ and finally.

$\phi(h, t/y_0)$ is the conditional probability that $x(t)$ is greater than the value $h$ at time $t$, given the same initial condition (It is assumed that $h$ is positive throughout). The distribution $\Phi(h, t/y_0)$ is therefore given by:

\[
\Phi(h, t/y_0) = \int_{-\infty}^{\infty} P(y, t/y_0) dy \ldots (A.2.2)
\]

where $P(y, t/y_0)dy$ is the conditional probability that $x(t)$ lies between $y$ and $y+dy$ at $t$, given $x(0)=y_0$.

Equation (A.2.1) simply classifies all the possible realisations of the function $x(t)$ which start from $y_0$ at $t=0$, and have values higher than $h$ at time $t$ according to the time $t_1 > 0$ at which they cross the value $h$ for the first time.
Next, this equation is solved when h is equal to the average value of x(t) which can be assumed zero without any loss of generality. The solution of this equation yields the probability f(0, t/y₀) which is related to the probability \( \mathcal{V}(\tau) \). Equation (A.2.1) with h=0 is given by

\[
f(0, t/y₀) = \phi(0, t/y₀) - \frac{1}{2} \int_0^t q(t₁, 0/y₀) dt₁ \quad \text{(A.2.3)}
\]

since

\[
\phi(0, t-t₁/y₀) = \int_0^\infty P(y, t-t₁/y₀) dy = \frac{1}{2} \quad \text{(A.2.4)}
\]

By the law of conditional probability, the probability \( \mathcal{V}(\tau) \) is related to the probability \( f(0, t/y₀) \) by

\[
\mathcal{V}(\tau) = \int_0^\infty P(y₀) f(0, t/y₀) dy₀ \quad \text{(A.2.5)}
\]

where \( P(y₀) \) is the probability density of the initial value \( y₀ \). Hence:

\[
\mathcal{V}(\tau) = \int_0^\infty P(y₀) \int_0^\infty P(y, t/y₀) dy dy₀ - \frac{1}{2} \int_0^t \int_0^\infty q(t₁, 0/y₀) P(y₀) dy₀ dt₁ \quad \text{(A.2.6)}
\]

which is obtained by substituting the R.H.S. of (A.2.3) for \( f(0, t/y₀) \) in (A.2.5) and then using (A.2.2).

Now the first term in the R.H.S. of (A.2.6) represents the probability that \( x(0) \) and \( x(t) \) are simultaneously positive. For Gaussian process this is equal to the value of the bivariate normal integral \( V₂ \) which is given by:

\[
V₂ = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho(\tau) \quad \text{(A.2.7)}
\]

where \( \rho(\tau) \) is the normalised auto-correlation function. Also, since \( x(t) \) must cross the zero value for the first time at some value of \( t \), greater than 0, it follows that:
\[
\int_0^\infty q(t_1,0/y_0) \, dt_1 = 1 - \int_t^\infty q(t_1,0/y_0) \, dt_1 \quad \ldots (A.2.8)
\]

But
\[
\int_t^\infty q(t_1,0/y_0) \, dt_1 = f(0,t/y_0) \quad \ldots (A.2.9)
\]

Hence the second term on the R.H.S. of (A.2.6) is given by:

\[
\frac{1}{2} \int_0^\infty p(y_0) \left(1-f(0,t/y_0)\right) \, dy_0 = \frac{1}{2} - \frac{1}{2} V(\zeta) \quad \ldots (A.2.10)
\]

Using (A.2.5)

Therefore using (A.2.7) and (A.2.10) we obtain for (A.2.6)

\[
V(\zeta) = \frac{1}{2\pi} \sin^{-1} \rho(\zeta) + \frac{1}{2} V(\zeta) \quad \ldots (A.2.11)
\]

Since, for a first order Markov process, \( \rho(\zeta) \) is given by \( e^{-|\zeta|} \), it follows that \( V(\zeta) \) for this process is given by:

\[
V(\zeta) = \frac{1}{2\pi} \sin^{-1} (e^{-|\zeta|}) \quad \ldots (A.2.12)
\]
APPENDIX THREE.

Numerical Integration and the solution of the second order Markovian Integral Equation.

In this appendix, the formulae used in evaluating the integrals (4.9.2) and (4.9.3), see Chapter 4, Section (4.9), are first derived. Later, the numerical method which was used to compute the solution of the integral equation (4.9.1) is discussed in detail.

A.3.1 Numerical Integration.

It is often desired that the integral \( \int_a^b f(x) \, dx \) can be approximated by the finite sum \( \sum_{k=0}^{n} \alpha_k f(x_k) \), where \( f(x_k) \), \( k=0, 1, 2, \ldots, n \), represent a series of ordinates or sample values, taken at the points \( x_0, x_1, \ldots, x_n \) and \( \alpha_k \) are the corresponding weighting coefficients. This may be written as:

\[
\int_a^b f(x) \, dx = \sum_{k=0}^{n} \alpha_k f(x_k) + E_n(f) \quad \text{...(A.3.1)}
\]

where \( E_n(f) \) is an error term which is a function of \( n \) and the function \( f(x) \).

A formula of this type is useful in situations where \( f(x) \) is only available as a series of sample values, e.g. in sampled data systems, or when \( f(x) \) is continuous but its integral with respect to \( x \) cannot be evaluated by analytical methods.

If \( f(x) \) is available at all values of \( x \) between \( a \) and \( b \), then a suitable formula can be derived by choosing the sample points \( x_0, x_1, \ldots, x_n \) as fixed numbers in the interval \( (b-a) \). The coefficients \( (\alpha_0, \ldots, \alpha_n) \) can then be selected...
such that the error $E_n(f)$ is zero for $f(x)=1, x, x^2, \ldots, x^n$. This results in $(n+1)$ simultaneous equations in the $(n+1)$ unknowns $\alpha_k$. Solving these equations yields the coefficients $\alpha_k$. This method is equivalent to approximating the function $f(x)$ by a polynomial $R_n(x)$ of degree $n$, chosen such that

$$R_n(x_k) = f(x_k), \quad 0 \leq k \leq n \quad \ldots (A.3.2)$$

It is possible to choose the sample points $x_k$ (hereafter referred to as nodes) so that the coefficients $\alpha_k$ are all equal. In general this results in a non-uniformly distributed nodes.

Integration formulae of the form (A.3.1) are normally referred to as interpolatory quadrature formulae. The accuracy of these formulae is expressed in what is termed 'degree of precision', which is defined as the maximum number $m$ so that the error $E_n(f=x^k)$ is zero for $k \leq m$, whereas $E_n(f=x^{m+1})$ is not equal to zero. Hence if a formula has a degree of precision $m$, then all polynomials of degree at most $m$ are integrated exactly by this formula. It follows that the formula (A.3.1) has a degree of precision of at least $n$.

The degree of precision of an interpolatory formula can be further increased if we allow the choice of both $n$ coefficients $\alpha_k$ and $n$ nodes $x_k$ in order that a formula with a maximum degree of precision is obtained. For this purpose the following theorem is given:

The quadrature formula

$$\int_a^b f(x) \, dx = \sum_{k=0}^n \alpha_k f(x_k) + E_n(f) \quad \ldots (A.3.1)$$

can have the maximum degree of precision $(2n-1)$. This is attained if, and only if, the $n$ nodes $x_j$ are the zeros of $R_n(x)$; the $n^{th}$ orthogonal polynomial over the interval
(a,b) and the formula is interpolatory.

The formulae determined by this theorem are called **Gaussian Quadrature Formulae**. It can be shown that the nodes \( x_k \) are all interior to the interval \((a,b)\) and the coefficients are given by

\[
\alpha_k = \frac{1}{R_n'(x_k)} \int_a^b \frac{R_n(x)}{x-x_k} \, dx \quad k=1,2, \ldots, n \quad \text{...(A.3.3)}
\]

Further, all \( \alpha_k \) are positive.

Gaussian quadrature formulae are generally associated with weighting functions \( \omega(x) \). In the formula \((A.3.1)\), \( \omega(x)=1 \). The weighted formulae take the form:

\[
\int_a^b \omega(x) f(x) \, dx = \sum_{k=0}^n \alpha_k \omega(x_k) + E_n(f) \quad \text{...(A.3.4)}
\]

This formula would have the maximum degree of precision if the nodes \( x_k \) are the roots of the polynomial \( R_n(x) \) which is assumed orthogonal over the interval \((a,b)\) with respect to the weighting function \( \omega(x) \). This means that \( R_n(x) \) must satisfy the relation:

\[
\int_a^b \omega(x) R_n(x) R_m(x) \, dx = \delta_{mn} \quad \text{...(A.3.5)}
\]

The coefficients \( \alpha_k \) are then given by:

\[
\alpha_k = \frac{1}{R_n'(x_k)} \int_a^b \frac{R_n(x)}{x-x_k} \omega(x) \, dx \quad \text{...(A.3.6)}
\]

The determination of a particular set of orthogonal polynomials \( R_n(x) \) whose roots are to form the nodes \( x_k \) for the quadrature formula, depends on the interval \((a,b)\) and the weighting function \( \omega(x) \). The following examples illustrate this:

1) \( a=-1, \ b=1; \ \omega(x)=1 \) corresponds to Legendre polynomials.
2) \( a=0, \ b=\infty; \ \omega(x)=e^{-x} \) corresponds to Laguerre polynomials.
3) \( a=-\infty, \ b=\infty; \ \omega(x)=e^{-x^2} \) corresponds to Hermite polynomials.
In fact only the interval of integration needs to be the decisive factor, since \( W(x) \) can be any function we like. This is so since

\[
\int_{a}^{b} f(x) dx = \int_{a}^{b} \frac{f(x)}{W(x)} W(x) dx = \int_{a}^{b} g(x) dx \quad \ldots (A.3.7)
\]

where \( g(x) = \frac{f(x)}{W(x)} \) \ldots (A.3.8)

We can then determine our formula with the maximum degree of precision to \((n-1)\) by dealing with \( g(x) \) instead of \( f(x) \). In this way, the weighting function is transformed from unity to \( W(x) \). This is chosen to be the weighting function that usually fits the given interval \((a,b)\).

### A.3.2 The Gauss-Laguerre Quadrature.

This refers to the integrating formula which corresponds to \( a=0, b=\infty \) and \( W(x) = e^{-x} \). The reason why this quadrature formula is singled out for discussion, is because it is needed for use in evaluating the integral:

\[
G(y,V) = \int_{\infty}^{0} dz P_{0}(z,V) (y,z,\mathcal{C}-V) \quad \ldots (4.9.2)
\]

The fact that the integration interval runs from \(-\infty\) to \(0\) instead of \(0 \rightarrow \infty\) should not make much difference, since as shown below it is possible to modify the Laguerre polynomials so that they become orthogonal over the interval \(-\infty \rightarrow 0\).

The \( n \)th degree Laguerre polynomial is defined:

\[
L_{n}(x) = (-1)^{n} e^{x} \frac{d^{n}}{dx^{n}} (xe^{-x}) \quad \ldots (A.3.9)
\]

It can be shown that the following relationships are true:

\[
L_{n}(x) = \sum_{r=0}^{n} \frac{(-1)^{r}}{r!} \left( \frac{n!}{(n-r)!} \right)^{2} x^{n-r} \quad \ldots (A.3.10)
\]

\[
L_{n}^{\prime}(x) = n(L_{n-1}(x) - L_{n-1}^{\prime}(x)) \quad \ldots (A.3.11)
\]

where \( L_{n}^{\prime} = \frac{d}{dx} L_{n} \)

and

\[
\int_{0}^{\infty} L_{m}(x) L_{n}(x) e^{-x} dx = (n!)^{2} S_{mn} \quad \ldots (A.3.12)
\]
(A.3.12) means that the Laguerre Polynomials are orthogonal with respect to the weighting function $e^{-x}$ over the interval $0 \rightarrow \infty$. Replacing $x$ by $(-x)$ in (A.3.12) we have:

$$\int_{-\infty}^{0} L_n(-x) L_m(-x) e^x dx = (n!)^2 \delta_{mn} \quad \text{(A.3.13)}$$

Hence by defining a new set of polynomials $q_n(x) = L_n(-x)$ (A.3.13) becomes:

$$\int_{-\infty}^{0} q_n(x) q_m(x) e^x dx = (n!) \delta_{mn} \quad \text{(A.3.14)}$$

and the set $q_n(x)$ is an orthogonal set over the interval $-\infty$ to 0 with respect to the weighting function $e^x$. Since

$$L_n(x) = \sum_{k=1}^{n} (x-x_k) \quad \text{(A.3.15)}$$

where $x_k$ are the roots of the polynomial $L_n(x)$, it follows that

$$q_n(x) = L_n(-x) = \sum_{k=1}^{n} (-x-x_k) \quad \text{(A.3.16)}$$

Therefore the roots of $q_n(x)$ are the negatives of those of the corresponding $L_n(x)$. The polynomial $q_n(x)$ may be normalised by dividing it by $n!$. Thus if

$$q_{n0}(x) = \frac{q_n(x)}{n!} \quad , \text{(A.3.14) becomes:}$$

$$\int_{-\infty}^{0} q_{n0}(x) q_{m0}(x) e^x dx = \delta_{mn} \quad \text{(A.3.17)}$$

and the polynomials $q_{n0}(x)$ are said to be orthonormal over the interval $(-\infty, 0)$ with respect to $e^x$. In terms of the roots of these polynomials, the quadrature formula is given by:

$$\int_{-\infty}^{0} f(x) dx = \sum_{k=1}^{n} \beta_k g(x_k) + E_n(g) \quad \text{(A.3.18)}$$

where $g(x) = \frac{f(x)}{e^x}$ and the coefficients $\beta_k$ are given by (29).

$$\beta_k = \frac{1}{n \cdot q_{n0}(x_k) q_{(n-1)0}(x_k)} \quad \text{(A.3.19)}$$

The formula (A.3.18) has the maximum degree of precision $(2n-1)$. 


A.33 Numerical Computation of the Solution of the Integral Equation.

The object of solving the integral equation (4.9.1) is to obtain the distribution \( P_0(\dot{y}, \tau) \) as a function of the slope \( \dot{y} \) at any given time. If we closely examine (4.9.1) it is clear that in order to obtain \( P_0(\dot{y}, \tau) \) it is necessary to know the value of this distribution as a function of \( \dot{y} \) at all times prior to \( \tau \). We therefore begin by replacing the continuous time scale with a discrete time scale which starts at zero and gets incremented in steps of \( \Delta \). The mesh width \( \Delta \) is chosen sufficiently small so that the probability of having two or more zero crossings with slopes of the same sign is negligible. In terms of the discrete time scale the integral equation (4.9.1) can be expressed as:

\[
P_0(\dot{y}, m \Delta) = Q(\dot{y}, m \Delta) - \sum_{k=0}^{m-1} \alpha_k G(\dot{y}, k \Delta) \quad \ldots (A.3.20)
\]

where

\[
G(\dot{y}, k \Delta) = \int_{-\infty}^{0} dz \ P_0(\dot{z}, k \Delta) \ P(\dot{y}, \dot{z}, (m-k) \Delta) \quad \ldots (A.3.21)
\]

The integral (A.3.21) can be evaluated numerically using the Gauss Laguerre quadrature discussed in the last section. For this purpose the distributions \( P_0, Q, G \& P \) need only be evaluated as functions of the slopes \( \dot{y} \) at the points determined by the roots of the polynomial \( q_{n0}(x) \) used in the quadrature formula. \( (A.3.20) \) \& \( (A.3.21) \) can now be written as:

\[
P_0(\dot{y}_j, m \Delta) = Q(\dot{y}_j, m \Delta) - \sum_{k=0}^{m-1} \alpha_k G(\dot{y}_j, k \Delta) \quad \ldots (A.3.22)
\]

and

\[
G(\dot{y}_j, k \Delta) = \sum_{l=1}^{n} \beta_1 P_0(\dot{z}_l, k \Delta) \ P(\dot{y}_j, \dot{z}_l, (m-k) \Delta) e^{\dot{z}_l} \quad \ldots (A.3.23)
\]

where \( j, l = 1, 2 \ldots n \). \( \beta_1 \) are the coefficients of the Gauss-Laguerre quadrature.

The quadrature \( \sum_{k=0}^{m-1} \alpha_k G(\dot{y}_j, k \Delta) \) integrates the function \( G \) which is given in the form of ordinates at
the points $0, \Delta \ldots (m-1)\Delta$ over the interval $(m-1)\Delta$. The coefficients $\alpha_k$ as explained earlier are functions of the number of ordinates $m$. In the course of solving the integral equation, the value of $m$ is incremented by one, every time the time scale is increased by $\Delta$. This makes it necessary to update the coefficients $\alpha_k$ as $m$ varies. As this is not convenient the following procedure is suggested. Since

$$\int_0^{(m-1)\Delta} G(\dot{y}_j, V) dV = \int_0^\Delta G(\dot{y}_j, V) dV + \int_0^{2\Delta} G(\dot{y}_j, V) dV + \int_0^{(m-2)\Delta} G(\dot{y}_j, V) dV$$

Then by evaluating each of the $(m-1)$ elementary integrals on the R.H.S. and adding them up together, we can obtain, numerically, the integral on the L.H.S. The dependence on $m$ is thus removed. Further, it is possible to use the same quadrature formula to evaluate each of the elementary integrals. The required quadrature formula can be based on $n$ points which means that a polynomial of degree $n$ is used to interpolate between $n$ ordinates. Two of these ordinates are provided by the two limits of the elementary integral whereas the remaining $(n-2)$ may be provided by ordinates immediately preceding and/or following these two. Once this is decided the quadrature formula can be worked out and the integral (A.3.24) may then be evaluated by the $(m-1)$ times applications of the quadrature formula.

The smallest value $n$ can assume is about 4 (in order to give a reasonable accuracy). This means that in the integral (A.3.24) $m$ should at least be 4, before we can attempt to solve equation (4.9.1) otherwise results are going to be in error and this will affect the whole solution $P_0(\dot{y}_j, m\Delta)$. The problem of obtaining $P_0(\dot{y}_j, m\Delta)$ for $m<4$ does not exist if the processes under consideration are not 'singular' in the sense explained in Chapter.
TWO processes, it is reasonable to assume that $P_0(\dot{y}_j, 0) = 0$ for all $j$ and $Q(\dot{y}_j, m\Delta)$ is a good approximation to $P_0(\dot{y}_j, m\Delta)$ if $m$ is small. The initial portion of the solution $P_0(\dot{y}_j, m\Delta)$ can thus be obtained without solving the integral equation. The remaining part of the solution can be determined successively as follows: Using the values of $P_0(\dot{z}_j, k\Delta)$ just determined for $l=1,2\ldots n; k=0,1\ldots m-1$ we can evaluate $G(\dot{y}_j, k\Delta)$ for $j=1,2\ldots n; k=0,1\ldots m-1$ using (A.3.23). The function $F(\dot{y}_j, \dot{z}_j(m-k)\Delta)$ being obtained first for $j, l=1,2\ldots n; k=0,1\ldots m-1$ using expression (4.7.5).

Next, the integral (A.3.24) is evaluated through the application of a suitable $n$ point formula. $P_0(\dot{y}_j, m\Delta)$ is then the difference between $Q(\dot{y}_j, m\Delta)$ and (A.3.24).

$m$ is now incremented by one and the same procedure is used to obtain $P_0(\dot{y}_j, (m+1)\Delta)$ for all $j$.

By repeating this procedure again and again, the solution $P_0(\dot{y}_j, m\Delta)$ is obtained as a function of $\dot{y}$ and $m\Delta$ for as high a value of $m$ as desired. Finally, since:

$$P_0(\dot{\zeta}) = \int_{-\infty}^{0} P_0(\dot{y}, \dot{\zeta}) \, d\dot{y} \quad \ldots (A.3.25)$$

we can obtain $P_0(m\Delta)$ by integrating $P_0(\dot{y}_j, m\Delta)$ using the Laguerre-Gauss quadrature (A.3.18), namely:

$$P_0(m\Delta) = \sum_{j=1}^{n} \beta_j P_0(\dot{y}_j, m\Delta) e^{-\dot{y}_j} \quad \ldots (A.3.26)$$

In this manner, a new approximation to the distribution $P_0(\dot{\zeta})$ can be obtained.

Unfortunately for processes which are truly 2nd order Markovian, i.e. those with power spectra of the form:

$$\phi(\omega) = \frac{1}{(1 + \omega^2)^2} \quad \ldots (A.3.27)$$

The third derivative of the auto-correlation function.
possesses a discontinuity at the origin and the initial behaviour of $P_0(y_j, m\Delta)$ is not the same as $Q(y_j, m\Delta)$. Processes with spectra such as these, form the singular case defined in Chapter 2. In this case, the initial portion of the solution $P_0(y, m\Delta)$ cannot be obtained with any reasonable degree of accuracy. Consequently, the latter part would be inaccurate and it is doubtful whether the overall result would be a better approximation than the solution of McFadden's integral equation based on the quasi-independence assumption.
APPENDIX FOUR.

A Brief Note on Digital Filtering.

A brief description of the techniques used to design the following power spectra by means of digital filtering is given.

1) \( \phi_{1,k}(\omega) = \frac{1}{(1 + \omega^2/\omega_0^2)^k} \) ...(A.4.1)

2) \( \phi_{2,k}(\omega) = \frac{1}{(1 + (\omega/\omega_0)^2)^k} \) ...(A.4.2)

These correspond to Group (i) and Group (ii) spectra respectively, as discussed in Chapter Six. The design techniques are described in detail in a paper by C.M. Rader and E. Gold (50).

In general:

\[ \phi(\omega) = H(j\omega) \cdot H(-j\omega) \] ...(A.4.3)

where \( H(j\omega) \) is the frequency spectrum which is assumed to have all its poles in the left side of the s plane.

For Group (i) spectra we have

\[ H_{1,k}(j\omega) = \frac{1}{(1 + j\omega/\omega_0)^k} \] ...(A.4.1)

Letting \( s=j\omega \) (A.4.1) becomes

\[ H_{1,k}(s) = \frac{1}{(1+s/\omega_0)^k} \] ...(A.4.2)

where \( H_{1,k}(s) \) is the transfer function \( \frac{Y(s)}{X(s)} \) of a linear system having input \( X(s) \) and output \( Y(s) \) in the frequency domains.

Using the transformation

\[ Z^{-1} = e^{-TS} \] ...(A.4.3)

maps the transfer function \( H_{1,k}(s) \) into the function \( G_{1,k}(Z^{-1}) \) in the \( Z \) plane. This transforms the continuous linear system into a sampled data system whose input \( X(Z^{-1}) \) and output \( Y(Z^{-1}) \) are related by:
\[
\frac{Y(z^{-1})}{X(z^{-1})} = G_{1,k}(z^{-1}) \quad \ldots (A.4.4)
\]

where
\[
G_{1,k}(z^{-1}) = (-1)^{k-1} \frac{\omega_0^k}{(k-1)!} \frac{\delta^{k-1}}{\delta \omega_0^{k-1}} \left( \frac{1}{1 - z^{-1} e^{-\omega_0 T}} \right) \quad \ldots (A.4.5)
\]

T being the sampling interval.

The L.H.S. of (A.4.5) may be written as the ratio of two polynomials of degrees \( r \) and \( m \) respectively in \( z^{-1} \). This, together with (A.4.4) results in

\[
Y(z^{-1})(b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}) = X(z^{-1})(a_0 + a_1 z^{-1} \ldots + a_r z^{-r}) \quad \ldots (A.4.6)
\]

where \( r < m \) and both are functions of \( k \).

Taking the inverse Z transform of (A.4.6) yields:

\[
\sum_{i=0}^{m} b_i Y [(n-i) T] = \sum_{i=1}^{r} a_i X [(n-i) T] \quad \ldots (A.4.7)
\]

which is the difference equation of the digital filter having a transfer function given by (A.4.5).

Hence to design a digital filter which simulates the analog transfer function (A.4.1) requires simply the evaluation of the corresponding function in the Z plane using (A.4.5).

This immediately yields the coefficients \( a_i \) and \( b_i \) required to set up the difference equation.

For group (ii) spectra if \( (\omega/\omega_0)^{2k} \) is replaced by \( \tan^{2k} \frac{\omega T}{2} / \tan^{2k} \frac{\omega_0 T}{2} \), then the resulting spectra will have frequency characteristics very similar to those given in (A.4.2). Further, this substitution also results in eliminating all the frequencies above that given by half the sampling frequency and consequently gets rid of aliasing errors which are inherent in sampled data systems.

Below this frequency the two spectra are very close.

Letting \( Z = e^{j\omega T} \) in the modified group (ii) spectra results in
Rader and Gold showed that the poles of \( \Phi_{2,k}(Z) \) in the Z plane are given by:

\[
U_m = \frac{1 - \tan^2 \frac{\omega_0 m}{Z} \cos \frac{\pi m}{k} + \tan^2 \frac{\omega_0 m}{Z}}{1 - 2 \tan \frac{\omega_0 m}{Z} \cos \frac{\pi m}{k} + \tan^2 \frac{\omega_0 m}{Z}} \quad \quad \text{... (A.4.9)}
\]

\[
V_m = \frac{2 \tan \frac{\omega_0 m}{Z} \sin \frac{\pi m}{k}}{1 - 2 \tan \frac{\omega_0 m}{Z} \cos \frac{\pi m}{k} + \tan^2 \frac{\omega_0 m}{Z}} \quad \quad \text{\( m = 0, 1 \ldots 2k-1 \)}
\]

where \( U_m \) and \( V_m \) are the real and imaginary co-ordinates of the \( m \)th pole. Half of these poles lie inside the unit circle \( Z = e^{j\omega T} \).

In addition, the function \( \Phi_{2,k}(Z) \) has a zero of order 2k at \( Z = -1 \).

The transfer function \( G_{2,k}(Z) \) corresponding to the power spectrum \( \Phi_{2,k}(Z) \) is synthesised from k zeros at \( Z = -1 \) and those \( k \) poles inside the unit circle, this gives:

\[
G_{2,k}(Z) = \frac{(Z+1)^k}{\prod_{i=1}^{r} (Z-Z_i)} \quad \quad \text{... (A.4.10)}
\]

where \( Z_i \) are the poles lying inside the unit circle. These poles are in general, complex which occur in conjugate pairs. Thus if we define:

\[
G_i(Z) = \frac{(Z+1)^2}{(Z-Z_i)(Z-Z_i^*)} \quad \quad \text{... (A.4.11)}
\]

where \( Z_i^* \) is the conjugate of the pole \( Z_i \), then (A.4.10) may be written as

\[
G_{2,k}(Z) = G_0(Z) \cdot \prod_{i=1}^{r} G_i(Z) \quad \quad \text{... (A.4.12)}
\]

where \( r = \frac{k}{2} \) if \( k \) is even, or \( \frac{k-1}{2} \) if \( k \) is odd. Further if
k is even, then \( G_0(Z) = 1 \), otherwise \( G(Z) \) is given by

\[
G_0(Z) = \frac{1+Z}{Z-a} \quad \text{(A.4.13)}
\]

where \( a \) is the only co-ordinate of the resulting real pole. By dividing both the numerator and the denominator of the L.H.S. of (A.4.11) by \( Z^2 \) we obtain an expression \( G_i(Z^{-1}) \) in \( Z^{-1} \). Taking the inverse \( Z \) transform of this expression yields the following difference equation:

\[
Y_i(nT) = X_i(nT) + 2X_i[(n-1)T] + X_i[(n-2)T] + 2\Re(Z_i)Y_i[(n-1)T] - \left|Z_i\right|^2 Y_i[(n-2)T] \quad \text{(A.4.14)}
\]

where \( \Re(Z_i) \) is the real co-ordinate of the conjugate pair \( Z_i \) and \( \left|Z_i\right|^2 \) is the modulus square.

Similarly for (A.4.13) we obtain the first order difference equation

\[
Y_0(nT) = X_0(nT) + X_0[(n-1)T] + aY_0[(n-1)T] \quad \text{(A.4.15)}
\]

A digital filter described by \( G_{2,k}(Z) \) can be realised by cascading a number of elementary quadratic sections \( G_i(Z) \) which can be each realised by setting up a difference equation of the form (A.4.14). If \( k \) is odd, then in addition an extra first order section described by the difference equation (A.4.15) is needed. The output \( Y_i(nT) \) of the \( i \)th quadratic section represents the input \( X_{i+1}(nT) \) of the following section whereas the input of the first section is either \( Y_0(nT) \) or \( X(nT) \) depending on whether \( k \) is odd or even. As shown above a pre-requisite for the design of the digital filter having a power spectrum \( \phi_{2,n}(Z) \) is the knowledge of the pole and zero pattern of this function in the \( Z \)-plane.
References. Part I


References. Part II.

3. Rice, S.O. 'Distribution of the duration of Fades in Radio transmission' Bell System Tech. J. 37, No. 3 (1958) Sections VI & IX.
10. Favreau, R.R. Low, H. & Pfeffer, I. 'Evaluation of Complex...
Statistical functions by an Analog Computer' I.R.E.

11. Longuet-Higgins, M.S. 'On the Intervals between Successive

12. Seigert, A.J.F. 'On the roots of Markoffian Random Functions'
Rand Memo,Sept.5,(1950).

13. Slepian, D. 'The One-Sided barrier problem for Gaussian

14. McFadden, J.A. 'The Axis Crossings of a Stationary Gaussian

15. Schaffli, L. 'On the Multiple Integral \( \int_{0}^{x} \int_{y}^{z} \cdots \int_{z}^{n} \)'
Quart.J.Pure App. Maths. Vol 2, pp. 269-301; and Vol.3
pp.56-68,97-108.

16. Al-Nuaimi, M.O. 'On the Use of Hermite Polynomials in
series expansions for Multi-Dimensional Gaussian
1970, Loughbrough University of Technology.

17. Aitken, & Turnbull 'Theory of Canonical Matrices' Blackie
(1931).

18. Moran, P. 'Rank Correlation and Product moment Correlation

19. Kendall, H.G. 'Proof of Relations connected with the
Tetrachoric series and its generalisation' Biometrika,32
(1941) pp.196-198.

20. Polya, G. 'Sur quelques points de la theorie des

21. McFadden, J.A. 'Urn Models of correlation under comparison


